الجمهورية العربية السورية
وزارة التعليم العالي
جامعة تشرين
كلية العلوم

مبادئ في الاحتمالات والإحصاء

الدكتور
مبارك اسبر ديب
مدرس في قسم الرياضيات

1429 - 1430
2008 - 2009

الجمهورية العربية السورية
وزارة التعليم العالي
جامعة تشرين
كليّة العلوم

مبادئ في الاحتمالات والإحصاء

الدكتور
مبارك اسبر ديب
مدرس في قسم الرياضيات

القسم: الرياضيات
السنة: الأولى

1429 - 1430 هـ
2008 - 2009 م

المحتويات

الفصل الثاني

مبادئ أساسيّة في جبر المجموعات

الفصل الثالث

بعض المفاهيم والقواعد الأساسيّة في الاحتمالات

الفصل الرابع

المتغيّرات العشوائيّة والتوزيعات الاحتماليّة

الفصل الخامس

توزيعات احتماليّة منفصلة

الفصل السادس

توزيعات احتماليّة مستمرة

الفصل السابع

التوزيع المشترك لجملة متغيّرات عشوائيّة

الفصل الثامن

توزيع العيّنة والتقدير (النقطي ـ المجالي)

بسم الله الرحمن الرحيم

مقدمة:

النقطة ـ الرمز والإشارة ـ أصل الكلمة، والكلمة أصل المعرفة، وتطور أي علمٍ من العلوم هو نتاج تاريخ طويل من المعارف البشرية، لا بدّ أن يعتمد ويحمل في طيّاته جوانب مختلفة من علوم مختلفة، ومن ثمّ سيدفع بدوره لتطورات جديدة في مجالات علميّة متنوعة لتُنشأ بذات الوقت اهتمامات أخرى تضع أمام الباحثين مسائل جديدة تنتظر بدورها تطور علوم وتكنولوجيا متقدّمة تُسهم في إيجاد الحلول المثلى للمسائل المطروحة... وهكذا تتطور العلوم... كل العلوم... بما فيها الاحتمالات والإحصاء.

تعود بداية علم الاحتمال إلى ألعاب الحظ في القرن السابع عشر، ثمّ تطور هذا العلم نتيجة تضافر جهودٍ مختلفة في مجالات علميّة متنوعة، وأصبح ميداناً رئيساً لكثيرٍ من العلوم الأخرى وخصوصاً الإحصاء وتطبيقاته العمليّة، الذي أصبح بدوره أساسياً وكثير الاستخدام في المجالات العلميّة المتنوعة، وهو يشكّل إحدى أهم الأدوات المعاصرة لاتخاذ القرارات المثلى في ظروف تتحكم فيها الصدفة (إحصاء استقرائي).

ويهتم الإحصاء بشكل أساسي بدراسة المجتمع الإحصائي (وهو مجموعة العناصر أو الوحدات التي تشكل هدف الدراسة) من خلال دراسة عيّنة أو عدة عيّنات عشوائيّة مأخوذة من ذات المجتمع رغم أنّه في معظم المسائل والأبحاث العلميّة لا ينصب الاهتمام على نتائج العيّنة أو على البيان الإحصائي المدروس، إنما تتأتى الأهميّة من كون نتائج العيّنة أو البيان الاحصائي ممثّلاً لمجموعة أكبر من المعلومات الإحصائيّة تشكل العيّنة أو البيان المدروس جزءاً منها. (العيّنة هي الجزء الذي يخضع عملياً للدراسة وبالطبع من ذات المجتمع الإحصائي غير المتجانس أصلاً والمُراد دراسته).

يتألف المقرر من ثمانية فصول أستعرضها فيما يلي بإيجاز:

يعالج **الفصل الأول** بعض طرق عرض البيانات الإحصائيّة وتوزيعاتها التكراريّة، بالإضافة لمقاييس النزعة المركزيّة والتشتّت.

ويتناول **الفصل الثاني** مفهوم المجموعة والعمليات عليها، بالإضافة لبعض الطرق الحسابيّة المعتمدة في طرق العد.

وتعرضنا في **الفصل الثالث** لمفهوم الحدث كعنصر في مجموعة أحداث ثم لتعريف الاحتمال وخواصه بالإضافة لبعض المفاهيم والنظريات المتعلقة بالأحداث ودرسنا التجارب المتكرّرة وسحب العيّنات.

أما **الفصل الرابع** فقد تناول مفهوم المتغيّر العشوائي بنوعيه: المنفصل والمستمر، بالإضافة لبعض الصفات المميّزة للمتغيّر العشوائي ومتراجحة تشيبتشيف.

وتعرفنا في **الفصل الخامس** إلى بعض التوزيعات الاحتماليّة المنفصلة المهمة، مع تبيان شروط كلٍ منها وقيمه العدديّة المميّزة.

وعرضنا في **الفصل السادس** بعض التوزيعات الاحتماليّة المستمرة الهامة، وتعرفنا إلى شروطها والقيم المميّزة الخاصة بها، كما تطرقنا وبشكل مختصر لتوزيعات (كوشي . فيشر . كاي . ريلاي . ويبل . باريتو).

ودرسنا في **الفصل السابع** مفهوم الشعاع العشوائي والتوزيعات المتعلقة به سواءً كان هذا الشعاع منفصلاً أم مستمراً، إضافة لمفهوم الاستقلال، وتوزيع شعاع عشوائي تابع لشعاع عشوائي آخر، كما تعرّفنا إلى بعض القيم المميّزة لهذه الأشعة العشوائيّة.

وبيّنا في **الفصل الثامن** أهمية توزيع العيّنة ودرسنا بعض المتغيّرات الإحصائيّة المهمة، بالإضافة لقانون الأعداد الكبيرة (نظرية تشيبتشيف) ونظريّة النهاية المركزيّة.

كما تعرضنا لمفهوم المقدّر الاحصائي وخواصه بالإضافة لنظرية كرامر ولطريقتي المعقوليّة العظمى والعزوم في التقدير النقطي ، كما تعرفنا إلى بعض مجالات الثقة في المجتمعات الطبيعية وغير الطبيعية.

وقد زوّدنا كل فصل من المقرّر بمجموعة من الأمثلة المحلولة والتمارين غير المحلولة، وألحقنا نهاية المقرّر بملحق لبعض الجداول الإحصائيّة الخاصة ببعض التوزيعات الاحتماليّة ، ثم المصطلحات والمراجع العلميّة العربيّة والأجنبيّة.

آمل من الله أن أكون وفقت في عملي المتواضع هذا، والكمال لله العلي العظيم... والله الموفق.

د. مبارك ديب

الفصل الأول

البيانات الإحصائيّة

(1- 1) مقدمة- *Introduction* :

الإحصاء هو فرع من فروع الرياضيات التطبيقيّة، وهو علم مهم وضروري وله تطبيقات واسعة جداً في شتى المجالات العلميّة والإداريّة والاقتصاديّة وغيرها.

عُرِّف الإحصاء قديماً بأنه مجرّد جمع وترتيب معلومات في جداول أو إظهارها بشكل رسوم بيانيّة أو أشكال أخرى، ثم تطوّر هذا المفهوم إلى المعنى الحديث حيث يعرَّف الإحصاء بأنّه: العلم الذي يبحث في جمع وتنظيم وعرض وتحليل البيانات ثم استقراء النتائج واتخاذ القرارات المثلى بناءً على المعطيات.

فجمع البيانات يعني: الحصول على المعلومات الخاصة بالقياسات أو التعدادات أو المشاهدات وما شابه، من قبل الإحصائي، وكلما كانت المعلومات دقيقة، كلما اعتُمِدَت الدراسة بشكل أوثق، وكانت النتائج أفضل، والعكس صحيح.

في حين تنظيم البيانات وعرضها يعني: عملية وضع البيانات في جداول منسّقة، ثم تمثيل هذه البيانات بطرق مناسبة (بيانيّاً أو هندسياً)، أو بشكل توزيعات تكرارية كما سنجد لاحقاً.

وأما تحليل البيانات فيعني: إيجاد مقاييس محدّدة من تلك البيانات نُريد اعتمادها في دراسة الفرضيّة أو المسألة المطروحة حيث يُفضّل دراسة وتلخيص البيانات بمؤشرات رقميّة، علماً أنّ تمثيل البيانات بشكل جدولي أو بياني يعطي فكرة عامة وسريعة في وصف الظاهرة المدروسة ، لكنها ليست كافية كونها ستختلف حتماً باختلاف الباحثين، وخصوصاً عندما تُمثَّل بطرق مختلفة.

وأخيراً تأتي مرحلة الاستقراء واتخاذ القرار الإحصائي المناسب وهي تأتي بشكل عام على شكل تنبؤ أو تقدير أو قرار (قبول أو رفض) الفرضيّة الإحصائيّة.

يقسم علم الإحصاء إلى قسمين: إحصاء وصفي (جمع وتنظيم وعرض وبعض التحليل)، وإحصاء استقرائي (تحليل واستقراء النتائج واتخاذ القرارات المناسبة للحالة الإحصائيّة أو للظاهرة المدروسة).

بما أنّ المعلومات مستمرة طالما ارتبطت بالزمن، وحواس الإنسان محدودة، بالتالي لا يمكن للإنسان أن يستوعب هذه المعلومات إلا بشكلها المتقطّع، ولما كانت البيانات هي جزء من تلك المعلومات، لذلك كان من الضروري عرض هذه البيانات بطرق محدّدة شيّقة وسهلة قدر الإمكان ليسهل فهمها وبالتالي المقارنة بين مفرداتها وتفسير الشاذ منها، ومن ثم اتخاذ القرار المناسب بشأنها.

(2-1) طرق عرض البيانات: *Methods of presenting Data*:

(1-2-1) طريقة الجداول (*Tables*):

حيث توضع البيانات في جداول تعبّر عن ارتباط الظواهر بعضها ببعض أو بالزمن.

مثال (1): لنفرض أنّ عدد عمال مؤسسة افتراضيّة كان بالشكل التالي:

الجدول (1): عدد عمال المؤسسة الافتراضيّة خلال الأعوام 1997 وحتى 2006 م

السنة	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
عدد العمال	200	220	280	300	400	525	600	675	700	750

يستحسن عند استعمال هذه الطريقة ذكر: عنوان الجدول، الوحدات المستخدمة، مصادر المعلومات أو البيانات (إن تمكنّا من ذلك)، وسبب شذوذ بعض البيانات إن وجدت.

(2-2-1) طريقة المستطيلات أو الأعمدة – *Bar Graph*:

وتستخدم هذه الطريقة للمقارنة بين قيم الظواهر حسب الزمن أو المسميّات، وفي هذه الطريقة يُعبَّر عن كل مسمّى بمستطيل أو عمود، بحيث يكون ارتفاعه ممثّلا للقيمة المقابلة لذلك المسمّى، وذلك باستعمال مقياس رسم مناسب.

مثال (2): إذا كان عدد خريجي أحد أقسام كليّة ما لعدة سنوات مُمثّلاً في الجدول التالي:

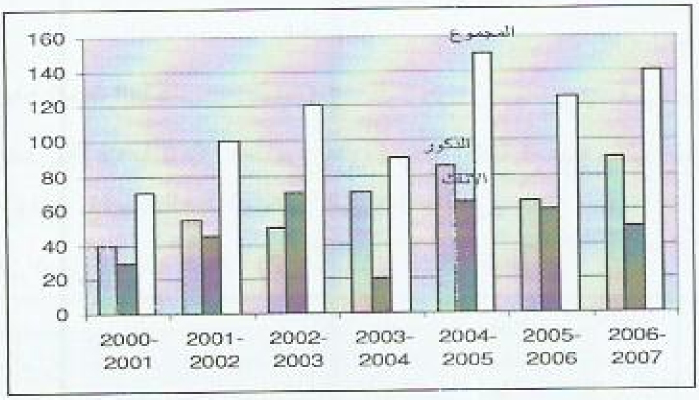
المطلوب: اعرض هذه البيانات بطريقة المستطيلات:

الجدول(2)

السنة	الذكور	الإناث	المجموع

70	30	40	2000-2001
100	45	55	2001-2002
120	70	50	2002-2003
90	20	70	2003-2004
150	65	85	2004-2005
125	60	65	2005-2006
140	50	90	2006-2007

إذا مثّلنا السنوات بخط أفقي، ورسمنا مقابل كل سنة ثلاثة أعمدة أو مستطيلات، بحيث يُعبّر المستطيل الأول منها عن عدد الذكور، ويُعبّر الثاني عن عدد الإناث، والثالث يمثّل المجموع ، فإنّنا نحصل على الشكل التالي:

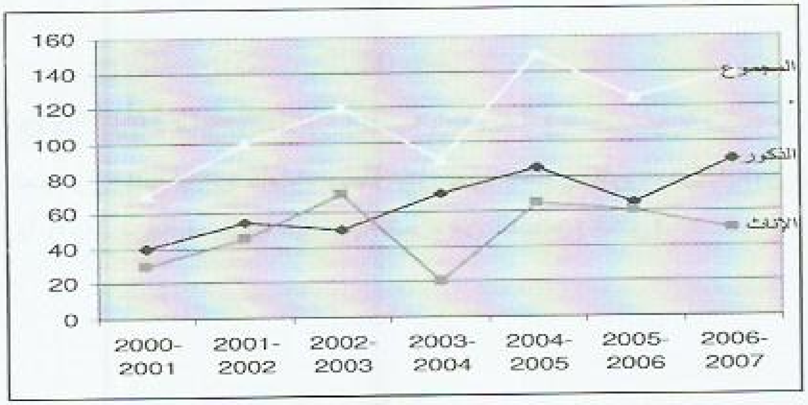


شكل (1)

(1 . 2 . 3) طريقة الخط المنكسر – *Broken Line Graph* :

وهي تعبّر عن تغيُّر ظاهرة أو عدة ظواهر مع الزمن أو مع مسمّيات أخرى.

مثال (3): لنعرض بيانات المثال (2) بطريقة الخط المنكسر وذلك برسم محورين متعامدين، وبمقياس رسم مناسب نحصل على الشكل التالي:

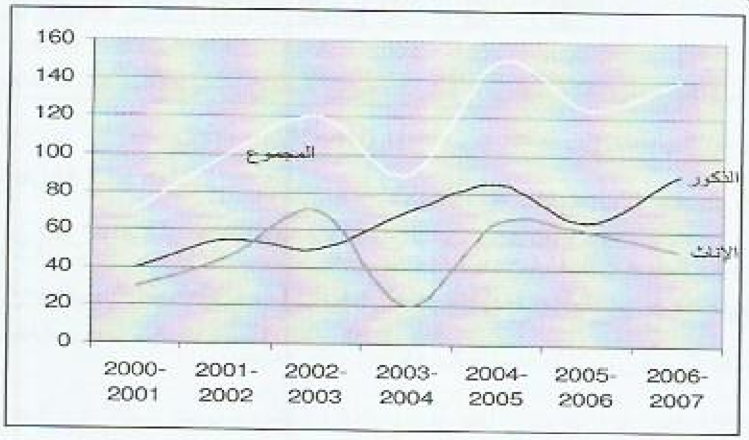


شكل (2)

(1-2-4) طريقة الخط المنحني – *Curve* :

وتستعمل هذه الطريقة عندما تتغيَّر الظاهرة على فترات زمنية قصيرة وكثيرة، وهي تماثل الطريقة السابقة وذلك بتقريب الخط المنكسر إلى خط منحني.

مثال (4): يمكن تمثيل بيانات المثال (2) كما في الشكل التالي:



شكل (3)

(1-2-5) طريقة الدائرة . *Pie Chart*:

وهنا يُمثَّل المجموع الكلي بدائرة كاملة، ويُمثَّل كل جزء بقطاع دائري بحيث يكون قياس زاويته مساوياً (360 درجة مضروباً في نسبة الجزء من المجموع الكلي).

مثال (5): إذا كان عدد مدرسي مادتي الرياضيات والفيزياء في خمسة من مراكز التعليم في المحافظة كما في الجدول التالي: اعرض هذه البيانات بطريقة الدائرة:

الجدول (3)

المركز	(1)	(2)	(3)	(4)	(5)
العدد	10	12	6	8	4

الحل: واضح أنّ عدد المدرسين هو: $40 = 4 + 8 + 6 + 12 + 10$

نحسب زاوية قطاع كل ثانوية، حيث:

زاوية قطاع الثانوية الأولى هي: $360\times\frac{10}{40}=90^{\circ}$

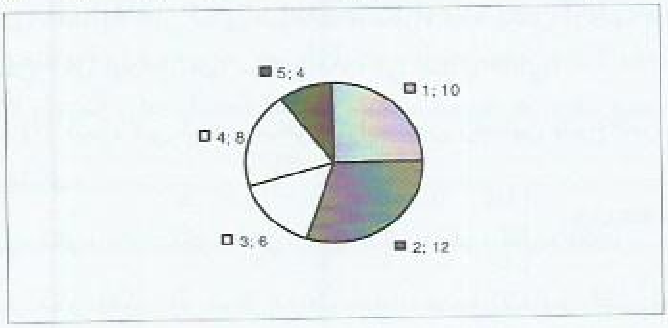
زاوية قطاع الثانوية الثانية هي: $360\times\frac{12}{40}=108^{\circ}$

زاوية قطاع الثانوية الثالثة هي: $360\times\frac{6}{40}=54^{\circ}$

زاوية قطاع الثانوية الرابعة هي: $360\times\frac{8}{40}=72^{\circ}$

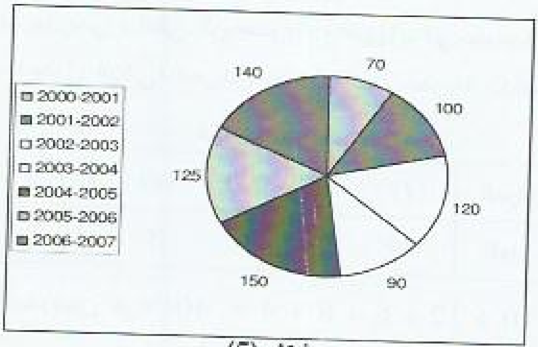
زاوية قطاع الثانوية الخامسة هي: $360\times\frac{4}{40}=36^{\circ}$

ثم نرسم دائرة، ونرسم القطاعات حسب زواياها، فنحصل على الشكل التالي:



شكل (4)

مثال (6): يمكن تمثيل مجموع الخريجين في بيانات المثال (2) بطريقة الدائرة بالشكل التالي:



شكل (5)

(3-1) التوزيع التكراري – *Frequency Distribution*:

إنّ البيانات التي نحصل عليها في أثناء جمع المعلومات الإحصائيّة، هي قياسات عدديّة (كميّة) أو وصفيّة، وبشكل عام: إنّ تأمُل بيانات كبيرة الحجم معروضة بطريقة الجداول أو المستطيلات بشكل غير منظم، لن يقدّم لنا إلا القليل عن مدلول هذه القياسات وتفسيرها، أو عن كيفية تغيّرها بالنسبة لبعضها البعض، ولما كان هذا من صلب اهتمامنا، لذا يتوجب علينا تصنيف هذه البيانات ثم تنظيمها، وذلك بتقسيم مدى هذه البيانات (المدى هو: أعلى قيمة في البيان مطروحاً منها أصغر قيمة) إلى فئات، ثم ذكر عدد البيانات الواقعة في كل فئة، وهذا سيساعدنا أكثر في فهمها وتفسيرها، وبالتالي إمكانية الحكم عليها. أما إذا كان مدى البيانات صغيراً، فإنّه يمكن ترتيبها تصاعديّاً أو تنازلياً، ثم نضع تكرار كل قيمة أمامها مباشرة، كما في المثال التالي:

مثال (7): سُئِلَت 10 نساء في أحد مشافي التوليد عن رغبتهن بنوع المولود القادم، وكانت أجوبتهن بالشكل:

أنثى، ذكر، ذكر، ذكر، أنثى، ذكر، ذكر، ذكر، أنثى، ذكر.

نلاحظ أنّ الاختيار ذكر تكرّر (أو تواتر) ست مرات، بينما الاختيار أنثى تكرّر ثلاث مرات.

يمكن ترتيب هذه المعلومات (البيانات) في جدول يوضّح كيف توزعت الأجوبة العشرة بين الاختيارين: ذكر، أنثى.

الجدول (4)

الاختيار	أنثى	ذكر
التكرار f	3	7

يسمّى هذا الترتيب للمعلومات توزيعاً تكرارياً، حيث رمزنا للتكرار بالرمز f وهذا التوزيع التكراري، كما هو واضح من الجدول، مؤلّف فقط من:

a . فئات قيم الشاهدات.

b . التكرارات الموافقة لكل مشاهدة أو فئة.

ومن الجدير ذكره أنّ البيانات الإحصائيّة، إما أن تكون مرتّبة / حيث ينتمي كل عنصر يخضع للتصنيف إلى صنف واحد منها فقط كالتصنيف: (ذكر، أنثى) أو (علمي، أدبي)/، أو مبوّبة (حيث كل عدد من الأعداد المختلفة يمكن أن يمثّل بأكثر من فئة وذلك عكس البيانات المرتّبة حيث كل عدد من الأعداد المختلفة في البيان يمثّل فئة بحد ذاته).

مثال (8): تمثّل البيانات التالية درجات خمسة عشر تلميذاً:

10 - 10 - 8 - 6 -10 – 7 - 9 - 8 - 6 – 4 –6 - 9 – 6 - 4 - 7

ولنعرض هذه البيانات في توزيع تكراري: قبل كل شيء نبحث عن المدى وهو الفرق بين أكبر قيمة وأصغر قيمة، وهنا نلاحظ أن المدى ($6 = 4 - 10$) وهو صغير، وبالتالي يمكن ترتيب هذه البيانات تصاعدياً ، فتبدأ بأصغر درجة وهي 4 وتنتهي بأعلى درجة وهي 10، ثم نضع تكرار كل درجة أمامها، وعندها نحصل على الجدول التالي:

الجدول (5)

الدرجة	4	6	7	8	9	10	مجموع التكرارات
التكرار	2	4	2	2	2	3	15

واضح أيضاً من هذا المثال أنّ التوزيع التكراري هو عبارة عن جدول مؤلّف من:

a – فئات قيم القياسات.

b – التكرارات الموافقة لكل قياس أو فئة.

(1-3-1) بناء التوزيع التكراري:

قلنا إنّه إذا كان مدى البيانات صغيراً، فإنّه يمكن استخدام جدول التوزيع التكراري السابق، ولكن إذا كان مدى البيانات كبيراً، فإنّه لبناء التوزيع التكراري ، يُستحسن تقسيم قيم البيانات إلى فئات، بالشروط التالية:

1 – أن يتراوح عدد الفئات من 5 إلى 20 فئة. (علماً أننا قد نخسر بعض المعلومات، إذا قلّ عدد الفئات عن خمس، وقد نحتفظ بما لا ضرورة له من التفاصيل، عندما يزيد عدد الفئات على عشرين).

2 – أن تكون الفئات منفصلة عن بعضها البعض ومتلاصقة مع بعضها (أي غير متقاطعة ولا تتضمن فراغاً بينها).

3 – الفئات متساوية في الطول ما أمكن.

وهذا الاختيار لتقسيم الفئات، يمكّننا من تفادي ظهور فئات خالية في أثناء عملية إفراغ البيانات أو ما يسمّى بعملية الفرز، بحيث إذا اخترنا أية قيمة من البيانات، فإنّها تكون موجودة حتماً في فئة واحدة وواحدة فقط . وبشكل عام ، إنّ عدد الفئات يتوقّف على عدد البيانات، صغيراً كان أم كبيراً ؟ وبعد تقسيم الفئات ، تُجمَع التكرارات المقابلة لكل فئة ، كما في المثال التالي :

مثال (9): تمثّل البيانات التالية درجات 50 طالباً في مقرر الاحتمالات والإحصاء:

الجدول (6)

85 – 94 – 79 – 77 – 70 – 91 – 50 – 61 – 60 – 72
57 – 63 – 66 – 72 – 81 – 64 – 79 – 66 – 76 – 80
95 – 87 – 57 – 64 – 74 – 91 – 72 – 80 – 73 – 68
55 – 59 – 66 – 89 – 71 – 62 – 56 – 75 – 84 – 98
64 – 73 – 87 – 66 – 80 – 75 – 53 – 69 – 83 – 76

والمطلوب: وضع هذه البيانات في توزيع تكراري .

الحل: نُرتبها تصاعدياً بالشكل:

50 – 53 – 55 – 56
57 – 57 – 59 – 60 – 61 – 62 – 63
64 – 64 – 64 – 66 – 66 – 66 – 66 – 68 – 69 – 70
71 – 72 – 72 – 72 – 73 – 73 – 74 – 75 – 75 – 76 – 76 – 77
79 – 79 – 80 – 80 – 80 – 81 – 83 – 84
85 – 87 – 87 – 89 – 91 – 91
94 – 95 – 98

ثم نفرزها ونوزعها (نصففها) بالشكل التالي:

الجدول (7)

5	0	3	5	6	7	7	9								
6	0	1	2	3	4	4	4	6	6	6	6	8	9		
7	0	1	2	2	2	3	3	4	5	5	6	6	7	9	9
8	0	0	0	1	3	4	5	7	7	9					
9	1	1	4	5	8										

وبتصفيفها بهذا الشكل (طريقة الجذع والورقة)، تكون البيانات قد توزعت إلى خمس فئات حيث نعتبر أنّ: نصيب الفئة الأولى هو كل القياسات التي تنتمي إلى المجال [50 ,60).

وهي بالضبط : 59 – 57 – 57 – 56 – 55 – 53 – 50

ونصيب الفئة الثانية هو كل القياسات التي تنتمي إلى المجال (70, 60] وهكذا ... حتى الفئة الخامسة التي تتضمن كل القياسات التي تنتمي إلى المجال (99, 91]

وهي بالضبط: 98 – 95 – 94 – 91 – 91

من الواضح في البيان السابق ، أنّ أكبر قياس في البيان هو 98، وأنّ أصغر قياس هو 50 ، وبالتالي يكون:

المدى = الفرق بين أكبر قياس وأصغر قياس أي : المدى = 48 = 50 – 98

لنقسم هذا المدى إلى عدد من الفئات نختاره بصورة كيفية، وليكن هنا، مثلاً سبع فئات، طول كل منها سبعة كما يلي: 98 – 92, 91 – 85 ,... ,63 – 57, 56 – 50

مع التأكيد على الشروط السابقة: الفئات منفصلة، أي غير متداخلة فيما بينها، ومتساوية في الطول ما أمكن، وكل قيمة في البيان الإحصائي يجب أن تقع في فئة واحدة وواحدة فقط.

ولنشكل جدولاً مؤلّفاً من عدة أعمدة مثل الجدول (8) التالي، حيث:

نضع في العمود الأول: حدود الفئات، ونضع في العمود الثاني (ويسمى عمود الفرز): (إفراغ البيانات)، حيث نضع بجانب كل فئة خطاً مائلاً مقابل كل قياس في البيان ينتمي إلى هذه الفئة، ونضع في العمود الثالث: تكرار الفئات (حيث تكرار الفئة i هو عدد القياسات في البيان الإحصائي التي تنتمي إلى الفئة i ونرمز له بالرمز f_i) وهذا العمود يبيّن عدد القياسات في البيان الإحصائي.

وعلى الرغم من أنّ العمودين: الأول والثالث، كافيان لتمثيل البيانات المذكورة في توزيع تكراري، إلا أننا سنضيف بعض الأعمدة الأخرى لحاجتها لاحقاً، حيث:

نضع في العمود الرابع: مركز الفئة، علماً أن:

مركز الفئة i يساوي مجموع حدّي الفئة i مقسوماً على 2 .

فمثلاً: مركز الفئة الأولى $= 53 = 2/(56 + 50)$ ، ومن الملاحظ أنه بإضافة طول الفئة إلى هذا المركز ، نحصل على مركز الفئة الثانية وهكذا ... ونعتبر مركز كل فئة ممثّلاً لجميع قياساتها . مثلاً : القياسات الفعليّة ضمن الفئة الأولى $56 - 50$ هي: 56 , 55 , 53 , 50 (انظر الجدول7) ومجموعها الفعلي هو 214 . في حين ، إذا أردنا حساب مجموع القياسات ضمن هذه الفئة، فإنّنا نتّخذ مركز الفئة، وهو هنا 53 ممثّلاً لجميع القياسات الأربعة، أي نفترض القياسات الأربعة في هذه الفئة كأنها: 53 , 53 , 53 , 53 . ونعتبر مجموع الفئة مساوياً لـ $212 = 53 \times 4$، ونلاحظ هنا أنّنا ارتكبنا خطأً بالنقصان مقداره 2 ، وهو الغرامة أو الثمن الذي ندفعه مقابل كفاءة العرض وسهولة سرعة الحسابات التي أردناها، وبالواقع إنّ الأخطاء في الفئات المختلفة لا تكون عادة في اتجاه واحد ، وبالتالي يعدّل بعضها بعضا، ويكون بشكل عام الخطأ الإجمالي تافهاً قياساً بالوفر الذي نكون حقّقناه.

كذلك يمكن أن نضع: في عمود خامس: التكرار النسبي (علماً أننا نرمز للتكرار النسبي للفئة i بالرمز $p_i = \frac{f_i}{n}$ ، أي : تكرار الفئة i مقسوماً على العدد الكلي للقياسات $n = 50$)، ونضع في عمود سادس: التكرار المئوي (حيث التكرار المئوي للفئة i يساوي التكرار النسبي للفئة i مضروباً بـ % 100)، وسنلاحظ في الجدول أنّ مجموع عمود التكرارات (وهو العمود الثالث) يجب أن يساوي 50 ، في حين أنّ مجموع عمود التكرارات النسبي يجب أن يساوي الواحد لأنّ:

$$\sum_i p_i = \frac{\sum_i f_i}{n} = \frac{n}{n} = 1$$

الجدول (8)

حدود الفئات	عمود الفرز	التكرار f_i	مركز الفئة	التكرار النسبي p_i	التكرار المئوي %
50-56	||||	4	(50+56)/2=53	4/50=0.08	8
57-63	~~||||~~ ||	7	60	0.14	14
64-70	~~||||~~ ~~||||~~	10	67	0.02	2
71-77	~~||||~~ ~~||||~~ ||	12	74	0.24	24
78-84	~~||||~~ |||	8	81	0.16	16
85-91	~~||||~~ |	6	88	0.12	12
92-98	|||	3	95	0.06	6
المجموع		50		1	

ولما كان التوزيع التكراري هو أي ترتيب للبيان الإحصائي، يُظهر توزُّع قياساته على فئات، فإنّه عند العرض النهائي للتوزيع التكراري يمكن أن نكتفي بكتابة العمودين الأول والثالث بالشكل:

الجدول (9)

حدود الفئات	50-56	57-63	64-70	71-77	78-84	85-91	92-98	مجموع التكرارات
التكرار f_i	4	7	10	12	8	6	3	50

ولكن، قد لا نهتم بمفردات البيان الإحصائي رغم أهميتها، إنما قد يقتصر اهتمامنا على معرفة كيفية توزُّعها على الفئات التي نحدِّدُها بدقّة، فكل الجداول صحيحة ، وأفضلها هو ذاك الذي يمكّننا بسهولة من تبيان النواحي المهمة فيه، وعلى سرعة الإجابة عن الأسئلة

المهمة المطروحة. مثلاً: ما نسبة القياسات التي تزيد أو تنقص عن قيمة معيّنة؟ ومن المعروف أنّ البيانات الإحصائيّة هي بشكل عام ، معلومات مأخوذة من عيّنة صغيرة نسبياً ، وذلك قياساً بحجم المجتمع الذي جاءت منه هذه المعلومات والذي نريد اتخاذ القرار بشأنه .

فمثلاً : ماذا لو غيّرنا عدد الفئات ، وبالتالي طول الفئة ؟ فهل سنحصل على توزيع تكراري مختلف ؟ وهل التوزيع التكراري الذي سنحصل عليه في هذه الحالة أفضل من التوزيع الآنف الذكر ؟ ولماذا ؟ أضف إلى أنّه ماذا لو حصل أحد الطلبة على إحدى الدرجات (56.5 أو 63.5 أو... أو 98.5) ؟ وعندئذٍ إلى أية فئة ستنتمي هذه الدرجات ؟

وبشكل عام ، إنّ التوزيع التكراري لبيان إحصائي على الفئات، يفترض أن يؤدي مجموعة أهداف مهمة منها:

- توزيع كل قياس على الفئة الواحدة هو توزيع منتظم.
- اعتبار مركز الفئة ممثلاً لجميع قياساتها (أي: كل قياس يمكن اعتباره مساوياً لمركز الفئة).
- يمكن بسهولة الإجابة بدقّة عن الأسئلة المطروحة.

ولهذه الاعتبارات وغيرها، وجد المهتمون الإحصائيون أنّه من المناسب تحديد عدد الفئات، ثم تعريف حدود الفئات بصورة واضحة تماماً، وهذا يتأتى من تحديد دقّة البيانات الإحصائيّة، أي: هل البيانات مقرّبة إلى عدد صحيح ؟ وتسمى عندئذٍ بالبيانات الإحصائيّة المنفصلة /مثل: عدد حالات الزواج، أو عدد حوادث المرور، أو عدد حالات النجاح أو الرسوب ... وما شابه/، أم هل هي مقرّبة إلى عدد عشري واحد أو عددين عشريين ؟ وتسمى عندئذ بالبيانات الإحصائيّة المستمرة / مثل: قياسات الطول أو درجات الحرارة أو الوزن ... وما شابه/ .

فعندما نقول إنّ طول طفل هو 80.5 سم، فهذا يعني أنّ القياس جرى بدقّة تصل إلى أقرب ملليمتر، وباستخدام جهاز آخر أكثر دقّة، يمكن أن نحصل على قياس يقع في

مكان ما بين 80.45 و 80.547 سم، وفي كل الأحوال سيدوِّر الرقم العشري لنحصل على العدد 80.5 سم.

انطلاقاً من هذا ، ولأسباب أخرى ، نأخذ الحدود المضبوطة لكل فئة وتسمى الحدود الفعليّة للفئة أو "حدي الفئة ". إذاً من دقّة البيانات تحدّد الحدود الفعليّة للفئات، وهذه الحدود الفعليّة ستضمَن حسابات أكثر دقّةً، وستعالج مشكلة وجود فراغات بين الفئات، حيث تضمن تطابق نهاية فئة مع بداية الفئة التي تليها مباشرة.

فمثلاً: من أجل الفئة 98 – 92 ، تكون الحدود الفعليّة لها هي: 98.5 – 91.5 أي: بطرح 0.5 من الحد الأدنى، وإضافة 0.5 إلى الحد الأعلى، علماً أنّ 0.5 هو نصف وحدة الدقّة.

أما إذا كانت البيانات الإحصائيّة مقرّبة إلى عدد عشري واحد، فإنّنا نحصل على الحدود الفعليّة للفئة بطرح 0.05 من الحد الأدنى، وإضافة 0.05 إلى الحد الأعلى، علماً أنّه يوجد طرق أخرى يمكن استخدامها للتعبير عن حدود الفئة، فمثلاً يمكن التعبير عن الفئة 98 – 92 بالشكل: – 92 ونعني بها: جميع الأعداد الواقعة ضمن المجال (98 – 92]، أي الأعداد بدءاً من 92 إلى أقل من 98.

ويبقى ثابتاً أنّ: طول الفئة يساوي الفرق بين حدّيها الفعليين، ومركز الفئة يساوي نصف مجموع حدّيها، ومن المستحسن أن يكون لكل الفئات طول واحد،

(مع ملاحظة أنّه في هذه الحالة يُحسَب طول الفئة بالشكل: طول الفئة = المدى مقسوماً على عدد الفئات، مقرّباً إلى أعلى).

فمثلاً: لو أردنا في مثالنا تقسيم المدى إلى خمس فئات، لكان: طول الفئة مقرباً إلى أعلى هو $10 = \frac{48}{5}$. وبشكل عام: في بيان إحصائي محدّد المدى، يكون طول الفئة متناسباً عكساً مع عدد الفئات، وبالتالي زيادة (أو نقصان) عدد الفئات سيقلّل (أو سيزيد) من طول الفئة، وبالعكس. وبناءً على ما تقدّم (آخذين بعين الاعتبار المراكز الفعليّة للفئات)، يكون التوزيع التكراري للمثال (9). كما في الشكل:

الجدول (10): التوزيع التكراري للمثال (9)

التكرار f_i (عدد القياسات في الفئة i)	مركز الفئة	الحدود الفعليّة للفئات

4	$(50+56)/2=53$	$49.5-56.5$
7	59	$56.5-63.5$
10	65	$63.5-70.5$
12	74	$70.5-77.5$
8	81	$77.5-84.5$
6	88	$84.5-91.5$
3	95	$91.5-98.5$
50		**المجموع**

مع ملاحظة أنّه في بعض الأحيان، يكفي أن نذكُر مركز الفئة فقط، ولا حاجة لذكر الحدود الفعليّة للفئات، كما يمكن وضع التكرار النسبي (وهو نسبة تكرار الفئة إلى مجموع التكرارات) والتكرار المئوي، مع مراكز الفئات (أو مع حدودها الفعليّة)، لنحصل على التوزيع التكراري النسبي والتكرار المئوي للمثال (9) كما في الشكل:

الجدول (11): التكرار النسبي والتكرار المئوي للمثال (9)

التكرار المئوي %	**التكرار النسبي** p_i	**الحدود الفعليّة للفئات**
8	0.08	$49.5-56.5$
14	0.14	$56.5-63.5$
2	0.02	$63.5-70.5$
24	0.24	$70.5-77.5$
16	0.16	$77.5-84.5$
12	0.12	$84.5-91.5$
6	0.06	$91.5-98.5$
50	1	**المجموع**

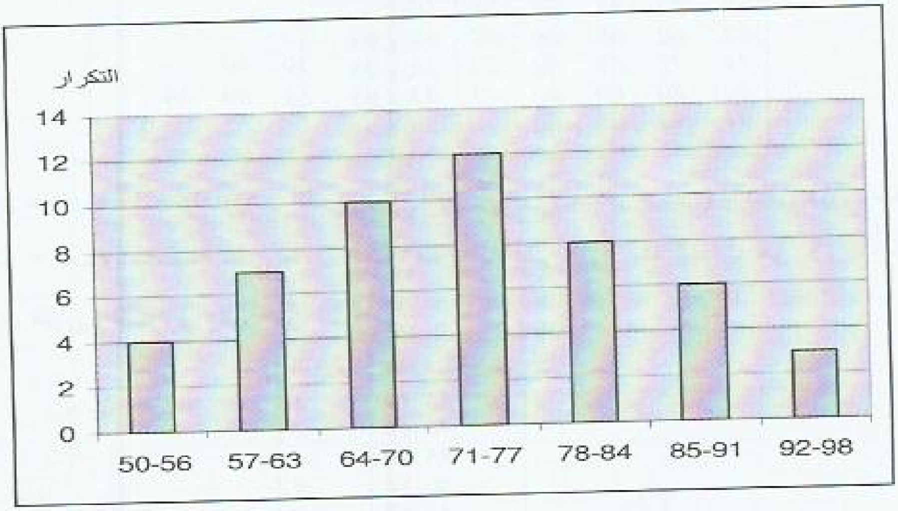
(1-4) التمثيل البياني للتوزيعات التكراريّة- *Graphic Presentation*:

يُستخدَم التمثيل البياني للمعلومات على نطاق واسع، وهو يمثّل أنموذج رسم، أو صورة هندسيّة لجملة معطيات (صورة مصغّرة عن الواقع)، وله أهميّة كبيرة في الاقتصاد وفي مراكز البحوث العلميّة والاجتماعيّة، وكذلك في مراكز اتخاذ القرار، وفي الصحف والمجلات...الخ ويستخدم بشكل شبه يومي حيث وسائل التعبير عنه كثيرة، وسنعتمد هنا أهم ثلاث طرق في التمثيل البياني في مجال الاستقراء الإحصائي والبيانات الإحصائيّة:

(1-4-1) المدرّج التكراري- *Frequency Histogram*:

نتّخذ المحور الأفقي لتمثيل الفئات، ونتّخذ المحور الشاقولي لتمثيل التكرارات، بحيث يمثّل تكرار كل فئة بمستطيل، قاعدته عبارة عن الحدود الفعليّة لتلك الفئة، وارتفاعه

يتناسب مع تكرار هذه الفئة (وهو يساوي بالضبط مقدار التكرار، إذا كانت الفئات متساوية الطول، وإذا لم تكن الفئات متساوية في الطول، يكون ارتفاع المستطيل مساوياً لحاصل قسمة تكرار الفئة على طول الفئة نفسها). وبناءً عليه، يمكن رسم المدرّج التكراري للمثال (9) حيث الفئات متساوية في الطول كما في الشكل التالي:



شكل (6)

مثال (10): فيما يلي درجات 40 طالباً في أحد المقررات:

الجدول (12)

69	56	81	75	52	65	49	80	67	59
52	76	66	54	73	69	83	62	53	79
88	80	44	71	72	87	91	82	89	79
42	88	37	75	98	93	73	62	96	80

والمطلوب:

1 . شكّل جدول التوزيع التكراري لهذا البيان مُستخدِماً الفئات:

35 – 39 ; 40 – 44 ;...

2 . ارسم مدرّج ومضلّع التكرار.

الحل: يمكن ترتيب البيانات تصاعدياً بالشكل:

الجدول (13)

37									
42	44	49							
52	52	53	54	56	59				
62	62	65	66	67	69	69			
71	72	73	73	75	75	76	79	79	
80	80	80	81	82	83	87	88	88	89
91	93	96	98						

المدى $= 61 = 37 - 98$ **(وهو الفرق بين أكبر قياس وأصغر قياس)** بالتالي فإنّ عدد الفئات = المدى مقسوماً على طول الفئة، مقرّباً للأعلى، أي عدد الفئات هو $13 \approx \frac{61}{5}$

وجدول التوزيع التكراري يكون من الشكل:

التكرار	حدود الفئات
1	35-39
2	40-44
1	45-49
4	50-54
2	55-59
2	60-64
5	65-69
4	70-74
5	75-79
6	80-84
4	85-89
2	90-94
2	95-99

نذكر أنّه يمكن كتابة الفئات السابقة بالشكل:

$35-, 40-, 45-, 50-, 55-, 60-, 65-, 70-, 75-, 80-, 85-, 90-, 95-$

(1-4-2) المدرّج التكراري النسبي:

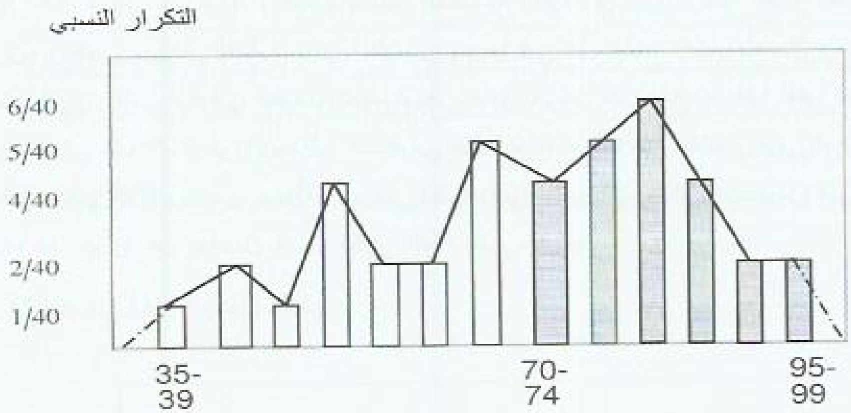
إنّ التوزيع التكراري النسبي: هو الجدول الذي يعطينا حدود الفئات أو مراكز الفئات مع تكراراتها النسبية، وطريقة رسم المدرّج التكراري النسبي تختلف عن مدرّج التكرار فقط، بارتفاع العمود أو المستطيل المقام فوق كل فئة، حيث ارتفاع العمود لكل فئة في هذه

الحالة، يمثّل بالتكرار النسبي لهذه للفئة، علماً أنّ التكرار النسبي لكل فئة هو نسبة تكرار الفئة إلى مجموع التكرارات.

مثال (11): (التوزيع . المدرّج) التكراري النسبي للمثال (10) هما كما يلي:

الجدول (15): التوزيع التكراري النسبي للمثال (10)

التكرارالنسبي	حدود الفئات
1/40	35-39
2/40	40-44
1/40	45-49
4/40	50-54
2/40	55-59
2/40	60-64
5/40	65-69
4/40	70-74
5/40	75-79
6/40	80-84
4/40	85-89
2/40	90-94
2/40	95-99



شكل (7): مدرّج التكرار النسبي للمثال (10)

وإذا اعتبرنا أنّ طول الفئة (وهي تساوي خمس وحدات) أصبحت وحدة قياس جديدة (أي طول الفئة بالوحدة الجديدة هو الواحد)، عندئذٍ تصبح مساحة المستطيل المُقام فوق الفئة، مساوية للتكرار النسبي الموافق لهذه الفئة، وستصبح المساحة الكليّة تحت مدرّج التكرار النسبي مساوية للواحد تماماً. لاحظ:

$(1 = \frac{40}{40} = \frac{6}{40} + \frac{10}{40} + \frac{12}{40} + \frac{10}{40} + \frac{2}{40})$ ، ولنسأل الآن، ما هي نسبة الطلبة الذين تزيد درجاتهم على 70 مثلاً ؟

بالعودة إلى مدرّج التكرار النسبي نرى أنّ هذه النسبة تشمل كل الفئات على اليمين من 70 . وبالاستفادة من الجدول السابق مباشرة (آخذين بالاعتبار الحدود الفعليّة للفئات) نجد أنّ 23 طالباً حصلوا على أكثر من 70 درجة، وبالحقيقة، إنّ هذه النسبة هي ذاتها المساحة تحت مدرّج التكرار النسبي والواقعة على يمين الدرجة 70. أي أنّ النسبة المطلوبة هي 23/40 ، وهي ذات المساحة المظللة على الشكل السابق.

(1-4-3) المضلّع التكراري – *Frequency Polygon :*

وهو عبارة عن مضلّع مغلق ، نحصل عليه بأخذ منتصفات الأضلاع العُلويّة للمستطيلات في المدرّج التكراري ، ثم نصل فيما بينها بخطوط مستقيمة (آخذين بعين الاعتبار أنّ هناك فئتين متطرفتين واحدة إلى أقصى اليسار والثانية إلى أقصى اليمين وتكرار كل منهما صفراً). وعندئذٍ سيبدأ المضلّع التكراري من مركز الفئة الصفرية اليسرى، وينتهي بمركز الفئة الصفرية اليمنى مروراً في كل منتصفات الأضلاع العُلويّة للمستطيلات. علماً أنّه في الفئات المتساوية يعتبر مركز كل فئة إحداثياً أفقياً لنقطة، ويعتبر تكرار هذه الفئة هو الإحداثي الشاقولي لتلك النقطة. أما في الفئات غير المتساوية فإنّ الإحداثي الأفقي يبقى ممثّلاً بمركز الفئة، في حين الإحداثي الشاقولي في هذه الحالة، هو عبارة عن حاصل قسمة تكرار الفئة على طولها.

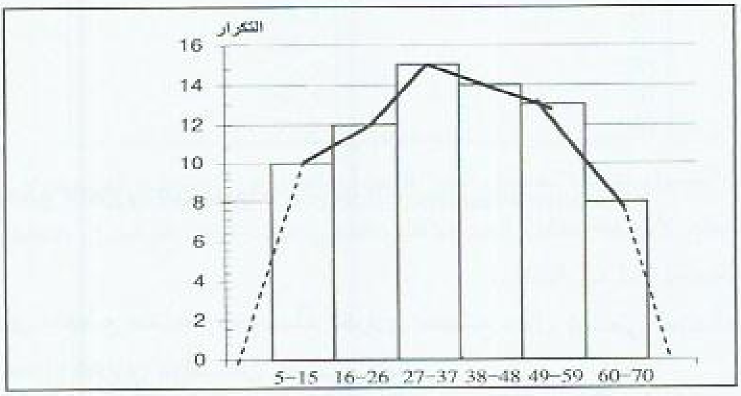
مثال (12): من أجل الجدول التالي:

الجدول (16)

التكرار	الحدود الفعليّة للفئات	حدود الفئات
10	4.5-15.5	5-15
12	15.5-26.5	16-26
15	26.5-37.5	27-37

38-48	37.5-48.5	14
49-59	48.5-59.5	13
60-70	59.5-70.5	8
المجموع		72

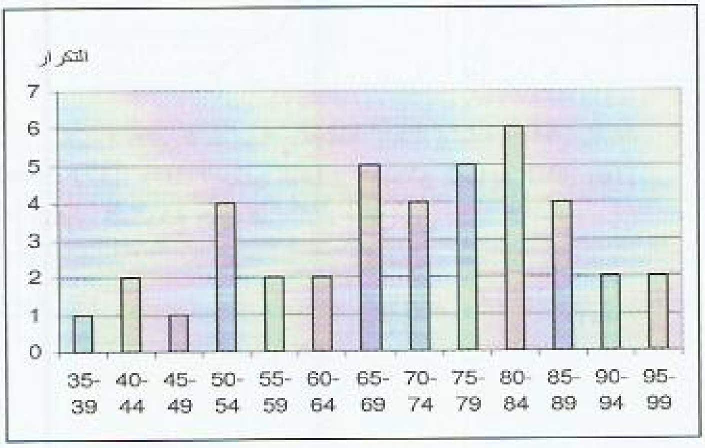
يكون (مدرّج ومضلّع) التكرار على الشكل التالي:



الشكل (8)

فإذا كانت الفئات متساوية تكون المساحة تحت المضلّع التكراري مساوية تماماً لمجموع مساحة المستطيلات (علماً أنّه لهذه المساحة دور كبير في الإحصاء) وإذا كانت الفئات غير متساوية، فإنّه يوجد فرق بين المساحتين، ويقل هذا الفرق كلما ازداد عدد الفئات وصَغُرَ طول الفئة.

مثال (13): المضلّع التكراري للمثال (10) كما في الشكل: (قارن مع الشكل (7) ولاحظ وحدة القياس على المحور الشاقولي):



الشكل (9)

(1−5) **المضلّع التكراري المتجمّع** – *mutative Frequency Polygon* :

وهو نوعان:

مضلّع تكراري متجمّع صاعد ، ومضلّع تكراري متجمّع نازل (يسمى أحياناً بالهابط)، ولكل منهما جدول تكراري مؤلّف من عمودين فقط.

أما بالنسبة للجدول التكراري المتجمّع الصاعد:

فإنّنا نضع في العمود الأول وعنوانه "أقل من" الحدود الفعليّة العليا للفئات، علماً أنّنا نبدأ بالحد الأدنى الفعلي للفئة الأولى حيث تكرارها صفر ، لأنّه لا توجد بيانات تقل قيمتها عن ذلك الحد أو تساويه)، ونضع في العمود الثاني وعنوانه "التكرار المتجمّع الصاعد"، عدد القياسات الموافق .

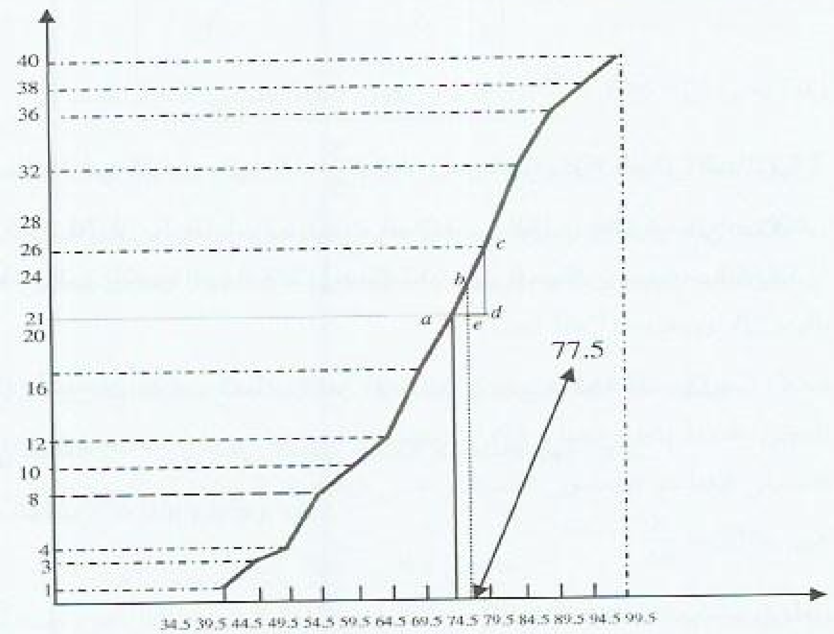
وبناءً عليه، فإنّ جدول التوزيع التكراري المتجمّع الصاعد للمثال (10) يكون على الشكل التالي:

الجدول (17): جدول التوزيع التكراري المتجمّع الصاعد للمثال (10)

التكرار المتجمّع الصاعد	أقل من
0	34.5
1	39.5
3	44.5

49.5	4
54.5	8
59.5	10
64.5	12
69.5	17
74.5	21
79.5	26
84.5	32
89.5	36
94.5	38
99.5	40

وبتمثيل هذا الجدول بيانياً، نحصل على المضلّع التكراري المتجمّع الصاعد (حيث يرصد التكرار المتجمّع لأي فئة مقابل الحد الأعلى الفعلي لهذه الفئة، ثم نصل النقاط فيما بينها بخطوط مستقيمة) كما في الشكل:



الشكل (10) : مضلّع التكرار المتجمّع الصاعد لدرجات الطلبة

ولما كان عدد القياسات هو $n = 40$، فإنّه للبحث عن القياس الذي يقع على اليسار منه 60% من القياسات، نبحث عن رتبة القياس المطلوب فنجد أنها: $n\times\frac{60}{100} = 40\times\frac{60}{100} = 24$ وهي تقع بين الرتبتين التصاعديتين 21 و26، وهذا يعني

أنّ القياس المطلوب ينتمي للمجال [79.5 , 74.5[. والمطلوب الآن هو تحديد قيمة هذا القياس؟ وهذا ممكن، حيث نجد من الجدول (17) أنّ:

21 قياساً من بين الأربعين قياساً أقل من 74.5.

26 قياساً من بين الأربعين قياساً أقل من 79.5.

وبتطبيق التناسب الطردي نقول: إنّه عندما زاد التكرار المتجمّع الصاعد بمقدار 5 (من 21 إلى 26) ، زاد القياس بمقدار 5 من (74.5 إلى 79.5) ، بالتالي ما هي قيمة الزيادة في القياس عندما يزداد التكرار المتجمّع بمقدار 3 (من 21 إلى 24) ؟ (من الواضح أنها ستكون 3) . أو بتطبيق التناسب الطردي نكتب :

زيادة القياس	زيادة التكرار
5	5
x	3

أي : الزيادة المطلوبة في القياس هي: $x = 3\times\frac{5}{5} = 3$ ، وعندئذٍ يكون القياس المطلوب هو : 77.5 = 3+74.5 أو بطريقة أخرى: (حيث العلاقة بين القياس والتكرار هي علاقة خطيّة، والقياسات التي تنتمي إلى فئة تتوزّع بانتظام عليها)، يمكن حسابه بالطريقة التالية:

من الشكل (10) يتّضح أنّ القياس المطلوب هو الإحداثي السيني للنقطة b، حيث:

الإحداثي السيني للنقطة b = الإحداثي السيني للنقطة a مضافاً إليه a e.

ولكن من تشابه المثلثين: a b e و a c d نجد:

$$\frac{ae}{ad} = \frac{eb}{dc} \Rightarrow \frac{ae}{5} = \frac{3}{5} \Rightarrow ae = 3$$

أي أنّ القياس المطلوب هو : $74.5 + 3 = 77.5$

وهو القياس الذي يقع على يساره 60% من القياسات. ولو حسبنا القياس الذي يقع على يساره 50% من القياسات (وهو ما سنطلق عليه اسم الوسيط، وسندرسه لاحقاً)، لكانت

رتبة القياس تساوي $20 = 40\times\frac{50}{100}$، وهي تقع بين الرتبتين التصاعديتين 17 و21، وهذا القياس ينتمي للمجال [69.5 , 74.5]، وبطريقة مشابهة نجد من جدول التكرار المتجمّع الصاعد أنّ: 17 قياساً من القياسات الأربعين أقل من 69.5 ، وأنّ 21 قياساً من القياسات الأربعين أقل من 74.5 .

وبتطبيق التناسب الطردي نقول : إنّه عندما زاد التكرار المتجمّع الصاعد بمقدار 4 (من 17 إلى 21)، زاد القياس بمقدار 5 (من 69.5 إلى 74.5)، بالتالي نحسب قيمة الزيادة في القياس عندما يزداد التكرار المتجمّع بمقدار 3 (من 17 إلى 20) كما يلي:

زيادة القياس	زيادة التكرار المتجمّع
5	4
x	3

وهذا يعني أنّ: $x = 3\times\frac{5}{4} = 3.75$ ، ويكون عندئذٍ القياس المطلوب هو :

$69.5 + 3.75 = 73.25$ وهو القياس الذي يقع على يساره 50% من القياسات.

وبشكل عام: يمكن تحديد القياس الذي يقع على يساره $k\%$ من القياسات، حيث k هو أي عدد صحيح بين الصفر والمئة. (يسمى مثل هذا القياس المئين k ونرمز له بالرمز P_k وسندرسه لاحقاً أيضاً).

فمثلاً: لمعرفة نسبة القياسات التي تقل عن 54.5 ، نرفع من النقطة 54.5 على المحور السيني عموداً يقطع مضلّع التكرار المتجمّع الصاعد في نقطة، نرسم منها موازياً للمحور السيني فيقطع المحور الصادي في النقطة 8 وعندئذٍ تكون النسبة المطلوبة هي: $\frac{8}{40} = 20\%$.

وبطرق مشابهة، يمكن الإجابة على كثير من الأسئلة من هذا النوع باستخدام المضلّع التكراري المتجمّع. **أما بالنسبة لجدول التوزيع التكراري المتجمّع النازل:**

فإنّنا نضع في العمود الأول وعنوانه "أكثر من" الحدود الفعليّة الدنيا للفئات (علماً أنّنا ننتهي بالحد الأعلى الفعلي للفئة الأخيرة حيث التكرار صفر، لأنّه لا يوجد بيانات تزيد قيمتها عن ذلك الحد أو تساويه).

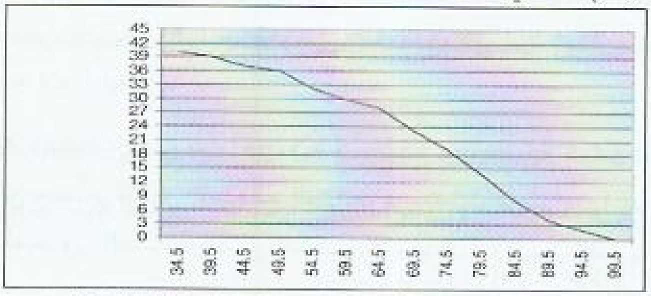
في حين نضع في العمود الثاني وعنوانه "التكرار المتجمّع النازل" عدد القياسات الموافق. ويمكن طرح أسئلة مشابهة لما سبق ، والإجابة عليها بطرق مماثلة.

مثال (14): الجدول التكراري المتجمّع النازل للمثال (10) كما يلي:

الجدول (18): جدول التوزيع التكراري المتجمّع النازل لدرجات الطلبة

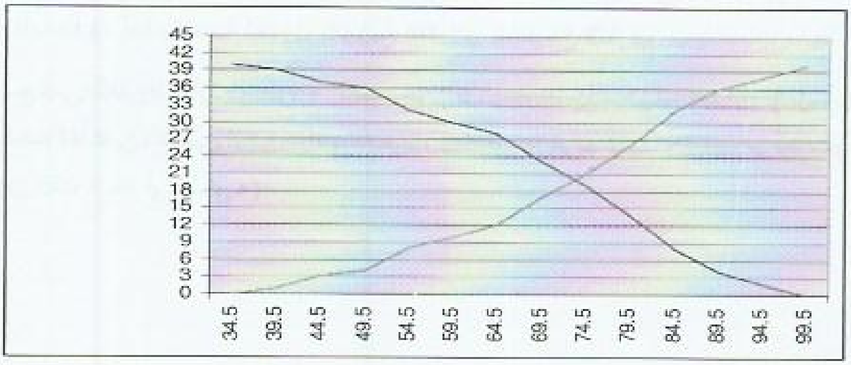
التكرار المتجمّع النازل	أكثر من
40	34.5
39	39.5
37	44.5
36	49.5
32	54.5
30	59.5
28	64.5
23	69.5
19	74.5
14	79.5
8	84.5
4	89.5
2	94.5
0	99.5

وبتمثيل هذا الجدول بيانياً، نحصل على المضلّع التكراري المتجمّع النازل (حيث يُرصد التكرار المتجمّع لأي فئة مقابل الحد الأدنى الفعلي لهذه الفئة، ثم نصل النقاط فيما بينها بخطوط مستقيمة) كما في الشكل:



الشكل (11): المضلّع التكراري المتجمّع النازل(للمثال 10)

وبرسم المضلّعين معاً (المتجمّع الصاعد والنازل) نحصل على الشكل التالي:



الشكل (12)

(1-6) المنحني التكراري *Frequency Curve* - :

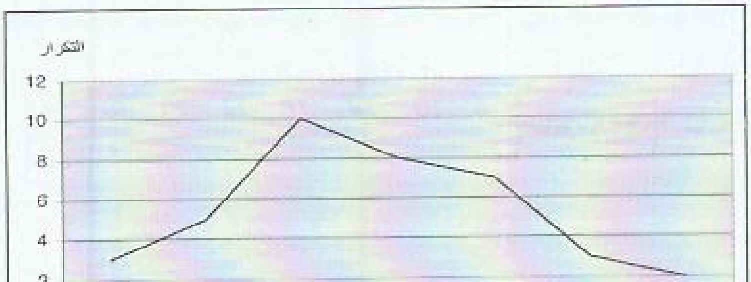
يمكن الحصول على المنحني التكراري إذا كان البيان الإحصائي كبيراً ومن النوع المستمر (قياسات حرارة، زمن، وزن، ...) حيث يكون عدد فئاته كبيراً، وطول الفئة صغير، وعندئذٍ يمكن تقريب المضلّع التكراري إلى منحنٍ تكراري، والمنحني التكراري هو صورة أو أنموذج لظاهرة ما يوحي بأنّ المسألة المطروحة تخضع لتوزيع ما، بمعنى آخر: المضلّع التكراري لظاهرة ما يوحي بشكل المنحني التكراري والذي بدوره يُعتَمَد لوصف أو لدراسة بعض أو كل عناصر تلك الظاهرة، وأهم ما يمكن الاستفادة منه، هو مقارنته (إن أمكن ذلك) بمنحنيات أخرى لتوزيعات احتمالية معروفة (سندرس بعضها لاحقاً) وبالتالي الاستفادة منها لتمثيل بيانات المسألة المدروسة ومعرفة بعض أو كل قيمها المميّزة.

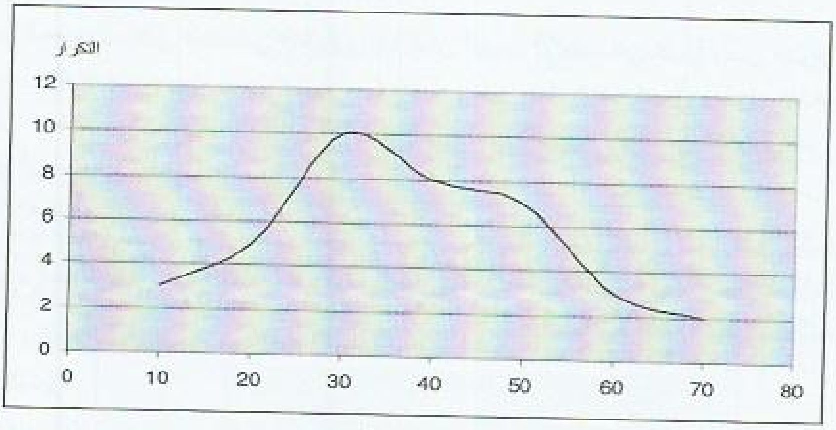
مثال (15): لنأخذ البيان الإحصائي التالي:

الجدول (18)

مركز الفئة	10	20	30	40	50	60	70
التكرار	3	5	10	8	7	3	2

ويكون المضلّع التكراري والمنحني التكراري له (وهو تقريب متواضع للمضلّع التكراري) كما في الشكل:





الشكل (13)

فالمنحني التكراري هو تقريب للمضلّع التكراري، ويمثله بشكل أفضل كلما كان عدد الفئات كبيراً وطول الفئة صغيراً، ولكن العدد القليل من القياسات قد يجعل العديد من الفئات خالية وتكرارها صفراً مما يُسَبِّب انقطاعات في المضلّع التكراري، وهذا بدوره سيؤدي إلى تمثيل سيئ للمنحنى التكراري، الأمر الذي يعني أنّنا لن نستطيع وصف أو دراسة الظاهرة بحيث نحصل على القرار الأمثل والمفضّل.

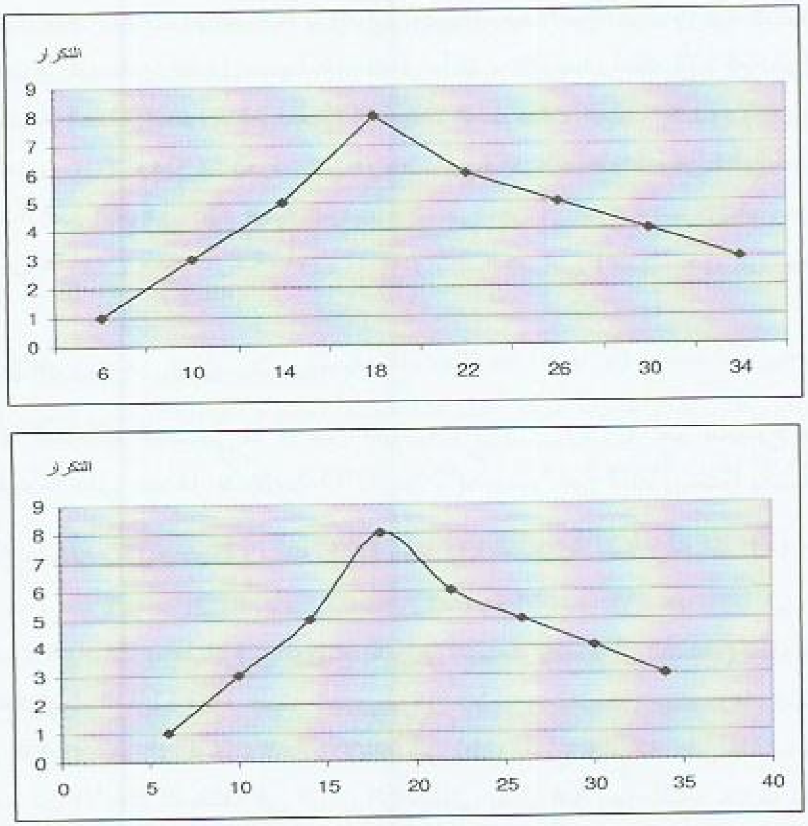
مثال (16): لنأخذ البيان الإحصائي التالي:

الجدول (19)

مركز الفئة	6	10	14	18	22	26	30	34

لتكرار	1	3	5	8	6	5	4	3

ويكون المضلّع التكراري والمنحني التكراري (المتواضع جداً) له كما في الشكل:



شكل (14)

لاحظ كيف يصعد بسرعة، ويتباطأ رويداً رويداً. هل يمكنك الإتيان بظاهرة مشابهة؟

والمنحني التكراري حيث يصبح أكثر تعبيراً ودقّة عندما يزداد عدد الفئات ويصغُر طول الفئة.

(7-1) مقاييس النزعة المركزيّة:

إنّ طرق تمثيل البيانات الإحصائيّة (جدوليّاً أو بيانيّاً) تعطي فكرة عامة وسريعة في وصف الظاهرة، وهي طرق مفيدة للغاية، لكنها ليست كافية، لأنّها تختلف باختلاف الباحثين، وخصوصاً عندما تُمثّل الظواهر أو البيانات بطرق مختلفة.

من المفضّل دراسة البيانات الإحصائيّة بمؤشرات رقميّة، بعيدة عن أهواء ورغبة الباحثين في اختيار طرق الوصف والتمثيل، حيث يحدث في أغلب التوزيعات التكراريّة أن تتمركز (تتجمّع) بيانات الظاهرة حول نقطة معيّنة، مثلاً: السلوكيات العامة للناس (الكرم – التعلّم – التعاون ...) وهذا ما يُعرَف في الإحصاء بظاهرة النزعة المركزيّة أي: نزعة القيم المختلفة إلى التمركز عند قيمة نموذجية تسمى مقياس النزعة المركزيّة، وهناك مجموعة من مقاييس هذه النزعة، أهمها: المتوسّط الحسابي (المتوسّط)، الوسيط ، المنوال .

- **المتوسّط الحسابي**– *Arithmetic Mean* :

تعريف (1): المتوسّط الحسابي لـ n من القياسات : $x_1, x_2, ..., x_n$ هو مجموع هذه القياسات مقسوماً على عددها n (نرمز له بالرمز $\overline{x}$)، ونعبّر عن ذلك رياضيّاً بالشكل:

$$\overline{x}=\frac{x_1+x_2+...+x_n}{n}=\frac{\sum_{i=1}^{n}x_i}{n}$$

أما إذا كانت البيانات معطاة في توزيع تكراري، حيث (القيم أو الفئات) معلومة، بالإضافة للتكرارات المقابلة لها، عندئذٍ نعتبر مركز الفئة، ممثّلاً لكل عناصرها. ونذكّر بأنّ التوزيع التكراري هو أي ترتيب للبيان الإحصائي يُظهِر توزُّع قياساته على فئات، وإذا كان كل عدد من الأعداد المختلفة في البيان الإحصائي يمثّل فئة بحد ذاته، عندها نقول إنّه لدينا بيان مُرتّب ، كالبيان الموضّح بالشكل التالي:

x_i (القيم المختلفة)	x_1	x_2 ...	x_m
f_i (التكرار)	f_1	f_2 ...	f_m

حيث يتّضح أنّه لدينا m قيمة أصلية منفصلة نرمز لها بـ $x_1, x_2, ..., x_m$ في البيان الإحصائي المؤلّف أصلاً من n قيمة، هي: $x_1, x_2, ..., x_n$ ، وبالتالي إذا رمزنا بـ $\sum_{i}^{n}x_i$ لمجموع قياسات البيان الإحصائي الأصلي ، فسيكون المجموع $\sum_{j}^{m}f_j x_j$ هو

عبارة عن قيمة أخرى للقيمة المضبوطة $\sum_{i}^{n} x_i$ ، بمعنى آخر: يمكن التعبير عن مجموع قيم البيان الإحصائي بشكلين متكافئين: $\sum_{i}^{n} x_i = \sum_{j}^{m} f_j x_j$ ، وممّا تقدّم يمكن القول: إنّ البيان المرتّب للقيم المتصلة هو بيان موزّع على فئات، بحيث كل عدد من الأعداد المختلفة فيه يمثّل فئة بحد ذاته، وفيما عدا ذلك، يسمى بالبيان المبوّب أو المصنف (والذي نفترض فيه أنّ كل القياسات التي تنتمي إلى فئة، تكون مساوية لمركز هذه الفئة)، كالبيان الموضّح بالشكل التالي:

(مركز الفئة) x'_i	x'_1	x'_2 ...	x'_m
(التكرار) f_i	f_1	f_2 ...	f_m

وهذا يعني أنّنا صنّفنا (أو بوّبنا) قيم البيان الإحصائي $x_1, x_2, ..., x_n$ في m قيمة هي: $x'_1\ ,\ x'_2\ ,\ ....\ ,\ x'_m$ واعتبرنا مركز كل فئة ممثّلاً لجميع القياسات التي تنتمي إلى هذه الفئة ، وهنا يكون المجموع $\sum_{j}^{m} f_j x_j$ تقريباً للمجموع $\sum_{i}^{n} x_i$.

تعريف (2): إذا كان لدينا توزيع تكراري عدد فئاته m ، وكانت مراكز الفئات (أو القيم) هي : $x_1\ ,\ x_2\ ,\ ....\ ,\ x_m$ ، وتكراراتها على الترتيب من الشكل: $f_1\ ,\ f_2\ ,\ ....\ ,\ f_m$ ، فإنّ المتوسّط الحسابي للقياسات المبوّبة يعرّف بالشكل التالي:

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i\, x_i}{\sum_{i=1}^{m} f_i}$$

أي أنّ المتوسّط الحسابي لـ n من القياسات المبوّبة $x_1, x_2, ..., x_n$ لظاهرة معيّنة، هو عبارة عن (مجموع حاصل ضرب مركز كل فئة في التكرار الموافق لها) مقسوماً على مجموع التكرارات . مع التذكير: أنّ التكرار f_i يقابل مركز الفئة i ، وأنّ : $n = \sum_{i=1}^{m} f_i$.

مثال (1): إنّ متوسط القيم: $6, 9, 2, 11, 4$ هو :

$$\bar{x} = \frac{6+9+2+11+4}{5} = \frac{32}{5} = 6.4$$

السؤال الثاني: قدّر المتوسّط الحسابي لبيانات الجدول التكراري التالي:

حدود الفئة	50-56	57-63	64-70	71-77	78-84	85-91	92-98
تكرار الفئة	4	7	10	12	8	6	3

الحل: لتطبيق العلاقة السابقة: $\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i}$ يلزم معرفة مراكز الفئات x_i، ولأجل ذلك نُشكّل الجدول التالي:

حدود الفئات	x_i (مراكز الفئات)	f_i	$x_i f_i$
50-56	53	4	212
57-63	60	7	420
64-70	67	10	670
71-77	74	12	888
78-84	81	8	648
85-91	88	6	528
92-98	95	3	285
$\sum$		50	3651

بالتالي يكون المتوسّط الحسابي المطلوب: $\bar{x} = \frac{\sum_{i=1}^{m} x_i f_i}{\sum_{i=1}^{m} f_i} = \frac{3651}{50} = 73.02$

وبإيجاد المتوسّط الحسابي مباشرة من البيانات الأوليّة، المعطاة في الجدول (6)، تكون النتيجة بالشكل:

$$\bar{x} = \frac{85+94+....+83+76}{50} = 72.9$$

وبمقارنة النتيجتين، تجد اختلافاً بينهما مقداره 0.12 ، وسبب ذلك الاختلاف يعود إلى أنّ المتوسّط الحسابي في التوزيعات التكراريّة المبوّبة إلى فئات هو تقريبي، ذلك لأنّنا فرضنا أنّ تكرار الفئة يقع على مركزها، مثلاً: فرضنا في الفئة الأولى أنّ العدد 53 (وهو مركز الفئة الأولى) قد تكرّر 4 مرات، بينما في الواقع التكرار 4 جاء من قيم مختلفة تتراوح ما بين 50 و56 وبالحقيقة أحياناً نهمل الفرق مقابل الوفرة في الجهود الحسابيّة اللازمة، بالإضافة إلى وضوح العرض وسرعة الإجابة عن أسئلة مطروحة، وكنا قد أشرنا إلى ذلك سابقاً.

س؟ : إذا كان متوسط 17 قياساً يساوي 9 ، فما مجموع هذه القياسات ؟

المتوسّط الحسابي المرجّح (الموزون) لأوساط حسابيّة:

تعريف: بفرض أنّه لدينا: مجموعة ذات n_1 من القياسات، متوسطها الحسابي $\bar{x}_1$ ، ومجموعة ثانية ذات n_2 من القياسات، متوسطها الحسابي $\bar{x}_2$ ، ومجموعة ثالثة ذات n_3 من القياسات، متوسطها الحسابي $\bar{x}_3$. عندئذٍ يكون المتوسّط الحسابي للمجموعة ذات $(n_1+n_2+n_3)$ من القياسات الناتجة من دمج المجموعات الثلاث هو :

$$\bar{x}=\frac{n_1\,\bar{x}_1+n_2\,\bar{x}_2+n_3\,\bar{x}_3}{n_1+n_2+n_3}=\frac{\sum_{i=1}^{3}n_i\,\bar{x}_i}{\sum_{i=1}^{3}n_i}$$

وهو ما يسمى بالمتوسّط الحسابي المرجّح (أو الموزون) لثلاثة متوسطات حسابية. وهو يقع حتماً بين المتوسّطات الحسابيّة التي حسب منها.

وفي الحالة الخاصة ، إذا كان: $n_1=n_2=n_3=n$ فإنّ:

$$\bar{x}=\frac{n\,(\bar{x}_1+\bar{x}_2+\bar{x}_3)}{3n}=\frac{\bar{x}_1+\bar{x}_2+\bar{x}_3}{3}$$

ويسمى عندئذٍ متوسط المتوسّطات.

ويمكن تعميم فكرة المتوسّط الحسابي المرجّح على أي عدد محدود من المجموعات مع بعضها البعض، بشرط أن يكون كل منها متوسطاً للعدد نفسه من القياسات، ومن الخطأ الشائع حساب متوسط المتوسّطات عند حساب متوسط عام.

مثال (3): إذا كان متوسط دخل ثلاث فئات من الموظفين شهرياً بالشكل :

الفئة	(1)	(2)	(3)
متوسط الدخل	6000	8000	10000

وكان عدد الموظفين في هذه الفئات هو على الترتيب : 12 , 10 , 7

احسب المتوسّط العام للدخل في الفئات الثلاث ؟

الحل:

إنّ إجمالي دخل الموظفين من الفئة الأولى هو : 72000=12 x 6000

وإنّ إجمالي دخل الموظفين من الفئة الثانية هو : 80000=10 x 8000

وكذلك إجمالي دخل الموظفين من الفئة الثالثة هو : 70000=7 x 10000

ويكون المتوسّط العام للدخل في الفئات الثلاث عندئذٍ :

$$\bar{x} = \frac{\sum_{i=1}^{3} n_i x_i}{\sum_{i=1}^{3} n_i} = \frac{72000 + 80000 + 70000}{12 + 10 + 8} = 7655.17$$

في حين نجد أنّ متوسط المتوسّطات يكون : $8000 = \frac{10000 + 8000 + 6000}{3}$ (وهو لا يطابق المتوسّط العام بالطبع) ، لأنّ عدد الموظفين ليس واحداً في الفئات الثلاث ، ولو كان عدد الموظفين في كل فئة متساوياً ويساوي مثلاً: 10 موظفين، عندئذٍ يكون:

دخل الموظفين من الفئات الثلاث هو بالترتيب : 100000 , 80000 , 60000

وحينئذٍ يصبح المتوسّط العام للدخل في الفئات الثلاث عبارة عن: $8000 = \frac{100000 + 80000 + 60000}{10 + 10 + 10}$. وهو يطابق متوسط المتوسّطات في هذه الحالة، والجواب عندئذ فقط يكون صحيحاً.

مثال (4): لنفترض أنّ ثلاث مجموعات من القياسات لها المتوسّطات: $\bar{x}_3 = 22$, $\bar{x}_2 = 20$, $\bar{x}_1 = 25$ ، تتضمّن 20 ، 25 ، و 30 قياساً بالترتيب. فما هو متوسطها بعد ضمّها في مجموعة واحدة ؟

الحل:

المتوسّط:
$$\bar{x}=\frac{n_1\,\bar{x}_1+n_2\,\bar{x}_2+n_3\,\bar{x}_3}{n_1+n_2+n_3}=\frac{\sum_{i=1}^{3}n_i\,\bar{x}_i}{\sum_{i=1}^{3}n_i}$$

$$=\frac{20\times25+25\times20+30\times22}{20+25+30}=\frac{500+500+660}{75}=\frac{1660}{75}=22.133$$

لاحظ أنّ متوسط المتوسّطات هو : $\frac{25+20+22}{3}=\frac{67}{3}=22.33$

(المتوسّط لا يُحسَب إلا من بيانات عدديّة).

خصـائص المتوسّـط الحسـابي: (اثـر الانسـحاب وتغييـر سلّم القيـاس فـي المتوسّـط الحسابي):

1 . يتأثر المتوسّط الحسابي بالعمليات الجبريّة $(+,-,\times,\div)$:

أي: إذا كانت $x_1, x_2, ..., x_n$ جملة من القياسات ، متوسطها $\bar{x}$ ، وإذا أضفنا لكل قياس فيها مقداراً ثابتاً a ، فإنّ المتوسّط يصبح : $\bar{x}+a$. ولبيان ذلك : نرمز لـ x_i+a بالرمز y_i عندئذٍ :

$$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{\sum_{i=1}^{n}(x_i+a)}{n}=\frac{\sum_{i=1}^{n}x_i}{n}+\frac{\sum_{i=1}^{n}a}{n}=\bar{x}+a$$

وكذلك بالنسبة لعملية الطرح أي: إذا طرحنا من كل قياس مقداراً ثابتاً a ، فإنّ المتوسّط يصبح : $\bar{x}-a$ ، ولبرهان ذلك نرمز لـ x_i-a بـ y_i فنجد أنّ :

$$\bar{y}=\frac{\sum_{i=1}^{n}y_i}{n}=\frac{\sum_{i=1}^{n}(x_i-a)}{n}=\frac{\sum_{i=1}^{n}x_i}{n}-\frac{\sum_{i=1}^{n}a}{n}=\bar{x}-a$$

وينتج عما تقدّم أنه إذا جمعنا أو طرحنا من القياسات x_i عدداً ثابتاً a ، فإنّ المتوسّط الجديد $\bar{y}$ يصبح: $\bar{y}=\bar{x}\pm a$ ، وتسمى مثل هذه العمليّة ، "عملية انسحاب" على المحاور الإحداثية .

وهكذا بالنسبة لعملية الضرب بعدد ثابت b، حيث نلاحظ:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} bx_i}{n} = \frac{b\sum_{i=1}^{n} x_i}{n} = b\overline{x}$$

وكذلك بالنسبة للقسمة على عدد $(b \neq o)$ ، حيث ينتج أنَّ: $\overline{y} = \frac{\overline{x}}{b}$ ، وتسمى عملية قسمة أو ضرب كل قياس بعدد ثابت ، تسمى عملية تغيير في سلم القياس ٥ .

وبشكل عام: إذا خضعت القياسات $x_1, x_2, ..., x_n$ لتحويل وفق العلاقة الخطيَّة: $y_i = b\,x_i + a$ ، أي إذا خضعت لعمليتين معاً هما: عملية تغيير في سلم القياس (الضرب بالعدد b) ، وعملية انسحاب (إضافة العدد الثابت a) ، فالعلاقة تبقى نفسها تربط بين المتوسّطين (القديم والجديد) أي: $\overline{y} = b\,\overline{x} + a$ ، ولبيان ذلك نلاحظ :

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} (bx_i + a)}{n} = \frac{\sum_{i=1}^{n} b\,x_i + \sum_{i=1}^{n} a}{n} = \frac{b\sum_{i=1}^{n} x_i}{n} + \frac{\sum_{i=1}^{n} a}{n} = b\overline{x} + \frac{na}{n} = b\overline{x} + a$$

وتستخدم عادة عمليتا الانسحاب والتغيير في سلم القياس (وخاصة عندما تكون الأرقام كبيرة) لتسهيل عملية الحسابات.

مثال (5): نفترض أنَّ خمسة قياسات للمتغيِّر X كانت من الشكل:

$$X \;:\; 3 \;,\; 4 \;,\; 6 \;,\; 10 \;,\; 14$$

المطلوب:

1- أوجد المتوسّط الحسابي.

2- اطرح المقدار الثابت 5 ، ثم اقسم الناتج على 2 ، وأوجد المتوسّط الجديد.

3- كيف يمكن الحصول على $\overline{X}$ من $\overline{Y}$ ؟

الحل:

1) واضح أنَّ: $\overline{X} = \frac{3+4+6+10+14}{5} = \frac{37}{5} = 7.4$

2) بطرح 5 من كل قياس من القياسات السابقة نحصل على القيم :

-2 , -1 , 1 , 5 , 9

ثم بقسمة هذه القيم على 2 نحصل على القيم الجديدة:

$$-1 \; , \; -0.5 \; , \; 0.5 \; , \; 2.5 \; , \; 4.5$$

ويصبح المتوسّط الجديد : $\overline{Y} = \dfrac{-1+(-0.5)+0.5+2.5+4.5}{5} = \dfrac{6}{5} = 1.2$

3) الحصول على $\overline{X}$ من $\overline{Y}$ يكمن بعملية عكسيّة : حيث أولاً : نضرب بـ 2 (لأنّ الضرب عكس القسمة) ، فنحصل على القيم : 9 , 5 , 1 , -1 , -2

وثانياً : نجمع المقدار الثابت 5 (لأنّ الجمع عكس الطرح) ، فنحصل على القيم :

14 , 10 , 6 , 4 , 3 وهي ذات قيم X ، وبالتالي يكون : $\overline{X} = 7.4$

من الواضح أنّ المثال المذكور ، يكافئ التحويل الخطي : $Y_i = \dfrac{X_i - 5}{2}$

(تحقّق من ذلك) .

وهكذا ، يمكن صياغة أمثلة مشابهة بإجراء تحويلات خطيّة أخرى .

2 . مجموع انحرافات جملة من القياسات عن متوسطها الحسابي يساوي صفراً، أي:

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0$$

لبيان ذلك : نرمز بـ : $x_i - \overline{x} = d$ (انحراف القياس x_i عن متوسطه الحسابي)

عندئذٍ:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} = n\overline{x} - n\overline{x} = 0$$

مثال (6): تحقّق من الخاصة: $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$ في المثال السابق:

في المثال السابق ، كان لدينا خمسة قياسات ، وكان: $\overline{X} = 7.4$

وحسب الجدول التالي:

i	1	2	3	4	5	$\sum$
x_i	3	4	6	10	14	37

$x_i - \bar{x}$	-4.4 -3.4 -1.4 2.6 6.6	0

يكون قد تحقّق المطلوب .

3- يتأثر المتوسّط الحسابي بوجود قيم شاذة (متطرّفة) بين القياسات.

مثال (7): إنّ متوسط القيم: 15 ,68 ,75 ,67 ,70 هو :

$$\bar{x} = \frac{70+67+75+68+15}{5} = \frac{295}{5} = 59$$

نلاحظ أنّ المتوسّط منخفض قياساً بمعظم القيم أعلاه ، وباستبعاد القيمة 15 ، يُصبح المتوسّط:

$\bar{x} = \frac{70+67+75+68}{4} = \frac{280}{4} = 70$ لذلك تسمى القيمة 15 وأمثالها من القيم، بالقيم المتطرّفة أو الشاذة ، ووجودها في البيان الإحصائي يؤثر كثيراً في المتوسّط ويجعله يفقد الموقع المركزي الذي يُفترَض أن يشغله ، وسبب ذلك أنّه سيميل حتماً نحو هذه القيم المتطرّفة سواءً كانت صغيرة جداً أم كبيرة جداً ، لذلك فهو شديد الحساسية بمثل هذه القيم .

وبما أنّ موضوعنا الآن هو المتوسّط ، فليس ضاراً أن نذكر المتوسّط الهندسي.

- **المتوسّط الهندسي:**

تعريف: المتوسّط الهندسي لـ n من الأعداد الموجبة غير المبوّبة : $x_1, x_2, ..., x_n$ (ونرمز له بـ $\overline{G}$) ، يعطى بالشكل:

$$\overline{G} = \sqrt[n]{x_1\, x_2 x_3 \ldots\; x_n} \equiv (x_1 x_2 x_3 \ldots x_n)^{\frac{1}{n}}$$

ومنه ينتج : $\log \overline{G} = \frac{1}{n}\sum_{i=1}^{n} \log x_i$

أما إذا كانت البيانات مبوّبة وكانت: $x_1, x_2, ..., x_k$ هي مراكز فئاته الموجبة وكانت تكراراتها: $f_1, f_2, ..., f_k$ ، فإنّ: $n = \sum_{i=1}^{k} f_i$: $\overline{G} = \sqrt[n]{x_1^{f_1}\;\; x_2^{f_2}\;\; x_k^{f_k}}$

وعندئذٍ يكون: $\log \overline{G} = \frac{1}{n}\sum_{i=1}^{k} f_i\, \log x_i$

مع ملاحظة أنّ التعامل مع اللوغاريتمات لإيجاد المتوسّط الهندسي، في كثير من الأحيان، هو الأسهل، (علماً أنّ المتوسّط الهندسي يستعمل في القياسات التي تكون ناتجة عن النسبة بين أي مقدارين متتاليين).

(2-7-1) الوسيط – *Median*:

تعريف: الوسيط لـ n من القياسات (البيانات) المرتَّبة ترتيباً تصاعدياً أو تنازلياً هو: العدد الأوسط منها (القياس الواقع في الوسط) إذا كان عددها فردياً، وهو المتوسّط (أو المتوسّط الحسابي) للعددين الأوسطين إذا كان عددها زوجياً . أي هو العدد الذي يقسم القياسات المرتَّبة إلى نصفين متساويين.

وإذا فرضنا أنّ الأعداد (أو القياسات) المرتَّبة هي: $x_1, x_2, ..., x_n$ ، عندئذٍ يكون الوسيط هو العدد: $X_{(n+1)/2}$ إذا كان n فردياً ، أي العدد الذي مرتبته $\frac{n+1}{2}$ في الترتيب التصاعدي، وهو العدد: $\frac{1}{2}[X_{n/2} + X_{(n/2)+1}]$ إذا كان n زوجياً ، أي: متوسط القياسين اللذين رتبتاهما $\frac{n}{2}$ و $\frac{n}{2}+1$ في الترتيب التصاعدي.

مثال (1): ما هو وسيط القياسات: $4 , 7 , 8 , 5 , 12 , 3 , 9 , 15 , 20$

الحل:

نرتب هذه القياسات تصاعدياً فنجد: $3 , 4 , 5 ,7 , 8 , 9 , 12 , 15 , 20$

والوسيط هو القياس الذي رتبته $5 = \frac{9+1}{2} = \frac{n+1}{2}$.

ولكن القياس الخامس في القياسات المرتَّبة أعلاه هو 8 ، وبالتالي تكون قيمة الوسيط المطلوبة 8 .

مثال (2): ما هو وسيط الأعداد: $12 , 8 , 24 , 6 , 36 , 48 , 15 , 64$

الحل:

بترتيب هذه القياسات تصاعدياً نجد: $6 , 8 , 12 , 15 , 24 , 36 , 48 , 64$

وبما أنّ عدد القياسات $n=8$ (زوجيّاً) ، فإنّ الوسيط هو: متوسط القياسين اللذين رتبتاهما: $\frac{n}{2}=4$ و $\frac{n}{2}+1=5$ الأمر الذي يعني أنّ قيمة الوسيط المطلوبة هي:

$$\frac{15+24}{2}=\frac{39}{2}=19.5$$

مثال (3): أوجد وسيط الاختبارات في الحالتين:

1 . ممتاز ، جيّد، ضعيف، جيّد جداً، مقبول، جيّد، جيّد جداً، مقبول، جيّد

2 . مقبول، مقبول، ممتاز ، جيّد، جيّد جداً، مقبول، جيّد، مقبول، ضعيف، ممتاز .

الحل:

1) نرتّب التقديرات فنجد: ضعيف، مقبول، مقبول، جيّد، جيّد، جيّد، جيّد جداً، جيّد جداً، ممتاز .

وهنا نلاحظ أنّ عدد الحالات يبلغ 9 حالات (فردي)، والوسيط هو التقدير المقابل للقياس الذي رتبته: $\frac{n+1}{2}=\frac{9+1}{2}=5$ ، أي المقابل للقياس الخامس هو جيّد ، وبالتالي يكون الوسيط في هذا البيان هو جيّد .

2) نُرتّب التقديرات فنجد: ضعيف، مقبول، مقبول، مقبول، مقبول، جيّد، جيّد، جيّد جداً، ممتاز ، ممتاز .

وهنا نلاحظ أنّ عدد الحالات يبلغ 10 حالات (زوجي)، والوسيط في هذه الحالة يقع بين القياسين اللذين رتبتاهما ($\frac{n}{2}=\frac{10}{2}=5$ و $\frac{n}{2}+1=5+1=6$)، ونلاحظ أنّ القياس الخامس هو مقبول ، والقياس السادس هو جيّد ، لذلك يكون الوسيط في هذا البيان هو: بين المقبول والجيّد . (لاحظ أنّه يمكن حساب الوسيط من بيانات عدديّة أو ترتيبية).

. الوسيط في حالة البيانات المبوّبة:

تعريف: الفئة الوسيطيّة (الفئة التي ينتمي إليها الوسيط): هي أول فئة يزيد تكرارها المتجمّع الصاعد عن $\frac{n}{2}$، حيث n هو مجموع التكرارات . ولإيجاد الوسيط للتوزيع التكراري ذي الفئات نتبع الخطوات التالية:

1. نكتب جدول التكرار المتجمّع الصاعد.

2. نحسب رتبة الوسيط $\frac{n}{2}$ ، وذلك سواءً كان عدد القياسات فردياً أم زوجياً .

3. نحدد رتبة الفئة الوسيطيّة التي ينتمي إليها الوسيط .

4. نفرض أنّ:

- ω طول الفئة الوسيطيّة .
- f_m التكرار المقابل للفئة الوسيطيّة في جدول التكرار المتجمّع الصاعد .
- f_p التكرار المقابل للفئة السابقة للفئة الوسيطيّة .
- l الحد الأعلى الفعلي المقابل للفئة السابقة .

عندئذٍ يُعطى الوسيط بالعلاقة التالية : $M = l + \frac{\frac{n}{2} - f_p}{f_m - f_p} \times \omega$

مع ملاحظة أنّ كل قياس ينتمي إلى فئة، هو حتماً أصغر من أي قياس ينتمي إلى فئة لاحقة، وأكبر من أي قياس ينتمي إلى فئة سابقة.

مثال (4): احسب وسيط البيان المعطى بالجدول التالي:

التكرار المتجمّع الصاعد	أقل من
0	34.5
1	39.5
3	44.5
4	49.5
8	54.5
10	59.5
12	64.5
17	69.5 (الفئة السابقة للفئة الوسيطيّة)
21	74.5 (الفئة الوسيطيّة)
26	79.5
32	84.5
36	89.5
38	94.5
40	99.5

الحل: حسب الخطوات السابقة نلاحظ:

1) إنّ الجدول المعطى مرتّب ترتيباً تصاعدياً .

2) إنّ رتبة الوسيط هي: $\frac{n}{2}=\frac{40}{2}=20$

3) إنّ الفئـة الوسـيطيّة هـي أول فئـة يزيـد التكـرار المتجمّـع المقابـل لهـا علـى 20 (موضّحة في الجدول)

4) من الجدول، واضح أنّ:

$$\omega = 74.5 - 69.5 = 5 \quad , \quad f_m = 21$$
$$l = 69.5 \quad , \quad f_p = 17$$

بالتالي يكون الوسيط:

$$M = l + \frac{\frac{n}{2} - f_p}{f_m - f_p} \times \omega = 69.5 + \frac{20-17}{21-17} \times 5 = 73.5$$

ملاحظة هامة: يمكن الحصول على النتيجة السابقة باستخدام مفهوم التناسب الطردي، وملاحظة ما يلي:

زيادة القياس   زيادة التكرار

$$\left\{\begin{array}{ll} f_m - f_p & \omega \\ \frac{n}{2} - f_p & ? \end{array}\right\}$$

(تناسـب طـردي) ، أي أنّ زيـادة القيـاس المطلـوب لبلوغ الوسيط هو: $\frac{\frac{n}{2} - f_p}{f_m - f_p} \times \omega = \frac{20-17}{21-17} \times 5 = 3.75$

وبالتالي يكون الوسيط عندئذٍ عبارة عن:

l (الحد الأعلى الفعلي للفئة السابقة) مضافاً إليه زيادة القياس المطلوبة لبلوغ الوسيط،

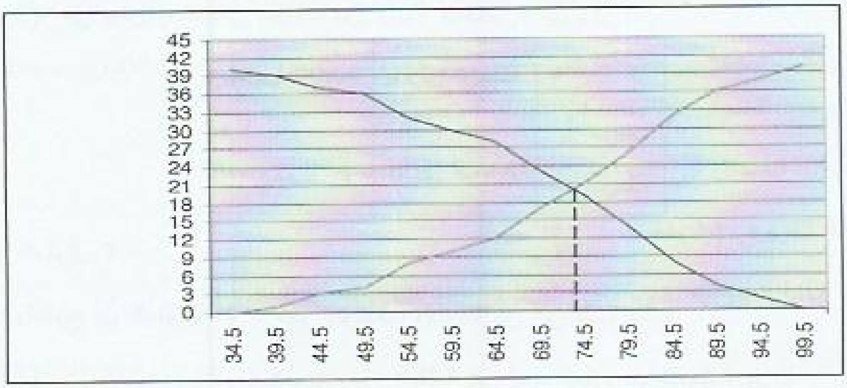
أي: $M = l + \frac{\frac{n}{2} - f_p}{f_m - f_p} \times \omega$ وبتطبيق قاعدة التناسب الطردي في مثالنا، يكون:

زيادة القياس   زيادة التكرار

$$21-17 \qquad\qquad 5$$
$$20-17 \qquad\qquad x$$

وهذا يعني أنَّ: $x = \frac{20-17}{21-17} \times 5 = 3.75$ وهي الزيادة المطلوبة لبلوغ الوسيط . وبالتالي قيمة الوسيط هي: $M = 69.5 + 3.75 = 73.25$ (نفس الجواب السابق) .

ويمكن أيضاً استنتاج قيمة الوسيط بيانيّاً ، وذلك برسم مضلعي التكرار المتجمّع الصاعد والمتجمّع النازل للمثال (10) ، فيكون الإحداثي السيني لنقطة تقاطع المضلّعين هو قيمة الوسيط . كما في الشكل التالي:



شكل (15)

ولحساب الوسيط مباشرة من الجدول المرتَّب (13) نجد:

بما أنَّ $n = 40$ (زوجي) ، فإنَّ الوسيط عبارة عن متوسط القياسين اللذين رتبتاهما $\frac{n}{2} = 20$] و $\frac{n}{2} + 1 = 21$] ي تكون قيمة الوسيط هي: $\frac{73+73}{2} = 73$

لاحظ الفرق بين النتيجتين، وبالطبع إنَّ الحساب المباشر من البيانات المبوّبة هو دائماً تقريبي، حيث تجاوزنا الدقّة التامة في حساب رتبة الوسيط واتخذناه على الدوام $\frac{n}{2}$ سواءً كان n زوجياً أم فرديّاً .

(3-7-1) المنوال – *Mode* :

تعريف: المنوال هو القياس الأكثر تكراراً في جملة من القياسات (القيمة التي تتكرّر أكثر من غيرها).

يمكن حساب المنوال في جميع أنواع البيانات الوصفيّة والعدديّة والترتيبيّة.

مثال (1): بفرض نتيجة اختبار 7 تلاميذ كانت بالشكل:

جيّد ، جيّد ، مقبول ، ممتاز ، جيّد جداً ، جيّد ، مقبول .

عندئذٍ : المنوال هو التقدير" جيّد" باعتباره القياس الأكثر تكراراً .

مثال (2): في بيان إحصائي مؤلّف من 100 شخص، تبيّن أنّ:

	يمارس الرياضة	لا يمارس الرياضة
موظف	30	12
غير موظف	18	40

ما هو المنوال ؟

الحل: واضح أنّ المنوال هو : "غير موظف ولا يمارس الرياضة".

مثال (3): إذا كانت درجات مجموعة من الطلبة في أحد المقررات العمليّة هي:

18 , 19 , 15 , 14 , 20 , 18 , 18 , 15 , 16 , 17

أوجد المنوال:

الحل:

واضح أنّ القيمة التي تكرّرت أكثر من غيرها هي 18، وبالتالي المنوال يساوي 18 .

. المنوال (في حالة التوزيعات التكراريّة):

نعتمد في التوزيعات التكراريّة ذات الفئات على المفاهيم والملاحظات التالية:

– الفئة المنواليّة: هي الفئة التي يقابلها أكبر تكرار .

- منوال البيان الإحصائي (أو المنوال التقريبي): هو مركز الفئة المنواليّة، وإنّ وجود عدة فئات منواليّة توافق وجود عدة منوالات تقريبيّة.

مع ملاحظة أنّ:

– التكرار الأعلى لا يعني المنوال، بل الصفة ذات التكرار الأعلى هي التي تقع بتواتر أكبر وبالتالي تكون هي المنوال.

– قد يوجد في البيان الإحصائي منوال واحد أو منوالان أو ثلاثة منوالات أو ...

ويسمى حينئذٍ: أحادي المنوال أو ثنائي المنوال أو ثلاثي المنوال أو ... على الترتيب.

- إذا كان لكل صفة أو صنف في البيان الإحصائي نفس العدد من التكرار ، عندها نقول: لا وجود للمنوال في مثل هذا البيان الإحصائي.

يمكن حساب المنوال في التوزيعات التكراريّة بعدة طرق ، منها:

1 . طريقة الفروق (طريقة بيرسون): حيث:

– نحدّد الفئة المنواليّة ، وبالتالي الفئة السابقة واللاحقة للفئة المتواليّة .

– نفرض أنّ :

w طول الفئة المنواليّة .

L_m الحد الأدنى للفئة المنواليّة .

f_2 تكرار الفئة المنواليّة .

f_1 تكرار الفئة السابقة للفئة المنواليّة .

f_3 تكرار الفئة اللاحقة للفئة المنواليّة .

ويعطى المنوال عندئذٍ (نرمز له بالرمز $\hat{X}$) بالعلاقة:

$$\hat{X} = L_m + \frac{f_2 - f_1}{2f_2 - f_1 - f_3} \times w$$

مثال (4): أوجد قيمة المنوال من البيانات التالية :

$\sum$	30-35	25-	20-	15-	10-	5-	**الفئات**

التكرار	60	90	125	105	95	75	550

الحل:

واضح أنَّ الفئة المنواليَّة هي : (20 - 15)، وبالتالي:

$$f_3 = 105 \ , \ f_1 = 90 \ , \ f_2 = 125 \ , \ L_m = 15 \ , \ w = 5$$

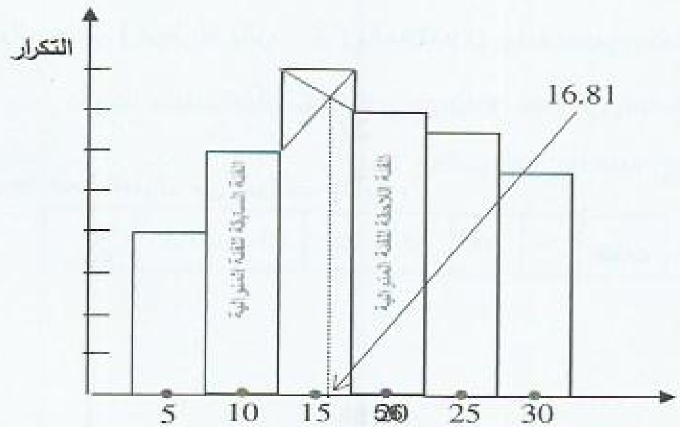
عندئذٍ : $\hat{X} = 15 + \dfrac{125 - 90}{250 - 90 - 105} \times 5 = 15 + \dfrac{175}{55} \approx 18.18$

2- الطريقة البياتيّة لحساب المنوال:

نرسم المدرَّج التكراري وبالضّبط: نأخذ ثلاثة أعمدة تمثِّل الفئات الثلاث (الفئة السابقة للفئة المنواليَّة ، الفئة المنواليَّة ، الفئة اللاحقة للفئة المنواليَّة)، ثم نصل الزاوية العليا اليمنى للفئة السابقة بالزاوية العليا اليمنى للفئة المنواليَّة، ونصل الزاوية العليا اليسرى للفئة اللاحقة بالزاوية العليا اليسرى للفئة المنواليَّة، ثم نسقط عموداً من نقطة التقاطع على المحور الأفقي، فتكون عندئذٍ قيمة المنوال هي الاحداثي الأفقي لنقطة التقاطع تلك .

مثال (5): احسب المنوال (بالطريقة البيانيَّة) في المثال (4):

الحل : برسم المدرَّج التكراري على ورق ملليمتري (علماً أنَّه يمكن أن نكتفي برسم المدرَّج الخاص بالفئات الثلاث: المنواليَّة، السابقة، اللاحقة) نحصل على الشكل التالي:



الفئات

إنّ الإحداثي الأفقي لنقطة التقاطع يمثل قيمة المنوال ، وتساوي تقريباً (الرسم على ورق ملليمتري): 18.18 وهي مطابقة للقيمة السابقة في المثال (4).

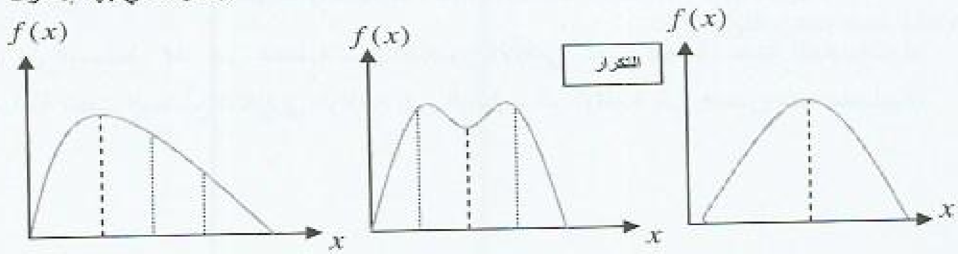
وبالواقع يلعب المنوال دوراً كبيراً في العلوم السلوكيّة باعتباره قابلاً للحساب في كل أنواع البيانات وفي الأعمال التجاريّة والتسويق والدعاية، حيث النوع الأكثر رواجاً في صناعة معيّنة يجذب المهتمين بها لصناعة المزيد منها، ونذكّر أنّه في البيانات الوصفيّة لا بديل عن المنوال.

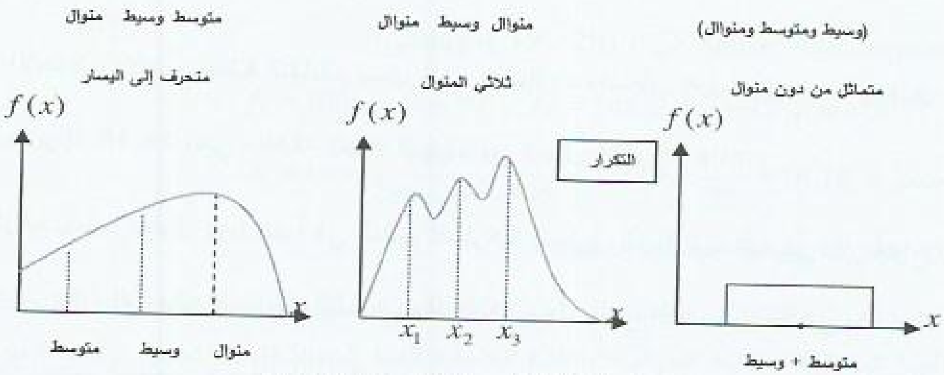
وبشكل عام: إنّ طبيعة الظاهرة وهدفها هما اللذان يحدّدان نوع المقياس المتوجب اعتماده في دراسة التوزيعات الإحصائيّة، علماً أنّ هذه التوزيعات كثيرة وأشكالها مختلفة (متناظرة، منحرفة) (أو ملتوية) إلى جهة، من دون منوال، أو ذات منوال واحد، أو متعددة المنوالات ... كما في الشكل التالي:

متناظر (وحيد المنوال) .

متماثل وثنائي المنوال .

منحرف (ملتو) نحو اليمين:
الوسيط هو الأقرب للمنوال ،
والمتوسّط في جهة الانحراف





شكل (16) أنواع مختلفة من التوزيعات

(1-8) مقاييس التشتت:

إنّ المقاييس السابقة لا تفي بالغرض لمعرفة كل التغيُّرات من مقياس إلى آخر ضمن البيانات الإحصائيَّة، علماً أنّ التغيُّر ظاهرة ملازمة لكل بيان إحصائي، وكذلك لا تفِ بالغرض لمعرفة مواقع القياسات عن مراكزها، وهناك معايير كميَّة خاصة لقياس شدة تبعثر القياسات في البيان الإحصائي، منها:

(1-8-1) المدى - *Rang*:

يعرّف المدى في البيانات الأوليّة بأنّه : الفرق بين أعلى قيمة وأصغر قيمة ، وفي البيانات المبوّبة بأنّه : الفرق بين الحد الأعلى للفئة العليا والحد الأدنى للفئة الدنيا .

واضح من تعريف المدى، أنّه لا يعتمد على كل القياسات، وبالتالي لن يعكس الطبيعة الحقيقيَّة للبيانات، لأنّه قد يكون لبيانين نفس المدى ، إلا أنّ الفارق كبير بين درجة تمركز كل منهما حول المتوسّط المشترك .

(1-8-2) الربيعات والعشيرات والمئينات - *Quartiles , Deciles, Percentiles* :

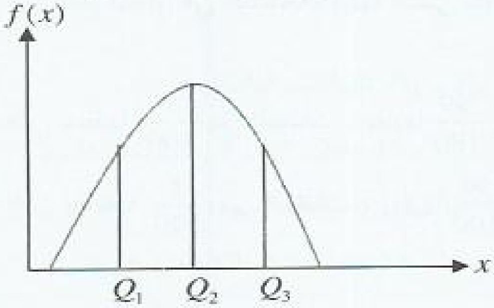
قلنا إنّ الوسيط M هو النقطة من المحور الأفقي التي تجعل عدد القياسات أو المساحة تحت المضلّع التكراري الواقعة إلى اليسار، ثم الواقعة إلى اليمين منه متساوية.

وانطلاقاً من هذا المفهوم، فإنّه يمكن وضع عدد من النقاط على المحور الأفقي، بحيث تقّسم هذه النقاط ، المساحة تحت المضلّع التكراري، إلى عدد من التقسيمات، وأهم هذه التقسيمات تحت المضلّع التكراري هي: تقسيمها إلى أربعة أقسام متساوية، تقسيمها إلى عشرة أقسام متساوية ، تقسيمها إلى مئة قسم متساوٍ.

فإذا فرضنا أنّ البيانات مرتبة ترتيباً تصاعدياً ، عندئذٍ:

أ . في حالة التقسيم إلى أربعة أقسام متساوية، فإنّها تسمى "الربيعات"، وعددها ثلاثة هي من اليسار إلى اليمين:

الربيع الأول (الأدنى) Q_1 ، الربيع الثاني Q_2 ، الربيع الثالث Q_3 انظر الشكل :



وعندئذٍ يمكن القول إنّ:

الربيع الأول (الأدنى) Q_1 : هو القيمة التي يسبقها ربع البيانات، ويليها ثلاثة أرباع البيانات.

الربيع الثاني Q_2 : هو الوسيط، وهو القيمة التي يسبقها نصف البيانات، ويليها النصف الآخر.

الربيع الثالث (الأعلى) Q_3 : وهو القيمة التي يسبقها $\frac{3}{4}$ البيانات، ويليها $\frac{1}{4}$ البيانات.

ب . وفي حالة التقسيم إلى عشرة أقسام متساوية في المساحة، فإنّها تسمى "عشيرات".

وبناءً عليه يمكن القول إنّ :

العشير الأول (الأدنى) D_1 : هو القيمة التي يسبقها $\frac{1}{10}$ من البيانات، ويليها $\frac{9}{10}$ من البيانات.

العشير الثاني D_2 : هو القيمة التي يسبقها $\frac{2}{10}$ من البيانات ، ويليها $\frac{8}{10}$ من البيانات وهكذا ...

العشير الخامس D_5 : هو الوسيط ، وهو القيمة التي يسبقها $\frac{1}{2}=\frac{5}{10}$ البيانات ، ويليها $\frac{5}{10}$ من البيانات .

ج . وأما في حالة التقسيم إلى مئة قسم متساوٍ في المساحة، فإنّها تسمى مئينات، وبالتالي يكون:

المئين الأول P_1 : هو القيمة التي يسبقها $\frac{1}{100}$ من البيانات ، ويليها $\frac{99}{100}$ من البيانات.

المئين الثاني P_2 : هو القيمة التي يسبقها $\frac{2}{100}$ من البيانات ، ويليها $\frac{98}{100}$ من البيانات وهكذا ...

المئين الخامس والعشرون P_{25} (الربيع الأول): هو القيمة التي يسبقها $25\% = \frac{25}{100}$ من البيانات ، ويليها 75% من البيانات .

المئين الخمسون P_{50} : هو الوسيط ، وهو القيمة التي يسبقها $\frac{1}{2}=\frac{50}{100}$ البيانات، ويليها $\frac{50}{100}$ من البيانات ، وبهذا يكون: $M = Q_2 = D_5 = P_{50}$ ، ويكون أكبر قياس في البيان الإحصائي هو المئون 100 (P_{100}) ، وأنّ أصغر قياس في البيان هو المئون الصفري P_0 ، وأن المدى هو : $P_{100} - P_0$ ، وبما أنّه يمكن تحويل أي ربيع أو عشير إلى مئين ، لذلك سنكتفي بحساب المئينات .

حساب المئينات في البيانات الأوليّة:

يُعرّف المئين k لبيان إحصائي بأنه القياس أو العدد الذي يقع على يساره $k\%$ من قياسات البيان الإحصائي .

وبشكل عام: نرمز للمئين k بالرمز P_k ، حيث k هو أي عدد صحيح بين الصفر والمئة . ولحساب المئينات في البيانات الأوليّة:

- نرتب البيانات ترتيباً تصاعديّاً ، وليكن عددها n

- لحساب المئين k، أي لحساب P_k : نوجد رتبة المئين k وهي: $n\frac{k}{100}$ ، فإذا كانت هذه القيمة كسراً، نقرّبه إلى أعلى، ويكون P_k مساوياً للعدد الذي ترتيبه القيمة المقرّبة، أما إذا كانت القيمة $n\frac{k}{100}$ عدداً صحيحاً وليكن b، فإنّ المئين k يساوي: $P_k = \frac{X_b + X_{b+1}}{2}$ حيث : X_b و X_{b+1} هما القياسان اللذان ترتيبهما هو b و $b+1$ على التوالي .

مثال (1): أوجد P_{60} ، P_{24} للبيانات التاليّة:

7 , 9 , 10 , 5 , 8 , 12 , 30 , 15 , 18 , 24 , 14

الحل : نرتب البيانات ترتيباً تصاعديّاً فنجد:

5 , 7 , 8 , 9 , 10 , 12 , 14 , 15 , 18 , 24 , 30

واضح أنّ: $n = 11$، بالتالي رتبة المئين هي: $n\frac{k}{100} = 11\frac{60}{100} = 6.6 \approx 7$ ، ومنه يكون: $P_{60} = 14$ (أي هو العدد الذي ترتيبه القيمة المقرّبة) .

ولحساب P_{24} : نجد: رتبة المئين هي: $11\frac{24}{100} \approx 3$ ، ومنه: $P_{24} = 8$ (لاحظ أنّ رتبة المئين كسريّة في هذا المثال) .

مثال (2): أوجد المئين العشرين والمئين السبعين في البيانات التالية:

1 , 3 , 5 , 8 , 12 , 6 , 4 , 20 , 15 , 40

الحل:

نرتب البيانات تصاعديّاً فنجد: 40 , 20 , 15 , 12 , 8 , 6 , 5 , 4 , 3 , 1

واضح أنّ: $n = 10$ ، وبالتالي رتبة المئين العشرين هي: $n\frac{k}{100} = 10\frac{20}{100} = 2$

وبما أنّ الناتج عدد صحيح ، فإنّ: $P_{20} = \frac{X_2 + X_3}{2} = \frac{3+4}{2} = 3.5$

وكذلك بالنسبة لحساب P_{70} حيث: $7 = \frac{70}{100} 10 = \frac{k}{100} n$، وبأخذ المتوسّط الحسابي للعددين اللذين رتبتاهما (7 و 8)، وهما (12 و 15) يكون: $P_{70} = \frac{12+15}{2} = 13.5$.

وهكذا بالنسبة لطلبات مشابهة ...

. المئينات للتوزيع التكراري:

تعريف: فئة المئين k هي أول فئة يزيد تكرارها المتجمّع عن $n\frac{k}{100}$ حيث n مجموع التكرارات .

ولإيجاد المئينات للتوزيعات التكراريّة ذات الفئات ، نتبع ما يلي :

a) نكتب جدول التكرار المتجمّع الصاعد .

b) نحدد رتبة المئين k ، وهي $n\frac{k}{100}$.

c) نحدد رتبة الفئة المئينيّة التي ينتمي إليها المئين .

d) نفرض أنّ :

- w طول الفئة المئينيّة .
- f_m التكرار المقابل للفئة المئينيّة في جدول التكرار المتجمّع الصاعد .
- f_p التكرار المقابل للفئة السابقة للفئة المئينيّة .
- l الحد الأعلى الفعلي المقابل للفئة السابقة .

وعندئذٍ يعطى المئين k بالعلاقة التالية:

$$P_k = l + \frac{\frac{nk}{100} - f_p}{f_m - f_p} \times w$$

ملاحظة: يمكن الحصول على العلاقة السابقة باستخدام مفهوم التناسب الطردي وبملاحظة أنّ:

زيادة القياس   زيادة التكرار

$$f_m - f_p \qquad\qquad w$$

$$\frac{nk}{100} - f_p \qquad\qquad ?$$

ومنه زيادة القياس المطلوب لبلوغ المئين k يساوي : $\dfrac{(\frac{nk}{100} - f_p)w}{f_m - f_p}$

ويكون المئون k عندئذٍ عبارة عن: l (الحد الأعلى للفئة السابقة) مضافاً إليه زيادة القياس المطلوبة لبلوغ المئين . أي: $P_k = l + \dfrac{\frac{nk}{100} - f_p}{f_m - f_p} \times w$

مثال(3): أوجد P_{25} للتوزيع التكراري التالي:

الفئات	10-15	16-21	22-27	28-33	**المجموع**
التكرار	7	13	6	8	**34**

الحل: حسب الخطوات السابقة:

أ . نشكل جدول التكرار المتجمّع الصاعد

أقل من	15.5	21.5	27.5	33.5
التكرار المتجمّع الصاعد	7	20	26	34

ب . رتبة المئين 25 هي: $n\dfrac{k}{100} = 34\dfrac{25}{100} = 8.5$

ج . الفئة المئينيّة هي أول فئة يزيد التكرار المتجمّع المقابل لها على 8.5 (وهي الفئة الثانية في الجدول).

د . من الجدول واضح أنّ: $w = 6$ ، $f_m = 20$ ، $l = 15.5$ ، $f_p = 7$ بالتالي:

$$P_{25} \equiv Q_1 = l + \frac{\frac{nk}{100} - f_p}{f_m - f_p} \times w = 15.5 + \frac{(8.5-7)\times 6}{20-7} = 15.5 + \frac{7.5}{13} = 16.5$$

وبإتباع الخطوات نفسها يمكن الحصول على P_{75} بالشكل:

رتبة المئين 75 هي: $n\frac{k}{100} = 34\frac{75}{100} = 25.5$

وفئة المئين 75 (الفئة المئينيَّة) هي أول فئة يزيد التكرار المتجمّع المقابل لها على 25.5 (وهي الفئة الثالثة في الجدول). بالتالي يكون :

$$P_{75} = 21.5 + \frac{(25.5-20)\times 6}{26-20} = 27$$

وهكذا ... بالنسبة لطلبات مشابهة.

ولكن، لمعرفة القيمة التي يسبقها نسبة مئويَّة معيَّنة من البيانات ويليها النسبة الباقية، أو لمعرفة النسبة المئويَّة للبيانات التي هي أقل من قيمة معيَّنة ، ندرس الرتب المئينيَّة:

. حساب الرتب المئينيَّة:

تعريف:

الرتبة المئينيَّة لقيمة ما (في بيان إحصائي) = $\frac{\text{التكرار المتجمّع لتلك القيمة} \times 100\%}{\text{مجموع التكرارات}}$

أي: هي النسبة المئويَّة لمجموع التكرارات للقيم التي هي أقل من تلك القيمة، بالنسبة إلى مجموع التكرارات الكلي.

لاحظ أنَّ المئين هو قيمة على المحور الأفقي، بينما الرتبة المئينيَّة هي نسبة مئويَّة (وهما وجهان لمسألة واحدة)، بمعنى: إذا عرفنا النسبة المئويَّة، فإنَّه يمكن معرفة القيمة على المحور الأفقي، بحيث تكون هذه النسبة المئويَّة مساوية لنسبة عدد البيانات التي هي أقل من هذه القيمة إلى عدد البيانات كلها، وبالعكس إذا عرفنا القيمة، فإنَّه يمكن إيجاد النسبة المئويَّة لتكرارات القيم التي هي أقل من هذه القيمة (أي يمكن معرفة الترتيب على شكل نسبة مئويَّة). ولحساب الرتب المئينيَّة لدينا حالتان:

a) **حساب الرتب المئينيَّة في البيانات الأوليَّة:** وفيها:

1) نرتب البيانات ترتيباً تصاعديّاً ، وليكن عددها n

2) نوجد عدد البيانات الأصغر من القيمة المفروضة ، وليكن m

3) تكون الرتبة المئينيّة عندئذٍ هي : $\frac{m}{n}\times 100\%$

مثال (1): أوجد الرتبة المئينيّة للقيمتين : 12 ، 16.5 في البيان الإحصائي التالي:

7 , 9 , 10 , 5 , 8 , 12 , 30 , 15 , 18 , 24

الحل: واضح أنَّ: $n = 10$ ، وبترتيب لقيم تصاعدياً نجد:

5 , 7 , 8 , 9 , 10 , 12 , 15 , 18 , 24 , 30

نلاحظ أنّه يوجد $(m = 5)$ قيماً أصغر من 12 ، بالتالي الرتبة المئينيّة للقيمة 12 هي:

$\frac{5}{10}\times 100\% = 50\%$ ، وبطريقة مشابهة نجد الرتبة المئينيّة للعدد أو للقيمة 16.5

حيث: $m = 7$ ، $n = 10$ بالتالي الرتبة المئينيّة للقيمة 16.5 تكون:

$\frac{7}{10}\times 100\% = 70\%$ ، أي أنّ القيمة 16.5 تقابل المئين السبعين . وهكذا ...

b) **حساب الرتب المئينيّة في التوزيعات التكراريّة:**

في هذه الحالة نوجد الرتبة المئينيّة باستعمال معادلة المئين k ، وهي:

$$P_k = l + \frac{\frac{nk}{100} - f_p}{f_m - f_p}\times w$$

بحيث:

1) تعيّن الفئة التي تقع فيها P_k ، ثم نعوض في معادلة المئين بحيث: نستبدل P_k بالقيمة المراد معرفة رتبتها المئويّة، ونستبدل القيمة $\frac{n\,k}{100}$ بالمجهول x .

2) نوجد قيمة x من الخطوة الأولى، ثم نبدل قيمتها في العلاقة: $k=\frac{x}{n}\times 100\,\%$

فنحصل على قيمة k (أي نحصل على الرتبة المئينيّة المطلوبة).

مثال (2): من أجل الجدول التالي:

أقل من	التكرار المتجمّع الصاعد
15.5	7
21.5	20
27.5	26
33.5	34

أوجد الرتب المئينيّة للقيمتين: 16 ، 28

الحل:

واضح أنّ 16 تقع في الفئة الثانية، وبالتعويض عن القيم (w, l, f_m, f_p) من الجدول المذكور يكون:

$$16=15.5+\frac{x-7}{20-7}\times 6 \Rightarrow x=27.5$$

إذاً: $$k=\frac{x}{n}\times 100\%=\frac{27.5}{34}\times 100\%=80.88\%$$

وكذلك، لحساب الرتبة المئينيّة للقيمة 28 نكتب:

$$28=15.5+\frac{x-7}{20-7}\times 6 \Rightarrow x=13.25$$

وبالتالي: $$k=\frac{x}{n}\times 100\%=\frac{13.25}{34}\times 100\%=38.97\,\%$$

ولو حسبنا الرتبة المئينيّة للقيمة 20 لكان الناتج: 72.2% وهكذا ... بالنسبة لطلبات مشابهة.

(1-8-3) متوسّط الانحراف *Mean Deviation*:

بفرض: $x_1, x_2, \dots, x_n$ جملة من القياسات متوسطها $\bar{x}$ ، ولنفرض أن: $d_i=x_i-\bar{x} \quad ; i=\overline{1, n}$ (انحراف كل قيمة من جملة القياسات عن المتوسّط). ولما

كانت الانحرافات قد تكون سالبة أو موجبة ، فإنّه بأخذ متوسط القيم المطلقة للانحرافات يكون لدينا التعريف التالي:

تعريف: يعرّف متوسّط الانحراف من أجل البيانات التالية: $x_1 , x_2 , ..., x_n$ بالشكل:

$$D = \frac{1}{n}\sum_{i=1}^{n}\left| x_i - \bar{x} \right| = \frac{1}{n}\sum_{i=1}^{n}\left| d_i \right|$$

وفي حالة التوزيعات التكراريّة يستخدم التعريف التالي:

$D = \frac{1}{n}\sum_{i=1}^{n}\left| x_i - \bar{x} \right| f_i = \frac{1}{n}\sum_{i=1}^{n}\left| d_i \right| f_i$ حيث f_i هو تكرار الفئة i ، الأمر الذي يعني أنّه لحساب متوسّط الانحراف لبيان إحصائي ، فإنّه يلزم معرفة كل مفردات ذلك البيان، وبالتالي يمكن القول: إنّ متوسّط الانحراف لبيان يعتمد على جميع مفرداته.

ولتلافي التعامل مع الانحرافات الموجبة والانحرافات السالبة ، فإنّه يتم أخذ مربّعات الانحرافات بدلاً من قيمها المطلقة وندرس ما يسمى بالتباين:

(4-8-1) التباين والانحراف المعياري – *Variance and Standard Deviation*:

يعرّف التباين في المجتمع بالشكل: $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

حيث: $\bar{x}$ هو المتوسّط الحسابي لـ n قياساً في المجتمع ونفرضها $x_1 , x_2 , ..., x_n$.

والتباين σ^2 موجب دوماً، لأنّه مجموع كميات موجبة ، علماً أنّه يساوي الصفر إذا كانت جميع القياسات متساوية أي: ($x_1 = x_2 = ... = x_n \Leftrightarrow \sigma^2 = 0$).

والانحراف المعياري هو الجذر الموجب للتباين أي: $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

وهو يعبّر عن تبعثر أو تباعد القيم عن متوسطها، ووحدة قياس التباين هي مربّع وحدة القياس المستخدمة في البيان الإحصائي، بينما وحدة قياس الانحراف المعياري هي بالضبط وحدة البيانات.

وبشكل عام: يمكن التعرّف على تباين مجتمع ما، من خلال معرفة تباين عيّنة (أو عدة عيّنات) من هذا المجتمع، بحيث نعتبر تباين هذه العيّنة هو الممثّل الفعلي لتباين

المجتمع الذي سُحبَت منه، تماماً كما اعتبرنا سابقاً أنّ مركز الفئة هو الممثّل الفعلي لكل قياساتها.

سنرمز لتباين العيّنة بالرمز S^2 (تمييزاً له عن σ^2) ويعرّف بالشكل:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{حيث:} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ .$$

ويكون الانحراف المعياري عندئذٍ للعيّنة بالشكل: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

علماً أنّه يمكن كتابة العلاقة السابقة S^2 بالشكل المختزل التالي:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

وهذه العلاقة أسهل وأسرع أثناء إجراء الحسابات.

مثال: لنأخذ جملة القياسات : 6 , 8 , 5 , 2 , 3

1. أحسب متوسط الانحراف ، التباين والانحراف المعياري .
2. أعد حساب التباين حسب الصيغة المختزلة.

الحل:

1) لدينا: $D = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$ ، لذلك نحسب المتوسّط :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{24}{5} = 4.8$$

وبالاستفادة من الجدول التالي:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$

3	−1.8	3.24
2	−2.8	7.84
5	0.2	0.04
8	3.2	10.24
6	1.2	1.44
24	0	22.8

نكتب مباشرة: $D=\frac{1}{5}(1.8+2.8+0.2+3.2+1.2)=\frac{9.2}{5}=1.84$

$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{22.8}{4}=5.7$ ومنه: $S=\sqrt{5.7}\approx 2.38$

2) لحساب الصيغة المختزلة $S^2=\frac{1}{n-1}[\sum_{i=1}^{n}x_i^2-n\bar{x}^2]$ نشكل الجدول التالي:

x_i^2	x_i	
9	3	
4	2	
25	5	
64	8	
36	6	
138	24	المجموع

بالتالي: $S^2=\frac{1}{4}(138-5(4.8)^2)=\frac{1}{4}(138-115.2)=5.7$

وفي التوزيعات التكراريّة ذات الفئات يعطى التباين بالعلاقة:

$$S^2=\frac{1}{n-1}[\sum_{i=1}^{n}x_i^2 f_i-n\bar{x}^2]$$

حيث: x_i ، f_i هما مركز وتكرار الفئة i على الترتيب .

وبالواقع، كلما نقص (زاد) التباين في العيّنة، كلما كانت العيّنة أكثر (أقل) تجانساً، وبالتالي تكون القياسات في هذه العيّنة أقل (أكثر) تغيُّراً. أي أنّ التباين يعبِّر بشكل عام عن خاصيّة التغيُّر في البيان الإحصائي سواء كان هذا البيان مأخوذاً مباشرة من المجتمع المدروس، أم مأخوذاً من العيّنة المسحوبة من ذات المجتمع.

وبما أن التباين وبالتالي الانحراف المعياري يعتمد على وحدة القياس المستخدمة في البيان الإحصائي، فهذا يجعله غير صالح للمقارنة بين عيّنتين من القياسات من حيث درجة التجانس في كلٍ منهما.

بمعنى آخر: التباين الأصغر يعني تجانس أكبر وبالعكس، وهذا محقّق (بشكل عام) فقط في ذات العيّنة أو المجتمع ذلك لأنّ طبيعة أو وحدة القياس هي ذاتها، ولكن هذا غير محقّق في مجتمعين أو عيّنتين من القياسات، ولتحقيق ذلك نستخدم مقياس آخر للتباين لا يعتمد على وحدة أو طبيعة القياس يسمى معامل التغيُّر:

(9-1) معامل التغيُّر (أحياناً يسمى معامل الاختلاف) – *Coefficient of Variation*:

تعريف: معامل التغيُّر (نرمز له بالرمز $c.v$) لجملة من القياسات متوسطها $\bar{x}$، وانحرافها المعياري S هو بالتعريف : $c.v = \frac{S}{\bar{x}} \times 100\ \%$

وهذا المعامل لا يعتمد على وحدة القياس المستخدمة ، وبالتالي يمكن استخدامه للمقارنة بين درجتي التجانس في عيّنتين من القياسات.

مثال: مجموعة (أ) من البيانات فيها: $\bar{x}_1 = 9\ ,\ S_1 = 2$

ومجموعة (ب) من البيانات فيها: $\bar{x}_2 = 12\ ,\ S_2 = 3$

أي من المجموعتين أكثر تغيُّراً ؟

الحل:

نحسب معامل التغيُّر (c . v) لكل من المجموعتين فيكون:

بالنسبة لـ (أ): $c.v = \frac{S_1}{\bar{x}_1} \times 100\ \% = \frac{2}{9} \times 100\,\% = 22.22\,\%$

وبالنسبة لـ (ب): $c.v = \frac{S_2}{\bar{x}_2} \times 100\ \% = \frac{3}{12} \times 100\,\% = 25\,\%$

الأمر الذي يعني أنّ المجموعة الثانية أكثر تغيُّراً من المجموعة الأولى .

وللمقارنة بين قياس من جملة أولى مع قياس من جملة ثانية ، ندرس :

(1-9-1) القيمة المعياريّة :

تُعرّف القيمة المعياريّة لأي قياس x من جملة قياسات (نرمز لها بالرمز Z) بالشكل:

$Z = \dfrac{x - \bar{x}}{S}$ وبشكل عام: إذا كانت جملة القياسات من الشكل: $x_1 , x_2 , \ldots, x_n$ فإنّ:

القيم $Z_i = \dfrac{x_i - \bar{x}}{S}$ تسمى قيماً معياريّة .

(بيّن أنّ: $S_z^{\ 2} = 1 \ , \ \bar{Z}_i = 0$ حيث: $S_z^{\ 2} \ , \bar{Z}_i$ هما : متوسط و تباين القيم المعياريّة).

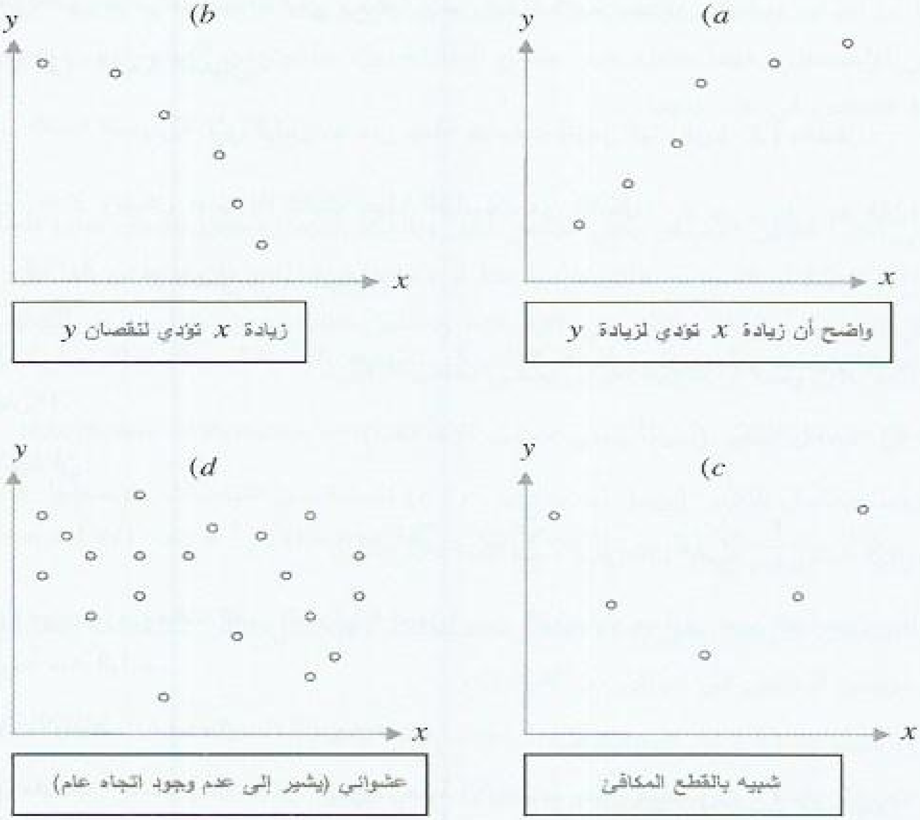
ثم لاحظ أنّ:

$$S_z^{\ 2} = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{z})^2 \Rightarrow \sum_{i=1}^{n}(Z_i - \bar{z})^2 = (n-1)\ S_z^{\ 2} = (n-1)\times 1 = n-1$$

أي أنّ مجموع مربّعات القيم المعياريّة لجملة من القياسات يساوي عدد القياسات n مطروحاً منه الواحد.

معامل الارتباط – *Correlation Coefficient* :

وهو مقياس عددي للتعبير عن درجة الارتباط بين متغيّرين . فإذا كانت $(x_i \ , \ y_i) \ ; \ i = \overline{1 \ , n}$ جملة من أزواج القياسات لظاهرتين أو لمتغيّرين x و y ، ومثّلنا هذه النقاط بيانيّاً، فإنّنا نحصل على ما يسمى: بمخطط الانتشار ، هذا المخطط يوحي أو يولّد نوعاً من الانطباع عن درجة الصلة أو الارتباط بين المتغيّرين x و y . والرسومات التالية توضّح بعضاً من أشكال هذه المخططات:



شكل (17)

وأهم مقاييس معامل الارتباط هو معامل بيرسون التالي:

(10-1) معامل بيرسون:

بفرض: $i=\overline{1\,,\,n}$; $(x_i\ ,\ y_i)$ جملة من أزواج القياسات للمتغيّرين x و y .

ولنفرض أنّ: $\overline{x}$ و S_x متوسط قيم المتغيّر x وانحرافها المعياري ، وأنّ: $\overline{y}$ و S_y

متوسط قيم المتغيّر y وانحرافها المعياري. وإذا رمزنا لقيمهما المعياريّة بالشكل:

$$Z_y = \frac{y_i - \overline{y}}{S_y} \;\; ; \; i = \overline{i\,,n} \qquad ، \; Z_x = \frac{x_i - \overline{x}}{S_x} \;\; ; \; i = \overline{i\,,n}$$

عندئذٍ يُعرّف معامل بيرسون للارتباط بين قيم x وقيم y (نرمز له بالرمز ρ) بالعلاقة:

$$\rho = \frac{1}{n-1} \sum_{i=1}^{n} Z_x \; Z_y$$

وفي الحالة الخاصة: إذا كان : $Z_x = Z_y$ فإنّ النقاط $(Z_n, Z_n^{\prime}),\dots,(Z_1, Z_1^{\prime})$ تقع على خط مستقيم هو منصف الربع الأول ، ويكون الارتباط في هذه الحلة ارتباطاً ايجابيّاً تامّاً ، وعندئذٍ يكون: $\rho = \frac{1}{n-1}\sum_{i=1}^{n} Z_x \; Z_y = \frac{1}{n-1}\sum_{i=1}^{n} Z_x^{\;2} = \frac{n-1}{n-1} = 1$

وإذا كان: $Z_x = -Z_y$ فإنّ: $R = -1$ والارتباط في هذه الحلة يكون ارتباطاً سلبياً تامّاً.

وبشكل عام فإنّ: $-1 \le R \le +1$ ، الأمر الذي يعني أنّه يوجد بين هاتين القيمتين عدد لا حصر له من درجات الارتباط بين الظاهرتين أو المتغيِّرين.

ملاحظة: يمكن صياغة معامل بيرسون بأشكال مختلفة منها:

$$(1) \; : \qquad \rho = \frac{1}{n-1} \sum_{i=1}^{n} Z_i \; Z_i^{\prime}$$

(وهي تتطلب جهداً بسبب رد القياسات إلى شكلها المعياري)

$$(2) \; : \;\; \rho = \frac{\sum_{i=1}^{n} X_i \; Y_i}{\sqrt{\sum_{i=1}^{n} X_i^{\;2} \; \sum_{i=1}^{n} Y_i^{\;2}}} \;\; ; \; X_i = x_i - \overline{x} \;\; , \; Y_i = y_i - \overline{y}$$

وهي أبسط في الحسابات من الصيغة الأولى.

$$(3) \; : \;\; \rho = \frac{n\sum_{i=1}^{n} x_i \; y_i - (\sum_{i=1}^{n} x_i)\;(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum_{i=1}^{n} x_i^{\;2} - (\sum_{i=1}^{n} x_i)^2]\;[n\sum_{i=1}^{n} y_i^{\;2} - (\sum_{i=1}^{n} y_i)^2]}}$$

وهي تعتمد على القياسات مباشرة.

(11-1) تمارين الفصل الأول:

1- من أجل الجدول التكراري التالي:

الفئة	10-15	16-21	22-27	28-33	34-39	40-45	46-51	Σ
التكرار	8	12	10	18	22	15	15	100

المطلوب:

1. احسب المنوال والمتوسّط الحسابي.

2. احسب التباين ومتوسط الانحراف.

2. متوسط 17 قياساً يساوي 9 . فما مجموع هذه القياسات ؟

3. ثبيّن المعلومات سجل لغياب العاملين في عدد من المراكز الوظيفيّة خلال فترة زمنيّة معيّنة:

عدد أيام الغياب x_i	0	1	2	3	4	5	6	7
التكرار f_i	5	15	23	22	17	10	6	3

احسب متوسط عدد أيام الغياب للعامل الواحد .

4- إذا كان الجدول التكراري لأعمار خمسين عاملاً في أحد القطاعات الخاصة بالشكل :

حدود الفئات	15-24	25-34	35-44	45-54	55-64
التكرار	8	11	25	4	2

احسب متوسط العمر للعمّال الخمسين في هذا القطاع .

5- بفرض الجدول التكراري التالي:

حدود الفئات	0.21-0.3	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70
التكرار	31	45	36	23	11

احسب المتوسّط والوسيط والمنوال .

6- من أجل القياسات التالية: 3 , 10 , 8 , 5 , 4 , 1 , 8 , 2 , 4

أ . احسب المدى والانحراف المعياري.

ب ـ تحقّق أنّك إذا أخذت متوسط مربّعات انحرافات كل قياس عن بقية القياسات، فإنّ الناتج يساوي ضعفي التباين .

7- ارسم مخطط الانتشار ، ثم احسب معامل بيرسون للارتباط من أجل أزواج القياسات التالية:

x	10	20	30	40	50	60	70
y	-4	-3	-2	0	3	6	7

8- من أجل الجدول التكراري التالي:

حدود الفئات	16-20	21-25	26-30	31-35	36-40	41-50	46-50	51-55	56-60
التكرار	12	27	101	180	120	50	25	30	55

المطلوب حساب:

1. المتوسّط والمنوال.
2. متوسط الانحراف والانحراف المعياري.
3. الربيع الأول والثالث.
4. المئين العاشر ثم التسعين.

9- بفرض البيان الاحصائي التالي:

24.5	23.6	24.1	25.0	22.9	24.7	23.8	25.2	24.9
24.1	23.7	24.4	24.7	23.9	25.1	24.6	23.3	24.3
24.8	22.8	24.6	23.9	24.1	24.4	24.5	25.7	23.6
24.0	24.7	23.1	23.9	24.2	24.7	24.9	25.0	24.8
24.5	23.4	24.6	25.3					

المطلوب:

1. لخّص هذا البيان الاحصائي في جدول توزيع تكراري مُتَّخِذاً الفئات:

$$22.5 - 22.9\ ,\ 23.0 - 23.4\ ,\ldots,\ 25.5 - 25.9$$

2. ارسم مدرّج التكرار، ومدرّج التكرار النسبي، ومضلّع التكرار.
3. اكتب جدول التكرار المتجمّع الصاعد.

4. ارسم مضلّع التكرار المتجمّع الصاعد.

5. ما عدد القياسات التي هي أقل من 23.75 ؟ أكثر من 23.45 ؟

6. ما القياس الذي يقل عنه خمسين بالمئة من القياسات ؟ خمس وعشرين بالمئة من القياسات ؟ وخمس وسبعين بالمئة من القياسات ؟

....

الفصل الثاني

مبادئ أساسيّة في جبر المجموعات

(2 . 1) المجموعات *Sets*:

تُستخدم المجموعة في الرياضيات لتدوين جملة من العناصر أو الأشياء ، ويفترض أن تكون عناصر هذه الجملة مختلفة مثنى مثنى وكذلك مختلفة عن كل العناصر التي لا تنتمي لهذه الجملة، وهي ضروريّة وذات أهميّة كبيرة لدراسة الاحتمالات، وبالحقيقة تعتبر المجموعة من المفاهيم الأكثر وضوحًا وبداهةً بحيث نستطيع القول إنّها لا تحتاج إلى تعريف.. فهي مفهوم عام لا يقتصر على علم دون سواه .. وبقدر بداهتها ووضوحها، بقدر ما نحن بحاجة لمفهومها.. وقد عرّفها بعضهم بأنّها (تجمُّع أي عدد من العناصر أو الأشياء). ولتمييز أو تسمية هذه المجموعات كان لابدّ من استخدام الرموز والإشارات التي ظهرت مع ظهور التفكير الإنساني، حيث لا يمكن للمرء أن يعيش إلا ضمن مجموعات... ولا نستطيع أن نستنتج شيئاً من آخر إلا حين يصبح الأخير إشارةً أو رمزاً للأول ... ولا علم من دون رموز ...

من أمثلة المجموعات:

مجموعة الأعداد الطبيعيّة.

مجموعة القيم المثلى لدالّة هدفيّه.

مجموع سكان قرية "المُعلّقة".

مجموعة الكتب في مكتبة ما.

مجموعة أولئك المسؤولين عمّا يقولون ...

مجموعة أولئك الذين يعتمدون في أثناء نقاشهم على الحجّة القويّة وليس على الحنجرة القويّة ...

يُرمَز عادةً للمجموعة بأحرف كبيرة $A, B, C ...$ ، ويُرمَز لعناصرها بأحرف صغيرة a,b,c... . فإذا كان a عنصرًا من المجموعة A نكتب: $a \in A$ ونقول: a ينتمي إلى A ، وبخلاف ذلك نكتب $a \notin A$ ، ونقول إنّ a لا تنتمي إلى A ، وتتحدّد المجموعة عادةً

إما بذكر كل عناصرها (طريقة العد)، وإما بذكر خاصة أو بعض الخواص التي تميّزها (طريقة القانون أو القائمة) فمثلاً:

مجموعة حلول المعادلة: $x^2-1=0$ هي: $\{1, -1\}$، والمجموعة التي لا تحوي أية عناصر نقول إنّها مجموعة خالية (*Empty Set*) ونرمز لها بالرمز Φ كالمجموعة:

$\Phi = \{x;$ سمكة لا تعيش في الماء $x\}$، أو المجموعة: $\{X+Y=1 \ \& \ 2X+2Y=1\}$

ومن أجل المجموعات:

$$A=\{1,2,3,4,5,8,10\} \quad , \quad B=\{a, b, c, d, e, f\}$$
$$D=\{y\,; P(y)\} \quad , \quad E=\{z\,; q(z) \ \& \ r(z)\}$$

$C=\{x\,;$ طالب جامعي $x\}$

واضح أنّ: A و B محدّدة العناصر تماماً، في حين ذُكِرت خاصة في المجموعة D ، بينما ذكرت خاصتان بالنسبة للمجموعة E (أي: المجموعة E هي تلك التي تحقّق خاصتين معاً). مما تقدّم ينتج أنّه عند دراسة أية مجموعة يجب التحقُّق من أنّها مُعرَّفة تماماً. بمعنى: يمكن معرفة انتماء أو عدم انتماء عنصرٍ ما لهذه المجموعة.

من المجموعات العددية التي سنتعامل معها: $\mathbb{N}$ ، $\mathbb{Z}$ ، $\mathbb{Q}$ حيث:

مجموعة الأعداد الطبيعية: $\mathbb{N} = \{n\} = \{1, 2, 3, \ldots, n\}$

مجموعة الأعداد الصحيحة: $\mathbb{Z} = \{n\} \cup \{0\} \cup \{-n\} = \{0 \pm 1, \pm 2, \ldots \pm n\}$

مجموعة الأعداد العادية: $\mathbb{Z} = \{\frac{a}{b}\,; a \ \& \ b \in \mathbb{Q}, b \neq 0\}$

. المجموعة الجزئية (*Subset*):

نقول إنّ المجموعة A هي مجموعة جزئية من المجموعة B، ونرمز لها بالرمز $A \subset B$، إذا كان كل عنصر في A منتمياً للمجموعة B، ونعبّر رياضياً بالشكل: $\forall x \in A \Rightarrow x \in B$ وتتساوى المجموعتان A و B (*Equality*) إذا تحقّق:

$$A=B \Leftrightarrow A \subset B \ \& \ B \subset A$$

وأما **المجموعة الكلية** (*Universal Set*) فهي تدرس جميع المجموعات قيد البحث باعتبارها مجموعات جزئية منها.

إذا كان أي عنصر من مجموعة A ينتمي لمجموعة أخرى B قلنا إنّ A مجموعة جزئية في B ونكتب : $A \subset B$ ، وبخلاف ذلك نكتب: $A \not\subset B$ كالمجموعة:

$[a,b] \not\subset]a,b[$.

واضح أنّ : $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ حيث: $\mathbb{R}$ مجموعة الأعداد الحقيقية و: $\mathbb{C}$ مجموعة الأعداد العقدية . ونقول عن المجموعة A إنّها قابلة للعد إذا أمكن مقابلة عناصرها بمجموعة جزئية من مجموعة الأعداد الطبيعة $\mathbb{N}$. كما نقول إنّها مجموعة منتهية إذا احتوت على عدد منته من العناصر فقط ، ومن المعلوم أنّه يمكن دوماً تمثيل المجموعات بمخططات (*Veen*) .

(2-2) العمليات على المجموعات: (*Set Operation*)

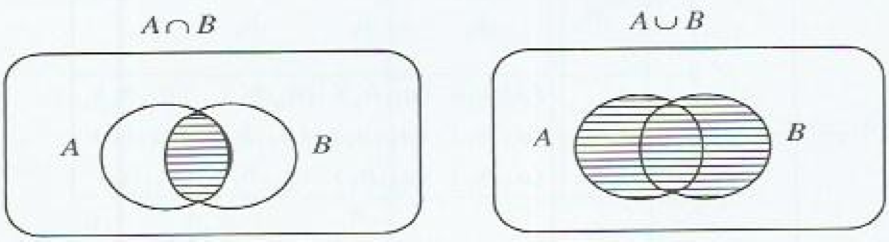
يُرمَز لاجتماع (*Union*) المجموعتين A و B بالرمز $\cup$ ونكتب ذلك كما يلي:

$$A \cup B = \{a \,;\, a \in A \ or \ a \in B\}$$

ويُرمَز لتقاطع (*Intersection*) المجموعتين A و B بالرمز $\cap$ ونكتبه كما يلي:

$$A \cap B = \{a \,;\, a \in A \ \& \ a \in B\}$$

ويمكن تمثيل هاتين العمليتين بوساطة مخططات فن (*veen*) كما يلي:

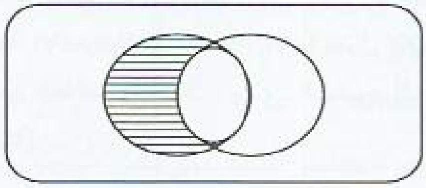


وإذا كان : $A \cap B = \Phi$ قلنا عن A و B إنّهما مجموعتان منفصلتان (*Disjoint*).

يمكن تعميم التقاطع والاجتماع على المجموعات المنتهية وغير المنتهية، أي من أجل المجموعات $A_1 , A_2 , A_3 , \ldots , A_k$ فان كلاً من اجتماعها $\bigcup_{k=1}^{\infty} A_k$ وتقاطعها $\bigcap_{k=1}^{\infty} A_k$ يُعرّف بطريقة مشابهة.

نرمز لفرق المجموعتين A و B ، (A , B *Difference Of*) بالرمز:

$$A - B = \{ x \, ; \, x \in A \ \& \ x \notin B \}$$



ونرمز للجداء الديكارتي للمجموعتين A و B بالرمز:

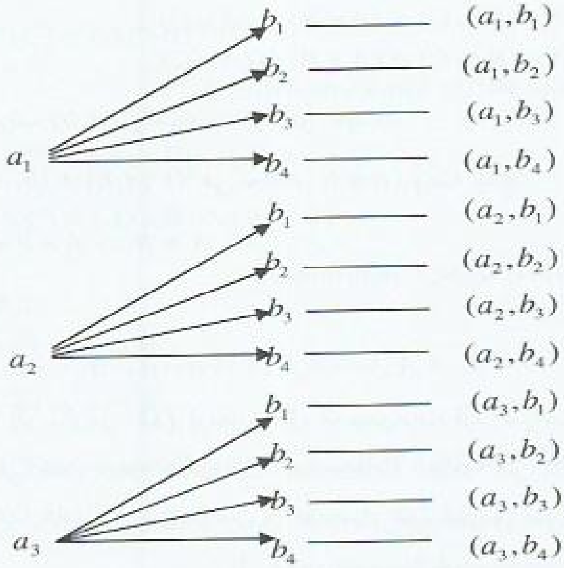
$$A \times B = \{ (a , b) ; a \in A \& b \in B \}$$

ومن أجل: $B = \{b_1 , b_2 , b_3 , b_4\}$ ، $A = \{a_1 , a_2 , a_3\}$ ، فإنّه يمكن تمثيل الجداء $A \times B$ بأشكالٍ مختلفةٍ منها:

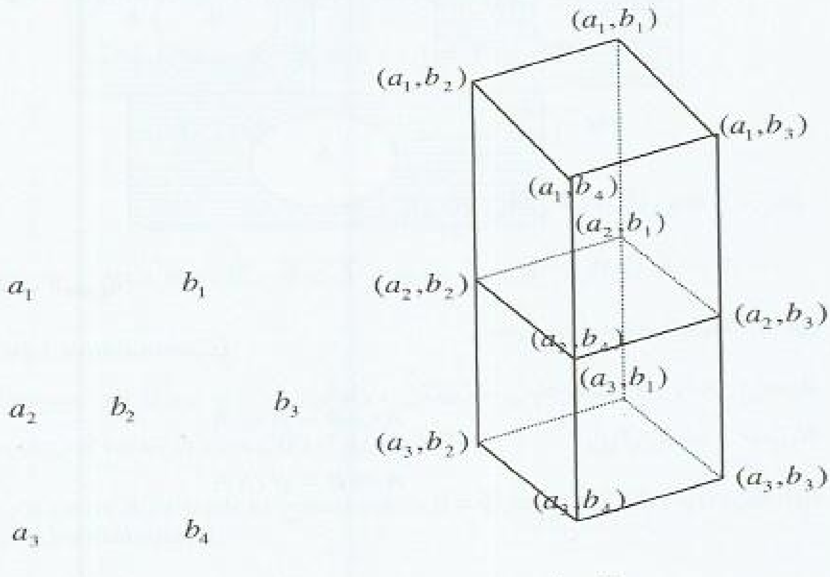
- **الشكل الجدولي:**

B / A	b_1	b_2	b_3	b_4
a_1	(a_1 , b_1)	(a_1 , b_2)	(a_1 , b_3)	(a_1 , b_4)
a_2	(a_2 , b_1)	(a_2 , b_2)	(a_2 , b_3)	(a_2 , b_4)
a_3	(a_3 , b_1)	(a_3 , b_2)	(a_3 , b_3)	(a_3 , b_4)

▪ المخطط الشجري :



▪ مخطط هاس: (يستخدم في الشبكات):



A B $A \times B$

أضف إلى أنه من أجل المجموعات : A, B, C يكون :

$$A\times(B\cup C)=(A\times B)\cup(A\times C)$$
$$A\times(B\cap C)=(A\times B)\cap(A\times C)$$

علماً أنه يجب التمييز بين الزوج المرتب (a,b) والمجموعة $\{a,b\}$ حيث:

$\{a,b\}=\{b,a\}$ دوماً، في حين: $(a,b)\neq(b,a)$ إلا في حالة: $a=b$ ، ومنه ينتج:

$$A\neq B\Rightarrow A\times B\neq B\times A \ .$$

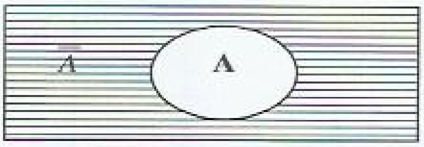
ويمكن تعميم الجداء الديكارتي بالشكل:

$$A_1\times A_2\times A_3\times\dots\times A_n=\{(a_1,a_2,a_3,\dots,a_n): a_i\in A_i\ ,\ i=\overline{1,n}\}$$

سنرمز للمجموعة الشاملة بالرمز E (أو بالرمز Ω) (وتضم كل المجموعات قيد الدراسة التي هي مجموعات جزئية من E) في حين نرمز لأسرة كل المجموعات الجزئية في E بالرمز : $P(E)=\{A; A\subseteq E\}$ ، فإذا كانت E مؤلفة من n عنصراً، فإنّ قدرتها هي: $P(E)=2^n$ ، وإذا كانت E مجموعة شاملة، عندئذٍ: $\forall A: A\cup E=E\ ,\ A\cap E=A$

ويعرّف **متمّم مجموعة** A ($Complement$) بالنسبة للمجموعة الكلية E بالشكل:

$$\overline{A}=\{x\ ;\ x\in E\ \ \&\ \ x\notin A\}$$



- **خواص العمليات الجبرية:**

خاصية التبديل: ($Commutative\ Law$)

$$A\cup B=B\cup A$$
$$A\cap B=B\cap A$$

خاصية التجميع: ($Associative\ Law$)

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

خاصية التوزيع: (*Distributive Law*)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

قانونا دي مورغان: *De Morgans*

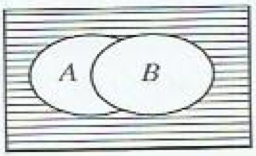
$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$

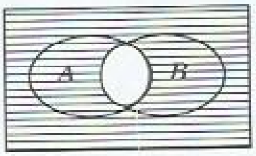
ويمكن تعميم هذه القوانين إلى n من المجموعات بدلاً من مجموعتين.

ومما سبق يتَّضح أنَّ: $A \cup \overline{A} = E \ , \ A \cap \overline{A} = \Phi \ , \ \overline{\Phi} = E \ , \ \overline{E} = \Phi$

$$\forall A , B \in E \ : \ \overline{(A \cup B)} = \overline{A} \cap \overline{B} \ , \overline{(A \cap B)} = \overline{A} \cup \overline{B}$$



$\overline{(A \cup B)}$



$\overline{(A \cap B)}$

وكلٍ من الشروط الآتية يكافئ $A \subset B$: :

$$A \cap B = A \ , \ A \cup B = B \ , \ \overline{B} \subset \overline{A} \ , \ A \cap \overline{B} = \Phi \ , \ B \cup \overline{A} = E$$

ويمكن تمثيلها بمخططات أيضاً.

بفرض: A و B مجموعتان منتهيتان ، وليكن: $n(A)$ ، $n(B)$ عدد عناصرهما على الترتيب . عندئذٍ يكون: $n(A \cup B) = n(A) + n(B) - n(A \cap B)$

فإذا كان: $n(A \cap B) = 0 \Leftarrow A \cap B = \Phi$ ، يصبح : $n(A \cup B) = n(A) + n(B)$

ومن أجل: $A\cap B\neq\Phi$ فإنّ عدد العناصر المشتركة يكون: $n(A\cap B)$، وبالتالي اجتماع هاتين المجموعتين يتشكل من عناصر A مضافاً إليه كل عناصر B التي لا تنتمي لـ A، وعدد هذه العناصر هو: $n(B)-n(A\cap B)$، ومنه يكون:

$$n(A\cup B)=n(A)+n(B)-n(A\cap B)$$

تطبيق:

مجموعة تضم 40 طالباً، منهم: 30 طالباً يجيد السباحة، 27 طالباً يجيد لعبة الشطرنج، 5 طلاب لا يجيدون السباحة ولا يجيدون الشطرنج. ما عدد الطلاب الذين يجيدون السباحة والشطرنج معاً ؟

الحل:

بفرض: A عدد الذين يجيدون السباحة أي : $n(A) = 30$

B عدد الذين يجيدون الشطرنج أي : $n(B) = 27$

C عدد الذين لا يجيدون السباحة ولا الشطرنج

عندئذٍ يكون المطلوب هو حساب $n(A\cap B)$ حيث:

$$\begin{aligned} n(A\cap B) &= n(A)+n(B)-n(A\cup B) \\ &= 30+27-35=22 \end{aligned}$$

بالواقع هناك بعض الخواص المشتركة بين (اجتماع وتقاطع المجموعات) من جهة، وبين (جمع وجداء) الأعداد من جهة أخرى.

فمثلاً:

في الأعداد	في المجموعات
$a+b=b+a$	$A\cup B=B\cup A$
$a.b=b.a$	$(A.B)=(B.A)$
$(a+b)+c=a+(b+c)$	$(A\cup B)\cup C=A\cup(B\cup C)$
$(a.b)c=a.(b.c)$	$(A\cap B)\cap C=A\cap(B\cap C)$
$a.(b+c)=a.b+a.c$	$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$

وبالحقيقة لا يمكن تعميم ذلك.

حيث في المجموعات لدينا:

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$
$$(A\cup A)=A$$
$$(A\cap A)=A$$

بينما نجد في الأعداد أنّ:

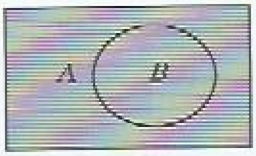
$$a+(b.c)\neq(a+b).(a+c)$$

$$a+a\neq a$$

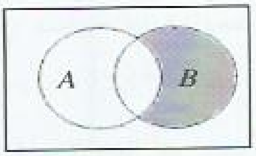
$$a\times a\neq a$$

أمثلة:

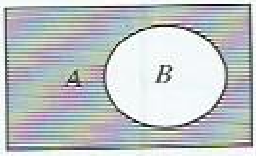
1ــ إذا كــان : $A\not\subset B$ فإنــه يمكــن تمثيـل المجموعــات: $\overline{A}\cap\overline{B}$, $B-A$, $\overline{A}\cup B$ بالشكل:



$\overline{A}\cup B$

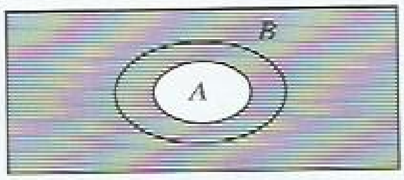


$B-A$

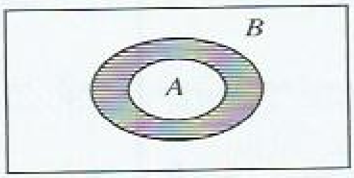


$\overline{A}\cap\overline{B}$

وإذا كان : $A\subset B$ يكون:

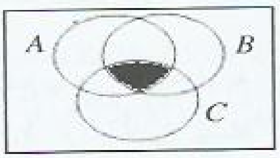


$\overline{A}\cup B$

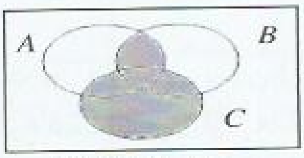


$B-A$

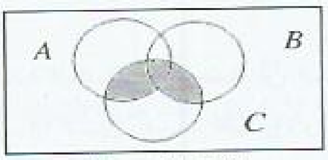
ومن أجل المجموعات: A,B,C يكون:



$A\cap B\cap C$



$(A\cap B)\cup C$



$(A\cup B)\cap C$

2 . من أجل:

$$A = \{x ; x = 0, 1, 2\}$$
$$B = \{x ; x = 2, 3, 4\}$$
$$C = [-1, 1]$$
$$D =]0, 3[$$

يكون:

$$A \cup B = \{x ; x = 0, 1, 2, 3, 4\}$$
$$A \cap B = \{x ; x = 2\}$$
$$C \cup D = [-1, 3[$$
$$C \cap D =]0, 1]$$

(2−3) . طرق حسابيّة . طرق العد: *Counting Methods*

هناك كثير من المسائل تعتمد مبدأ العد ، وبما أنّه لا توجد طريقة عامة تُطبَّق على كل المسائل، لذلك كان لا بدّ من إيجاد طرق حسابية تفيد في بعض المسائل الخاصة ومنها:

(2 . 3 . 1) . المبدأ الأساسي في العد:

بفرض A حدثاً يمكن أن يقع بـ m طريقة، B حدثا آخر يمكن أن يقع بـ n طريقة.

بالتالي:

a . إذا فرضنا أنّ: A و B لا يمكن أن يقعا معاً، فإن وقوع A أو B يتم بـ $(m+n)$

طريقة .(وهو ما يسمى: **قاعدة الجمع** − *Addition Rule*).

b . إذا فرضنا أنّ: A و B يمكن أن يقعا معاً عندئذٍ: وقوع A و B يتم بـ $(m.n)$

طريقة .(وهو ما يسمى: **قاعدة الضرب** − *Rule of multiplication*).

وبالحالة العامة: إذا فرضنا أنّ الحدث $A_i \; , \; i=\overline{1 \, , \, n}$ يمكن أن يقع بـ : $k_i \, , i=\overline{1,n}$ طريقة، فإنّ عدد طرق وقوع الحدث: A_1 أو A_2 أو A_3 أو... أو A_n يتم بـ: $k_1 + k_2 + k_3 + ... + k_n$ طريقة. **(حالة عامة لقاعدة الجمع)**.

بينما يتم وقوع الحدث: A_1 و A_2 و A_3 و...و A_n بـ : $k_1 . k_2 . k_3 ... k_n$ طريقة. (حالة عامة لقاعدة الضرب).

فمثلاً:

إذا كان بالإمكان السفر من الموقع A إلى الموقع B بـ5 طرق ، ومن الموقع B إلى الموقع C بـ3 طرق ، فإنّه بالإمكان السفر من الموقع A إلى الموقع C بعدد من الطرق مقداره : $5\times3=15$ طريقة.

مثال (1):

على طالب أن يسجل مقرّر واحد في أحد الفصول، وكان لديه الاختيارات التالية: 3 مقرّرات جبر ، مقرّرين في التحليل ، 3 مقرّرات في اللغة.

فما عدد الاختيارات التي تُمكّن الطالب من تسجيل مقرّر واحد فقط ؟

الحل:

عدد الاختيارات أو الطرق هو : 8 =3+2+3 طرق.

مثال (2):

يراد تركيب شبكة خطوط هاتف في مدينة ما، فإذا علمت أنّ رقم الهاتف مكون من 6 أرقام أولها من اليسار 4 أو 3 . فكم خط هاتفي يمكن تركيبه في المدينة ؟

الحل:

يمكن ملء الخانة اليسرى لرقم الهاتف بطريقتين (حيث يوجد الرقم 4 أو 3) ، وأما باقي الخانات الخمس فيمكن ملء كل منها بـ 10 طرق هي الأرقام:

(0,1,2,3,4,5,6,7,8,9) وبالتالي حسب قاعدة الضرب يمكن تركيب عدد من خطوط الهاتف مقداره :000 200 خط.

(2-3-2) المتبادلات: *Permutation*

وهي طرق ترتيب كل أو بعض عناصر مجموعة ما، أي: إنّ تبديل عدد من الأشياء هو تنظيم هذه الأشياء في ترتيب محدّد.

إنّ المبدأ السابق a – هو بالحقيقة حالة خاصة من الجداء الديكارتي للمجموعات ويستخدم بالواقع لحساب عدد الترتيبات (أو المتبادلات كما تسمى عادةً) لمجموعة من الأشياء.

فمثلاً:

إذا أردنا ترتيب الأحرف a , b , c (يمكن اعتبارها أحداثاً) أي: ما عدد تباديل حروف المجموعة $\{a,b,c\}$؟

عدد التباديل هو عدد الطرق التي يتم فيها ترتيب الحروف a,b,c (واضح هنا أنّ عدد الأماكن يساوي عدد العناصر)، بالتالي:

يمكن ملء الموضع الأول بثلاث طرق، وملء الموضع الثاني بطريقتين، وملء الموضع الثالث بطريقه واحدة ، وبالتالي حسب المبدأ الأساسي في العد يكون عدد طرق ترتيب الأحداث أو الأحرف السابقة هو: (1 . 2 . 3 = 6 طرق).

وهي: a b c , c b a , c a b , b a c , b c a , a c b وكل منها يسمى ترتيبة أو متبادلة، في حين نحتاج لكتابة (720 =6.5.4.3.2.1) متبادلة بصدد متبادلات ستة أحرف وهذا أمرٌ شاقٌ بالطبع. وبشكل عام ، يمكن ترتيب n من العناصر بعدد من الطرق مقداره: $n(n-1)(n-2)...3.2.1 \equiv n!$. أما إذا أردنا أن نرتب بعض عناصر مجموعة: فهذا يعني أنّ عدد الأماكن أقل من عدد العناصر ، وهذا ما يسمى بعدد تباديل n عنصراً متميزاً بحيث نأخذ منها m عنصراً في كل مرة.

وبالحالة العامة: عدد متبادلات n من الأشياء المختلفة مأخوذاً منها m شيئاً $(m \le n)$ في وقت واحد ، يُرمّز ويعطى بالشكل:

$$p_m^n = n(n-1)(n-2)(n-3)...(n-m+1)$$

وبضرب الطرف الأول بـ : $1 \equiv \dfrac{(n-m)!}{(n-m)!}$ يكون : $p_m^n = \dfrac{n!}{(n-m)!}$

ومن أجل: $m = n$ يكون : $p_n^n = \dfrac{n!}{0!} = n! = n\,(n-1)(n-2)\,...3\,.\,2\,.1$

(حيث: $0!=1$).

فمثلا:

يمكن ترتيب 3 كتب مأخوذة من بين 8 كتب مختلفة بعدد من الطرق مقداره:

$$P_3^8 = \frac{8!}{5!} = 8\times7\times6 = 356$$

ويُعرّف الرمز: $\binom{n}{n_1, n_2, \dots, n_k}$ بأنه : $\frac{n!}{n_1! n_2! \dots n_k!}$ ، ونرمّزه بالشكل :

$$P^n_{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \dots n_k!}$$

حيث: $n_1, n_2, \dots, n_k$ أعداد صحيحة غير سالبة ، و: $n = n_1 + n_2 + \dots + n_k$ ، وهو يمثل عدد متبادلات n من العناصر المختلفة التي تحوي:

n_1 عنصراً متماثلاً من نوع أول ، n_2 عنصراً متماثلاً من نوع ثان ، ...،

n_k عنصراً متماثلاً من النوع k.

أمثلة:

1 – عدد طرق تقسيم 7 دمى بين 3 أطفال على أن يأخذ: الطفل الأول 3، والثاني 2، والثالث 2 هو : $p^7_{3,2,1} = \frac{7!}{3!.2!.2!}$

2 – إنّ عدد الكلمات المؤلفة من 3 أحرف مختلفة والتي يمكن تشكيلها من الأبجدية العربية هو : $p_3^{28} = 28\times27\times26 = 19656$

3 – إنّ عدد طرق اختيار ثلاث لجان من عشرة طلاب ، بحيث تضم اللجنة الأولى 5 طلاب، والثانية 3 طلاب ، والثالثة طالبين هو : $252 = \frac{10!}{5!\,3!\,2!} = P^{10}_{5,3,2}$.

4 – ما عدد المتبادلات التي يمكن تشكيلها من حروف اسمك ؟

5 – إذا اشتركت ثمانية جياد في سباق لإحراز المراتب الأربعة الأولى ، فما هو عدد الطرق الممكنة للفوز في هذا السباق؟

واضح أنه للترتيب أهمية، وبالتالي المطلوب يكون:

$$P_4^8 = \frac{8!}{(8-4)!} = \frac{8!}{4!} = 8\times7\times6\times5 = 1680 .$$

مما سبق يتضح أنّ التبديلة a b c تختلف عن التبديلة b c a ، وإذا أهملنا الترتيب تكون بصدد المتوافقات:

(4-3-2) المتوافقات (التوافيق) *Combination*:

وهي الطرق التي نختار بها عدداً معيَّناً من عناصر مجموعة معيَّنة دون النظر إلى الترتيب، وبمعنى آخر: هي عدد الطرق التي نقسم بها مجموعة من العناصر إلى مجموعتين تحتوي إحداهما عددا معيناً من العناصر وتحتوي الأخرى بقية العناصر دون النظر إلى ترتيب تلك العناصر في أي من المجموعتين، وهي بصورة عامة، تُمكّننا من الإجابة على السؤال المهم التالي: بكم طريقة يمكن أن نختار m شيئاً من أصل n شيئاً ؟

إنّ الرمز p_m^n يمثل عدد الاختيارات الممكنة وعدد المتبادلات المختلفة في كل اختيار، ويسمّى كل اختيار (متوافقة).

مثلاً: a b d ،a b c هما متوافقتان مختلفتان من ثلاثة أحرف، في حين: c a b ،a b c هما متبادلتان مختلفتان لنفس المتوافقة.

نرمز لعدد توافيق n من الأشياء مأخوذاً منها m شيئاً (m ≤ n) في الوقت نفسه بالرمز C_m^n.

نظرية:

$$C_m^n = \frac{n!}{m!(n-m)!}$$

البرهان:

كل متوافقة من m شيئاً تعطي: $P_m^m = m!$ متبادلة، وبالتالي: C_m^n متوافقة سيقابلها $C_m^n . P_m^m$ متبادلة . وهذا العدد هو بالحقيقة عدد المتبادلات لـ n شيئاً مأخوذاً منها m شيئاً في الوقت نفسه، أي: هذا العدد عبارة عن: P_m^n . بمعنى: $P_m^n = P_m^m . C_m^n$

ومنه:

$$C_m^n = \frac{P_m^n}{P_m^m} = \frac{n!}{m!(n-m)!}$$

ملاحظة:

يمكن إعطاء الرمز C_m^n تفسيراً مختلفاً ، وهو بالواقع يمثل عدد الطرق التي يمكن أن نقسّم فيها n شيئاً إلى مجموعتين: إحداهما تحوي m شيئاً، والأخرى تحوي الـ ($n-m$) شيئاً الباقية.

لنفرض الآن أننا نريد تقسيم الـ n شيئاً إلى ثلاث مجموعات تحوي على الترتيب: n_1, n_2, n_3 من الأشياء بحيث: $n = n_1 + n_2 + n_3$. في هذه الحالة ، نقسم أولاً الـ n شيئاً إلى مجموعتين: تحوي المجموعة الأولى: n_1 شيئاً، وتحوي المجموعة الثانية: $n_2 + n_3$ شيئاً الباقية، وهذا يمكن أن يحدث بـ: $C_{n_1}^n$ طريقة. وبعد ذلك نقسم المجموعة الثانية إلى مجموعتين: تحوي الأولى: n_2 شيئاً، وتحوي الثانية: n_3 شيئاً، وهذا التقسيم يتم بـ: $C_{n_2}^{n_2+n_3}$ طريقة، وحسب المبدأ الأساسي للعد، يكون عدد طرق التقسيمين معاً هو : $C_{n_1}^n \, . \, C_{n_2}^{n_2+n_3} = \dfrac{n!}{n_1! \, . \, n_2! \, . \, n_3!}$

وبالحالة العامة: يمكن تقسيم n شيئاً إلى k مجموعة تحوي على الترتيب $n_1 , n_2 , n_3 , ..., n_k$ شيئاً (حيث: $n = n_1 + n_2 + n_3 + ... + n_k$) بعدد من الطرق ممثلاً بالرمز :

$$C_{n_1\,,\,n_2\,,\,n_3\,,\,....\,,\,n_k}^n = \frac{n!}{n_1! \, . \, n_2! \, . \, n_3! ... n_k!}$$

وهذا العدد يمثل أيضاً عدد المتبادلات المختلفة لـ n من الأشياء التي تحوي:

n_1 شيئاً متشابهاً من نوع أول.

n_2 شيئاً متشابهاً من نوع ثان.

n_3 شيئاً متشابهاً من نوع ثالث. و... n_k شيئاً متشابهاً من النوع رقم k .

ويمكن أن نكتب:

$$P_{n_1\,,\,n_2\,,\,n_3\,,\,....\,,\,n_k}^n \equiv C_{n_1\,,\,n_2\,,\,n_3\,,\,....\,,\,n_k}^n = \frac{n!}{n_1! \, . \, n_2! \, . \, n_3! ... n_k!}$$

حيث: $n = n_1 + n_2 + n_3 + ... + n_k$

أمثلة:

1- بكم طريقة يمكن توزيع هدايا رمزية على 8 طلاب ثبتَ أنّهم جديرون بالاحترام، وذلك في الحالتين التاليتين:

1) إذا كان يوجد 4 هدايا ، ونريد إعطاء كل طالبين هدية واحدة ؟

2) إذا كان يوجد 8 هدايا ، ونريد إعطاء كل طالب هدية واحدة ؟

الحل:

في الحالة الأولى: $C_2^8.C_2^6.C_2^4.C_2^2 = \frac{8!}{6!.2!}\frac{6!}{4!.2!}\frac{4!}{2!.2!}\frac{2!}{0!.2!} = 2520$ طريقة.

في الحالة الثانية: $P_8 = 8! = 40320$ طريقة.

2 - إنّ عدد المصافحات عندما يلتقي 3 أشخاص بأن واحد هو: $C_2^3 = \frac{3!}{1!3!} = 3$

3 - لديك 8 زملاء ، وفي إحدى المناسبات دعت الضرورة لدعوة 5 زملاء منهم فقط لحضورها.

المطلوب:

1 . بكم طريقة يمكن أن تتم هذه الدعوة إذا كان اثنان من الزملاء لا يمكن أن يحضرا إلا بصحبة صديقين آخرين ؟ (للضرورة أحكام).

2 . بكم طريقة يمكن أن تتم هذه الدعوة إذا كان اثنان من الزملاء لا يمكن أن يحضرا سوية لظروف ما ؟

الحل:

1) لدينا الحالات التالية:

الحالة الأولى: يمكن أن يتم الاختيار بدعوة الزميلين مع صديقيهما، ثم اختيار زميل خامس ويتم ذلك بـ: $C_2^2\, C_1^6 = 6$ طرق.

الحالة الثانية: يمكن أن يتم الاختيار بدعوة واحد من الزميلين المرتبطين بصديقيهما، ثم اختيار 3 زملاء آخرين، ويتم ذلك بـ: $C_1^2\, C_3^6 = 40$ طريقة.

الحالة الثالثة: يمكن أن يتم الاختيار دون دعوة أي من الزميلين المرتبطين بصديقيهما، ويتم ذلك بـ: $C_5^6 = 6$ طرق . وبالتالي فإنّ عدد طرق الاختيار هو مجموع الإمكانات السابقة، أي : $6 + 40 + 6 = 52$ طريقة.

2) الجواب هو : $C_5^6 + C_1^2 . C_4^6 = 36$ طريقة.

4 – بكم طريقه يمكن اختيار لجنة مؤلفة من 3 طلاب و 4 طالبات من بين 5 طلاب و 6 طالبات ؟

الحل: عدد الطرق الممكنة لاختيار 3 طلاب من بين 5 طلاب هو: C_3^5

عدد الطرق الممكنة لاختيار 4 طالبات من بين 6 طالبات هو : C_4^6

بالتالي وحسب المبدأ الأساسي في العد ، فإنّ عدد الطرق الممكنة لتشكيل اللجنة هو:

$$C_3^5 . C_4^6 = \frac{5!}{2!3!} \frac{6!}{2!4!} = 150$$

5 – على طالب أن يجيب في الامتحان على 8 أسئلة من 10 أسئلة.

والمطلوب:

1 . كم اختياراً لديه؟

2 . كم اختياراً لديه إذا كان عليه أن يجيب عن الـ 3 أسئلة الأولى؟

3 . كم اختياراً لديه إذا كان عليه أن يجيب عن أربعة على الأقل من الأسئلة الخمسة الأولى؟

الحل:

1) عدد الاختيارات هو : $C_8^{10} = 45$ اختيار .

2) إذا أجاب الطالب عن الأسئلة الثلاثة الأولى، فبإمكانه بعد ذلك اختيار 5 أسئلة من الـ7 أسئلة المتبقية، أي لديه في هذه الحالة : $C_5^7 = 21$ طريقة.

3) إذا أجاب الطالب عن الأسئلة الخمسة الأولى، فبإمكانه بعد ذلك اختيار 3 أسئلة من الـ 5 أسئلة المتبقية بعدد من الطرق مقداره : $C_3^5 = 10$ طرق.

ومن ناحية أخرى، إذا أجاب الطالب عن 4 أسئلة من الأسئلة الخمسة الأولى، فبإمكانه اختيار هذه الأسئلة الأربعة بـ : $C_4^5 = 5$ طرق ، وبإمكانه اختيار الأسئلة الأربعة الأخرى بـ: $C_4^5 = 5$ طرق ، وبالتالي يمكنه اختيار الأسئلة الثمانية بـ (25 = 5 . 5) طريقة.
ويكون عندئذٍ لدى الطالب (35 = 10 + 25) اختيار.

6- برهن أنّ: $C_k^n + C_{k-1}^n = C_k^{n+1}$

الحل:

$$C_k^n + C_{k-1}^n = \frac{n!}{(n-k)!\,k!} + \frac{n!}{(n-k+1)!\,(k-1)!}$$
$$= \frac{n!}{(n-k)!\,k!} + \frac{k\,.\,n!}{(n+1-k)\,(n-k)!\,k!}$$
$$= \frac{n!}{(n-k)!\,k!}[1 + \frac{k}{(n+1-k)}]$$
$$= \frac{n!}{(n-k)!\,k!}[\frac{n+1}{n+1-k}] = \frac{(n+1)!}{(n+1-k)!\,k!} = C_k^{n+1}$$

(2-3-4) صيغة نيوتن:

نعلم أنّ:

$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

أو بشكل آخر:

$$(a+b)^2 = C_0^2\, a^2 + C_1^2 ab + C_2^2\, b^2$$
$$(a+b)^3 = C_0^3 a^3 + C_1^3 a^2 b + C_2^3\, ab^2 + C_3^3\, b^3$$

والسؤال المطروح: هل يمكن وباستخدام التوافيق أيضاً التعبير عن $(a+b)^n$ (حيث: a, b اختياريه و n طبيعي) بصيغة جبرية مشابهة ؟

فمثلاً: بضرب العلاقة الأخيرة بـ: $(a+b)$ نحصل على:

$$(a+b)^4 = C_0^3\, a^4 + (C_1^3 + C_0^3)a^3 b + (C_2^3 + C_1^3)a^2 b^2 + (C_3^3 + C_2^3)ab^3 + C_3^3 b^4$$

ولكن:

$$C_0^3=C_0^4\ ,\ C_1^3+C_0^3=C_1^4\ ,\ C_2^3+C_1^3=C_2^4\ ,$$
$$C_3^3+C_2^3=C_3^4\ ,\ C_3^3=C_4^4$$

بالتالي:
$$(a+b)^4=C_0^4\,a^4\ +\ C_1^4\,a^3b+C_2^4\,a^2b^2+C_3^4\,ab^3+C_4^4\,b^4$$

وبالحالة العامة لدينا النظرية التالية:

نظرية:
$$(a+b)^n=C_0^n\,a^n\ +\ C_1^n\,a^{n-1}b+\ldots+C_k^n\,a^{n-k}b^k+\ldots+C_n^n\,b^n$$

أو بالشكل: $(a+b)^n=\sum_{k=0}^{n}C_k^n\,a^{n-k}b^k$ ، (حيث: $a\ ,\ b$ اختياريه و n طبيعي)

البرهان:

من أجل: $n=1\Rightarrow(a+b)^1=C_0^1a+C_1^1b=a+b\ ;\ C_0^1=C_1^1=1$

لنفرض أنّ الصيغة صحيحة من أجل : $n=m$ أي: $(a+b)^m=\sum_{k=0}^{m}C_k^m\ a^{m-k}b^k$

عندئذٍ:
$$(a+b)^{m+1}=(a+b)\sum_{k=0}^{m}C_k^m\,a^{m-k}b^k$$
$$=\sum_{k=0}^{m}C_k^m a^{m+1-k}b^k+\sum_{k=0}^{m}C_k^m a^{m-k}b^{k+1}$$
$$=C_0^m a^{m+1}+\ \sum_{k=1}^{m}\ C_k^m\,a^{m+1-k}b^k+\sum_{k=1}^{m}C_{k-1}^m\ a^{m+1-k}b^k+C_m^m b^{m+1}$$
$$=C_0^m a^{m+1}+\sum_{k=1}^{m}(C_k^m+C_{k-1}^m)a^{m+1-k}b^k+C_m^m b^{m+1}$$

وبما أنّ: $C_0^m=C_0^{m+1}\ ,\ C_k^m+C_{k-1}^m=C_k^{m+1}\ ,\ C_m^m=C_{m+1}^{m+1}$

يكون: $(a+b)^{m+1}=\sum_{k=0}^{m+1}C_k^{m+1}\ a^{m+1-k}b^k$ ، وبما أنّ الصيغة المعطاة صحيحة من أجل: $n=m$ فهي صحيحة من أجل: $n=m+1$ ، ولطالما هي صحيحة من أجل: $n=1$ ، فإنّه حسب مبدأ الاستقراء الرياضي تكون الصيغة المعطاة صحيحة من أجل أي عدد طبيعي n. أي: $(a+b)^n=\sum_{k=0}^{n}C_k^n\ a^{n-k}b^k$ وهو المطلوب.

أمثلة:

1 – أوجد حاصل الجمع : $C_0^n + C_1^n + C_2^n + ... + C_n^n$

لدينا: $(a+b)^n = \sum_{k=0}^{n} C_k^n \, a^{n-k} b^k$ ، وبالتالي من أجل : $a=b=1$ يكون:

$$(1+1)^n = \sum_{k=0}^{n} C_k^n = C_0^n + C_1^n + C_2^n + ... + C_n^n = 2^n$$

(عدد المجموعات الجزئية التي يمكن تشكيلها من مجموعة مكونة من n عنصر).

2 – باستخدام صيغة نيوتن يمكن كتابة العدد المركب $(2+i)^6$ بالصيغة الجبرية التالية: $(2+i)^6 = \sum_{k=0}^{6} C_k^6 \, 2^{6-k} \, i^k = -117 + 44i$

3 – اثبت أنَّ: $C_k^n = C_{n-k}^n$:

$$C_{n-k}^n = \frac{n!}{(n-k)!\,(n-(n-k))!} = \frac{n!}{(n-k)!\,k!} = C_k^n$$

4 – إذا كان: $P_r^n = 12$, $C_r^n = 6$ أوجد n و r.

الحل:

$$\left.\begin{aligned} P_r^n &= \frac{n!}{(n-r)!} = 12 \\ C_r^n &= \frac{n!}{(n-r)!\,r!} = 6 \end{aligned}\right\} \Rightarrow \frac{12}{r!} = 6 \Rightarrow r = 2$$

وبالتبديل يكون : $\frac{n!}{(n-r)!} = 12 \Rightarrow n^2 - n - 12 = 0$

والحل هو : ($n=-3$ مرفوض ، $n=4$ مقبول).

(4-2) تمارين الفصل الثاني:

1 . بفرض المجموعات:

$A =$ {مجموعة قواسم العدد 12 }

$B =$ { مجموعة جذور المعادلة $x^2 - 6x + 5 = 0$ }

$C = \{x ; 3 \leq x \leq 12$; x فردي }

أوجد: $A \cup B$, $B \cap C$, $(A \cup B) \cap C$, $A \cap B \cap C$

2 – هات أمثلة تحقّق العلاقات:

$A \cup B = R \ \& \ A \cap B = \Phi$ (حيث R مجموعة الأعداد الحقيقيّة)

$A \cup B = A \ \& \ A \cap B = B$

3 – تَجمُّع سكني يضم 1400 شخص ، منهم : 1250 يمارس العمل x ، 952 يمارس العمل y ، 60 لا يمارس أي من العملين x أو y .

ما عدد أولئك الذين يمارسون العملين x و y ؟

4 – إذا كان:

$A = [0,3]\ ,\ B =]1,5[\ \ ,\ C =]-2,0]$

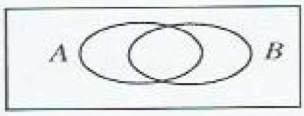
$F = \{1,2\}\ ,\ E = \{a,b\}\ ,\ G = \{c,d\}$

أوجد:

1– $A \cup B,\ A \cap B\ ,\ A \cap C\ ,\ A \cap B \cap C$

2– تحقّق أنّ: $F \times (E \cup G) = (F \times E) \cup (F \times G)$

5 – لأجل A و B الممثلتين بالشكل التالي:



مثّل المجموعات التالية: $A \cup \overline{B}\ ,\ \overline{A \cup B}\ ,\ \overline{A} \cup \overline{B}\ ,\ \overline{A} \cap B$

$\overline{A} \cup B\ ,\ \overline{A \cap B}\ ,\ (\overline{A} \cap B) \cup (A \cap \overline{B})$

6 – أوجد n إذا كان:

$$\frac{P_{n+5}^{n+5}}{P_{n-k}^{n-k}} = 240\ P_{k+3}^{n+3}$$

7 – أوجد مجموعة حلول المتراجحة: $C_{x-1}^{10} > 2.C_x^{10}$

8 – أوجد n في الحالات الثلاث التالية:

$$C_{n+1}^{n+4} - C_n^{n+3} = 15(n+2)$$

$$\frac{1}{C_n^4} = \frac{1}{C_n^5} + \frac{1}{C_n^6}$$

$$5\,C_3^n = C_4^{n+2}$$

9 – أوجد مجموعة قيم الدالتين التاليتين:

$$f_1(x) = P_{x-3}^{7-x}$$
$$f_2(x) = C_{2x-8}^{x+1}$$

....

الفصل الثالث

بعض المفاهيم والقواعد الأساسيَّة في الاحتمالات

(3-1) تعاريف ومبادئ أوليَّة:

التجربة العشوائيَّة: *Random Experiment:*

هي كل عمليَّة تعطي مشاهدة أو قياساً لظاهرة ما بحيث لا نستطيع التنبؤ بنتيجتها تماماً وإن كنّا نعلم مسبقاً جميع نتائجها الممكنة، والتجربة العشوائيَّة بشكل عام هي كل عمل فيزيائي أو كيميائي نقوم به للرد على سؤال محدَّد ، فاختيار ورقة من مجموعة أوراق اللعب، ورمي حجر النرد أو قطعة نقود، واختبار فعاليَّة دواء ما، وتسجيل عدد السيارات المارَّة في اتجاه معيَّن خلال فترة زمنيَّة محدَّدة ... كلّها أمثلة لتجارب عشوائيَّة.

فضاء العيِّنة: *Sample Space:*

لكل تجربة نتائج، وفضاء العيِّنة هو مجموعة كل النتائج الممكنة Ω لتجربة عشوائيَّة ما (يرمز له أحياناً S)، وقد يكون منتهياً أو غير منته ولكنه معدود.

فإذا كانت التجربة هي إلقاء حجر النرد يكون: $\Omega = \{1, 2, 3, 4, 5, 6\}$، مع ملاحظة أنَّه قد يكون للتجربة العشوائيَّة أكثر من فضاء عيِّنة وهذا يتوقف على الطريقة التي ننظر بها إلى النتائج.

مثلا: إذا انصبَّ اهتمامنا على عدد الصور التي تظهر على الوجه العلوي أثناء رمي قطعتي نقود، فإنَّ فضاء العيِّنة يكون: $\Omega = \{0,1,2\}$، وبشكل عام: لا تتحدَّد التجربة العشوائيَّة تماما إلا بتحديد فضاء العيِّنة المرتبط بها، وقد يكون منتهياً أي محدوداً، أو غير منتهي بمعنى غير محدود، وهو (من حيث عدد العناصر) ثلاثة أنواع:

- فضاء محدود (النرد): *Finite Sample Space*
- فضاء لانهائي معدود: *Countable infinity's* S يحتوي على عدد لا نهائي من العناصر لكنه قابل للعد (يوجد تقابل بينه وبين Z^+ أو أي مجموعة جزئية منها)
- فضاء لا نهائي: *Infinity's* S . ونقول عن فضاء إنَّه فضاء منفصل S – *Discrete* إذا كان محدوداً أو لا نهائياً معدوداً ويمكن عندئذٍ ربط عناصره $(1-1)$ مع Z^+.

مثلا:

. ارم قطعة نقود حتى تظهر الصورة H .

في هذه الحالة يكون: $\Omega = \{H, TH, TTH, TTTH, ...\}$ ، حيث يدل الرمز H على ظهور الصورة والرمز T على ظهور الكتابة، ويمكن عندئذ ربط الحدث H مع 1 والحدث TH مع 2 والحدث TTH مع 3 والحدث $TTTH$ مع 4 وهكذا ...

أي: قد نقوم بإعادة التجربة مرة واحدة أو مرتين أو ثلاث أو أربع أو ...

. عند اختيار نقطة داخل دائرة يكون: $\Omega = \{(x, y) : x^2 + y^2 \leq a^2\}$ ، وهو فضاء غير محدود.

وأحيانا ينصبُّ الاهتمام على بعض عناصر الفضاء دون بعضها الآخر، وهذا يدفعنا لدراسة ما يسمّى بالحادث أو الحدث.

. **الحدث** *Event*:

نسمي بالحدث أو الحادث: كل مجموعة جزئية من Ω .

فإذا كان A هو حدث ظهور الأعداد الفردية في تجربة رمي حجر النرد فإنّ:

$A = \{1, 3, 5\}$ ، وإذا احتوت هذه المجموعة الجزئية عنصراً واحداً فقط قلنا إنّ الحدث بسيط أو أولي، كحادث ظهور الرقم 1 مثلاً عند إلقاء حجر النرد، وإذا احتوت على نتيجتين أو أكثر (عنصرين أو أكثر) قلنا إنّه حدث أو حادث مركّب.

مثلاً:

. عند رمي قطعة نقود مرة واحدة ، فإنّ فضاء العيّنة يكون: $\Omega = \{H, T\}$ والحدث يكون بسيطا.

. عند رمي زهرة النرد مع قطعة النقود فإنّ:

$$\Omega = \{(T,1), (T,2), ((T,3), (T,4), (T,5), (T,6)(H,1), (H,2), (H,3), (H,4), (H,5), (H,6)\}$$

وكل حدث فيه يكون مركّباً.

الحدث المستحيل Φ : *Impossible Event* :

نؤكّد استحالة ظهور الرقم 7 في تجربة إلقاء حجر النرد ، وبالتالي يمكن القول: إنّ الحدث المستحيل يمثل الحالة التي لا يكون فيها للتجربة نتائج ، ونؤكّد بنفس الوقت أنّه سيظهر رقم أقل من 7 في التجربة وهذا ما يسمّى:

الحدث الأكيد (أو المؤكّد): *Sure Event*:

وهو يمثّل مجموعة كل النتائج الممكنة Ω (حتماً سيقع أحد الأرقام من 1 إلى 6 في تجربة النرد) ، ونسمي بفضاء الحوادث : لمجموعة الحوادث التي يمكن تشكيلها من فضاء العيّنة بمعنى: فضاء الحوادث هو: مجموعة كل المجموعات الجزئية التي يمكن تشكيلها أو تكوينها من Ω .

مثلا:

- عند رمي قطعة نقود يكون: $\Omega = \{H,T\}$ (حيث تشير H للصورة و T تشير للكتابة) وفضاء الحوادث يكون: $P(\Omega) = \{\Phi, \{T\}, \{H\}, \Omega\}$

– ومن أجل: $\Omega = \{a,b,c\}$ يكون:

$$P(\Omega) = \{\Phi, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \Omega\}$$

وبالحالة العامة: إذا كان Ω مؤلفًا من n حدثا ، فإنّ : $P(\Omega) = 2^n$

- **الأحداث المتنافية:** *Mutually Exclusive Events* :

نقول عن الحدثين A و B من Ω إنّهما متنافيان إذا كان وقوع أحدهما يمنع أو ينفي وقوع الحدث الآخر ، وبالتالي: $A \cap B = \Phi$.

ولما كانت الأحداث تمثّل مجموعات جزئيّة من Ω ، فإنّه يمكن الربط بينها لتكوين أحداث جديدة ، وبالتالي يمكن استخدام جميع مفاهيم نظريّة المجموعات والتعامل مع ما يسمّى" بجبر الأحداث" كتسمية موافقة لجبر المجموعات.

(2-3) الاحتمال: *Probability*:

إذا كـان Ω فضـاء عيّنـة (فضـاء إمكانـات) لتجربـة عشـوائيّة E، وكانـت $S(\Omega)$ هـي مجموعة كل الأحداث المعرّفة على Ω، فإنّه يرافق كل حدث $A\in S(\Omega)$ عدد حقيقي $P(A)\in[0,1]$، ويسمى هذا العدد احتمـال وقـوع الحـدث A، إذا تحقّقت المسلمات الثلاث التالية:

a- $\forall A\subset \Omega \Rightarrow \exists P(A)\ ;\ P(A)\geq 0\ \ or\ \ 0\leq P(A)\leq 1$

b- $P(\Omega)=1$.

c- إذا كان A و B حدثين متنافيين، فإنّ: $P(A\cup B)=P(A)+P(B)$

وبالحالة العامة: إذا كانت $A_i\ ,\ i=\overline{1\ ,\ n}$ أحداث متنافية مثنى مثنى، فإنّ:

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i),i=\overline{1,n}\ .$$

- التعريف الكلاسيكي للاحتمال: *Classical Definition of Probability* :

إذا كانت الأحداث متكافئة (نفس الفرصة في الوقوع) أي إذا كان لكل نقاط Ω نفس الاحتمال، عندئذٍ نقول عن الفضاء Ω إنّه فضاء منتظم، وإذا كان هذا الفضاء منتهياً، فإنّه يمكن أن نُعرّف (من وجهة نظر كلاسيكية) احتمال وقوع الحدث A بالشكل:

$$\boxed{P(A)=\frac{N(A)}{N(\Omega)}=\frac{\text{عدد العناصر في } A}{\text{عدد العناصر في } \Omega}=\frac{\text{عدد الطرق التي يمكن أن يقع بها الحث } A}{\text{عدد الطرق التي يمكن أن يقع بها الحث } \Omega}}$$

وهـو تعريـف يقتصـر علـى دراسـة التجـارب ذات الفضـاء المنتهـي وذات الإمكانـات المتساوية.

مثلا:

- إنّ احتمال ظهور الصورة $P(A)$ عند رمي قطعة نقود منتظمة هو:

$P(A)=\dfrac{n(A)}{n(\Omega)}=\dfrac{1}{2}$ (رمزنا لحدث ظهور الصورة بالرمز A) حيث:

$n(A)=1$ (عدد الطرق التي يمكن أن تظهر بها الصورة على السطح العلوي)

$n(\Omega)=2$ (عدد الطرق التي يمكن أن يظهر بها السطح العلوي)

– إنّ احتمال أن يظهر رقم أقل من 3 في أثناء رمي قطعة نرد منتظمة هو : $\frac{2}{6}$

واحتمال أن يظهر رقم ≤ 4 في نفس التجربة هو : $\frac{3}{6}$

– عند رمي قطعة نقود مع زهرة النرد، فإنّ احتمال كل نتيجة بسيطة هو : $\frac{1}{12}$

واحتمال ظهور الحدث $A = \{(H,5),(T,4),(H,2)\}$ في نفس التجربة هو : $\frac{3}{12}$

التعريف الإحصائي (التجريبي) للاحتمال *Experimental Definition of Probability :*

عندما تكون الأحداث غير متكافئة، وإذا تكررت التجربة في نفس الشروط عدداً كبيراً من المرات مقداره n مرة، ووقع الحدث A بمقدار f مرة، عندئذٍ نسمي النسبة $\frac{f}{n}$ بالتردد النسبي أو التكرار النسبي للحدث A، ويمكن اعتبارها مساويه تقريباً $P(A)$، ومن هنا يمكن إعطاء التعريف الإحصائي للاحتمال بالشكل : $P(A) = \lim\limits_{n \to \infty} \frac{f}{n}$، ولكن من وجهة نظر إحصائيّة، لا يمكن الوصول إلى نهاية النسبة $\frac{f}{n}$ وذلك عندما تسعى n إلى اللانهاية، أضف إلى أنّه يوجد عدد كبير من التجارب لا يمكن تكرارها عدداً كبيراً من المرات في ذات الشروط.

واضح أنّ هذا التعريف لا يشترط تساوي فرص ظهور النتائج أي: لا يشترط الإمكانات المتساوية ، بل هو مبني على إجراء التجربة عدداً كبيراً من المرات.

أمثلة:

1 . أجرى مهندس زراعي ($n = 300$) عملية تطعيم نباتي، فنجح منها ($m = 280$) عملية. ما احتمال نجاح عملية التطعيم ؟

إذا أشرنا لنجاح عملية التطعيم بالرمز A ، يكون: $P(A) = \frac{m}{n} = \frac{280}{300} \approx 0.9333$

2 . إذا كان من بين 2000 مصباح يوجد 60 مصباحاً غير صالح للاستعمال، عندئذٍ: احتمال وجود مصباح جيّد هو: $\frac{1940}{2000} = 0.97$ ، واحتمال وجود مصباح تالف هو

$\frac{60}{2000}=0.03$. من الملاحظ أنّه يوجد عيوب في التعريفين السابقين، حيث اشترطنا في التعريف الكلاسيكي تساوي فرص ظهور النتائج الممكنة للتجربة العشوائيّة، وفي التعريف الإحصائي يُشترَط تكرار التجربة عدداً كبيراً من المرات، وهذا غالباً ما يكون متعذراً.

ولتفادي هذه العيوب طُوِّر هذا التعريف بدوره إلى تعريف أدق للاحتمال يُعرَف باسم العالم الروسي كولموجوروف (التعريف الرياضي للاحتمال)، هذا التعريف المبني على أساس بعض المسلمات والمسماة: مسلمات نظريّة الاحتمالات -Axioms of pr. Theory- التي شكّلت أساس نظريّة الاحتمالات، بل شكّلت أساساً لعلم الاحتمالات بشكل عام، كما النقطة والمستقيم أساس الهندسة المستوية.

التعريف الرياضي للاحتمال (تعريف كولموجوروف): *Mathematical Def. of Prob.*

بفرض E تجربة عشوائيّة فضاء إمكاناتها Ω ، وليكن A حدثاً مرتبطاً بها. إذا رمزنا بـ $S(\Omega)$: لأسرة المجموعات الجزئية في Ω ، عندئذ يعرَّف الاحتمال بأنّه التطبيق: $P : S(\Omega) \to [0,1]$ المحقّق للشروط التالية:

. $P(\Omega)=1$

. (أحداث متنافية مثنى مثنى) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

. وبشكل خاص: $P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad ; \; A_1 \cap A_2 = \Phi$

(3-3) خواص الاحتمال:

من المسلمات السابقة يمكن اقتطاف وإثبات الخواص التالية:

1 - $P(\Phi)=0$

$$P(\Phi) + P(\Omega) = P(\Phi \cup \Omega) = P(\Omega)$$
$$\Rightarrow P(\Phi) = P(\Omega) - P(\Omega) = 0$$

2 - $P(\overline{A}) = 1 - P(A)$

$$P(A)+P(\overline{A}) = P(A \cup \overline{A}) = P(\Omega) = 1$$
$$\Rightarrow P(\overline{A}) = 1 - P(A)$$

إذا كان احتمال أن تقابل صديقا هو: 0.8 ، فإنّ احتمال عدم مقابلته هو:

(1-0.8=0.2)

3 -

$$A \subset B \Rightarrow \begin{cases} P(B-A) = P(B) - P(A) \\ P(A) \le P(B) \end{cases}$$

$$B = [A \cup (B - A)]$$
$$P(B) = P(A) + P(B - A)$$
$$P(B) - P(A) = P(B - A) \ge 0$$
$$P(A) \le P(B)$$

كذلك: $\forall A : A \subseteq \Omega \Rightarrow P(A) \le P(\Omega) = 1$

ومن التعريف واضح أنَّ: $P(A) \ge 0$ ، وبالتالي: $\forall A : 0 \le P(A) \le 1$

وبالحالة العامة إذا كان: $A_1 \subset A_2 \subset A_3 \subset \ldots \subset A_n$

فإنَّ: $P(A_1) \le P(A_2) \le P(A_3) \le \ldots \le P(A_n)$

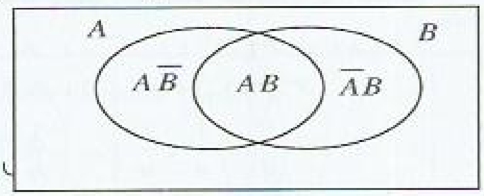
4 - $\forall A, B : P(A \cup B) = P(A) + P(B) - P(A \cap B)$

لنأخذ: $Z = B - (A \cap B) \;\; ; A \cap Z = \Phi \;\; \& \;\; A \cup Z = A \cup B$

عندئذٍ:

$$\begin{aligned} P(A \cup B) = P(A \cup Z) &= P(A) + P(Z) \\ &= P(A) + P[B - (A \cap B)] \\ &= P(A) + P(B) - P(A \cap B) \quad ; \quad A \cap B \subseteq B \end{aligned}$$

وبطريقة أخرى لدينا: $A \cup B = A\overline{B} \cup AB \cup \overline{A}B$ (لاحظ المخطط)



بالتالي: $P(A$

$$= P(A\overline{B}) + P(AB) + P(\overline{A}B) \; ; \; AB = A \cap B$$

$$= P(A) - P(AB) + P(AB) + P(B) - P(AB)$$

$$= P(A) + P(B) - P(AB)$$

إنَّ الرسم يساعد على البرهان (لكنه لا يُعتَمد كبرهان) حيث اعتبرنا الحادث وكأنَّه مساحة ، وبالتالي ما يصح على (جمع ، فرق ، جداء) المساحات ، هو صحيح على احتمالات الأحداث الموافقة لهذه المساحات على الرسم أو المخطط .

وتعرف هذه الخاصة بقاعدة الجمع، ويمكن تعميمها على n حدث بواسطة الاستقراء الرياضي كما يلي:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) - P(A_1A_2) - \ldots$$
$$- P(A_{n-1}A_n) + P(A_1A_2A_3) + \ldots + P(A_{n-2}A_{n-1}A_n)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$+ (-1)^{n-1} P(A_1A_2A_3 \ldots A_n)$$

فمثلاً:

- من أجل $n=3$ يكون:

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_1A_3) - P(A_2A_3) + P(A_1A_2A_3)$$

وإذا كانت: $A_i \ ; \ i = \overline{1\,,\,n}$ أحداث متنافية مثنى مثنى في تجربة ذات فضاء منتظم (أي ذات إمكانات متساوية)، فإنَّ:

$$P(A_1) + \ldots + P(A_n) = P\left(\bigcup_{i=1}^{n} A_i\right) = P(\Omega)$$
$$\Rightarrow n\,P(A_i) = 1 \ ; \ P(A_i) = \frac{1}{n} \ , \ i = \overline{1\,,\,n}$$

- إذا كان: A حدثاً مرتبطاً بالتجربة، وعدد عناصره a، عندئذٍ:

$$\begin{aligned} P(A) &= P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_a) \\ &= P(A_1) + P(A_2) + P(A_3) + \ldots + P(A_a) \\ &= \frac{1}{n} + \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} = \frac{a}{n} \equiv \frac{N(A)}{N(\Omega)} \end{aligned}$$

وهذا يطابق التعريف الكلاسيكي للاحتمال.

وكحالة خاصة:

إذا كان: A_1 , A_2 حدثين متنافيين، وكان الحدث: $A = (A_1 \cup A_2)$ عندئذٍ: إذا تكررت التجربة عدداً من المرات وبشكل مستقل، فإنّه في كل مرة سيظهر الحدث A ، وبالتالي سيظهر أحد الحدثين: $A_1 \ or \ A_2$.

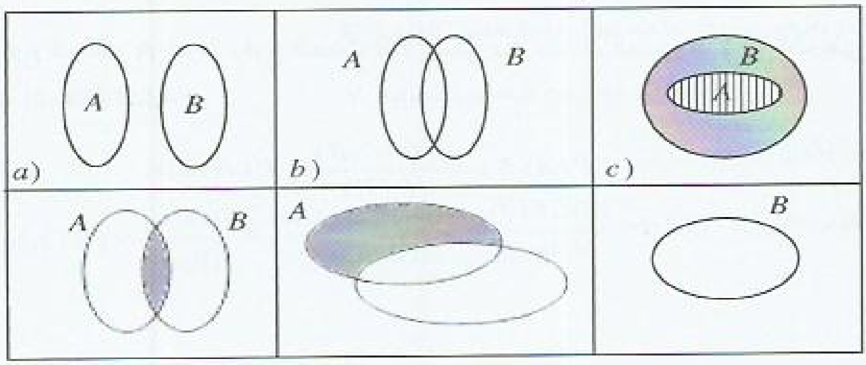
فإذا فرضنا أنّ:

- n عدد مرات تكرار التجربة.
- $n(A)$ عدد مرات ظهور الحدث A .
- $n(A_1)$ عدد مرات ظهور الحدث A_1 .
- $n(A_2)$ عدد مرات ظهور الحدث A_2 .

عندئذٍ : $n(A) = n(A_1) + n(A_2) \Rightarrow n(A)/n = n(A_1)/n + n(A_2)/n$

وعندما تكون n كبيرة بشكلٍ كافٍ، فإنّه من الناحية العملية ستتطابق النسب السابقة / الترددات/ مع الاحتمالات الموافقة لكلٍ منها، ويكون: $P(A) = P(A_1) + P(A_2)$ وهي حالة خاصة من قانون جمع الاحتمالات، وكما نوّهنا سابقاً، فإنّه يمكن تمثيل المجموعات (الأحداث) بمخططات، وهذه المخططات تُظهر صحة العلاقات بين المجموعات، ولكن لا تثبتها، وعلى الرغم من ذلك، يمكن القول: إنّ الحيّز أو المساحة التي تشغلها الأحداث بعلاقاتها في أثناء تمثيلها بمخططات، هذه المساحة تناظر تماماً خواص الاحتمال الموافقة علماً أنّ المخطط يلعب دور وسيلة الإيضاح التي تُيسّر متابعة وفهم خطوات البرهان، ولكنّه لا يشكل جزءاً من البرهان كما أشرنا.

فمثلاً من أجل المخطط التالي:



B A

$d)$ $e)$ $f)$

وعلى سبيل المثال:

حسب الشكل e: $P(A-B)=P(A)-P(A\cap B)$

حسب الشكل c: $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

حسب الشكل d: $P(A\cap B)=P(A)+P(B)-P(A\cup B)$

وهكذا.... وتبقى المساحة المشغولة على المخطط مناظرة لخواص الاحتمال الموافق أي: كلما اتسع الحدث وتضمن عدداً أكبر من النقاط ، كلما ازداد احتمال وقوعه... (كما سلوك الإنسان ينسجم مع المساحة الفكريّة والأخلاقيّة التي يشغلها في حياته كما يفترض أن يكون).

(3-4) نظريّة جداء وجمع الاحتمالات:

الاحتمال المشروط *Conditional Probability* **وقاعدة الضرب:** *Multiplication Rule*

في الواقع الحياتي، أو عند دراسة ظاهرة ما، وبالملاحظة يُطرح سؤال مهم:

كيف يتأثر وقوع حدث ما بوقوع أو عدم وقوع حدث آخر؟

فقد يتأثر وقوع الحدث A بوقوع الحدث B ، وقد يؤدي وقوع A لاستحالة وقوع B .

وبشكل عام ما تأثير وقوع أحد الحدثين على الآخر؟

تعريف:

احتمال وقوع الحدث A شرط وقوع الحدث B ، يسمّى بالاحتمال المشروط (أو الشّرطي)، ويُعرّف بالعلاقة التالية:

$$P(A\setminus B)=\frac{P(A\cap B)}{P(B)}\ ;\ P(B)\neq 0$$

حيث يتضح أنّ وقوع الحدث A يتم وفق شروط جديدة . بمعنى آخر : إنّ دراسة الاحتمال المشروط توضّح أنّه بإضافة معلومات جديدة تتغيّر قيم احتمالات الأحداث .

ومن خواصه:

$$0 \le P(A \setminus B) \le 1$$

$$B \subseteq A \Rightarrow P(A \setminus B) = 1$$

$$A \cap B = \Phi \Rightarrow P(A \setminus B) = 0$$

وإذا كانت: $A_k ; k = \overline{1,n}$ أحداث متنافية مثنى مثنى ، وكان: $A = \underset{k}{\cup} A_k$

فإنّ: $P(A \setminus B) = \sum_{k=1}^{n} P(A_k \setminus B)$ وهو يحقّق جميع خواص الاحتمال السابقة.

مثال (1):

في تجمّع سكني ما ، تبيّن أنّ: 15% من السكان يمارسون الهواية X ، 35% من السكان يمارسون الهواية Y ، 10% من السكان يمارسون الهواية X و Y.

المطلوب:

1- من بين أولئك الذين لا يمارسون الهواية Y، ما نسبة السكان الذين يمارسون الهواية X ؟

2- ما النسبة المئوية للسكان الذين لا يمارسون أي من الهوايتين المذكورتين ؟

3- من بين أولئك الذين يمارسون الهواية X ، ما هي نسبة السكان الذين يمارسون الهواية Y ؟

الحل: بفرض: A حدث ممارسة شخص للهواية X

B حدث ممارسة شخص للهواية Y

عندئذٍ يكون:

المطلوب الأول: $p(A \setminus \overline{B}) = \dfrac{p(A\overline{B})}{p(\overline{B})} = \dfrac{p(A) - p(AB)}{1 - p(B)} = \dfrac{0.15 - 0.1}{1 - 0.35} = 0.077$

أي: 7.7% فقط من السكان الذين لا يمارسون الهواية Y، هم ممّن يمارس الهواية X.

والمطلوب الثاني: $P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)$

$$=1-[P((A)+P(B)-P(AB)$$

$$=1-0.15-0.35+0.1=0.6$$

بمعنى: 60% من السكان لا يمارسون أياً من الهوايتين.

والمطلوب الثالث: $p(B\setminus A)=\dfrac{p(AB)}{p(A)}=\dfrac{0.1}{0.15}=0.66$

أي: 66% من السكان الذين يمارسون الهواية X , هم ممّن يمارس الهواية Y .

مثال (2):

رمينا زهرتي نرد مرّة واحدة ، فإذا كان مجموع النقط 6 .

فما احتمال أن يظهر على إحدى الزهرتين الرقم 1 ؟

الحل: واضح أنّ: $N(\Omega)=6\times6=36$

وإذا فرضنا أنّ: A حدث إحدى الزهرتين تحمل الرقم 1

B حدث مجموع النقط على سطحي الزهرتين هو 6

عندئذ يكون المطلوب هو حساب $P(A\setminus B)$ حيث:

$$A=\{(1,2),(1,3),(1,4),(1,5),(1,6)(2,1),(3,1),(4,1),(5,1),(6,1)\}$$

$$B=\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

$$A\cap B=\{(1,5),(5,1)\}$$

وبما أنّ: $N(A\cap B)=2$ ، $N(B)=5$ ، $N(A)=10$

يكون: $P(A\setminus B)=\dfrac{P(AB)}{P(B)}=\dfrac{\frac{N(AB)}{N(\Omega)}}{\frac{N(B)}{N(\Omega)}}=\dfrac{\frac{2}{36}}{\frac{5}{36}}=\dfrac{2}{5}$

مثال (3):

عائلة لديها طفلان . ما احتمال أن يكونا ذكرين علماً أنّ:

1 . الطفل الأكبر ذكر.

2 . واحد من الطفلين على الأقل ذكر.

الحل:

إذا رمزنا للذكر بـ b وللأنثى بـ g عندئذٍ يكون:

$$\Omega=(bb,bg,gb,gg)\Rightarrow N(\Omega)=4$$

وبفرض:

B حدث كون الطفل الأكبر ذكراً أي: $B=(bb,bg)\Rightarrow N(B)=2$

A هو حدث كون الطفل الأصغر ذكراً أي: $A=(bb,gb)\Rightarrow N(A)=2$

بالتالي يكون المطلوب الأول هو: $P(AB\setminus B)=\dfrac{P(AB)}{P(B)}=\dfrac{\frac{N(AB)}{N(\Omega)}}{\frac{N(B)}{N(\Omega)}}=\dfrac{\frac{1}{4}}{\frac{2}{4}}=\dfrac{1}{2}$

حيث: $AB=(bb)\Rightarrow N(AB)=1$

والمطلوب الثاني هو: $P(AB\setminus A\cup B)=\dfrac{P(AB)}{P(A\cup B)}=\dfrac{\frac{N(AB)}{N(\Omega)}}{\frac{N(A\cup B)}{N(\Omega)}}=\dfrac{1}{3}$

حيث: $A\cup B=\{(bb),(bg),(gb)\}\Rightarrow N(A\cup B)=3$

مثال (4):

سلة تحوي 12 زهرة ، منها 5 تالفة . اختيرت عشوائياً وعلى التوالي 3 زهرات، ما احتمال أن تكون جميعها جيّدة (غير تالفة) ؟

الحل: بفرض : A حدث كون الزهرة الأولى جيّدة

B حدث كون الزهرة الثانية جيِّدة

C حدث كون الزهرة الثالثة جيِّدة

عندئذٍ يكون المطلوب هو: $P(ABC) = P(A)P(B \setminus A)P(C \setminus AB)$

$$= \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10} = \frac{210}{1320} = 0{,}159$$

. الاستقلال – الحوادث المستقلة: *Independent Events*

تعريف:

نقول عن الحدثين A و B إنهما مستقلان بالتبادل (كلاهما مستقل عن الآخر) إذا كان احتمال وقوع أحدهما لا يتعلق بوقوع أو عدم وقوع الحدث الآخر، بمعنى:

$P(A) = P(A \setminus B) = P(A \setminus \overline{B})$، أو: $P(B) = P(B \setminus A) = P(B \setminus \overline{A})$

وبخلاف ذلك يكون A و B مرتبطين .

نتيجة:

$$P(AB) = P(B)P(A \setminus B)$$
$$P(AB) = P(A)P(B \setminus A)$$

فإذا فرضنا أنّ: $N(A) = m\,,\ N(AB) = k,\ N(\Omega) = n$ (حيث: Ω المجموعة الشاملة أو الكلية) عندئذٍ: $P(A) = \frac{m}{n}$ ، $P(AB) = \frac{k}{n}$ ، ولكن شرط وقوع A يختزل الفضاء إلى $N(AB) = k$ ، وبالتالي يكون : $P(B \setminus A) = \frac{k}{n}$ وتبقى العلاقة:

$P(AB) = P(A)P(B \setminus A)$ صحيحة.

فإذا كان A مستقلاً عن B ، عندئذٍ يمكن أن يُكتب شرط الاستقلال أيضاً بالشكل:

$$P(AB) = P(A)P(B) \quad ; P(AB) \equiv P(A \cap B)$$

وبالحالة العامة ،العكس ليس بالضرورة أن يكون صحيحاً.

نتائج:

. الأحداث المستقلة عن نفسها هي: Ω و Φ فقط.

. الحدثان: Ω و Φ هما الحدثان الوحيدان المستقلان عن كل حدث.

. إذا كان A و B حدثين مستقلين بالتبادل فإنّ كلاً من:

$\overline{A} \,\&\, B$ ، $A \,\&\, \overline{B}$ ، $\overline{A} \,\&\, \overline{B}$ مستقلان بالتبادل

. الحدثان المتنافيان لا يمكن أن يكونا مستقلين / *برهن هذه النتائج* /.

مثال (1):

بفرض E تجربة عشوائيّة ما، وليكن A و B حدثين مرتبطين بهذه التجربة.

إذا فرضنا أنّ: $P(A)=\dfrac{3}{10}$, $P(B)=\dfrac{3}{7}$, $P(A\cup B)=\dfrac{6}{10}$

بيّن ما إذا كان الحدثان A و B مستقلين أم لا ؟

الحل:

أحد أشكال شروط الاستقلال: $P(AB)=P(A).P(B)=\dfrac{3}{10}.\dfrac{3}{7}=\dfrac{9}{70}$

ومن ناحية أخرى لدينا:

$$P(AB)=P(A)+P(B)-P(A\cup B)=\frac{3}{10}+\frac{3}{7}-\frac{6}{10}=\frac{9}{70}$$

والحدثان مستقلان.

مثال (2):

سُحبَت ورقة من مجموعة أوراق اللعب . فإذا كان: A حدث كون الورقة المسحوبة تحمل الرقم 1 ، B حدث كون الورقة المسحوبة قلب . فهل الحدثان مستقلان ؟

الحل:

من الواضح أنّ: $P(A)=\dfrac{4}{52}$ ، $P(B)=\dfrac{13}{52}$ ، $P(AB)=\dfrac{1}{52}$

وبما أنّ: $\dfrac{4}{52}\,\dfrac{13}{52}=\dfrac{1}{52}$ أي: $P(A\,B)=P(A)\,P(B)$ والحدثان مستقلان.

مثال (3):

اختير عشوائياً 3 أشخاص من بين 4 ذكور و 3 إناث. ما احتمال أن يكون من بينهم:

1 . ذكران ؟

2 . ذكران على الأقل ؟

الحل:

بفرض : m ترمز لاختيار ذكر ، w ترمز لاختيار أنثى . عندئذٍ:

$\Omega = \{mmm, mmw, mwm, wmm, wwm, wmw, mww, www\}$ ، حيث:

$n(\Omega) = 2^3 = 8$

لحساب احتمالات هذه العناصر نأخذ عنصر (وليكن mwm) ونحسب احتمال الحصول عليه: $p(mwm) = p(m)\, p(w \setminus m)\, p(m \setminus mw)$

$$= \frac{4}{7} \times \frac{3}{6} \times \frac{3}{5} = \frac{6}{35}$$

وبحساب احتمالات عناصر فضاء العيّنة نحصل على الفضاء الاحتمالي التالي:

$$P = [\ \frac{1}{35}, \frac{4}{35}, \frac{4}{35}, \frac{4}{35}, \frac{6}{35}, \frac{6}{35}, \frac{6}{35}, \frac{4}{35}\ ]$$

وإذا فرضنا أنّ:

A هو حدث اختيار ذكرين يكون: $A = (mmw, mwm, wmm)$

B هو حدث اختيار ذكرين على الأقل يكون: $B = (mmw, mwm, wmm, mmm)$

بالتالي: $p(A) = p(mmw) + p(mwm) + p(wmm) = \frac{6}{35} + \frac{6}{35} + \frac{6}{35} = \frac{18}{35}$

وبالمثل يكون: $p(B) = \frac{22}{35}$.

(3–5) الأحداث الشاملة:

- إذا كانت: $A_i\ ,\ i=\overline{1\ ,\ n}$ أحداث (ليست مستحيلة) متنافية مثنى مثنى في فضاء إمكانات تجربة عشوائيّة ما، وكان: $A=\cup_{i=1}^{n} A_i$ ، عندها نقول إنّ الأحداث $A_i\ ,\ i=\overline{1\ ,\ n}$ تشكّل تجزئة للحدث A .

- نقول عن أسرة أو صف الأحداث $A_i\ ,\ i=\overline{1\ ,\ n}$ إنّها تُشكّل جماعة شاملة (أو جماعة تامة) إذا كانت تُشكّل تجزئة للحدث Ω (حيث Ω فضاء العيّنة).

واضح أنّ الحدثين: A و $\overline{A}$ يشكلان جماعة شاملة حيث:

$$A \cup \overline{A}=\Omega\ ,\ A \cap \overline{A}=\Phi$$

. نظريّة الأحداث الشاملة /صيغة الاحتمال التام/:

إذا كانت: $B_k\ ;k=\overline{1,n}$ تجزئة للحدث الأكيد Ω (أي جماعة شاملة من الأحداث) واحتمالاتها معلومة، وكان: A حدثاً مرتبطاً بالتجربة نفسها (بمعنى يقع فقط مع واحد من الأحداث B_k) فإنّ: $P(A)=\sum_k P(B_k)P(A\backslash B_k)$

(هو ما يسمّى بنظريّة الأحداث الشاملة أو صيغة الاحتمال التام).

البرهان:

لدينا: $A=A\cap\Omega=A\cap(B_1\cup B_2\cup B_3\cup\ldots\cup B_k)$

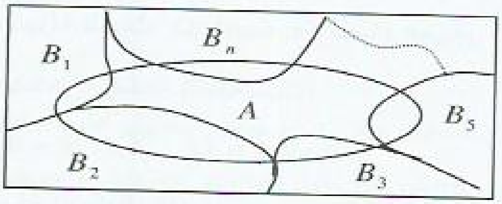
$$=AB_1\cup AB_2\cup AB_3\cup\ldots\cup AB_k$$

وبما أنّ الأحداث $B_k\ ,\ k=\overline{1\ ,\ n}$ تشكل جماعة شاملة ، فهي متنافية ، وبالتالي الأحداث $AB_1\ ,\ AB_2\ ,\ AB_3\ ,\ \ldots.\ AB_k$ متنافية أيضا ، وبالاعتماد على قانوني جمع وجداء الاحتمالات يكون:

$$\begin{aligned}P(A)&=P(AB_1)+P(AB_2)+P(AB_3)+\ldots+P(AB_k)\\&=P(A\backslash B_1)P(B_1)+P(A\backslash B_2)P(B_2)+\ldots+P(A\backslash B_k)P(B_k)\\&=\sum_k P(B_k)P(A\backslash B_k)\end{aligned}$$

. صيغة بايز: *Bayes*

إذا كانـت: $B_k ; k=\overline{1,n}$ تجزئـة للحدث الأكيد Ω (أي جماعـة شاملة من الأحـداث) واحتمالاتها معلومة ، وكان : A حدثاً مرتبطاً بالتجربة نفسها (بمعنى يقع فقط مع واحد من الأحداث B_k ، انظر الشكل) فإنّ : $P(B_k \backslash A)=\dfrac{P(B_k)\,P(A\backslash B_k)}{\sum_k P(B_k)P(A\backslash B_k)}$



لبيان ذلك:

لدينا حسب نظريّة جداء الاحتمالات:

$$P(AB_k)=P(A)P(B_k \backslash A)=P(B_k)P(A\backslash B_k)$$

ومنه: $$P(B_k \backslash A)=\frac{P(B_k)\,P(A\backslash B_k)}{P(A)}$$

ولكن: $$P(A)=\sum_k P(B_k)P(A\backslash B_k)$$

بالتالي: $P(B_k \backslash A)=\dfrac{P(B_k)\,P(A\backslash B_k)}{\sum_k P(B_k)P(A\backslash B_k)}$ وتسمى هذه الصيغة بصيغة بايز.

مثال (1):

ثلاثة صناديق متماثلة تحوي كرات بيضاء وسوداء بحيث : يحوي الصندوق الأول 5 بيضـاء و 5 سوداء ، ويحوي الثاني 8 بيضـاء وكرتين سوداويين ، ويحوي الثالث 4 بيضاء و 6 سوداء. سُحِبَ صندوق عشوائيّاً وسُحِبَتْ منه كرة دون تحيُّز.

المطلوب : ا . ما احتمال أن تكون هذه الكرة بيضاء ؟

ب . إذا علمت أنّها بيضاء ، ما احتمال أن تكون من الصندوق الثاني؟

الحل:

لنرمز بـ : B_i لحدث سحب الصندوق رقم : ($i\ ;\ i=\overline{1,3}$)

A لحدث سحب كرة بيضاء .

ولكي نستطيع تطبيق نظريّة الأحداث الشاملة أو صيغة بايز ، فإنّه يجب التحقُّق أنّ الأحداث المفترضة تُشكّل جماعة شاملة ، ثم هل الحدث A يقع فقط مع واحد من الأحداث ؟ (واضح أنّ ما ذكرناه محقّق في المثال المطروح) لذلك يكون مباشرةً :

المطلوب الأول: $$P(A)=\sum_{i=1}^{3}P(A\setminus B_i)P(B_i)=\frac{5}{10}\cdot\frac{1}{3}+\frac{8}{10}\cdot\frac{1}{3}+\frac{4}{10}\cdot\frac{1}{3}=\frac{17}{10}$$

المطلوب الثاني: $$P(B_2\setminus A)=\frac{P(A\setminus B_2).P(B_2)}{P(A)}=\frac{\frac{8}{10}\cdot\frac{1}{3}}{\frac{17}{10}}=\frac{8}{17}$$

مثال (2):

ثلاث آلات تنتج مصابيح بالشكل التالي:

إنتاج الآلة الأولى 30% ونسبة المُعيب فيها 1% .

إنتاج الآلة الثانية 36% ونسبة المُعيب فيها 2% .

إنتاج الآلة الثالثة 34% ونسبة المُعيب فيها 2% .

1- اخترنا مصباحاً بشكل عشوائي ، ما احتمال أن يكون هذا المصباح مُعيباً ؟

2- إذا كان المصباح مُعيباً ، فما احتمال أن يكون من إنتاج الآلة الأولى ؟

الحل :

إذا رمزنا بـ : $i=\overline{1\,,3}\ ;\ A_i$ لحدث كون المصباح من إنتاج الآلة رقم i ،

: B لحدث كون المصباح مُعيباً. عندئذٍ:

واضـح أنّ الأحـداث $A_i\ ;\ i=\overline{1\ ,\ 3}$ تشـكّل جماعـة شـاملة ، وبالتـالي حسـب نظريّـة الأحداث الشاملة يكون المطلوب الأول:

$$P(B)=P(B\setminus A_1)\ P(A_1)+P(B\setminus A_2)\ P(A_2)+P(B\setminus A_3)\ P(B_3)$$

$$=(0.01)\ .\ (0.30)\ \ +\ \ (0.02)\ .\ (0.36)\ \ +\ \ (0.02)\ .\ (0.34)=0.017$$

والمطلوب الثاني:

$$P(A_1\setminus B)=\frac{P(A_1\ B)}{P(B)}=\frac{P(B\setminus A_1)\ P(A_1)}{P(B)}=\frac{(0.01)\ .\ (0.30)}{0.17}=\frac{1}{17}$$

(6-3) التجارب المتكرّرة:

تطرح المسألة بالشكل التالي: E تجربة عشوائيّة ، A حدث مرتبط بها، فإذا تكررت E عـــدداً مـــن المـــرات مقـــداره n مـــرة فـــي نفـــس الشـــروط، وكـــان:

$$P(A)=p\ ,\ P(A')=q$$

المطلوب:

ما احتمال ظهور الحدث A عدداً من المرات مقداره $(r\ ;\ r\le n)$ مرة ؟ (يمكن التعميم).

من الواضح أنّ ظهور الحدث A عدداً من المرات مقداره r مرة ، يكافئ القول:

إنّ الحث A' ظهر $(n-r)$ مرة . بمعنى آخر : سنحصل على قائمة من الشكل:

$$L_i=\{A,A,A',A,A',A',....,A,A,A',...,A\}$$

بحيث يتكرر A عدداً من المرات مقداره r مرة ، ويتكرر A' عدداً من المرات مقداره $(n-r)$ مرة ، واحتمال الحصول على مثل هذه القائمة هو :

$$P[\,L_i\,]=\underbrace{p.p.p...p}_{r}\ \underbrace{q.q.q...q}_{(n-r)}=p^r\ .q^{n-r}$$

حيث التجارب المتكرّرة مستقلة عن بعضها.

ولكن هذه النتيجة يمكن أن ترد بأية قائمة أخرى تحوي نفس العدد من الرموز A أو A' وبترتيب مختلف، وبالتالي يكون المطلوب هو احتمال الحصول على إحدى هذه القوائم، أي هو : $P[\bigcup_i L_i]=\sum_{i=1}^{?} P(L_i)$ ، وأصبح من الضروري معرفة عدد هذه القوائم (علماً

أنّ القائمتين L_i و L_j تمثلان حدثين متنافيين، وتختلفان فقط بترتيب أماكن الرموز A و A')، وعدد القوائم هذه يكافئ عدد الطرق التي يمكن فيها أن نقسم مجموعة مؤلفة من n عنصراً إلى مجموعتين:

[تحوي الأولى r عنصراً ، وتحوي الثانية $n-r$ عنصراً] ، وهذا يمثله العدد C_r^n كما نعلم.

بمعنى: $P[\bigcup_i L_i] = \sum_{i=1}^{C_r^n} P(L_i) = C_r^n \cdot p^r \cdot q^{n-r}$ ، أي: إذا رمزنا بـ A_r لحدث وقوع A عدداً من المرات مقداره r ، ولاحتمال وقوعه بالرمز $P(A_r)$ فإنّ: $P(A_r) = C_r^n \cdot p^r \cdot q^{n-r}$ وهو: **قانون التجارب المتكرّرة (برنولي)**.

وإذا كان المطلوب أن يكون: $l \le r \le m$ ، فإنّ الاحتمال السابق يصبح بالشكل:

$$P[\bigcup_{r=l}^{m} A_r] = \sum_{r=l}^{m} P(A_r) = \sum_{r=l}^{m} C_r^m \cdot p^r \cdot q^{m-r}$$

وذلك لأنّ الأحداث: $A_l, A_{l+1}, \ldots, A_r, \ldots, A_{m-1}, A_m$ متنافية ، والمطلوب هو الحصول على أحدها.

حاله خاصة:

من أجل: [$l=0$ ، $m=n$] يكون: (حدث أكيد): $\sum_{r=0}^{n} C_r^n\, p^r\, q^{n-r} = (p+q)^n = 1$

وفي مسألة التجارب المتكرّرة ، إذا اعتمدنا الرموز التالية:

$$A_1 = A \ ,\ A_2 = A'$$
$$P_1 = p \ ,\ P_2 = q$$
$$n_1 = r \ ,\ n_2 = n - r$$

بمعنى آخر: إذا اعتبرنا أنّ: A_1 و A_2 هما الحدثان الوحيدان المرتبطان بالتجربة العشوائيّة E (وهما يشكلان جماعه شاملة بالطبع)، وإذا تكررت التجربة n مرة تحت شروط واحدة ، فإنّ احتمال: { ظهور A_1 عدداً من المرات مقداره n_1 مرة ، وظهور A_2 عدداً من المرات مقداره n_2 مرة } هو:

$$P(n_1, n_2) = \frac{n!}{n_1! n_2!} p_1^{\,n_1} p_2^{\,n_2}$$

الحالة العامة:

إذا فرضنا أنَّ: $A_1 , A_2 , A_3 , ..., A_k$ جماعة شاملة من الأحداث المرتبطة بـ E ، واحتمالات وقوعها: $p_1 , p_2 , p_3 , ..., p_k$ على الترتيب . فإذا تكررت التجربة n مرة تحت شروط واحدة ، فإنَّ احتمال:

{ ظهور A_1 عدداً من المرات مقداره n_1 مرة ، وظهور A_2 عدداً من المرات مقداره n_2 مرة ، وظهور A_3 عدداً من المرات مقداره n_3 مرة ، و ... وظهور A_k عدداً من المرات مقداره n_k مرة } يعطى بالعلاقة التالية:

$$P(n_1, n_2, ..., n_k) = \frac{n!}{n_1! n_2! .... n_k!} P_1^{n_1} P_2^{n_2} ... P_k^{n_k}$$

وهو : **الحالة العامة لقانون التجارب المتكررة** ويسمى أحياناً بقانون الأحداث المركَّبة.

أمثلة:

1- أُلقيتْ قطعة نردٍ 5 مرات. والمطلوب:

a. ما احتمال ظهور الرقم 3 على سطحها العلوي 4 مرات ؟

b. ما احتمال ظهور الرقم 3 على سطحها العلوي 4 مرات على الأقل ؟

الحل:

$$P_1 = C_r^n \; p^r \; q^{n-r} \equiv C_4^5 \, (\frac{1}{6})^4 (\frac{5}{6})^{5-4} \approx 0.0032 \;\; ; \; p = \frac{1}{6} \, , \, q = \frac{5}{6}$$

$$P_2 = \sum_{r=4}^{5} C_r^n \; p^r \; q^{n-r} \equiv C_4^5 \; p^4 \; q^1 + C_5^5 \; p^5 \; q^0 \approx 0.0030$$

2- يرمي رامٍ على هدف، ولنفرض أنَّ احتمال إصابة قذيفة واحدة لمركز الهدف هو 0.2 وأنَّ احتمال أن تصيب بقية الهدف هو 0.5 .

المطلوب: عند إطلاق 10 قذائف، ما احتمال أن تكون 4 منها في مركز الهدف و 4 منها في الجزء المتبقي من الهدف ؟

الحل:

بفرض: A_1 حدث إصابة القذيفة لمركز الهدف.

A_2 حدث إصابة القذيفة لبقية أجزاء الهدف.

A_3 حدث عدم إصابة الهدف.

وهي تشكل جماعه شاملة كما هو واضح.

لدينا: $$P(n_1,n_2,n_3)=\frac{n!}{n_1!n_2!n_3!}p_1^{n_1}p_2^{n_2}p_3^{n_3}$$

حيث: $p_1=P(A_1)=0.2\ ,\ p_2=P(A_2)=0.5\ ,\ p_3=P(A_3)=0.3$

بالتالي: $$P(4,4,2)=\frac{10!.}{4!4!2!}(0.2)^4(0.5)^4(0.3)^2=0.028$$

3- أراد X بناء علاقات صداقه في المرحلة الجامعية. فإذا فرضنا أنّ الاحتمال كي يجد صديقاً مقرّباً هو : 0.02 واحتمال أن يجد صديقاً عادياً هو : 0.6 . فإذا تعرّف X على 10 زملاء ، ما احتمال أن يجد منهم صديقاً واحداً مقرّباً و 5 عاديين ؟

الحل:

بفرض الأحداث:

A_1 حدث أن يجد صديقاً مقرّباً .

A_2 حدث أن يجد صديقاً عادياً .

A_3 حدث عدم وجود صديق .

واضح أنّ: $$\left.\begin{array}{l}P(A_1)=0.02\\ P(A_2)=0.6\end{array}\right\}\Rightarrow P(A_3)=0.38$$

والاحتمال المطلوب هو : $$P_{(1,5,4)}=\frac{10!}{1!\,5!\,4!}(0.02)\,(0.6)^5(0.38)^4=0.04$$

(3-7) سحب العيّنات:

المسألة المطروحة بالشكل التالي: مجتمع A (كلمة مجتمع قد تعني: مجموعة بشريّة، مجموعة أشياء ، مجوعة نتائج تجربة...) يحتوي على N عنصراً ، وهذه العناصر مقسمة إلى k نوعاً بحيث:

عدد عناصر النوع A_i هو : N_i ، $i=\overline{1,k}$ ، و : $N = N_1 + N_2 + ... + N_k$

سُحِبَتْ وبشكل عشوائي عيّنة من هذا المجتمع حجمها n (أي عدد عناصرها n).

المطلوب:

ما احتمال أن تحوي هذه العيّنة على n_i عنصراً من النوع A_i ، $i=\overline{1,k}$ علماً أنّه لكل عناصر A حظاً متساوٍ في الظهور؟

لكن، وبما أنّ السحب قد يكون (مع إعادة أو دون إعادة)، فالجواب يتأتّى من حالتين:

a- السحب مع إعادة:

والجواب يكون: $P = \dfrac{n!}{n_1!n_2!...n_k!}(\dfrac{N_1}{N})^{n_1}(\dfrac{N_2}{N})^{n_2}...(\dfrac{N_k}{N})^{n_k}$

ولبيان ذلك نلاحظ ما يلي:

1- إنّ عدد الطرق الممكنة لتشكيل العيّنة الكيفيّة هو: $\underbrace{(N.N...N)}_{n} = N^n$ طريقة. (وهو يمثّل عدد الحالات الكليّة) ، وبما أنّه نريد أن تحوي هذه العيّنة الكيفيّة على :

n_1 من النوع A_1 .

n_2 من النوع A_2 .

... .

n_k من النوع A_k .

فهذا يكافئ القول: إنّ العيّنة الكيفيّة n يجب أن تنقسم إلى k مجموعة ، بحيث تحوي المجوعة رقم i على n_i عنصراً من النوع A_i ، وهذا التقسيم يتم بـ : $N_i^{n_i}$ طريقة.

لاحظ مثلاً:

عدد طرق تشكيل عيّنة حجمها n_1 من النوع A_1 هو : $N_1^{n_1} = \underbrace{N_1.N_1...N_1}_{n_1}$ طريقة.

عدد طرق تشكيل عيّنة حجمها n_2 من النوع A_2 هو : $N_2^{n_2} = \underbrace{N_2.N_2...N_2}_{n_2}$ طريقة.

....

عدد طرق تشكيل عيّنة حجمها n_k من النوع A_k هو : $N_k^{n_k} = \underbrace{N_k.N_k...N_k}_{n_k}$ طريقة.

وهكذا... وبالتالي فإنّ عدد طرق تشكيل هذه العيّنات هو : $N_1^{n_1} \cdot N_2^{n_2} \ldots N_k^{n_k}$.

ولكن عدد طرق تقسيم العيّنة ذات الحجم n إلى k مجموعه (كما نعلم) هو :

$$C^n_{n_1,n_2,\ldots,n_k} = \frac{n!}{n_1!.n_2!\ldots n_k!}$$

2 – مما تقدّم ، وحسب مبدأ الـ (n . m) يكون عدد طرق تشكيل عيّنة ملائمة لطلبنا هو :

$$C^n_{n_1,n_2,\ldots,n_k} . N_1^{n_1} N_2^{n_2} \ldots N_k^{n_k} = \frac{n!}{n_1!.n_2!\ldots n_k!} N_1^{n_1} N_2^{n_2} \ldots N_k^{n_k}$$

ويكون الاحتمال المطلوب:

$$P = \frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الكليّة}} = \frac{\frac{n!}{n_1!.n_2!\ldots n_k!} N_1^{n_1} N_2^{n_2} \ldots N_k^{n_k}}{N^n = N^{n_1+n_2+\ldots n_k}}$$

وبما أنّ: $N^n = N^{n_1+n_2+\ldots+n_k}$ يكون:

$$P = \frac{n!}{n_1!n_2!\ldots n_k!} (\frac{N_1}{N})^{n_1} (\frac{N_2}{N})^{n_2} \ldots (\frac{N_k}{N})^{n_k}$$

وهو : **قانون السحب مع إعادة .**

مثال:

شعبة متفوقين تضم: 8 طلاب بمعدّل جيّد جداً ، 20 طالب بمعدّل ممتاز.

اختيرت عشوائيّاً عيّنة مع إعادة حجمها: $n = 6$ طلاب.

ما احتمال أن تضم هذه العيّنة 4 طلاب بمعدّل ممتاز ؟

الحل:

لدينا: $N = 28 \ , \ N_1 = 8 \ , \ N_2 = 20$

$n = 6 \ ; (\ n_1 = 2 \ , \ n_2 = 4 \ )$

والاحتمال المطلوب: $P = \dfrac{n!}{n_1!\, n_2!} (\dfrac{N_1}{N})^{n_1} (\dfrac{N_2}{N})^{n_2} = \dfrac{6!}{2!4!} (\dfrac{8}{28})^2 (\dfrac{20}{28})^4$

b- السحب من دون إعادة:

والجواب يكون: $$P = \frac{C_{n_1}^{N_1} \cdot C_{n_2}^{N_2} \cdots C_{n_k}^{N_k}}{C_n^N}$$

لنبيّن ذلك: واضح أنّه يمكن اختيار العنصر الأول من العيّنة بـ N طريقة ، ويمكن اختيار العنصر الثاني من العيّنة بـ $(N-1)$ طريقة

....

ويمكن اختيار العنصر ذي الرقم n من العيّنة بـ $(N-n+1)$ طريقة .

بالتالي وحسب قاعدة الـ $(n . m)$ يكون:

عدد الطرق الممكنة لتشكيل العيّنة الكيفيّة التي حجمها n هو: $N(N-1)(N-2)(N-3)...(N-n+1) = P_n^N$ (وهو يمثل عدد الحالات الكليّة)

بالتالي لدينا: $P_{n_1}^{N_1}$: طريقة لاختيار عيّنة حجمها n_1 من النوع A_1

$P_{n_2}^{N_2}$: طريقة لاختيار عيّنة حجمها n_2 من النوع A_2

....

$P_{n_k}^{N_k}$: طريقة لاختيار عيّنة حجمها n_k من النوع A_k

وبما أنّ عدد طرق تقسيم العيّنة ذات الحجم n إلى k مجموعه هو: $C^{n}_{n_1,n_2,\dots,n_k}$ ، فإنّ عدد طرق تشكيل عيّنة ملائمة لطلبنا وحسب قاعدة الـ (n . m) يكون:

$$C^{n}_{n_1,n_2,\dots,n_k}\, P^{N_1}_{n_1} P^{N_2}_{n_2} \dots P^{N_k}_{n_k}$$

وبالتالي ، الاحتمال المطلوب هو:

$$P=\frac{C^{n}_{n_1,n_2,\dots,n_k} P^{N_1}_{n_1} P^{N_2}_{n_2}\dots P^{N_k}_{n_k}}{P^{N}_{n}}=\frac{C^{N_1}_{n_1} C^{N_2}_{n_2}\dots C^{N_k}_{n_k}}{C^{N}_{n}}$$

وهو: **قانون السحب من دون إعادة.**

مثال:

وُزِّعَ ورق اللعب على 4 لاعبين، ما احتمال أن يكون لدى لاعب معيّن: 6 صور و 3 آسات ؟

الحل:

لدينا: $N=52$ (أوراق اللعب) ، $N_1 = 12$ (صوره) ، $N_2 = 4$ (آسات)،

$N_3 = 36$ (أوراق أخرى) ، $n = 13$ (حجم العيّنة) ، $n_1 = 6$ (صور)،

$n_2 = 3$ (آسات) ، $n_3 = 4$ (أوراق أخرى).

بالتالي:

$$P=\frac{C^{N_1}_{n_1} C^{N_2}_{n_2} C^{N_3}_{n_3}}{C^{N}_{n}}=\frac{C^{12}_{6} C^{4}_{3} C^{36}_{4}}{C^{52}_{13}}=\frac{3267}{9529015}\approx 0.0003$$

وهو المطلوب.

(3 . 8) تمارين الفصل الثالث:

1- إذا كان: $P(\overline{A}B) = 0.1$, $P(A\overline{B}) = 0.4$, $P(\overline{AB}) = 0.6$

أوجد: $P(A)$, $P(B)$, $P(A \cup B)$, $P(\overline{A} \cup B)$

2- ثلاث عائلات: A , B , C دُعيت للعشاء إلى مكانٍ ما، فإذا علمت أنَّ احتمال حضورهم للعشاء كان على النحو التالي: $P(A) = 0.8$, $P(B) = 0.6$, $P(C) = 0.9$

المطلوب:

1 . ما احتمال أن تحضر العائلات الثلاث إلى العشاء ؟

2 . ما احتمال ألاَّ تحضر أي من العائلات الثلاث ؟

3 . ما احتمال أن تحضر عائلة واحدة على الأقل ؟

3 . لنأخذ طلاب أحد المقرَّرات ولنفرض أنَّ:

40% من الطلبة يحصلون على درجات جيِّده في المقرَّر .

60% من الطلبة يثابرون على الدراسة.

30% من الطلبة يثابرون على الدراسة ودرجاتهم جيِّده في المقرَّر .

اخترنا طالباً بشكلٍ عشوائي:

المطلوب:

1 . ما احتمال أن تكون درجة هذا الطالب إما جيِّده أو أنَّه مثابر على الدراسة؟

2 . إذا علمت أنَّ درجته كانت جيِّده، فما احتمال أن يكون مثابراً على الدراسة ؟

4 - إذا كان: $P(A) = P(B) = \dfrac{1}{2}$, $P(A \cup B) = \dfrac{2}{3}$

المطلوب:

1 . بيِّن ما إذا كان A و B مستقلين أم لا ؟

2 . احسب الاحتمالات: $P(\overline{A}B)$, $P(\overline{AB})$, $P(A \setminus B)$

3 . بيِّن أنَّ: $P(AB) = P(\overline{AB})$

5 – صندوقان: يحوي الأول (5 كرات بيضاء و 3 كرة سوداء)، ويحوي الثاني (2 كرة بيضاء و 2 كرة سوداء) .

سُحِبَت عشوائيّاً كرة من الصندوق الأول ووضعت في الثاني، ثمّ سُحِبَت كرة من الصندوق الثاني وبشكلٍ عشوائي أيضاً.

المطلوب:

1 . ما احتمال أن تكون الكرة المسحوبة من الصندوق الثاني بيضاء ؟

2 . إذا علمت أنَّ الكرة المسحوبة من الصندوق الثاني كانت بيضاء ، ما احتمال أن تكون الكرة المسحوبة من الصندوق الأول بيضاء أيضاً ؟

6 – يتقدّم طالب للامتحان بـ 5 مقرّرات، فإذا كان احتمال نجاحه في كل مقرّر هو $\frac{2}{3}$، احسب احتمال نجاحه بأربع مقرّرات على الأقل ؟

7 – صندوق يحوي: 6 كرات بيضاء و 4 حمراء (نفترض دوماً أن الكرات متماثلة). سحبنا على التوالي وبشكل عشوائي كرتين دون إعادة.

المطلوب:

1 . ما احتمال أن تكون الكرتين المسحوبتين من اللون الأبيض ؟

2 . ما احتمال كون الكرة الثانية من اللون الأحمر ؟

8 - لدى A و B إحدى عشر قطعة حلوى صغيرة. أكل A أربع قِطع منها، وأكل B ست قِطع، ولديهم كلب أكل القطعة الباقية، ولكن تبيَّن أخيراً أن ثلاث قطع من القطع المأكولة كانت فاسدة.

المطلوب:

1 . ما احتمال ألا يكون قد تسمَّم الكلب ؟

2 – ما احتمال أن يكون A و B قد تسمّما معاً شريطة الكلب لم يتسمّم ؟

3 . ما احتمال أن يكون A و B قد تسمّما معاً، والكلب لم يتسمّم ؟

4 . ما احتمال أن يكون قد تسمّم الثلاثة ؟

9 - وعاء يحوي: M كرة بيضاء و N - M كره سوداء، نسحب n كرة (دون إعادة).

المطلوب:

1 . ما احتمال أن تظهر كرة بيضاء في السحب k ؟

2 . ما احتمال أن تظهر كرتان بيضاويتان في السحبين k و L .

3 . ما احتمال أن تظهر m كرة في نهاية السحب ؟

10 . أعد طلبات المسألة السابقة بافتراض أنّ السحب يتم مع إعادة.

11- في تجربة إلقاء قطعة نرد غير منتظمة لمرة واحدة . إذا كان احتمال ظهور أي رقم يتناسب مع مربع هذا الرقم.

المطلوب:

1 . أوجد الفراغ الاحتمالي للتجربة.

2 . ما احتمال ظهور رقم فردي ؟

3 . ما احتمال ظهور رقم أقل من 3 ؟

....

الفصل الرابع

المتغيّرات العشوائيّة والتوزيعات الاحتماليّة

(4-1) المتغيّر العشوائي: *Random Variable:*

ليس بالضرورة أن تكون نتائج التجربة العشوائيّة Ω ممثّلة دائماً بأعداد (علماً أنّه يمكن ذلك اصطلاحاً) حيث تختلف نتائج التجربة دوماً باختلاف التجربة ذاتها، وبالحالة العامة: يمكن دوماً نقل نتائج التجربة إلى مجموعة عددية ، وبالتالي يُعرَّف المتغيّر العشوائي X بأنّه التطبيق: $X : \Omega \to R$ بحيث تكون الصورة العكسيّة لأي مجال من مجموعة الأعداد الحقيقيّة R حدثاً ينتمي إلى Ω .

(4-2) تابع التوزيع:

إذا كان X متغيّراً عشوائياً ، فإنّه في كثيرٍ من الأحيان يُطلب معرفة قيمة الاحتمال $P(X \leq x)$.

لنفرض أنّنا بصدد تجربة إلقاء قطعة النرد حيث: $\Omega = \{1, 2, 3, 4, 5, 6\}$

وبالتالي إذا كان:

x عبارة عن ظهور الحدث $\{1\}$ فإنّ: $P(X \leq x) = \dfrac{1}{6}$

x عبارة عن ظهور الحدث $\{1, 2\}$ فإنّ: $P(X \leq x) = \dfrac{1}{3}$

x عبارة عن ظهور الحدث $\{1, 2, 3\}$ فإنّ: $P(X \leq x) = \dfrac{1}{2}$

أي أنّ: $P(X \leq x)$ يتغيّر مع تغيُّر x، وبالتالي يكون تابعاً لـ x، ولذلك نرمز له بالرمز: $F(x) = P(X \leq x)$ ويسمّى تابع توزيع المتغيّر العشوائي X .

- **خواص تابع التوزيع:**

1- $0 \leq F(x) \leq 1$

2- $\forall\, a, b \in R : a \leq b \Rightarrow F(a) \leq F(b)$

حيث الحدث: $X \le b$ يتضمّن الحدث: $X \le a$

بمعنى: $(X \le a) \subset (X \le b)$

بالتالي: $P(X \le a) \le P(X \le b)$

ومنه: $F(a) \le F(b)$ وبالتالي $F(x)$ غير متناقص .

3 - $P(a < X \le b) = F(b) - F(a)$

لدينا: $(X \le b) = (X \le a) \cup (a < X \le b)$

وبما أنّ الحدثين: $(a < X \le b)$, $(X \le a)$ متنافيان فإنّ:

$$P(X \le b) = P(X \le a) + P(a < X \le b)$$

$$F(b) - F(a) = P(a < X \le b)$$

4- كذلك: $F(-\infty) = 0$, $F(+\infty) = 1$ لأنّ:

$$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \le x) = P(X < -\infty) = 0$$

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} P(X \le x) = P(X < \infty) = 1$$

تُصنّف المتغيّرات العشوائيّة إلى نوعين: منفصلة (متقطعة) ومستمرة.

(3-4) المتغيّرات العشوائيّة المنفصلة *Discrete Random Variables* :

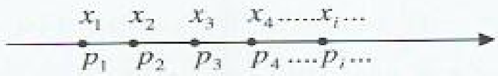
. التابع الاحتمالي وتابع التوزيع:

تعريف:

نقول عن المتغيّر العشوائي X (أو عن تابع توزيعه) إنّه من النوع المنفصل إذا كانت قيمه معدودة (أي: كان عدد قيمه الممكنة محدوداً، أو أن يكون غير محدود إلا أنّه معدود).

وباختصار نكتب: $X : x_1 , x_2 , x_3 , \dots , x_n$

فإذا رمزنا بالرمز: $P_i = P(X = x_i) \; ; \; i = \overline{1, n}$ وقرنّا كل نقطه x_i من نقاط محور الإحداثيات بالكتلة النقطيّة P_i حيث: $P_1 + P_2 + \dots + P_i + \dots = 1$ كما في الشكل:



عندئذٍ تكون الكتلة الكليَّة الموزعة على طول المحور مساوية للواحد (تمثيل فيزيائي)، في حين نجد أنَّ: $P[x_l < X < x_j] = \sum_{x_l \le x_i \le x_j} P_i$ ممثّلاً لمجموع الكتل الواقعة بين النقطتين : x_l و x_j .

وعندها نسمّي الجدول التالي:

X	x_1	x_2	x_3	...	x_n
P_i	p_1	p_2	p_3	...	p_n

بجدول التوزيع الاحتمالي للمتغيِّر العشوائي X ، حيث: P_i يحقّق الشرطين:

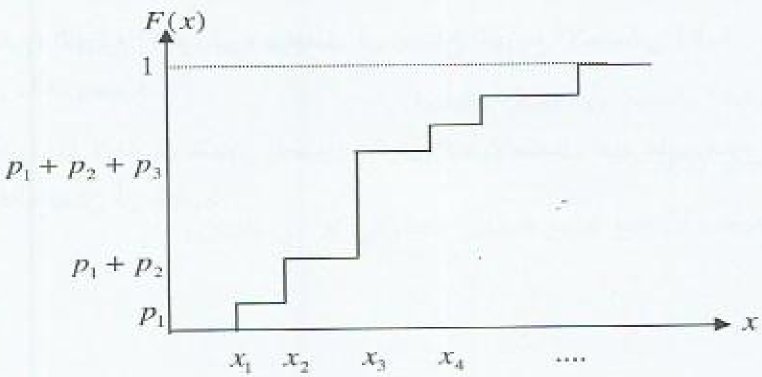
$$\forall i \;:\; P_i \ge 0 \quad , \quad \sum_{i=1}^{n} P_i = 1$$

وهو يرمز في الصورة الميكانيكيَّة إلى كمية الكتلة الموجودة في النقطة x_i .

وعندئذ يعرّف تابع توزيع المتغيِّر العشوائي X بالشكل:

$$F(x)=P(X \le x)=\sum_{x_i \le x} P_i = \begin{cases} 0 & , \quad \forall : -\infty \le x < x_1 \\ P_1 & , \quad \forall : \; x_1 \le x < x_2 \\ P_1 + P_2 & , \quad \forall : \; x_2 \le x < x_3 \\ \cdots\cdots\cdots & \\ P_1 + P_2 + \ldots + P_i = 1 \; , & \forall : x_i \le x < \infty \end{cases}$$

وبرسم بيان التابع $F(x)$ نجد أنّه درجي الشكل:



مثال (1):

لنفرض أنَّ التجربة العشوائيّة هي رمي قطعتي نقود . إذا رمزنا بـ X لعدد مرات ظهور الصورة ، وباعتبار أنَّ: $P_i = P[X = x_i]$ يكون لدينا التوزيع الاحتمالي التالي:

$$X : 0 , 1 , 2$$

$$P_1 = P(X = 0) = 1/4$$
$$P_2 = P(X = 1) = 2/4$$
$$P_3 = P(X = 2) = 1/4$$

أما تابع التوزيع فيأخذ الشكل التالي: $F(x) = P(X \leq x) = \sum_{x_i \leq x} P_i$

وقيمه على الشكل التالي:

$$F(0) = P(X \leq 0) = P_0 = 1/4$$
$$F(1) = P(X \leq 1) = P_0 + P_1 = P(X = 0) + P(X = 1) = 3/4$$
$$F(2) = P(X \leq 2) = P_0 + P_1 + P_2 = 1$$

وهكذا من معرفة P_i يمكن معرفة $F(x)$ وبالعكس، حيث نلاحظ أنَّ العكس يُحسَب كما يلي:

$$P_0 = F(0) = 1/4$$
$$P_1 = F(1) - F(0) = 2/4$$
$$P_2 = F(2) - F(1) = 1/4$$

وتبقى العلاقة: $P_i = P(X = x_i) = F(x_i) - F(x_{i-1})$ صحيحة .

تسمّى: P_i (دالّة أو تابع) التوزيع الاحتمالي، ويفضّل أن يصاغ التوزيع الاحتمالي (كلما أمكن ذلك) على شكل دالّة رياضيّة:

في مثالنا السابق: إذا رمزنا لاحتمال ظهور الصورة بالرمز P ، ولاحتمال عدم ظهورها بالرمز $q = 1 - p$ ، فإنّه يمكن أن نكتب:

$$P_0 = P(X=0) = C_0^2\ p^0\ q^2 = 1.1.(\frac{1}{2})^2 = \frac{1}{4}$$

$$P_1 = P(X=1) = C_1^2\ p.q = 2.(\frac{1}{2})(\frac{1}{2}) = \frac{2}{4}$$

$$P_2 = P(X=2) = C_2^2\ p^2\ q^0 = 1.(\frac{1}{2})^2.1 = \frac{1}{4}$$

وبالحالة العامة: إذا رمينا n من قطع النقود في تجربتنا السابقة، فإنّ دالّة الاحتمال تأخذ الشكل: $P_i = C_i^n\ p^i\ q^{n-i}$ وسنتعرّف لاحقاً على هذا التوزيع .

مثال (2):

إذا كانت القيم الممكنة للمتغيّر العشوائي X من الشكل: $X : a , b , c$

حيث: $a < b < c$ ، وكان: $P(X = x_i) = \frac{1}{3} , \forall\ x_i$ عندئذ يكون جدول X من الشكل:

X	a	b	c
P_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

أما تابع التوزيع فيأخذ الشكل التالي:

$$F(x) = \begin{cases} 0 , -\infty < x < a \\ \frac{1}{3} , a \le x < b \\ \frac{2}{3} ,b \le x < c \\ 1 , c \le x < \infty \end{cases}$$

وهكذا بالنسبة لأية مسألة مشابهة.

مثال (3):

بفرض أنّ تابع توزيع المتغيّر العشوائي X من الشكل:

$$F(x)=\begin{cases}0 & ;x<1\\ \frac{1}{3}+\frac{2}{3}(x-1) & ;1\le x\le 2\\ 1 & ;x>2\end{cases}$$

أوجد:

$$P(\frac{4}{3}<X<\frac{5}{3})\ ,\ P(X>\frac{3}{2})\ ,\ p(X\le\frac{5}{3})$$

الحل:

$$P(X\le\frac{5}{3})=F(\frac{5}{3})=\frac{1}{3}+\frac{2}{3}(\frac{5}{3}-1)=\frac{7}{9}$$

$$p(X>\frac{3}{2})=1-P(X\le\frac{3}{2})=1-F(\frac{3}{2})=\frac{1}{3}$$

$$P(\frac{4}{3}<X<\frac{5}{3})=F(\frac{5}{3})-F(\frac{4}{3})=\frac{2}{9}$$

(4-4) المتغيّرات العشوائيّة المستمرة *Continues Random Variables*:

. دالّة الكثافة وتابع التوزيع:

تعريف:

نقول عن المتغيّر العشوائي X أو عن تابع توزيعه إنّه من النوع المستمر إذا كان X يأخذ قيمه بشكل مستمر/ بمعنى آخر: إذا كانت واحدة الكتل في الصورة الميكانيكيّة (التمثيل الفيزيائي) موزّعة على طول محور الإحداثيات وبكثافة مقدارها $f(x)$ في النقطة x /.

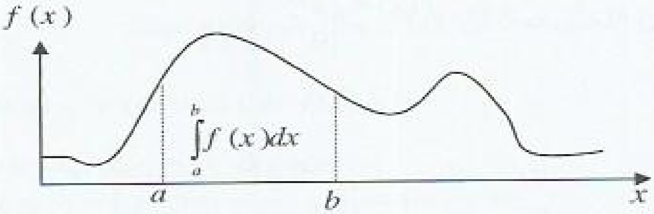
يسمّى $f(x)$ تابع كثافة المتغيّر العشوائي X ، وهو يحقّق الشرطين التاليين:

. $f(x)\ge 0\ ,\ \forall\ x\in R$ ومستمر

. $\int_{-\infty}^{\infty} f(x)dx=1$

وكل تابع حقيقي يحقّق الشرطين السابقين، يُحدِّد توزيعاً مستمراً لـ X الذي كثافته $f(x)$ ،

والتكامل $\int_a^b f(x)dx$ يمثل المساحة تحت المنحني للتابع $f(x)$ بين a و b كما في الشكل:



وهذه المساحة تمثل الاحتمال: $P(a < X < b) = \int_a^b f(x)dx$ ، وبالتالي تابع التوزيع الذي عرّفناه سابقاً بالشكل العام : $F(x) = P(X \leq x)$ سيُمثّل في هذه الحالة بالمساحة تحت المنحني للتابع $f(x)$ والواقعة إلى يسار النقطة x .

أي:
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$

وحسب النظرية الأساسيّة في الحساب التكاملي يكون: $f(x) = \frac{d}{dx} F(x) = F'(x)$

ملاحظه:

إذا كان X مستمراً فإنّ:
$$P(X = x) = P(x \leq X \leq x) = \int_x^x f(x)\,dx = 0$$

وهذا متوقع، لأنه لا توجد مساحة عند نقطة، ويكون الاحتمال في هذه الحالة صفراً.

وبالتالي: $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

بمعنى: الإشارتان /أصغر، أصغر أو يساوي/ متكافئتان في حالة التوزيعات المستمرة.

نشير أخيراً أنّ الكمية $f(x)$ ليست احتمالاً، وإنما تمثّل كثافة احتمال، وهذا المفهوم يطابق تماماً مفهوم كثافة الكتلة التي لا تمثّل كتلة في لغة الميكانيك، في حين:

$P(x < X < x + dx) = f(x)\,dx$ تمثّل احتمالاً لا متناهياً في الصغر يسمّى عنصر الاحتمال.

مثال (1):

بفرض أنّ:
$$f(x) = \begin{cases} \dfrac{a}{e^{2x}} & ; x > 0 \\ 0 & ; x \le 0 \end{cases}$$

المطلوب:

1. أوجد قيمة الثابت a حتى يكون $f(x)$ تابع كثافة.
2. أوجد: $F(x)$ ، ثم احسب الاحتمال: $P(1 \le X \le 2)$.

الحل:

لدينا:
$$1 = \int_{-\infty}^{\infty} f(x)dx = a\int_{0}^{\infty} e^{-2x}dx = \frac{a}{2} \Rightarrow a = 2$$

$$F(x) = \int_{-\infty}^{x} f(x)dx = 2\int_{0}^{x} e^{-2x}dx = -[e^{-2x}]_0^x = 1 - e^{-2x}$$

$$P(1 \le X \le 2) = F(2) - F(1) = (e^{-2} - e^{-4})$$

مثال (2):

ليكن X متغيّراً عشوائياً مستمراً تابع توزيعه من الشكل:

$$F(x) = c\,x^3 \;\; ; \; 0 \le x \le 3$$

المطلوب:

1. أوجد: تابع الكثافة $f(x)$ / مع حساب قيمة الثابت c /.
2. احسب: $P(2 \le X \le 3)$.

الحل:

1. نعلم أنّ: $f(x) = F'(x) = 3c\,x^2$

ومن ناحية أخرى لدينا: $a \le X \le b \Rightarrow F(b) = 1 \,,\ \forall\ x \ge b$

بالتالي:

$$F(3)=27c=1 \Rightarrow c=\frac{1}{27}$$

$$\Rightarrow f(x)=\frac{1}{9}x^2$$

ويمكن التحقُّق فعلاً أنَّ $f(x)$ يمثل تابع كثافة حيث:

$$\int_0^3 f(x)\,dx \equiv 1\ ,\ \forall x\in[0\,,3]\ \&\ f(x)\geq 0$$

.2
$$P(2\leq X\leq 3)=\int_2^3 f(x)dx=\frac{1}{27}\,x^3\}_2^3=1-\frac{8}{27}=\frac{19}{27}$$

أو مباشرة:
$$P(2\leq X\leq 3)= \;=F(3)-F(2)=1-\frac{8}{27}=\frac{19}{27}$$

مثال (3):

بفرض تابع كثافة المتغيِّر العشوائي المستمر X من الشكل:

$$f(x)=\begin{cases} k(x^2+2x+2)\ ; & x\in[0\,,4] \\ 0 & ; x\notin[0\,,4]\end{cases}$$

أوجد: $F(x)$ ، $F(2)$ ، $P(1\leq X\leq 3)$.

الحل:

لحساب $F(x)$ يلزم حساب الثابت k حيث:

$$\int_0^4 f(x)dx=1\ ,\Rightarrow k=\frac{3}{136}$$

$$\Rightarrow F(x)=\int_{-\infty}^{x} f(x)\,dx=\int_0^x f(x)\,dx=....=\frac{3}{136}(\frac{x^3}{3}+x^2+2x)$$

$$F(2)=\frac{3}{136}(\frac{8}{3}+4+4)=\frac{4}{17}$$

$$P(1\leq X\leq 3)=F(3)-F(1)=\frac{72}{136}-\frac{10}{136}=\frac{62}{136}$$

أو بالشكل:
$$=\int_1^3 f(x)dx=...=\frac{62}{136}$$

(4-5) توزيع متغيّر عشوائي تابع لمتغيّر عشوائي آخر:

بفرض: X متغيّر عشوائي معلوم التوزيع ، Y متغيّر عشوائي آخر مجهول التوزيع يرتبط مع X بالعلاقة القابلة للعكس: $Y=\varphi(X)\ ;\ X=\psi(Y)$.

عندئذٍ لمعرفة توزيع Y نميّز حالتين:

a) . X منفصل:

أي إذا كانت قيم X الممكنة من الشكل : $X : x_1\ ,\ x_2\ ,\dots,\ x_n$

وكان: $Y=\varphi(X)\ ;\ X=\psi(Y)$ بحيث كل قيمة لـ Y تقابلها قيمة وحيدة لـ X . عندئذٍ يكون لـ Y توزيعاً منفصلاً بالشكل:

$$Y : y_1\ ,\ y_2\ ,\dots,\ y_i\ ,\dots\quad ;\ y_i=\varphi(x_i)$$

ويكون: $P(Y=y_i)=P(\varphi(X)=y_i)=P(X=\psi(Y))=P(\psi(y_i))$

مثال:

إذا كان: $X : 0\ ,\ 1\ ,\ 2\ ,\ 3\ \ ;\ P(X=x_i)=\dfrac{1}{4}\ ,\ (i=\overline{0\ ,\ 3})$

وكان: $Y=|X-1|$. أوجد: (جدول و تابع) توزيع Y .

الحل:

واضح أنّ جدول توزيع X من الشكل:

X	0	1	2	3
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

وقيم Y الممكنة هي: $Y : 0\ ,\ 1\ ,\ 2$

بالتالي:

$$P(Y=0)=P(X=1)=\frac{1}{4}$$

$$P(Y=1)=P(X=0)+P(X=2)=\frac{2}{4}$$

$$P(Y=2)=P(X=3)=\frac{1}{4}$$

وجدول توزيع Y من الشكل:

Y	0	1	2
	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

وإذا رمزنا لتابع توزيع Y بالرمز $G(y)$ ، فإنّ:

$$G(y)=P(Y\le y)=\begin{cases} 0 & , -\infty \le y<0 \\ \frac{1}{4} & , \ 0\le y<1 \\ \frac{3}{4} & , \ 1\le y<2 \\ 1 & , \ 2\le y<\infty \end{cases}$$

b) . إذا كان X مستمراً وتابع كثافته $f(x)$:

عندئذٍ من أجل: $Y=\varphi(X)\ ;\ X=\psi(Y)$ (ψتابع عكسي لـ φ)، فإنّه يكون لـ Y توزيع مستمر وتابع كثافته يعطى بالعلاقة التالية:

$$g(y)= f(\psi(y)).\left|\psi'(y)\right|$$

للبرهان على ذلك نميِّز حالتين:

الحالة الأولى : ψ متزايد:

عندئذٍ: $$\psi'(y)>0 \Rightarrow \psi'(y)=\left|\psi'(y)\right|$$

وإذا رمزنا لتابع توزيع Y بالرمز $G(y)$، فإنّه حسب تعريف تابع التوزيع يكون:

$$G(y)=P(Y\le y)=P(\varphi(X)\le y))=P(X\le\psi(y))=\int_{-\infty}^{\psi(y)}f(t)\,dt$$

وباشتقاق الطرفين بالنسبة لـ y (حسب النظرية الأساسيّة للتكامل) يكون:

$$g(y)=G'(y)=f(\psi(y)).\left|\psi'(y)\right|$$

الحالة الثانية: ψ متناقص:

عندئذٍ: $$\psi'(y)<0 \quad \psi'(y)\ \left|\psi'(y)\right|=--\Rightarrow$$

ومنه يمكن أن نكتب:

$$G(y)=P(Y\leq y)=P(X\geq\psi(y))=1-P(X\leq\psi(y))=1-\int_{-\infty}^{\psi(y)}f(t)\,dt$$

وباشتقاق الطرفين بالنسبة لـ y يكون: $g(y)=G'(y)=f(\psi(y)).\left|\psi'(y)\right|$

(وتبقى العلاقة صحيحة أياً كان ψ متزايداً أم متناقصاً).

ملاحظة:

يمكن كتابة العلاقة السابقة أيضاً بالشكل: $g(y)=f(x)\left|\frac{dx}{dy}\right|_{x\to y}$

حيث:

1 . نأخذ القيمة الموجبة للمشتقة التفاضلية الأولى $\frac{dx}{dy}$ (وذلك من العلاقة التي تربطهما).

2 . نضرب الناتج السابق بـ $f(x)$.

3 . نعوض قيمة x بدلالة y / وتُحسَب حدود y من العلاقة التي تربط x بـ y /.

مثال (1):

إذا كان تابع كثافة X من الشكل: $f(x)=\begin{cases}\frac{1}{2a} & ; X\in[-a\ ,\ a]\\ 0 & ; X\notin[-a\ ,\ a]\end{cases}$

أوجد تابع كثافة Y في الحالتين:

1 . إذا كان: $y=\frac{x+a}{2a}$

2 . إذا كان: $y=\frac{x+a}{x-a}$

الحل:

(1 $\quad y=\frac{x+a}{2a}\Rightarrow dy=\frac{dx}{2a}\ ;\ \frac{dx}{dy}=2a$

ومنه: $$g(y) = f(x)\left|\frac{dx}{dy}\right|_{x\to y} = \frac{1}{2a}.2a = 1 \; ; 0 \le Y \le 1$$

2) $$y = \frac{x+a}{x-a} \Rightarrow dy = \frac{(x-a)-(x+a)}{(x-a)^2}dx = \frac{-2a}{(x-a)^2}dx \; ; \frac{dx}{dy} = \frac{(x-a)^2}{-2a}$$

بالتالي: $$g(y) = f(x)\left|\frac{dx}{dy}\right|_{x\to y} = \frac{1}{2a}.\frac{(x-a)^2}{2a} = \frac{(x-a)^2}{4a^2}$$

وبحساب x بدلالة y من العلاقة: $y = \dfrac{x+a}{x-a}$ ، ينتج : $x - a = \dfrac{2a}{y-1}$

وأخيراً، وبالتبديل في تابع كثافة Y يكون: $$g(y) = \frac{1}{(y-1)^2} \; ; 0 \le Y \le \infty$$

مثال (2):

بفرض: X متغيّر عشوائي مستمر تابع كثافته $f(x)$.

إذا كان: $Y = aX + b \; ; a,b \in R$

أوجد: $g(y)$ (تابع كثافة Y).

الحل:

لدينا: $$g(y) = f(\psi(y)).\left|\psi'(y)\right|$$

ولكن: $$Y = aX + b \equiv \varphi(X) \Rightarrow X = \frac{Y-b}{a} \equiv \psi(Y)$$

إذن: $$g(y) = f\left(\frac{Y-b}{b}\right).\frac{1}{|a|}$$

- التحويل: $Y = X^2$ $.Transforman$

لنفرض أنّ العلاقة بين X و Y من الشكل: $Y = X^2$ ، واضح أنّ التابع في هذه الحالة ليس وحيداً (لكل قيمة لـ y توجد قيمتان لـ x)، وفي هذه الحالة يمكن إيجاد تابع كثافة Y من العلاقة التالية: $$g(y) = f(x)\left|\frac{dx}{dy}\right|_{x\to y} + f(-x)\left|\frac{dx}{dy}\right|_{x\to y}$$

وإذا كانت $f(x)$ داله متماثلة، أي: $f(x) = f(-x)$ ، فإنّ العلاقة السابقة تُكتب بالشكل:

$$g(y) = 2f(x)\left|\frac{dx}{dy}\right|_{x\to y}$$

مثال (3):

بفرض: $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \; ; -\infty \le x \le \infty$

والمطلوب:

1 . إذا كان: $Y = X^2$. أوجد: $g(y)$ (تابع كثافة Y).

2 . ومن أجل: $Z = \frac{1}{2} X^2$ أوجد: $f(z)$ (تابع كثافة Z).

الحل:

1) واضح أنَّ $f(x)$ متماثلة، وبالتالي يمكن أن نكتب مباشرة:

$$g(y) = 2f(x)\left|\frac{dx}{dy}\right|_{x\to y}$$

حيث: $y = x^2 \Rightarrow \frac{dx}{dy} = \frac{1}{2x} \; ; x = \sqrt{y}$

ومنه: $g(y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-y}{2}} y^{\frac{-1}{2}} \; , 0 \le y \le \infty$

2) كذلك من أجل: $z = \frac{1}{2}x^2 \Rightarrow \frac{dx}{dz} = \frac{1}{x} \; ; x = \sqrt{2z}$

سينتج: $f(z) = \frac{1}{\sqrt{\pi}} e^{-z} z^{-\frac{1}{2}}$ وهو تابع كثافة Z (تحقَّق من ذلك).

(مع ملاحظة أننا سنعود إلى هذا المثال لاحقا في أثناء دراسة التوزيعات المستمرة).

(4-6) الصفات العددية المميّزة للمتغيّرات العشوائيّة:

نعلم أنَّ المتغيّر العشوائي إما أن يكون منفصلاً، وإما أن يكون مستمراً، والصفات العدديّة (تسمى أحياناً بالصفات المميّزة أو بالمقاييس العدديّة) تتبع هذين النوعين من المتغيّرات

وهي تعطي وصفاً بسيطاً وسريعاً لملامح التوزيع بحيث تمكّننا من تمييز توزيع عن آخر بشكلٍ عام ، ومن هذه المقاييس نتعرّف إلى ما يلي:

1 ـ التوقع الرياضي: *Math - Expectation*

نُذَكّر أنّه يكون للمتغيّر العشوائي X توزيعاً احتماليّاً (تابع احتمال) P_i إذا كان X منفصلاً، أو تابع كثافة $f(x)$ إذا كان X مستمراً.

بفرض X متغيّر عشوائي . عندئذٍ يعرّف التوقع الرياضي لـ X بالشكل:

$$E(X) = \sum_i x_i P_i \quad ; \quad X \text{ متغيّر منفصل}$$

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx \quad ; \quad X \text{ متغيّر مستمر}$$

علماً أنّ معظم نقاط التوزيع تميل إلى التمركز حول التوقع الرياضي.

ـ خواص التوقع الرياضي:

a - $E(C) = C \;;\; C$ ثابت

أي بما أنّ X يأخذ قيمة واحدة فقط فإنّ: $P(X = C) = 1$

بالتالي: $E(C) = C.1 = C$

b - $E(CX) = C.E(X)$

c - $\forall X \in [a, b] \Rightarrow E(X) \in [a, b]$

$$\forall x_i \in [a, b];\; i = \overline{1, n} \;\Rightarrow\; a \le x_i \le b$$

بالتالي :

$$a.\sum_i P_i \le \underbrace{\sum_i x_i P_i}_{E(X)} \le b\sum_i P_i \;;\; \sum_i P_i = 1$$

$$\Rightarrow a \le E(X) \le b \Rightarrow E(X) \in [a, b]$$

والبرهان مشابه إذا كان X مستمراً.

d - $E(X+Y) = E(X) + E(Y)$ (شريطة وجود التوقع الرياضي لكلٍ منهما).

e - من أجل : $Y = \varphi(X)$ يكون:

إذا كان X منفصلاً: $E(Y) = E(\varphi(X)) = \sum_i \varphi(x_i).P_i$

إذا كان X مستمراً: $or: = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\, dx$

علماً أنّ شرط وجود $E(X)$ مرهوناً بتقارب السلسلة $\sum_i x_i\, P_i$ أو بتقارب التكامل $\int_{-\infty}^{+\infty} f(x)\, dx$.

مثال (1):

إذا كان: $X \;:\; x_i \equiv 3^i \;,\; i = 1, 2, 3, \ldots$ ، وكان: $P_i = P(x_i) = (\frac{1}{2})^i$ ، عندئذٍ السلسلة: $\sum_i x_i\, P_i = \sum_i (\frac{3}{2})^i$ غير متقاربة ، وبالتالي لا يمكن إيجاد $E(X)$ في هذه الحالة.

مثال (2):

إذا كان لـ X تابع كثافة من الشكل: $f(x) = \frac{1}{\lambda(1+x^2)} \;;\; x \in R$

عندئذٍ التكامل: $\int_{-\infty}^{+\infty} x\, f(x)\, dx$ غير متقارب، وبالتالي لا يمكن حساب $E(X)$ أيضاً.

حالة خاصة:

إذا كان: $X : x_1 , x_2 , x_3 , \ldots , x_n$

وكان: $P(X = x_1) = P(X = x_2) = \ldots = P(X = x_n) = p$

أي: $P_i = p \;;\; i = \overline{1, n}$

عندئذٍ: $\sum_i P_i = 1 \equiv \underbrace{p + p + \ldots + p}_{n} = n\, p \;\Rightarrow\; p = \frac{1}{n}$

بالتالي: $$E(X)=\sum_{i=1}^{n} x_i\, p_i=\frac{x_1+x_2+\ldots+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n} x_i$$

ويسمّى في هذه الحالة بالمتوسط الحسابي أو المتوسط فقط.

ملاحظة:

عندما لا نستطيع حساب $E(X)$، فإنّنا نقيس التمركز في هذه الحالة بمقياس آخر يسمّى بالوسيط أو **(القيمة النصفية لـ X)**:

وهي تعريفاً النقطة التي لأجلها يكون:

$$P(X\le x)=P(X\ge x)=\frac{1}{2}\ \ ,\ or: F(x)=\frac{1}{2}$$

. مبرهنة (مبرهنة ماركوف):

بفرض أنّ X متغيّر عشوائي موجب ، توقعه الرياضي $E(X)$ محدود، عندئذٍ:

$$\forall \varepsilon>0: P(X>\varepsilon)\le\frac{E(X)}{\varepsilon}$$

البرهان:

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} xf(x)dx = \\
&= \int_{0}^{\varepsilon} xf(x)dx + \int_{\varepsilon}^{\infty} xf(x)dx \\
&\ge \int_{\varepsilon}^{\infty} xf(x)dx \ge \varepsilon\int_{\varepsilon}^{\infty} f(x)dx = \varepsilon.P(X>\varepsilon) \\
&\Rightarrow P(X>\varepsilon)\le\frac{E(X)}{\varepsilon}
\end{aligned}$$

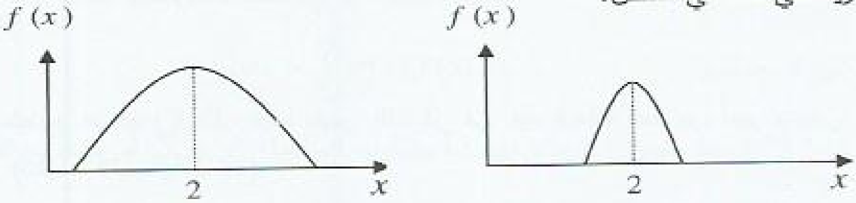
2 . التباين - *Variance*:

نرمز لتباين المتغيّر العشوائي X بالرمز: $\sigma^2(X)$ or $Var(X)$، وإذا فرضنا أنّ: $\mu=E(X)$ عندئذٍ يعرّف التباين كما يلي:

X منفصل: $$\sigma^2(X)\ =\ \sum_i (x_i-\mu)^2\, P_i$$

X مستمر: $$or:\ =\int_{-\infty}^{+\infty}(\,x-\mu\,)^2\, f(x)\,dx$$

وهو يبيّن كيفية انتشار نقاط التوزيع حول توقعه الرياضي، علماً أنّه قد يوجد توزيعان لهما نفس التوقع الرياضي، ويختلفان بكيفية انتشار أو تباعد نقاطهما عن التوقع الرياضي كما في الشكل:



ولكن، وبما أنّ تباعد نقاط أي توزيع عن توقعه الرياضي يقاس بوحدة طول (وليس بوحدة مساحة)، لذلك كان لزاماً علينا دراسة الجذر الموجب للكمية $\sigma^2(X)$، وهو ما يسمّى بالانحراف المعياري، أي: $\sigma(X) = \sqrt{\sigma^2(X)}$

ويمكن أن يكتب التباين بدلالة التوقع الرياضي مباشرة بالشكل:

$$\sigma^2(X) = E(X-\mu)^2 = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2$$

- خواص التباين:

a – إذا كان $X = c$ فإنّ: $\sigma^2(c) = 0 \;;\; c$ ثابت

وذلك لأنّ:

$$\sigma^2(c) = E(c^2) - [E(c)]^2 = 0$$

b - أيضا:

$$\sigma^2(cX) = c^2\,\sigma^2(X)$$

حيث:

$$\begin{aligned}\sigma^2(cX) &= E(cX)^2 - [E(cX)]^2 \\ &= c^2 E(X^2) - c^2[E(X)]^2 \\ &= c^2\{E(X^2) - [E(X)]^2\} = c^2\sigma^2(X)\end{aligned}$$

c -

$$\sigma^2(cX \mp b) = c^2\,\sigma^2(X)$$

d -

$$\sigma^2(X \mp Y) = \sigma^2(X) + \sigma^2(Y) \quad ; \quad X\,,\,Y \text{ مستقلين}$$

e - وإذا كان X و Y غير مستقلين يكون:

$$\sigma^2(X \mp Y) = \sigma^2(X) + \sigma^2(Y) + 2[E(XY) - E(X)E(Y)]$$

وفي حالة خاصة: من أجل: $Z=\dfrac{X-\mu}{\sigma} \Rightarrow \begin{cases} E(Z)=0 \\ \sigma^2(Z)=1 \end{cases}$

وذلك لأنّ:

$$E(Z)=E[\frac{X-E(X)}{\sigma(X)}]=\frac{1}{\sigma(X)}[E(X)-E(X)]=0$$

$$\sigma^2(Z)=\sigma^2[\frac{X-E(X)}{\sigma(X)}]=\frac{1}{\sigma^2(X)}.\sigma^2(X)=1$$

ويسمّى Z في هذه الحالة **بالقيمة المعيارية** لـ X (أو **بالمتغيّر المنتظم** كما يسمّى في بعض المراجع) وكذلك كل متغيّر يحقّق ذات الخاصة.

مثال (1):

إذا كانت التجربة العشوائيّة هي رمي حجر النرد ، ورمزنا بـ X لعدد النقاط التي تظهر على الوجه العلوي لحجر النرد.

عندئذٍ:

$$E(X)=\sum_{i=1}^{6}x_i\,P_i=\frac{1+2+3+4+5+6}{6}=\frac{21}{6}$$

$$E(\frac{3X+4}{7})=E(\frac{3X}{7}+\frac{4}{7})=\frac{3}{7}E(X)+\frac{4}{7}=\frac{3}{7}\;\frac{21}{6}+\frac{4}{7}=\frac{29}{14}$$

$$\sigma^2(X)=E(X^2)-[E(X)]^2=$$

$$=\sum_{i=1}^{6}x_i^2P_i-[E(X)]^2=\frac{1+4+9+16+25+36}{6}-(\frac{21}{6})^2=$$

$$=\frac{91}{6}-\frac{441}{36}=\frac{100}{36}$$

$$\sigma^2(3X+5)=9\sigma^2(X)=\frac{900}{36}$$

مثال (2):

إذا كان: جدول توزيع X من الشكل:

X	-2	2	3
	0.4	0.5	0.1

وكان: $Y = X^2$ ، عندئذٍ يكون جدول توزيع Y من الشكل:

Y	4	9
	0.9	0.1

$$\sigma^2(Y) = E[Y^2] - [E(Y)]^2 = \{4^2(0.9) + 9^2(0.1)\} - \{4(0.9) + 9(0.1)\}^2 = 2.25$$

ولكي نبيِّن أهمية الانحراف المعياري في قياس تباين متغيِّر عشوائي X عن توقعه الرياضي $m = E(X)$ ندرس:

. متراجحة تشيبتشف:

بفرض: X متغيِّر عشوائي توقعه الرياضي m ، وانحرافه المعياري σ.

عندئذٍ: $$P[|X - m| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2} \quad ; \quad \varepsilon \in R^+$$

أي: احتمال أن يأخذ X قيماً تبعد عن توقعه الرياضي بمقدار أكبر أو يساوي ε، هذا الاحتمال محدود من الأعلى، وقيمته أصغر أو تساوي $\frac{\sigma^2}{\varepsilon^2}$.

بمعنى آخر: أنَّ احتمال عدم انتماء X للمجال $[m - \varepsilon , m + \varepsilon]$ هو محدود من الأعلى، وقيمة هذا الاحتمال أصغر أو تساوي $\frac{\sigma^2}{\varepsilon^2}$، وهذا يعني من وجهة نظر ميكانيكيَّة أنَّ مجموع الكتل الواقعة خارج المجال المذكور أصغر أو يساوي النسبة المذكورة.

البرهان:

لدينا: $$|X - m| \geq \varepsilon \Rightarrow (X - m)^2 \geq \varepsilon^2$$

ومنه: $(X - m)^2 f(x) \geq \varepsilon^2 f(x) \; ; \; X$ تابع كثافة $f(x)$

بالتالي:

$$\int_{-\infty}^{m-\varepsilon}(x-m)^2 f(x)dx \geq \varepsilon^2 \int_{-\infty}^{m-\varepsilon} f(x)dx$$

$$\int_{m-\varepsilon}^{m+\varepsilon}(x-m)^2 f(x)dx \geq 0$$

$$\int_{m+\varepsilon}^{\infty}(x-m)^2 f(x)dx \geq \varepsilon^2 \int_{m+\varepsilon}^{\infty} f(x)dx$$

وبالاستفادة من العلاقات السابقة يمكن أن نكتب:

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty}(x-m)^2 f(x)dx \\
&= \int_{-\infty}^{m-\varepsilon}(x-m)^2 f(x)dx + \int_{m-\varepsilon}^{m+\varepsilon}(x-m)^2 f(x)dx + \int_{m+\varepsilon}^{\infty}(x-m)^2 f(x)dx \\
&\geq \varepsilon^2 [\int_{-\infty}^{m-\varepsilon} f(x)dx + 0 + \int_{m+\varepsilon}^{\infty} f(x)dx]
\end{aligned}$$

هذا من ناحية، ومن ناحية أخرى، لدينا:

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{m-\varepsilon} f(x)dx + \int_{m-\varepsilon}^{m+\varepsilon} f(x)dx + \int_{m+\varepsilon}^{\infty} f(x)dx$$

$$\begin{aligned}
\Rightarrow \int_{-\infty}^{m-\varepsilon} f(x)dx + \int_{m+\varepsilon}^{\infty} f(x)dx &= 1 - \int_{m-\varepsilon}^{m+\varepsilon} f(x)dx \\
&= 1 - P[m-\varepsilon \leq X \leq m+\varepsilon] \\
&= 1 - P[-\varepsilon \leq X - m \leq \varepsilon] \\
&= 1 - P[|X-m| < \varepsilon] \\
&= P[|X-m| \geq \varepsilon]
\end{aligned}$$

بالتعويض يكون: $\sigma^2 \geq \varepsilon^2 P[|X-m| \geq \varepsilon]$

ومنه: $P[|X-m| \geq \varepsilon] \leq \dfrac{\sigma^2}{\varepsilon^2}$ وهو المطلوب .

ملاحظـة: من الممكن أيضـاً برهـان متراجحـة تشيبتشيفَ ومباشرةً اعتمـاداً علـى مبرهنـة ماركوف السابقة بالشكل التالي:

بما أنّ: $(X-m)^2 \geq 0$ فإنّ: $P[|X-m|^2 > \varepsilon^2] \leq \dfrac{E[(X-m)^2]}{\varepsilon^2} = \dfrac{\sigma^2}{\varepsilon^2}$

ومنه: $P[|X-m| > \varepsilon] \leq \dfrac{\sigma^2}{\varepsilon^2}$ وهو المطلوب.

ولهذه المتراجحة أشكال أخرى مكافئه، فمثلاً:

إذا فرضنا أنّ: $\varepsilon = k\sigma \;;\; k \in R^+$ فإنّ متراجحة تشيبتشف تأخذ الشكل:

$$P[|X-m| > k\sigma] \leq \frac{1}{k^2} \;\;;\;\; k \in R^+$$

أو الشكل: $P[|X-m| < k\sigma] \geq 1-\dfrac{1}{k^2}$ وجميعها أشكال متكافئة.

وعلى الرغم من أنّ متراجحة تشيبتشف لم تعط تماماً قيمة الاحتمال المطلوب، إلا أنّها تؤكّد أنّه محدود من الأعلى بالمقدار $\dfrac{\sigma^2}{\varepsilon^2}$، ومن هنا تأتي أهميتها.

فمثلاً: من أجل:

$$k = 2 \Rightarrow P[|X-m| \geq 2\sigma] \leq \frac{1}{4}$$

$$or \;:\;\; P[|X-m| < 2\sigma] \geq 1-\frac{1}{4} = \frac{3}{4}$$

بمعنى: $\dfrac{3}{4}$ الكتلة على الأقل يقع في المجال $]m-2\sigma \,,\, m+2\sigma[$.

3 . العزوم المركزيّة والابتدائيّة : *Moments*

تعرّف العزوم بدلالـة التوقـع الرياضـي لمتغيّر عشـوائي، وهـي بـالواقع تطبيقـات مباشـرة للتوقع الرياضي.

فإذا رمزنا بالرمز : $E(X) = m$ ، وكانت c نقطه من المحور الحقيقي، عندئذٍ:

1 . نسمّي الكميّة: $E[(X-c)^i]\ ;\ i=0\ ,1\ ,2\ ,...$ بالعزم المركزي من المرتبة i للمتغيّر X بالنسبة للنقطة c.

2 . من أجل:

$$c=0 \Rightarrow \alpha_i = E(X^i) = \sum_i x^i P_i$$

$$or \quad = \int_{-\infty}^{+\infty} x f(x)\, dx$$

(ويسمّى بالعزم الابتدائي أو الأولي من المرتبة i)

3 . ومن أجل:

$$c=m \Rightarrow \mu_i = E\left[(X-m)^i\right] = \sum_i (x_i-m)^i P_i$$

$$or = \int_{-\infty}^{+\infty} (x-m)^i f(x)\, dx$$

ويسمى بالعزم المركزي من الرتبة i.

مما تقدّم ينتج:

$$\alpha_0 = 1\ ,\ \alpha_1 = E(X)\ ,\ \alpha_2 = E(X^2)\ ,\ \alpha_3 = E(X^3), ...$$

$$\mu_0 = 1\ ,\ \mu_1 = 0\ ,\quad \mu_2 = \sigma^2(X) = \alpha_2 - \alpha_1^2$$

وبحساب μ_3 و μ_4 نحصل على:

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3$$

$$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4$$

ويكتفى عادةً بحساب العزوم الأربعة الأولى سواءً حول المبدأ أو حول التوقع الرياضي، وهي كافية لتعريف المقاييس المختلفة الشائعة الاستخدام، مع ملاحظة أنّ العلاقة بين العزوم المركزيّة والعزوم الابتدائيّة (الأوليّة) يمكن أن تُكتب بالشكل التالي:

$$\mu_r = E\left[(X-\alpha_1)^r\right] = \sum_i \left[(x_i-\alpha_1)^r\right] P_i$$

$$= \sum_i \left[\sum_{k=0}^{r} C_k^r (-\alpha_1)^k x_i^{\,r-k} \right] P_i$$

$$= \sum_{k=0}^{r} (-1)^k\ C_k^r\, \alpha_1^{\,k} \left(\sum_i x_i^{\,r-k}\ P_i \right)$$

$$= \sum_{k=0}^{r} (-1)^k\ C_k^r\, \alpha_1^{\,k}\, \alpha_{r-k}$$

حيث: α_{r-k} العزم الأولى من الرتبة $(r-k)$. بالتالي، وبعد التبديل في العلاقة الأخيرة يكون:

$$r = o \Rightarrow \mu_0 = \alpha_0 = 1$$
$$r = 1 \Rightarrow \mu_1 = \alpha_1 - \alpha_1 = 0$$
$$r = 2 \Rightarrow \mu_2 = \alpha_2 - \alpha_1^{\,2}$$
$$r = 3 \Rightarrow \mu_3 = \alpha_3 - 3\alpha_1\,\alpha_2 + 2\alpha_1^{\,3}$$

وهكذا ... يمكن حساب العزوم المركزيّة من كافة الرتب.

4 . الدالّة المولّدة للعزوم: *Moment Generation Function*

تعريف:

تعرّف الدالّة المولّدة للعزوم/ الدالّة المولّدة للعزوم حول الصفر/ والتي نرمز لها بالرمز: $\Psi_X(t)\ ,\ t \in R$ بالشكل:

$$\Psi_X(t) = E(e^{tX}) = \sum_i e^{tx_i} P_i \qquad ; \quad X \text{ منفصل}$$

$$= \int_{-\infty}^{+\infty} e^{tx} f(x)\,dx \equiv \sum_{i=0}^{\infty} \frac{\alpha_i}{i!} t^i \qquad ; \quad X \text{ مستمر}$$

لنبرهن حالة الاستمرار:

لدينا حسب منشور (ماك . لوران) للتوابع: $\forall\ x : e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$

بالتالي:

$$\Psi_X(t) = E(e^{tX}) = E\left[\sum_{i=0}^{\infty}\frac{t^i}{i!}X^i\right]$$

$$= \sum_{i=0}^{\infty}\frac{t^i}{i!}E(X^i) = \sum_{i=0}^{\infty}\frac{\alpha_i}{i!}t^i$$

علماً أنّـه يمكن البرهان أنّ: $\alpha_i = \Psi_X{}^{(i)}(0)$ ، (حيث : $\Psi_X{}^{(i)}(t)$ هو المشتق من المرتبة i للتابع $\Psi_X(t)$ بالنسبة للمتغيّر t) بالشكل:

من دراسة السلاسل الصحيحة، إذا كان: $f(x) = \sum_{i=0}^{\infty} c_i x^i$ فإنّ الأمثال c_i تعطى بالعلاقة: $c_i = \dfrac{f^{(i)}(0)}{i!}$; $i = 0, 1, 2, \ldots$ وبالمقارنة مع العلاقة السابقة يكون:

$$\frac{\alpha_i}{i!} = \frac{\Psi_X{}^{(i)}}{i!} \Rightarrow \alpha_i = \Psi_X{}^{(i)}(0)$$

ومنها يمكن استنتاج كل العزوم من مختلف الرتب / كما هو واضح من التسمية/.

الآن، إذا كان X منفصلاً:

(سنبيّن دور الرمز t الوارد في الدالّة المولّدة للعزوم من خلال التالي):

لدينا:

$$\Psi_X(t) = E(e^{tX}) = \sum_x e^{tX}P(x)$$

بما أنّ:

$$e^{tX} = \sum_{r=0}^{\infty}\frac{(tx)^r}{r!} = 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \ldots + \frac{(tx)^r}{r!} + \ldots$$

يكون:

$$\Psi_X(t) = \sum_x\left\{1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \ldots + \frac{(tx)^r}{r!} + \ldots\right\}P(x)$$

$$= \sum_x P(x) + \frac{t}{1!}\sum_x x\,P(x) + \frac{t^2}{2!}\sum_x x^2P(x) + \ldots + \frac{t^r}{r!}\sum_x x^rP(x) + \ldots$$

وبملاحظة أنّ:

$$\sum_x P(x) = \sum_x x^0P(x) = \alpha_0 = \frac{t^0}{0!}\sum_x x^0P(x) = 1$$

يمكن أن نكتب: $\Psi_X(t) = \alpha_0 + \frac{t}{1!}\alpha_1 + \frac{t^2}{2!}\alpha_2 + \ldots + \frac{t^r}{r!}\alpha_r + \ldots$

وهذه المعادلة تعطي كل العزوم دفعة واحدة حيث:

α_1 العزم الابتدائي من المرتبة الأولى هو: أمثال $\frac{t}{1!}$

α_2 العزم الابتدائي من المرتبة الثانية هو : أمثال $\frac{t^2}{2!}$

....

α_r العزم الابتدائي من الرتبة r هو: أمثال $\frac{t^r}{r!}$. وهكذا ...

ـ خواص الدالّة المولّدة للعزوم:

a. $t = 0 \Rightarrow \Psi_X{}^{(i)}(0) = 1$

b. $\alpha_i = \Psi_X{}^{(i)}(0)$

c. $Y = aX + b \Rightarrow \Psi_Y(t) = e^{bt}\,\Psi_X(at)$

حيث:

$$\Psi_Y(t) = \Psi_{aX+b}(t) = E\left[e^{atX+bt}\right] = e^{bt}\ E\left[e^{atX}\right] = e^{bt}\,\Psi_X(at)$$

وبالتالي من أجل: $a=1$, $b=-\alpha_1$ يكون:

$$\Psi_{X-\alpha_1}(t) = e^{-t\alpha_1}\,\Psi_X(t) \equiv e^{-t\alpha_1}\sum_i e^{tx_i}P_i = \sum_i e^{t(x_i-\alpha_1)}P_i$$

وهي تعطي الدالّة المولّدة للعزوم حول التوقع الرياضي حيث : $\alpha_1 = E(X)$.

بمعنى آخر: لحساب الدالّة المولّدة للعزوم حول التوقع الرياضي لتوزيع ما، نقوم بحساب الدالّة المولّدة للعزوم $\Psi_X(t)$ حول الصفر، ثمّ نضربها بالمقدار $e^{-t\alpha_1}$.

d. إذا كان: X و Y مستقلين فإنّ: $\Psi_{(X+Y)}(t) = \Psi_X(t).\Psi_Y(t)$ حيث:

$$\Psi_{(X+Y)}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX}.e^{tY}\right] = E(e^{tX}).E(e^{tY}) = \Psi_X(t).\Psi_Y(t)$$

أي: الدالّـة المولّـدة للعـزوم لمجمـوع متغيّرين عشـوائيين مسـتقلين تسـاوي حاصـل ضـرب الدالّة المولّدة للعزوم لكليهما.

وبالحالة العامة:

من أجل n من المتغيّرات المستقلة يكون:

$$\Psi_{\sum_i X_i}(t) = E\left[e^{t\sum_i X_i}\right] = \prod_{i=1}^{n}\Psi_{X_i}(t)$$

وإذا كان لكل من : $X_i\ ;\ i=\overline{1\ ,\ n}$ نفس الدالّة المولّدة للعزوم فإنّ:

$$\Psi_{\sum_i X_i}(t) = \left[\Psi_X(t)\right]^n$$

ومن أجل المتوسط (الوسط الحسابي): مستقلة $X_i\ \ ;\ \ \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

يكون:

$$\Psi_{\overline{X}}(t) = \Psi_{\frac{1}{n}\sum_i X_i}(t) = E\left[e^{\frac{t}{n}\sum_i X_i}\right] = \prod_{i=1}^{n} E\left(e^{\frac{t}{n}X_i}\right) = \prod_{i=1}^{n}\Psi_X\left(\frac{t}{n}\right) = \left[\Psi_X\left(\frac{t}{n}\right)\right]^n$$

مثال (1):

إذا كانت القيم الممكنة للمتغيّر العشوائي المنفصل X من الشكل : $X\ :1\ ,2\ ,3$

بحيث: $P(X=1)=P(X=2)=P(X=3)=a$ ، وكان : $Y=2X$. عندئذٍ:

$$\Psi_Y(t) = \Psi_{2X}(t) = E(e^{2tX}) = (e^{2t}+e^{4t})\ a + e^{6t}\ (1-2a)$$

5 ـ التناظر: *Correspondence*

ذكرنا سابقاً أنّه إذا وجد تناظر فسوف يكون حول التوقع الرياضـي حتماً، وسينطبق مركز الكتل في الصـورة الميكانيكيّة على نقطة التناظر تلك ، لذلك يعرّف التناظر دائماً بالنسبة للتوقع الرياضـي كما يلي:

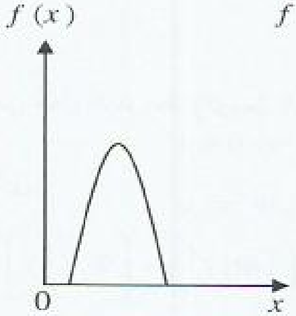
تعريف:

بفرض: X متغيّر عشوائي ما، توقعه الرياضي m . عندئذٍ يكون X متناظراً بالنسبة لـ m إذا كان:

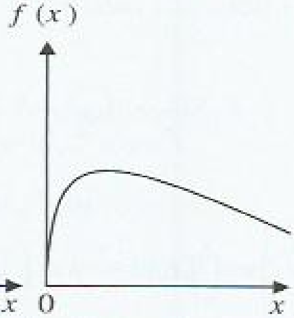
X منفصلاً: $\forall\, x \in R \;:\; P(X = m + x) = P(X = m - x)$

X مستمراً: $\forall\, x \in R \;:\; f(m+x) = f(m-x)$

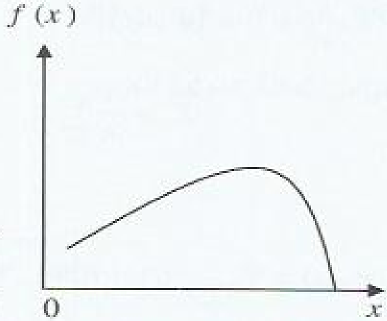
وإذا لم يكن التوزيع متناظراً، فسيكون منحرفاً (أو ملتوياً) إما إلى اليمين وإما إلى اليسار كما في الشكل التالي:



متناظر



غير متناظر (منحرف لليمين)



غير متناظر (منحرف لليسار)

نظرية:

إذا كان X متناظراً بالنسبة لتوقعه الرياضي m ، فإنّ العزوم المركزيّة من المراتب الفرديّة لـ X تكون كلها معدومة ، أي: $\mu_{2i+1} = 0 \;,\; i = 0,1,2,3,...$

البرهان:

لنبرهن النظرية بفرض X متغيّر عشوائي مستمر وتابع كثافته $f(x)$ ، (ويوجد برهاناً مشابهاً إذا كان X منفصلاً):

لدينا:

$$\mu_{2i+1} = E[(X-m)^{2i+1}] = \int_{-\infty}^{\infty} (x-m)^{2i+1} f(x)dx$$

$$= \int_{-\infty}^{m} (x-m)^{2i+1} f(x)dx + \int_{m}^{\infty} (x-m)^{2i+1} f(x)dx$$

نفرض: $x-m=y \Rightarrow dx=dy$

ومنه: $$\mu_{2i+1}=\underbrace{\int_{-\infty}^{0} y^{2i+1}f(y+m)dy}_{*}+\underbrace{\int_{0}^{\infty} y^{2i+1}f(y+m)dy}_{**}$$

مع ملاحظة أنَّ:

$$x\in[-\infty,m]\Rightarrow y\in[-\infty,0]\ ,\ x\in[m,\infty]\Rightarrow y\in[0,+\infty]$$

وإذا بدلنا في ** (أي في التكامل: $\int_{0}^{\infty} y^{2i+1}f(y+m)dy$) كل (y بـ u -) يكون:

$$\int_{0}^{\infty} y^{2i+1}f(y+m)dy=\int_{0}^{-\infty} u^{2i+1}f(m-u)du=-\int_{-\infty}^{0} u^{2i+1}f(m-u)du$$
$$=-\int_{-\infty}^{0} y^{2i+1}f(m-y)dy=-\int_{-\infty}^{0} y^{2i+1}f(y+m)dy$$

بالتالي، وبعد ملاحظة أنَّ: $f(m-y)=f(m+y)$ ، يكون:

$$\mu_{2i+1}=*+**=0\ ,\ i=0,1,2,...$$

. عامل التناظر:

بما أنَّ العزوم المركزيّة الفرديّة الرتبة معدومة ، فإنّه من المنطقي اختيار أحدها كي نقيس به عدم التناظر في التوزيع، وبما أنّ ($\mu_1=0$ دوماً)، فقد اختير μ_3 كقياس لواقع التناظر حول التوقع الرياضي، ولكون μ_3 من الدرجة الثالثة بالنسبة لوحدات المتغيّر X ، فقد جرت العادة على اعتبار الكمية: $\gamma=\frac{\mu_3}{\sigma^3};\sigma\neq 0$ مقياساً لواقع التناظر، أي:

عامل التناظر هو: $\gamma=\frac{\mu_3}{\sigma^3};\sigma\neq 0$. فإذا كان:

- . $\gamma=0$ يكون التوزيع متناظراً بالنسبة لتوقعه الرياضي $m\equiv E(X)$.
- . $\gamma>0$ يكون التوزيع ملتوياً (منحرفاً) نحو اليمين، ويزداد هذا الانحراف بزيادة γ .
- . $\gamma<0$ يكون التوزيع ملتوياً (منحرفاً) نحو اليسار، ويزداد هذا الانحراف كلما ابتعدت γ عن الصفر.

ملاحظة مهمة:

إذا كان: $\sigma = \sigma(X) = 0$ فإنّه ينتج من متراجحة تشيبتشف أنّ:

$$P[|X-m|>0] \le \frac{1}{k^2} \;, \forall k \in R^+$$

$$k \to \infty \Rightarrow P[|X-m|>0] = 0$$

وهذا يعني أنّ الحدث: $|X-m|>0$ مستحيل الوقوع، وبالتالي لدينا حتماً: $|X-m|=0$، أي: $X=m$ دوماً، وهذا يعني بدوره أن المتغيّر X يأخذ قيمة واحدة فقط هي m والتوزيع متناظر بالطبع في هذه الحالة.

مثال: بفرض تابع كثافة X من الشكل:

$$f(x) = \begin{cases} \frac{1}{2} e^{-\frac{x}{2}} & ,\forall x > 0 \\ 0 & ,\forall x \le 0 \end{cases}$$

أوجد: $\Psi_X(t)$ ، $\sigma^2(X)$ ، γ (عامل التناظر) .

الحل:

$$\Psi_X(t) = E[e^{tX}] = \frac{1}{2}\int_0^{\infty} e^{tx} . e^{-\frac{x}{2}} dx = \frac{1}{2}\int_0^{\infty} e^{(t-\frac{1}{2})x} dx$$

$$= \frac{1}{2(t-\frac{1}{2})}[e^{(t-\frac{1}{2})x}]_0^{\infty} = \frac{-1}{2(t-\frac{1}{2})} = (1-2t)^{-1} \;; t < \frac{1}{2}$$

ومنه:

$$\left.\begin{array}{l} \alpha_0 = \Psi_X(0) = 1 \quad , \alpha_1 = \Psi'_X(0) = 2 \\ \alpha_2 = \Psi''_X(0) = 8 \; , \alpha_3 = \Psi_X'''(0) = 48 \end{array}\right\} \Rightarrow \mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = 16$$

بالتالي:

$$\sigma^2 = \alpha_2 - \alpha_1^2 = 4$$

$$\gamma = \frac{\mu_3}{\sigma^3} = \frac{16}{8} = 2$$

والتوزيع غير متناظر، وهو ينحرف نحو اليمين بمقدار 2 .

(4 - 7) تمارين الفصل الرابع:

1 - إذا كان جدول توزيع X من الشكل:

X	2	4	8	16
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

أوجد: $F(x)$, $\sigma^2(X)$, $\Psi_X(t)$, $E(2^{\frac{X}{2}})$, $E(\frac{1}{X})$, $E(X^2), E(X)$

2 - إذا كان: $P(A)=\int_A f(x)dx$ ، حيث : $f(x)=\begin{cases}6x(1-x) \ ; x\in(0,1)\\ 0 \quad ; x\notin(0,1)\end{cases}$

وكانت: $A_1=\{x\ ;-2<x<1\}$, $A_2=\{x\ ;0<x<3\}$

أوجد: $P(A_2)$, $P(A_1)$

3 - بفرض تابع كثافة X من الشكل:

$$f(x)=\begin{cases}2(1-x)\ , x\in(0,1)\\ 0 \quad , x\notin(0,1)\end{cases}$$

أوجد: $E(X)$ ، $E(X^2)$ ، $\sigma^2(X)$ ، $F(x)$ ، $E[(X+10)^2]$ ، $E(\frac{1}{1-X})$.

4 - بفرض X متغيّر عشوائي منفصل يأخذ القيم : $X: 0 , 1 , 2 , 3$

إذا كان: $\begin{Bmatrix}P(X=1)=P(X=2)=\frac{3}{8}\\ P(X=0)=P(X=3)\end{Bmatrix}$ ، وكان: $Y=2X+1$.

أوجد: جدول توزيع Y ، $\Psi_Y(t)$ ، $E(\frac{1}{Y})$.

5 - بفرض: $f(x)=\begin{cases}c\,x\ , x\in(0,1)\\ 0 \quad , x\notin(0,1)\end{cases}$

المطلوب:

1. أوجد قيمة الثابت c لكي يكون $f(x)$ تابع كثافة .
2. إذا كان: $Y=\sqrt{X}$ أوجد: $g(y)$, $G(y)$.

6 . بفرض X متغيّر عشوائي مستمر تابع كثافته من الشكل:

$$f(x)=\begin{cases} a & ,x\in[-1,1] \\ 0 & ,x\notin[-1,1] \end{cases}$$

المطلوب:

1. أوجد قيمة الثابت a
2. أوجد تابع التوزيع $F(x)$
3. إذا كان : $Y=\dfrac{X-3}{2}$ أوجد $g(y)$ (تابع كثافة Y) .
4. أوجد: $E(Y)$, $\Psi_Y(t)$, $E(X^2)$, $P(X<\frac{1}{2})$

7 . إذا كان تابع توزيع المتغيّر العشوائي X من الشكل:

$$F(x)=\begin{cases} 0 & ,x<0 \\ 1-e^{-0.4x} & ,x\geq 0 \end{cases}$$

احسب: $P(X\leq 3)$, $P(X>4)$, $P(3\leq X\leq 4)$, $P(X=2)$

8 . بفرض: Z,Y,X متغيّرات عشوائيّة مستقلة.

إذا كان: $E(Y-Z)=-3$, $E(X-Z)=-4$, $E(X+Y)=1$

$$\sigma^2(X)=4\ ,\ \sigma^2(Y)=3\ ,\ \sigma^2(Z)=12$$

المطلوب:

1. أوجد: $E(X)$, $E(Y)$, $E(Z)$
2. أوجد: $\sigma^2(\frac{1}{2}X-Y)$, $\sigma^2(X\pm Y\pm Z)$

9 . إذا كان: $f(x)=\begin{cases} \dfrac{x^2}{9} & ,x\in[0,3] \\ 0 & ,x\notin[0,3] \end{cases}$ وكان : $Y=X^3$

المطلوب:

1. أوجد: $g(y)$ ، $G(y)$

2. احسب الاحتمال: $P(Y>1)$

3. بيّن ما إذا كان Y متغيّراً منتظماً أم لا ؟

10 . إذا كان تابع توزيع X من الشكل: $F(x)=a(x+1)\ ,x\in[-1,1]$

المطلوب:

1. أوجد قيمة الثابت a

2. أوجد: $\sigma^2(X)$

3. من أجل: $Y=2X+3$ أوجد : $g(y)$.

11 . إذا كان تابع كثافة X من الشكل:

$$f(x)=\begin{cases}\dfrac{x-55}{100} & ,\ 55\le x\le 65\\[2mm] \dfrac{75-x}{100} & ,\ 65\le x\le 75\end{cases}$$

المطلوب:

1. أوجد $F(x)$.

2. إذا كان: $Y=\frac{5}{9}(X-32)$ أوجد $G(y)$ (تابع توزيع Y).

12 . نلقي ثلاث قطع من النقود ولنرمز بـ Y لعدد الصور التي تظهر على الأوجه،
فإذا فرضنا أنّ X متغيّر عشوائي آخر يرتبط مع Y بالعلاقة: $2Y-X+1=0$.

المطلوب:

1. أوجد جدول توزيع كلٍ من X و Y .

2. أوجد: $\Psi_{2X-1}(t)$.

13 . إذا كان جدول توزيع Y من الشكل:

Y	-1	0	1
	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$

تحقّق من صحة متراجحة تشيبتشف بالنسبة للمتغيّر Y وذلك من أجل: $k = 2$.

14 . إذا كان X متغيّراً عشوائياً مستمراً تابع كثافته $f(x)$ ، وإذا فرضنا أنّ:

$(E(X) = b$ ، $E(X^2) = b^2 + \sigma^2)$، وكان: $Y = \dfrac{X - b}{\sigma}$.

أوجد: $E[\dfrac{Y-2}{3}]^2$ ، $g(y)$ (تابع كثافة Y).

....

الفصل الخامس
توزيعات احتماليّة منفصلة

(5-1) التوزيع على نقطة:

وهو أبسط أنواع التوزيعات المنفصلة حيث إنّ جدول توزيعه من الشكل:

X	m
	1

$$E(X)=m$$
$$\sigma^2(X)=0$$
$$F(x)=\begin{cases}0\ ,\forall x<m\\1\ ,\forall x\geq m\end{cases}$$

(5-2) التوزيع المنفصل المنتظم *Uniform Discrete Distribution*:

إذا كانت قيم المتغيّر العشوائي X من الشكل: $X : x_1, x_2,\ldots, x_n$

وكان: $P_i=P(X=x_i)=\dfrac{1}{n}\ ;\ \forall i=\overline{1\,n}\quad,\quad n\in Z^+$

عندئذٍ يسمّى توزيع X بالتوزيع المنفصل المنتظم، ويكون لـ X التوقع الرياضي والتباين التاليين: $E(X)=\dfrac{n+1}{2}$ ، $\sigma^2(X)=\dfrac{n^2-1}{12}$

مثال:

في تجربة إلقاء قطعة النرد المنتظمة مرّة واحدة : إذا فرضنا أنّ X يمثل العدد الظاهر على السطح العلوي لقطعة النرد، حيث فضاء العيّنة: $\Omega=\{1,2,3,4,5,6\}$ ، عندئذٍ يكون لـ X توزيع منفصل منتظم ويكون:

$$P_i=P(X=i)=\frac{1}{6}\ ,\ i=\overline{1,n}$$

$$E(X)=\frac{n+1}{2}=\frac{6+1}{2}=\frac{7}{2}$$

$$\sigma^2(X)=\frac{n^2-1}{12}=\frac{36-1}{12}=\frac{35}{12}$$

علماً أنّه يمكن حساب $E(X)$ و $\sigma^2(X)$ من تعريفهما أيضاً، حيث:

$$E(X)=\sum_{i=1}^{6} x_i p_i=\frac{7}{2}$$

$$\sigma^2(X)=E(X^2)-(E(X))^2=\frac{91}{6}-\frac{49}{4}=\frac{35}{12}$$

(5-3) التوزيع على نقطتين (توزيع برنولي):

بفرض X عدد مرات وقوع الحدث A في التجربة العشوائيّة. عندئذٍ يوصف هذا التوزيع بالشكل التالي:

X	0	1
P_i	q	p

أي: إما أن يقع الحدث A باحتمال P، وإما ألا يقع باحتمال $q=1-p$ ومنه:

$$E(X)=\sum_i x_i P_i=p$$

$$\sigma^2(X)=E(X^2)-\{E(X)\}^2=p-p^2=pq$$
$$\Rightarrow \sigma(X)=\sqrt{pq}$$

$$F(x)=\begin{cases}0\ , & x<0\\ q\ , & 0\le x<1\\ 1\ , & 1\le x<\infty\end{cases}$$

$$\Psi_X(t)=E(e^{tX})=\sum_i e^{tx_i}P_i=q+pe^t$$

$$\gamma=\frac{\mu_3}{\sigma^3}\ ;\ \mu_3=\alpha_3-3\alpha_1\alpha_2+2\alpha_1^3$$
$$=pq(q-p)\ ;\ \alpha_i=\Psi_X^{(i)}(0)$$

بالتالي: $\gamma=\dfrac{pq(q-p)}{pq\sqrt{pq}}=\dfrac{q-p}{\sqrt{pq}}$

فإذا كان:

- (التوزيع متناظراً بالنسبة لـ m) $p=q\Rightarrow\gamma=0\Rightarrow$

- (التوزيع منحرفاً إلى اليسار $\Leftarrow 0 > \gamma \Leftarrow p>q$)

- (التوزيع منحرفاً إلى اليمين) $\Leftarrow 0 < \gamma \Leftarrow p<q$

(5-4) توزيع ثنائي الحد:

يستخدم هذا التوزيع في التجارب العشوائيّة التي تتكرّر n مرة والتي ينتج عن أي منها إحدى نتيجتين: إما ظهور الحدث (النجاح) باحتمال ثابت: $P(A) = p$، وإما عدم ظهور الحدث (إخفاق أو فشل) باحتمال: $q = 1 - p$. مثل: الولادة ذكر أم لا؟ الطالب مجتهد أم غير مجتهد ؟ أُصيبَ الهدف أم لا ؟

شروط التوزيع:

أ . أن تكون التجارب مستقلة، ولنفرض أن عددها هو n.

ب . الاحتمال ثابت من تكرار إلى آخر ويساوي p .

ج . إذا رمزنا بـ X لعدد مرات ظهور الحدث A (النجاح) عند تكرار التجربة العشوائيّة عدداً من المرات مقداره n، فإنّ قيم X الممكنة تكون عندئذٍ من الشكل:

$$X : 0\ ,1\ ,2\ ,3\ ,....,\ i\ ,....,\ n$$

لنحاول استنتاج الشكل العام للتوزيع من خلال الملاحظات التالية:

. من أجل: $n=2$ ، وليكن المطلوب هو: النجاح مرّة واحدة ، وإخفاق مرّة واحدة، عندئذٍ:

إما أن تكون: التجربة الأولى هي النجاح والثانية إخفاق، وإما أن تكون: التجربة الأولى هي إخفاق والثانية النجاح، أي:

$$P(X=1)=p\,q+q\,p=2\,p\,q=2\,p^1q^1=C_1^2\;p^1q^1$$

. من أجل: $n=3$، وليكن المطلوب هو: النجاح مرتين، وإخفاق مرّة واحدة، عندئذ:

إما أن يكون: (نجاحاً، نجاحاً، فشلاً) أو (نجاحاً، فشلاً، نجاحاً) أو (فشلاً، نجاحاً، نجاحاً) أي: $P(X=2)= ppq+pqp+qpp=3\,p^2q=C_2^3\;p^2q^1$

وبالحالة العامة: إذا أجرينا التجربة n مرة، وكان المطلوب تحديد احتمال ظهور الحدث A في التجربة عدد من المرات مقداره x مرّة ، وبالتالي عدم ظهوره $(n-x)$ مرّة ،

حيث: n عدد مرات تكرار التجربة، عندئذٍ يكون: $P(X=x)=C_x^n p^x q^{n-x}$ وهو الشكل العام للتوزيع .

وممّا تقدّم يكون جدول توزيع X من الشكل:

X	0	1	$2 \ldots$	$x \ldots$	$i \ldots$	n
P_i	p_0	p_1	$p_2 \cdots$	$p_x \cdots$	$p_i \cdots$	p_n

حيث التوزيع الاحتمالي يأخذ الشكل التالي: $P_i=C_i^n\ p^i q^{n-i}$

ويكون لدينا أيضاً: $P_i \geq 0\ \ ,\ \forall\ i \leq n$

$$\sum_{i=0}^{n} C_i^n\ p^i q^{n-i} = (\ p+q\ )^n = 1$$

أضف إلى أنّ المتغيِّرات $X_i\ ;\ (i=\overline{1\ ,\ n})$ مستقلة مثنى مثنى، وكلٍ منها يتوزع على نقطتين (أي: إما أن يقع، أو لا يقع)، أضف إلى أن:

$$X = X_1 + X_2 + \ldots + X_n$$

يرمز عادةً لتوزيع ثنائي الحد بالرمز: $X : B(n, p)$ حيث يعتمد على n و p، لذلك سمّي بتوزيع ثنائي الحد، لأن الصورة العامة للتوزيع تأخذ شكل "الحد العام" في مفكوك ذي الحدين.

- القيم المميّزة لتوزيع ثنائي الحد:

$$- \ E(X)=\sum_i E(X_i)=\underbrace{p+p+p+\ldots+p}_{n}=np$$

$$-\sigma^2(X)=\sum_i \sigma^2(X_i)=\underbrace{pq+pq+\ldots+pq}_{n}=npq \Rightarrow \sigma(X)=\sqrt{npq}$$

$$-\Psi_X(t)=E(e^{tX})=E(\prod_{i=1}^{n} e^{tX_i})=\prod_{i=1}^{n} E(e^{tX_i})=\prod_{i=1}^{n} \Psi_{X_i}(t)=\underbrace{(q+pe^t)\ldots(q+pe^t)}_{n}=\left[q+pe^t\right]^n$$

$$- \qquad \gamma=\frac{\mu_3}{\sigma^3}\ ;\ \mu_3=\alpha_3-3\alpha_1\alpha_2+2\alpha_1^3$$

ومن العلاقة: $\Psi_X(t)=(\ q+pe^t\ )^n$; $\alpha_i=\Psi_X^{(i)}(0)$ ينتج:

$\alpha_1 = np \equiv \Psi'_X(0)$

$\alpha_2 = n(n-1)p^2 + np \equiv \Psi''_X(0)$

$\alpha_3 = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np \equiv \Psi'''_X(0)$

ومنه: $\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = npq(q-p)$ ، بالتالي يكون:

$$\gamma = \frac{\mu_3}{\sigma^3} = \frac{q-p}{\sqrt{npq}}$$

فإذا كان:

- $p = q$ يكون التوزيع متناظراً بالنسبة لتوقعه الرياضي.
- $p > q$ يكون التوزيع منحرفاً نحو اليسار.
- $p < q$ يكون التوزيع منحرفاً نحو اليمين.

مثال (1):

لتكن E تجربة إلقاء قطعة نقود متجانسة 6 مرات، فإذا رمزنا بـ X لعدد مرات ظهور الشعار خلال التجربة. أوجد: $\Psi_X(t), \sigma(X), E(X)$

الحل:

نرمز بـ A لحدث ظهور الشعار في كل مرّة من مرات تكرار التجربة. عندئذٍ يكون الاحتمال ثابتاً في كل تجربة ويساوي:

$$p = p(A) = \frac{1}{2}\ ,\ q = p(A') = \frac{1}{2}$$

$$E(X) = n.p = 6\frac{1}{2} = 3$$

$$\sigma^2(X) = n.p.q = 6.\frac{1}{2}.\frac{1}{2} = \frac{6}{4} = 1.5$$

$$\Rightarrow \sigma(X) = \sqrt{1.5} \approx 1.2247$$

$$\Psi_X(t) = (q + pe^t)^n = (\frac{1}{2} + \frac{1}{2}e^t)^6$$

وبما أنّ: $p = q$ فالتوزيع متناظر حتماً.

مثال (2):

يطلق رامٍ على هدف 10 طلقات، فإذا رمزنا بـ X لعدد الطلقات التي تصيب الهدف خلال هذه التجربة، وكان احتمال إصابته للهدف في كل مرّة هو $\frac{1}{4}$. والمطلوب:

1 . ادرس المتغيِّر X .

2 . ما احتمال أن يُصاب الهدف بطلقتين على الأقل ؟

الحل:

1. بفرض A حدث إصابة الهدف، عندئذٍ:

$$p(A) = p = \frac{1}{4} \ , \ p(A') = q = \frac{3}{4}$$

بما أنّه يتم إطلاق 10 طلقات، فهذا يعني أن التجربة تتكرّر $n = 10$ مرات، وبالتالي فإنّ جدول توزيع X من الشكل:

X	0	1	2	...	i	...	10
	p_0	p_1	p_2	...	p_i	...	p_{10}

والمسألة هي مسألة تجارب متكرِّرة تخضع للتوزيع الثنائي: $P_i = C_i^n p^i q^{n-i}$، بالتالي ومباشرةً يكون:

$$E(X) = n.p = 10(\frac{1}{4}) = 2.5$$

$$\sigma^2(X) = n.p.q = 10(\frac{1}{4})(\frac{3}{4}) = \frac{30}{16} \Rightarrow \sigma(X) = \sqrt{\frac{30}{16}}$$

$$\Psi_X(t) = (q + pe^t)^n = (\frac{3}{4} + \frac{1}{4}e^t)^{10}$$

.2 $P(X \geq 2) = 1 - P(X < 2) = 1 - p_0 - p_1 = 1 - (\frac{3}{4})^{10} - 10(\frac{1}{4})(\frac{3}{4})^9 = 0.756$

(5-5) توزيع بواسون:

نقول عن المتغيّر العشوائي المنفصل X إنّه يخضع لتوزيع بواسون بالوسيط λ إذا وفقط إذا كان له جدول التوزيع التالي:

X	0	1	2	$\dots$	i	$\dots$
	p_0	p_1	p_2	$\dots$	p_i	$\dots$

حيث: $i = 0,1,2,3,\dots,\ ;\quad p_i = \dfrac{e^{-\lambda}\,.\,\lambda^i}{i!}$ ، و: $\lambda \in R^+$ (مقدار ثابت وهو معدّل عدد النجاحات في فترة زمنية معيّنة أو منطقة محدّدة)، أضف إلى أن X هو عدد مرات وقوع الحدث المفترض في نفس الفترة الزمنية أو المنطقة المحدّدة.

من الواضح أنّ: $\sum_i p_i = e^{-\lambda} \sum_{i=0}^{\infty} \dfrac{\lambda^i}{i!} = e^{-\lambda}\, e^{\lambda} = 1$ ، وهذا متوقع بالطبع.

. شروط التوزيع:

1 . أن تكون التجارب مستقلة.

2 . أن يكون الاحتمال ثابتاً من محاولة إلى أخرى وصغيراً بشكل عام.

3 . أن تكون قيم X الممكنة هي: $X:\ 0\ ,\ 1\ ,\ 2\ ,\dots,\ i\ ,\ \dots.$

وهذا التوزيع يصلُح بشكل عام للأحداث النادرة الوقوع أو للمتغيّرات التي تحدث في أزمنةٍ عشوائيةٍ معلومة. مثلاً: عدد المكالمات الهاتفيّة خلال فترة محدّدة، عدد الأخطاء المطبعية في صفحة كتاب ما، عدد السيارات المارة في مكان ما خلال فترة زمنية محدّدة، عدد الزلازل السنويّة، عدد الحوادث الأسبوعيّة، عدد المصابيح التالفة المُنتَجة خلال فترة زمنيّة محدّدة...

. القيم المميّزة لتوزيع بواسون:

لنحسب التوقع الرياضي والتباين باستخدام الدالّة المولّدة للعزوم حيث:

حسب (ماك . لوران) في النشر يمكن أن نكتب:

$$\Psi_X(t) = E(e^{tX}) = \sum_{i=0}^{\infty} e^{ti} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$$(\Psi_X(t))' = \lambda e^t (e^{\lambda(e^t-1)})$$

ومنه ينتج:

$$E(X)=\alpha_1=(\Psi_X(0))'=\lambda$$

$$\alpha_2=\lambda^2+\lambda$$

$$\Rightarrow \sigma(X)=\sqrt{\lambda}\ \ \sigma^2(X)=\alpha_2-\alpha_1^{\ 2}=\lambda \equiv \mu_2$$

$$\alpha_3=\lambda^3+3\lambda^2+\lambda$$

$$\mu_3=\lambda$$

$$\gamma=\frac{\mu_3}{\sigma^3}=\frac{1}{\sqrt{\lambda}}$$

. تقريب توزيع ثنائي الحد لتوزيع بواسون:

إذا اعتبرنا n كبيرة بشكلٍ كافٍ ($n \geq 50$) و p صغيرة ($p<0.1$) بحيث يبقى الجداء $n\,p$ مقداراً ثابتاً وليكن λ، عندئذٍ يمكن تقريب التوزيع الثنائي إلى توزيع بواسون الحالي بالوسيط $\lambda=n\,p$.

مثال (1):

إذا كان نسبة وجود أنثى لا تكثر الكلام في مدينة ما هي: 0.001، ما احتمال عدم وجود أي أنثى تكثر الكلام في حي يضم 4000 أنثى؟

الحل: بفرض X عدد الإناث اللواتي يكثرن من الكلام في الحي عندئذٍ:

$X : B(4000,0.001)$ بالتالي:

$$P(X=0)=C_0^{4000}(0.001)^0(0.999)^{4000}=0.0182$$

وحسب توزيع بواسون حيث: ($\lambda=n\,p=4000\times 0.001=4$) يكون:

$P(X=0)=\dfrac{e^{-4}4^0}{0!}=0.0183$ واضح أنّ الفرق صغير والتقريب جيِّد جداً.

مثال (2):

في اختبار لفعالية دواء ما، إذا فرضنا أنّه ضار باحتمال 0.001 ، وأنّه أعطي لـ 2000 شخص من المتبرعين، والمطلوب (باستخدام تقريب بواسون):

1 . ما احتمال أن يتضرّر أكثر من شخصين نتيجة تلقيّهما هذا الدواء ؟

2 . ما احتمال أن يتضرّر شخصان على الأقل ؟

الحل:

1 . لنرمز لحدث ضرر شخص بالرمز A ، وبـ X لعدد الأشخاص المتضررين باللقاح . إن إعطاء الدواء لـ 2000 شخص يعني أن التجربة تتكرّر $n = 2000$ مرّة، وبالتالي يكون المطلوب:

$$P(X>2)=1-P(X\le 2)=1-(P_0+P_1+P_2)$$
$$=1-\{P(X=0)+P(X=1)+P(X=2)\}$$

ولكن: $P_i = P(X=i)=\dfrac{e^{-\lambda}\,\lambda^i}{i!}\ ;\ \lambda = n\,p = 2000\times 0.001 = 2$

بالتالي:

$$P(X=0)=\frac{e^{-\lambda}\,\lambda^0}{0!}=e^{-\lambda}=e^{-2}\approx 0.135$$

$$P(X=1)=\frac{e^{-\lambda}\lambda^1}{1!}=2e^{-2}\approx 0.270$$

$$P(X=2)=\frac{e^{-\lambda}\lambda^2}{2!}=2e^{-2}$$

ومنه يكون:

$$P(X>2)=1-[e^{-2}+2e^{-2}+2e^{-2}]=1-5\,e^{-2}\approx 0.323$$

2 . لدينا:

$$X \;\Big|\; 0 \;\; 1 \;\; 2 \;\; \dots \;\; i \;\; \dots \;\; 2000$$

$$P_x \mid P_0 \;\; P_1 \;\; P_2 \ldots \;\; P_i \ldots \;\; P_{2000}$$

حيث: $P_i = P(X=i) = \dfrac{e^{-\lambda}\ \lambda^i}{i!}$ ، بالتالي يكون الاحتمال المطلوب هو :

$$P(X \geq 2) = 1 - P(X < 2) = 1 - (P_0 - P_1) = 1 - \{P(X=0) + P(X=1)\}$$
$$= 1 - (e^{-2} + 2e^{-2}) = 0.593$$

(6-5) التوزيع الهندسي: *Geometric Distribution*

بفرض أنّ E تجربة عشوائية ثنائية ، A حدث مرتبط بها بحيث:

$$.\quad p(A) = p \;,\; p(A') = q$$

وأنّنا سنقوم بتكرار التجربة عدداً من المرات X حتى يقع الحدث A لأول مرّة (أي نوقف التجربة عندما يقع الحدث A للمرّة الأولى) مثلاً: عدد مرات تقديم مقرّر امتحاني حتى النجاح في ذلك المقرّر، عدد الولادات حتى تُرزق سيّدة ما بأنثى...

*** شروط التوزيع:**

. التجارب مستقلة.

. $p \;,\; q$ كلاهما ثابت.

ـ إذا رمزنا بـ X لعدد مرات تكرار التجربة حتى وقوع الحدث A لأول مرّة، فإنّ قيم X تكون من الشكل: $X : 1 \;,\; 2 \;,\; 3 \;, \ldots , i , \ldots , \infty$ ، وإذا فرضنا أنّ A قد وقع لأول مرّة في التجربة رقم i ، فهذا يعني أنّه لم يقع خلال $(i-1)$ مرّة (أي يكون قد وقع A' خلال الـ $(i-1)$ مرّة ، ثمّ وقع A في التجربة رقم i) ، وبالتالي:

$$P(X=i) = \underbrace{q \,.\, q \,.\, q \ldots q}_{(i-1)} \,.\, p = p \,.\, q^{i-1}$$

وبشكل عام نجد أنّ: $p_i = p \,.\, q^{i-1} \;,\; i = 1 \;,\; 2 \;,\; 3 \;,\; \ldots$

وهي الصورة العامة للتوزيع .

وقد سمّي بالتوزيع الهندسي لأن الاحتمالات الموافقة لقيم المتغيّر العشوائي توافق حدود متوالية هندسيّة . ويكون جدول توزيع X من الشكل:

X	1	2	...	i	...
	p	pq	...	pq^{i-1}	...

واضح أنَّ: p_i تشكل متوالية هندسيّة حدّها الأول p وأساسها q وبالتالي:

$$\sum_i p_i = \sum_{i=1}^{\infty} p\, q^{i-1} = \frac{p}{1-q} = \frac{p}{p} = 1$$

ونرمز للتوزيع الهندسي بالشكل: $X:G(p) \Leftrightarrow X$ يخضع لتوزيع هندسي بالوسيط p .

. القيم المميّزة للتوزيع الهندسي:

$$E(X) = \sum_{i=1}^{\infty} x_i\, p_i = \sum_i i\, p\, q^{i-1} = p \sum_{i=1}^{\infty} (q^i)'$$

وبما أنَّ: $\sum_{i=1}^{\infty} q^i$ سلسلة هندسيّة حدّها الأول q وأساسها q يكون:

$$E(X) = p\left(\frac{q}{1-q}\right)' = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

$$\Psi_X(t) = E(e^{tX}) = \sum_{i=1}^{\infty} e^{ti}\, p\, q^{i-1} \;\; ; \; x_i \equiv i$$

$$= \frac{p}{q}\sum_{i=1}^{\infty} e^{ti}\, q^i = \frac{p}{q}\sum_{i=1}^{\infty} (q e^t)^i = \frac{p}{q}\, \frac{q e^t}{1-qe^t} = \frac{p e^t}{1-qe^t}$$

$$\alpha_i = \Psi_X^{(i)}(0)$$

$$\Psi_X'(t) = \frac{pe^t}{(1-qe^t)^2}$$

$$\alpha_1 = \Psi_X'(0) = \frac{1}{p}$$

$$\alpha_2 = \Psi_X''(0) = \frac{1+q}{p^2}$$

$$\sigma^2(X) = \alpha_2 - \alpha_1{}^2 = \frac{q}{p^2} \equiv \mu_2$$

$$\Rightarrow \sigma(X) = \frac{\sqrt{q}}{p}$$

مثال:

صندوق يحوي 10 كرات حمراء و 6 كرات خضراء . نسحب كرة من الصندوق مع إعادة ونعيد السحب حتى تظهر كرة خضراء لأول مرّة . والمطلوب:

1 . ادرس المتغيّر العشوائي X المُعبّر عن عدد مرات تكرار التجربة.

2 . ما احتمال أن تُكرّر التجربة 6 مرات ؟

الحل:

1) نفرض:

A حدث سحب كرة خضراء ، فيكون: $p(A) = p = \frac{6}{16} \; ; \; q = 1 - p = \frac{10}{16}$

X عدد مرات تكرار التجربة حتى ظهور الحدث المفترض.

من الواضح أنّ : $X : G(\frac{6}{16})$ وبالتالي يكون:

$$\Psi_X(t) = \frac{pe^t}{1 - qe^t} = \frac{\frac{6}{16}e^t}{1 - \frac{10}{16}e^t} \; ; \alpha_i = \Psi_X^{(i)}(0)$$

يمكن حساب (α_1 و α_2) بالاشتقاق، ثم نحسب $\sigma^2(X)$، أو يمكن أن نكتب مباشرة:

$$\alpha_1 = E(X) = \frac{1}{p} = \frac{16}{10}$$

$$\sigma^2(X) = \frac{q}{p^2} = \frac{\frac{10}{16}}{(\frac{6}{16})^2} = \frac{160}{36}$$

ومنه: $\sigma(X)=\dfrac{4\sqrt{10}}{6}$

2) $P(X=6)=(\frac{6}{16})(\frac{10}{16})^{6-1}=(\frac{6}{10})(\frac{10}{16})^{5}=0.057$ ، حيث: $P(X=i)=pq^{i}$

ملاحظة:

إذا كان المطلوب هو تكرار التجربة إلى أن يظهر الحدث A ، r مرّة ، نكون بصدد توزيع آخر يسمى: التوزيع الثنائي السالب.

(7-5) التوزيع الثنائي السالب: *Negative Binomial Distribution* :

(ويسمى أحياناً توزيع باسكال)

بفرض E تجربة عشوائية ، A حدث مرتبط بها بحيث:

$\{P(A)=p$ (النجاح) و: $P(A')=q$ (إخفاق)$\}$ أي التجربة لها نتيجتان فقط هما: نجاح أو إخفاق .

تُكرّر التجربة في نفس الشروط حتى يقع الحدث A عدداً من المرات مقداره r مرّة.

فإذا رمزنا بـ X لعدد مرات التكرار حتى ظهور A ، r مرّة.

المطلوب: دراسة توزيع X ؟

مما تقدم يتضح أنّ: r مقدار ثابت، X متغيّر وقيمه الممكنة من الشكل:
$X: x_0, x_1, x_2, ..., x_i, ...\infty$ ، حيث: $x_i=r+i \ , \ i=0,1,2,3,...$.

*** شروط التوزيع:**

- التجارب كلها مستقلة.
- r, q, p ثوابت.
- $X: x_0, x_1, x_2, ..., x_i, ...$ حيث: $x_i=r+i \ , \ i=0,1,2,3,...$.

. الصورة العامة للتوزيع:

واضح أنه إذا وقع الحدث A ، r مرّة في التكرار رقم: x_i (لنرمز لهذا الحدث بالرمز B)، فهذا يعني أنّ A يكون قد وقع $(r-1)$ مرّة في التكرار رقم (x_i-1)، (ولنرمز لهذا الحدث بالرمز C) ، ومنه بناءً على الاستقلال وعلى التجارب المتكرّرة، يكون:

$$P_i = P(X=i) = P(BC) = \underbrace{P(B)}_{p} \, . \, \underbrace{P(C)}_{C_{r-1}^{x_i-1} p^{r-1} q^{x_i-r}} = C_{r-1}^{x_i-1} p^r q^{x_i-r}$$

واضح (حسب النشر وفق سلسلة صحيحة) أنّ:

$$\sum_{i=0}^{\infty} P_i = \sum_{i=0}^{\infty} C_{r-1}^{x_i-1} p^r q^{x_i-r} = p^r \sum_{i=0}^{\infty} C_{r-1}^{x_i-1} q^{x_i-r} = p^r (1-q)^{-r} = \frac{p^r}{p^r} = 1$$

. القيم المميّزة للتوزيع الثنائي السالب:

$$\Psi_X(t) = E(e^{tX}) = \sum_{i=0}^{\infty} e^{tx_i} P_i = \sum_{i=0}^{\infty} e^{tx_i} C_{r-1}^{x_i-1} p^r q^{x_i-r}$$

$$= p^r \sum_{i=0}^{\infty} (qe^t)^{x_i-r} . \, e^{tr} = p^r . e^{tr} \sum_{i=0}^{\infty} (qe^t)^{x_i-r}$$

$$= p^r e^{tr} (1-qe^t)^{-r} \quad = \quad [\frac{pe^t}{1-qe^t}]^r$$

$$\Psi'_X(t) = r[\frac{pe^t}{1-qe^t}]'[\frac{pe^t}{1-qe^t}]^{r-1} = r\frac{pe^t - pqe^{2t} + pqe^{2t}}{(1-qe^t)^2}[\frac{pe^t}{1-qe^t}]^{r-1}$$

$$= \quad r\frac{(pe^t)^r}{(1-qe^t)^{r+1}}$$

ومنه:

$$E(X) = \Psi'_X(0) = r\frac{p^r}{(1-q)^{r+1}} = r\frac{p^r}{p^{r+1}} = \frac{r}{p}$$

$$\sigma^2(X) = \alpha_2 - \alpha_1^2 = \ldots = r\frac{q}{p^2}$$

واضح أنه من أجل: $r=1$ نكون في حالة التوزيع الهندسي (تحقّق من ذلك وقارن بينهما).

مثال (1):

بفرض كررنا تجربة إلقاء قطعة نرد 5 مرات. ما احتمال ظهور الرقم 3 في نهاية التجربة 3 مرات ؟

الحل: (يمكن تطبيق التوزيع الثنائي السالب مباشرةً)

لدينا بالحالة العامة: $P(X=k)=C_{r-1}^{k-1}p^r q^{k-r}$ حيث: $\{k=5 \ , \ r=3\}$.

ومنه: $$P(X=5)=C_2^4 p^3 q^2=\frac{4!}{2!\,2!}(\frac{1}{6})^3(\frac{5}{6})^2\approx 0.115$$

مثال (2):

إذا كان احتمال أن تجد شخصاً في مكانٍ ما وبشروط محدّدة هو: 60% فما احتمال أن تذهب إلى ذات المكان (وبذات الشروط) 10 مرات، لتجده 3 مرات ؟

الحل:

لدينا: $$P(X=k)=C_{r-1}^{k-1}\ p^r\ q^{k-r}\ ;\begin{cases} p=\frac{6}{10}\ ,\ q=\frac{4}{10} \\ k=10\ \ ,\ \ r=3 \end{cases}$$

بالتالي يمكن أن نكتب مباشرة : $P(X=10)=C_2^9\ p^3\ q^7\approx 0.012$

(5–8) التوزيع فوق الهندسي *Hyper geometric Distribution* :

(ويسمى أحياناً بالتوزيع التوافقي)

المسألة المطروحة بالشكل: صندوق يحوي N شيئاً مقسماً إلى نوعين:

A_1 عدد عناصره N_1 .

A_2 عدد عناصره N_2 .

نُسحب عيّنة حجمها n (دون إعادة)، ولنرمز بـ: X لعدد عناصر النوع A_1 التي تظهر في العينة، وبالتالي سيكون عدد عناصر النوع A_2 في العينة هو: $n-X$.

والمطلوب دراسة توزيع X :

نقدّم (للإيضاح) المثال البسيط التالي:

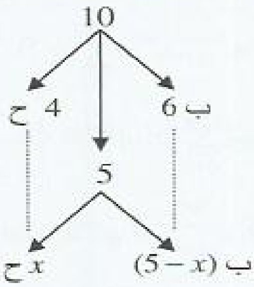
مثال:

لنفرض وعاء يحوي 10 كرات متماثلة، منها: 4 لونها أحمر (ح)، والأخرى لونها أبيض (ب) . سُحِبَتْ 5 كرات من الوعاء (بدون إعادة). ما احتمال أن يكون من بينها x كرة لونها أحمر؟

الحل:

واضح أنّ سحب x كرة حمراء من بين 5 كرات حمراء يتم بـ C_x^5 طريقة. وسحب $(5-x)$ كرة بيضاء من بين 6 كرات بيضاء يتم بـ C_{5-x}^6 طريقة.

وبالتالي (حسب المبدأ الأساسي في العد) يكون عدد طرق وقوع الحدثين معاً هو: $C_x^5 . C_{5-x}^6$ طريقة . (وهو يمثل عدد الحالات الملائمة أو المطلوبة).



ومنه الاحتمال المطلوب:

$$P= \frac{\text{عدد الحالات الملائمة}}{\text{عدد الحالات الكلية}} = \frac{C_x^5 C_{5-x}^6}{C_5^{10}} \; ; \; x = 0,1,2,3,4,5$$

وبالعودة إلى التوزيع المذكور وبالحالة العامة يمكن أن نكتب: $X : 0 , 1 , 2 ,..., n$

$$P_i = P(X = i) = \frac{C_i^{N_1} . C_{n-i}^{N_2}}{C_n^N}$$

وهي الصيغة العامة للتوزيع.

واضح أنّ: $\forall i : P_i \geq 0$ أضف إلى أنّ:

$$\sum_{i=0}^{n} P_i = \sum_{i=0}^{n} \frac{C_i^{N_1} \,.\, C_{n-i}^{N_2}}{C_n^N} = \frac{1}{C_n^N} \underbrace{\sum_{i=0}^{n} C_i^{N_1} \,.\, C_{n-i}^{N_2}}_{C_n^N} = \frac{C_n^N}{C_n^N} = 1$$

ملاحظة مهمة:

يمكن رد التوزيع فوق الهندسي إلى التوزيع الثنائي بالوسيطين: $\{ n$ و $P = (\frac{N_1}{N}) \}$.

ولتبيان ذلك:

لدينا:

$$C_n^N = \frac{N!}{n!(N-n)!} = \frac{N(N-1)(N-2)...(N-(n-1))}{n!}$$
$$= \frac{N^n}{n!}(1-\frac{1}{N})(1-\frac{2}{N})...(1-\frac{n-1}{N})$$

وبناءً على ذلك، نكتب العلاقة $P_i = \dfrac{C_i^{N_1} \,.\, C_{n-i}^{N_2}}{C_n^N}$ بالشكل:

$$P_i = \frac{\frac{N_1{}^i}{i!}(1-\frac{1}{N_1})(1-\frac{2}{N_1})...(1-\frac{i-1}{N_1})\frac{N_2{}^{n-i}}{(n-i)!}(1-\frac{1}{N_2})...(1-\frac{n-i-1}{N_2})}{\frac{N^n}{n!}(1-\frac{1}{N})(1-\frac{2}{N})...(1-\frac{n-1}{N})}$$

حيث نلاحظ أنّه عندما: N, N_1, N_2 كبيرة جداً (بحيث يمكن اعتبارها تسعى إلى ∞) و n صغيرة جداً يكون:

$$P_i = \frac{\frac{N_1{}^i}{i!}\,\frac{N_2{}^{n-i}}{(n-i)!}}{\frac{N^n}{n!}} = \frac{n!}{i!(n-i)!}\frac{N_1{}^i N_2{}^{n-i}}{N^n} \quad ; N^n = N^i N^{n-i}$$

$$= C_i^n (\frac{N_1}{N})^i (\frac{N_2}{N})^{n-i} \equiv C_i^n p^i q^{n-i} \quad ; \quad p = \frac{N_1}{N}\ ,\ q = \frac{N_2}{N}$$

وبالمقارنة مع توزيع ثنائي الحد، حيث كان:

$$X \;:\; 0\,,1\,,2\,,\ldots,\,i\,,\ldots,\,n$$

$$P_i = P(X=i) = C_i^n\, p^i q^{n-i}$$

يكون: $p = (\dfrac{N_1}{N})$. وهكذا نجد أنّه يمكن فعلاً رد التوزيع فوق الهندسي إلى التوزيع الثنائي بالوسيطين: n و $p = (\dfrac{N_1}{N})$.

وبالواقع يلعب التوزيع فوق الهندسي دوراً هاماً في الإحصاء من خلال الاستقراءات المبنية على دراسة العيّنة أو العيّنات العشوائيّة المسحوبة من المجتمع الإحصائي ومن ثمّ تعمّم هذه الدراسة على كل المجتمع، ونظراً لهذه الأهميّة فإنّه توجد جداول خاصة تعطي قيم P_i في التوزيع المذكور.

. من الصفات المميزة للتوزيع فوق الهندسي:

* **التوقع الرياضي:** $E(X) = n\dfrac{N_1}{N}$

حيث:

$$E(X) = \sum_{i=0}^{n} i\; P_i = \sum_i i\,\frac{C_i^{N_1} C_{n-i}^{N_2}}{C_n^N} = \frac{1}{C_n^N}\sum_i i\,\frac{N_1!}{i!(N_1-i)!}C_{n-i}^{N_2}$$

$$= \frac{1}{C_n^N}\sum_i \frac{1}{(i-1)!}\frac{N_1!}{(N_1-i)!}C_{n-i}^{N_2} = \frac{N_1}{C_n^N}\sum_i \frac{1}{(i-1)!}\frac{(N_1-1)!}{(N_1-i)!}C_{n-i}^{N_2}$$

$$= \frac{N_1}{C_n^N}\underbrace{\sum_{i=0}^{n} C_{i-1}^{N_1-1} C_{n-i}^{N_2}}_{C_{n-1}^{N-1}} = \frac{N_1}{C_n^N}C_{n-1}^{N-1} = \ldots = n\frac{N_1}{N}$$

** **التباين:** وهو يعطى بالعلاقة: $\sigma^2(X) = \dfrac{n\,N_1 N_2 (N-n)}{N^2 (N-1)}$ (نقبلها دون برهان).

علماً أنّه لبرهانها يُعتمد على الخاصة:

$$.\sum_{i=0}^{s} C_i^k C_{s-i}^l = C_s^{k+l} \quad ; \; 0 \le i \le k \;\;,\; 0 \le s-i \le l$$

مثال:

سلة تحوي 10 زهرات منها 4 حمراء. سُحبت عشوائياً وبدون إعادة 3 زهرات: المطلوب:

1 . ما التوزيع الاحتمالي لعدد الزهور الحمراء المسحوبة (ما الدالّة الاحتماليّة) ؟

2 . ما احتمال أن يكون من بين الزهور المسحوبة زهرتان بيضاويان ؟

3 . أوجد التوقع الرياضي والتباين.

الحل:

1) إذا رمزنا بـ X لعدد الزهور البيضاء، فإنّ التوزيع الاحتمالي هو فوق الهندسي .

وبالتالي:

$$P_x = \frac{C_x^{N_1}\ C_{n-x}^{N_2}}{C_n^N} \quad ; \begin{cases} N = 10 \ , \ N_1 = 4 \\ n = 3 \ ; \ N_2 = 6 \end{cases}$$

(2
$$P_2 = P(X = 2) = \frac{C_2^4\ C_1^6}{C_3^{10}} = \frac{3}{10}$$

(3
$$E(X) = n\frac{N_1}{N} = 3\frac{4}{10} = 1.2$$

$$\sigma^2(X) = \frac{nN_1N_2(N-n)}{N^2(N-1)} = \ldots = 0.56$$

(5-9) تمارين الفصل الخامس:

1 . في تجربة إلقاء قطعة نرد منتظمة 4 مرات، ما احتمال عدم ظهور الرقم 4 فيها، ثم ما احتمال ظهوره أربع مرات؟

2 . إذا كان معدّل حوادث السير في مكان معيّن 4 حوادث في الشهر الواحد:

أ . ما احتمال عدم ظهور أي حادث في شهر معيّن؟

ب . ما احتمال حدوث 3 حوادث أو أقل في شهر معيّن؟

3 . إذا كان X يتبع توزيع بواسون وكان: $P(X=1)=3P(X=0)$

أوجد: $P(X=2)$.

4 . إذا كان: $X : B(20,0.1)$

احسب $P(X=3)$ باستخدام: توزيع ثنائي الحد، ثم توزيع بواسون.

5 . حزمة أقلام تضم 10 أقلام ، منها 4 أقلام سيّئة الصنع ، اخترنا عشوائياً عيّنة مؤلَّفة من 3 أقلام من هذه الحزمة. المطلوب:

أ . حدّد التوزيع الاحتمالي لعدد الأقلام السيّئة في العيّنة (الدالّة الاحتماليّة للتوزيع).

ب . ما احتمال أن يكون قلمان في العيّنة المسحوبة سيئي الصنع؟

6 . قرّر شخص ترك العمل الحر إذا وفّق بـ 3 مشاريع ناجحة . فإذا كان احتمال نجاح هذا الشخص في كل مشروع هو 0.4 .

أوجد التوزيع الاحتمالي لعدد مرات الإخفاق (الفشل) ، ثم أوجد التوقع الرياضي والانحراف المعياري لهذا التوزيع.

الفصل السادس

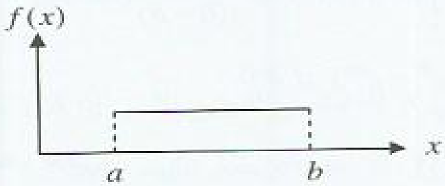
التوزيعات الاحتماليّة المستمرّة

(1-6) التوزيع المنتظم *Uniform Distribution*:

(يسمى أحياناً بالتوزيع المستطيل*Rectangular*): وهو أبسط أنواع التوزيعات المستمرّة حيث الصيغة العامة للتوزيع هي من الشكل:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & , \quad a\le x\le b\\ 0, & \text{فيما عدا ذلك}\end{cases}$$

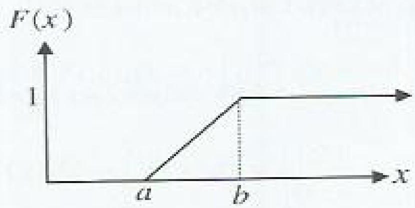
واضح أنّ دالّة (أو تابع) الكثافة مقدار ثابت ، ويمكن توضيحها بيانيّاً بالشكل:



دالة كثافة الاحتمال للتوزيع المنتظم

ومن الرسم يتّضح لماذا سميّ بالتوزيع المستطيل، وتابع توزيعه من الشكل:

$$F(x)=\begin{cases}0 & ; \quad x<a\\ \dfrac{x-a}{b-a} & ; \; a\le x<b\\ 1 & ; \quad x\ge b\end{cases}$$



ـ القيم المميّزة للتوزيع المنتظم:

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\frac{1}{b-a}\int_{a}^{b}xdx=\frac{a+b}{2}$$

$$E(X^2)=\frac{1}{b-a}\int_{a}^{b}x^2dx=\frac{1}{3}\ \frac{b^3-a^3}{b-a}$$

وبالتالي:

$$\sigma^2(X)=E(X^2)-(E(X))^2=\frac{1}{3}(b^2+ab+a^2)-(\frac{a+b}{2})^2$$

$$=\frac{(b-a)^2}{12}$$

$$\Psi_X(t)=E(e^{tX})=\int_{a}^{b}e^{tx}f(x)dx=\frac{1}{t(b-a)}[e^{tx}]_a^b$$

$$=\frac{1}{t(b-a)}(e^{tb}-e^{ta})\ \ ;t\neq 0$$

ويمكن كتابتها أيضاً بالشكل التالي:

$$\Psi_X(t)=\frac{1}{t(b-a)}(e^{tb}-e^{ta})$$

$$=\frac{1}{t(b-a)}\{[1+\frac{tb}{1!}+\frac{(tb)^2}{2!}+...]-[1+\frac{ta}{1!}+\frac{(ta)^2}{2!}+...\}$$

$$=\frac{1}{t(b-a)}\{\frac{t(b-a)}{1!}+\frac{t^2}{2!}(b^2-a^2)+\frac{t^3}{3!}(b^3-a^3)+...\}$$

$$=1+\frac{t}{2!}(b+a)+\frac{t^2}{3!}(b^2+ab+a^2)+...$$

ومنها يمكن استنتاج كل العزوم ومن المراتب كافة كما وجدنا سابقاً في أثناء دراسة الدالّة المولّدة للعزوم، حيث:

$$\alpha_i=\ \Psi_X^{(i)}(0)$$

$$\alpha_1=\frac{a+b}{2}$$

$$\alpha_2 = \frac{a^2+ab+b^2}{3}$$

$$\alpha_3 = \frac{a^3+ab^2+a^2b+b^3}{4}$$

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = 0$$

$$\gamma = \frac{\mu_3}{\sigma^3} = 0$$

(2-6) التوزيع الأسي *Exponential Distribution*:

(ويسمى أحياناً توزيع لابلاس Laplace).

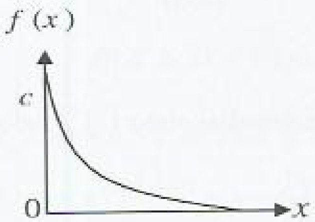
وهو من التوزيعات المهمّة في الإحصاء، وله تطبيقات كثيرة في الاقتصاد .

الشكل العام للتوزيع:

نقول عن المتغيّر العشوائي X إنّه يخضع للتوزيع الأسي إذا كان تابع كثافته من الشكل:

$$f(x) = \begin{cases} ce^{-cx} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$ ، حيث c مقدار ثابت موجب .

وبيانه من الشكل:

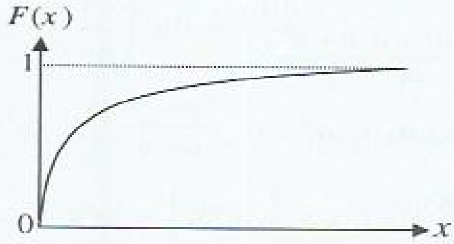


. من الواضح أنّ: $f(x) \geq 0$ (لاحظ أنّ c مقدار ثابت موجب).

. كذلك: $\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} f(x)dx = c\int_{0}^{\infty} e^{-cx}dx = -e^{-cx}]_0^{\infty} = 1$

وأنّ تابع توزيعه هو : $F(x) = \int_{-\infty}^{x} f(x)dx = \int_{0}^{x} f(x)dx = \begin{cases} 1-e^{-cx} & ; x \geq 0 \\ 0 & ; x < 0 \end{cases}$

وبيانه من الشكل:



ومن الظواهر التي يصفها هذا التوزيع: زمن الانتظار لوقوع حادثة ما، الوقت اللازم لتعطيل بعض الأجهزة الالكترونيّة ، عمر مصباح كهربائي ...

ـ القيم المميّزة للتوزيع الأسي:

لدينا:

$$\alpha_r = E[X^r] = \int_0^{\infty} x^r f(x)dx = c\int_0^{\infty} x^r e^{-cx} dx$$

وإذا فرضنا أنّ: $y = cx$ فإنّ : $dy = c\,dx \Rightarrow dx = \frac{dy}{c}$ وبالتبديل يكون:

$$\alpha_r = \frac{1}{c^r}\int_0^{\infty} y^r e^{-y} dy$$

ومن تعريف دالة جاما (Γ) وخواصها(سنتطرق لها لاحقاً) نجد أنّ:

$$\alpha_r = \frac{1}{c^r}\int_0^{\infty} y^r e^{-y} dy = \frac{\Gamma(r+1)}{c^r} = \frac{r!}{c^r}$$

ومنه:

$$\alpha_0 = 1 \quad \Rightarrow \mu_0 = \alpha_0 = 1$$

$$\alpha_1 = \frac{1}{c} \quad \Rightarrow \mu_1 = \alpha_1 - \alpha_1 = 0$$

$$\alpha_2 = \frac{2}{c^2} \quad \Rightarrow \mu_2 = \alpha_2 - \alpha_1^2 = \frac{1}{c^2} \equiv \sigma^2$$

$$\alpha_3 = \frac{6}{c^3} \quad \Rightarrow \mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = \frac{2}{c^3}$$

$$\Rightarrow \gamma = 2$$

$$\Psi_X(t) = E(e^{tX}) = \int_0^{\infty} e^{tx} f(x)dx = c\int_0^{\infty} e^{tx} e^{-cx} dx$$

$$= c\int_0^{\infty} e^{-x(c-t)} dx = c[\frac{-1}{c-t} e^{-x(c-t)}]_0^{\infty} = \frac{c}{c-t}$$

ومنه يمكن الحصول أيضاً على العزوم السابقة حيث: $\alpha_r = \Psi_X^{(r)}(0)$ (تحقّق من ذلك).

مثال:

إذا فرضنا أنّ عمر المصابيح الكهربائية يتبع التوزيع الأسي بمتوسط 1000 ساعة:

1- ما احتمال أن يعمل أحد المصابيح أكثر من 2000 ساعة ؟

2- ما احتمال أن يحترق أحد المصابيح خلال 100 ساعة ؟

الحل:

إذا فرضنا أنّ X هو عمر المصباح بالساعات، وأنّ μ هو متوسط العمر، عندئذٍ:

بما أنّ: $\mu = \frac{1}{c} = 1000 \Rightarrow c = \frac{1}{1000} \Rightarrow f(x) = \frac{1}{1000} e^{-\frac{x}{1000}} \quad ; x \geq 0$

وحيث: $P(X \leq x) = F(x) = 1 - e^{-cx}$

فإنّ: $P(X > x) = e^{-cx} \Rightarrow P(X > 2000) = e^{-\frac{2000}{1000}} = e^{-2} \approx 0.135$

ويكون المطلوب الثاني: $P(X \leq 100) = 1 - e^{-100/1000} = 1 - e^{-0.1} = 1 - 0.9 = 0.1$

(3-6) توزيع جاما *Gamma Distribution*:

قبل الدخول في توزيع Γ، نرى أنّه من المفيد التعرّف على دالّة Γ كدالّة رياضية:

دالّة جاما *gamma function*:

نرمز لها بالرمز $\Gamma(n)$، وتعرّف رياضيّاً بالشكل:

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$$

عندما تكون n موجبة يكون التكامل متقارباً، ويكون:

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$
$$\Gamma(1) = 1$$
$$\Gamma(n) = (n-1)\Gamma(n-1)$$

أما عندما تكون n عدداً صحيحاً موجباً فيكون:

$$\Gamma(n) = (n-1)(n-2) \ldots 3.2.1 = (n-1)!$$

$$\Gamma(n+1) = n!$$

. الشكل العام لتوزيع Γ :

نقول عن المتغيّر العشوائي X إنّه يخضع لتوزيع جاما Γ ونرمّزه بالشكل $X : \Gamma(n,c)$ ، إذا كان له تابع الكثافة التالي:

$$X : \Gamma(n,c) \ f(x) = \frac{c^n}{\Gamma(n)} e^{-cx} . x^{n-1} \quad ; \ 0 \le x \le \infty \Leftrightarrow$$

حيث: c و n عددان حقيقيان موجبان، وهو الشكل العام لتوزيع Γ .

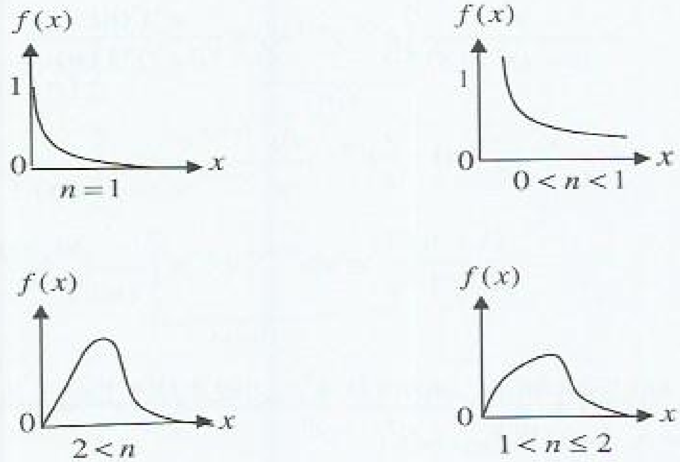
من الواضح أنّ:

$$f(x) \ge 0 \ , \ \forall x \in [0, \infty]$$

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x) dx &= \frac{c^n}{\Gamma(n)} \int_0^{\infty} e^{-cx} x^{n-1} dx \\
&= \frac{c^n}{\Gamma(n)} \int_0^{\infty} e^{-y} \frac{y^{n-1}}{c^{n-1}} \frac{dy}{c} \quad ; \ y = cx \\
&= \frac{1}{\Gamma(n)} \underbrace{\int_0^{\infty} e^{-y} . y^{n-1} \ dy}_{\Gamma(n)} = \frac{\Gamma(n)}{\Gamma(n)} = 1
\end{aligned}$$

ومن الشكل العام للتوزيع نجد أنّه من أجل: $n = 1$ يأخذ توزيع Γ الشكل: $f(x) = ce^{-cx}$ وهو يمثل الصيغة العامة للتوزيع الأسي (لابلاس) كما نلاحظ ومن أجل: $c = 1$ يأخذ الصيغة البسيطة التالية: $f(x) = \frac{1}{\Gamma(n)} e^{-x} x^{n-1}$

وأنَّ بيانه يعتمد على n ، وله أحد الأشكال التالية:



نهاية عظمى عند $x = n - 1$
نهاية صغرى عندما $x \to \infty$

نهاية عظمى عند: $x = n - 1$
نهاية صغرى عندما: $x = 0$

. القيم المميّزة لتوزيع Γ:

$$\Psi_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x)dx$$

$$= \frac{c^n}{\Gamma(n)} \int_0^\infty e^{tx}\, e^{-cx}\, x^{n-1} dx$$

$$= \frac{c^n}{\Gamma(n)} \int_0^\infty e^{-x(c-t)}\, x^{n-1}\, dx$$

وبفرض أنَّ: $x(c-t) = y \Rightarrow x = \dfrac{y}{c-t} \; , \; dx = \dfrac{dy}{c-t}$

نجد أنَّ:

$$\Psi_X(t) = \frac{c^n}{\Gamma(n)} \int_0^\infty e^{-y} \frac{y^{n-1}}{(c-t)^{n-1}} \frac{dy}{c-t}$$

$$= \frac{c^n}{(c-t)^n \Gamma(n)} \underbrace{\int_0^\infty e^{-y} y^{n-1} dy}_{\Gamma(n)} = \frac{c^n \Gamma(n)}{(c-t)^n \Gamma(n)}$$

$$= \frac{c^n}{(c-t)^n} = (1 - \frac{t}{c})^{-n}$$

أي: $\Psi_X(t) = (1 - \frac{t}{c})^{-n}$

ويمكن كتابتها بالشكل:

$$\Psi_X(t) = 1 + \frac{n}{c}\frac{t}{1!} + \frac{n(n+1)}{c^2}\frac{t^2}{2!} + \frac{n(n+1)(n+2)}{c^3}\frac{t^3}{3!} + \dots$$

بحيث يمكن الحصول منها على العزوم دفعة واحدة (كما ورد في دراسة الدالّة المولّدة للعزوم) بمعنى:

$\alpha_1 = E(X) = \frac{n}{c}$: وهي عبارة عن أمثال الحد $\frac{t}{1!}$

$\alpha_2 = \frac{n(n+1)}{c^2}$: وهي عبارة عن أمثال الحد $\frac{t^2}{2!}$

$\alpha_3 = \frac{n(n+1)(n+2)}{c^3}$: وهي عبارة عن أمثال الحد $\frac{t^3}{3!}$

وهكذا...

- التباين:

$$\sigma^2(X) = \alpha_2 - \alpha_1^2 = \frac{n}{c^2}$$

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = \frac{2n}{c^3}$$

$$\gamma = \frac{\mu_3}{\sigma^3} = \frac{2}{\sqrt{n}}$$

$$\alpha_i = E(X^i) = \int_0^\infty x^i f(x)dx = \frac{c^n}{\Gamma(n)}\int_0^\infty x^i e^{-cx} x^{n-1}dx$$

$$= \frac{c^n}{\Gamma(n)}\int_0^\infty e^{-cx} x^{n+i-1}dx \quad : \quad cx = y$$

$$= \frac{c^n}{\Gamma(n)}\int_0^\infty e^{-y} \frac{y^{n+i-1}}{c^{n+i-1}}\frac{dy}{c}$$

$$= \frac{c^n}{c^{n+i}\Gamma(n)}\underbrace{\int_0^\infty e^{-y} y^{n+i-1}dy}_{\Gamma(n+i)} = \frac{\Gamma(n+i)}{c^i\,\Gamma(n)}$$

ومن هذه العلاقة (وبالاستفادة من خواص التابع Γ) يمكن مباشرةً حساب الصفات القياسيّة أو المميّزة للتوزيع. فمثلاً:

$$\alpha_1 = E(X) = \frac{\Gamma(n+1)}{c\,\Gamma(n)} = \frac{n\Gamma(n)}{c\Gamma(n)} = \frac{n}{c}$$

$$\alpha_2 = \frac{\Gamma(n+2)}{c^2\,\Gamma(n)} = \dots = \frac{n(n+1)}{c^2}$$

نتيجة:

إذا كان: X_1 و X_2 متغيّرين عشوائيين مستقلين بحيث:

$$X_2 : \Gamma(n_2, c) \ , \ X_1 : \Gamma(n_1, c)$$

فإنَّ: $X = X_1 + X_2 : \Gamma(n_1 + n_2 \ , c)$

لدينا:

$$\Psi_{X_1}(t) = (1-\frac{t}{c})^{-n_1} \quad , \quad \Psi_{X2}(t) = (1-\frac{t}{c})^{-n_2}$$

وبالتالي يكون:

$$\Psi_X(t) = \Psi_{X_1+X_2}(t) = \Psi_{X_1}(t).\Psi_{X_2}(t)$$
$$= (1-\frac{t}{c})^{-n_1}.(1-\frac{t}{c})^{-n_2} = (1-\frac{t}{c})^{-(n_1+n_2)}$$

ومنه: $X : \Gamma(n_1 + n_2 \ , c)$

ويمكن تعميم هذه النتيجة لأي عدد من المتغيِّرات المستقلة.

مثال:

إذا كانت أعمار القطع الإلكترونية X بالسنين تتبع التوزيع الاحتمالي التالي :

$$f(x) = \lambda \ x \ e^{-2x} \ ; \ x \geq 0 \ ، \ \lambda \ \text{ثابت}$$

المطلوب:

1- أوجد قيمة الثابت λ.

2- عند الاختيار عشوائياً لقطعة الكترونية، ما احتمال أن تعمل أقل من 4 شهور ؟

الحل: من الواضح أنَّ X يتبع توزيع Γ بالوسيط $c = 2$ و: $n = 2$

حيث: $f(x) = \dfrac{c^n}{\Gamma(n)} e^{-cx} . x^{n-1} \quad ; \ 0 \leq x \leq \infty$

وبالمقارنة، نجد المطلوب الأول بالشكل:

$$\begin{cases} n-1 = 1 \Rightarrow n = 2 \\ c = 2 \end{cases} \Rightarrow \left\{ \lambda = \frac{c^n}{\Gamma(n)} = \frac{4}{\Gamma(2)} = \frac{4}{1} = 4 \right.$$

والمطلوب الثاني (حيث 4 شهور تعادل $\frac{1}{3}$ سنة) يكون:

$$P(X \leq \frac{1}{3}) = 4 \int_0^{\frac{1}{3}} x e^{-2x} dx = \ldots$$

(4-6) توزيع بيتا (β) *Beta Distribution*:

سنشير في البداية إلى ما يسمى بدالّة بيتا وتعريفها الرياضي:

دالّة β: *beta function*:

تُرمَز وتعرَّف رياضياً بالشكل: $\beta(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1} dx \quad ; \ n, m > 0$

وأنّ العلاقة بين β و Γ هي كما في الشكل: $\beta(m,n) = \dfrac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$

أضف إلى أنّ: $\beta(m,n) = \beta(n,m)$

. الشكل العام لتوزيع β:

نقول عن المتغيّر العشوائي X إنّه يخضع لتوزيع بيتا في المجال $[0,1]$ إذا كان تابع كثافته من الشكل:

$$f(x) = \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1} \quad ; \begin{cases} 0 \le x \le 1 \\ n, m > 0 \end{cases}$$

ويسمى: توزيع بيتا من النوع الأول.

ونرمز للمتغيّر X الذي يخضع لتوزيع بيتا بالرمز $X : \beta(m,n)$

*** الصفات المميّزة لتوزيع بيتا:**

. العزم الأولي من الرتبة i :

$$\alpha_i = E(X^i) = \int_0^1 x^i f(x)dx = \frac{1}{\beta(m,n)} \int_0^1 x^i x^{m-1}(1-x)^{n-1} dx$$
$$= \frac{1}{\beta(m,n)} \int_0^1 x^{m+i-1}(1-x)^{n-1} dx = \frac{\beta(m+i,n)}{\beta(m,n)}$$

ومنه ينتج:

$$\alpha_1 = E(X) = \frac{\beta(m+1,n)}{\beta(m,n)} = \frac{\Gamma(m+1)\Gamma(n)\Gamma(m+n)}{\Gamma(m+n+1)\Gamma(m)\Gamma(n)} ; \begin{cases} \Gamma(n) = (n-1)! \\ \Gamma(n+1) = n! \end{cases}$$
$$= \frac{m!(m+n-1)!}{(m+n)!(m-1)!} = \frac{m}{m+n}$$

$$\alpha_2 = \frac{m(m+1)}{(m+n)(m+n+1)}$$

$$\sigma^2(X) = \alpha_2 - \alpha_1^2 = \frac{m.n}{(m+n)^2(m+n+1)}$$

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = \frac{2mn(m-n)}{(m+n)^3(m+n+1)(m+n+2)}$$

$$\gamma = \frac{\mu_3}{\sigma^3} = \frac{2(n-m)\sqrt{m+n+1}}{(m+n+2)\sqrt{mn}}$$

ومن أجل: $m = n \Leftarrow \gamma = 0$ والتوزيع متناظر.

$$\psi_x(t) = E(e^{tX}) = \int_0^1 e^{tx} f(x)dx = \frac{1}{\beta(m,n)} \int_0^1 e^{tx} x^{m-1}(1-x)^{n-1} dx$$

$$= \frac{1}{\beta(m,n)} \int_0^1 x^{m-1}(1-x)^{n-1}[1 + \frac{tx}{1!} + \frac{t^2x^2}{2!} + ...]dx$$

$$= \frac{1}{\beta(m,n)} \{\int_0^1 x^{m-1}(1-x)^{n-1} dx + \frac{t}{1!}\int_0^1 x^m (1-x)^{n-1} dx + \frac{t^2}{2!}\int_0^1 x^{m+1}(1-x)^{n-1} dx + ...\}$$

$$= \frac{1}{\beta(m,n)} \{\beta(m,n) + \frac{t}{1!}\beta(m+1,n) + \frac{t^2}{2!}\beta(m+2,n) + ...\}$$

وبالاستفادة من خواص الدالّة β يمكن أن نكتب:

$$\psi_X(t) = 1 + \frac{m}{m+n}\frac{t}{1!} + \frac{m(m+1)}{(m+n)(m+n+1)}\frac{t^2}{2!} + \frac{m(m+1)(m+2)}{(m+n)(m+n+1)(m+n+2)}\frac{t^3}{3!} + ...$$

ونذكّر (من دراسة الدالّة المولّدة للعزوم) أنّ:

$\alpha_1 = E(X) = \dfrac{m}{m+n}$ أمثال الحد $;\dfrac{t}{1!}$

$\alpha_2 = \dfrac{m(m+1)}{(m+n)(m+n+1)}$ أمثال الحد $;\dfrac{t^2}{2!}$

$\alpha_3 = \dfrac{m(m+1)(m+2)}{(m+n)(m+n+1)(m+n+2)}$ أمثال الحد $;\dfrac{t^3}{3!}$

وهكذا... يمكن استنتاج جميع العزوم دفعة واحدة كما أكّدنا سابقاً.

مثال:

إذا كان للمتغيّر العشوائي المستمر X تابع الكثافة التالي: $f(x) = \lambda x^2(1-x) \;\; ; \;\; 0 \le x \le 1$

المطلوب:

1 . أوجد قيمة λ.

2 . أوجد: $E(X)$ ، $\sigma^2(X)$.

الحل: واضح أنَّ X يتبع توزيع β بالوسيطين m و n حيث دالة كثافته:

$$f(x)=\frac{1}{\beta(m,n)}x^{m-1}(1-x)^{n-1}\quad ;0\le x\le 1$$

بالمقارنة ينتج: $\begin{cases} n-1=1 \Rightarrow n=2 \\ m-1=2 \Rightarrow m=3 \end{cases}$

ويكون المطلوب الأول: $\lambda=\dfrac{1}{\beta(m,n)}=12$

والمطلوب الثاني: $P(X\le 0.6)=12\int\limits_0^{0.6}x^2(1-x)dx=[4x^3-3x^4]_0^{0.6}=0.475$

$$E(X)=\frac{m}{m+n}=\frac{3}{3+2}=\frac{3}{5}$$

$$\sigma^2(X)=\frac{mn}{(m+n)^2(m+n+1)}=\frac{6}{25\times 6}=\frac{1}{25}$$

(5-6) التوزيع الطبيعي *Normal Distribution* :

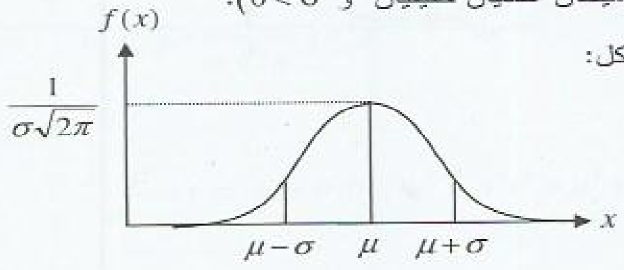
يعتبر التوزيع الطبيعي من الأعمدة الأساسيَّة في الإحصاء ، وهو أهم التوزيعات المستمرَّة على الإطلاق من حيث تطبيقاته الكثيرة والمتنوعة في مختلف المجالات، والعديد من المجتمعات الاحصائيَّة هي بالواقع مجتمعات طبيعيَّة.

يعطى تابع كثافة التوزيع الطبيعي بالشكل:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\quad ;\ -\infty\le x\le\infty$$

حيث: (σ,μ وسيطان عدديان حقيقيان و $0<\sigma$).

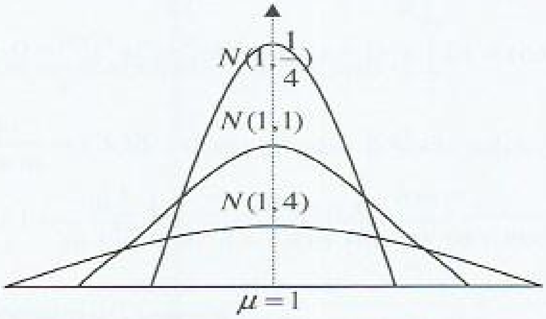
وبيانه كما في الشكل:



وهو متناظر حول μ (أي: $f(\mu+\sigma)=f(\mu-\sigma)$) ويأخذ شكل جرس (أو ناقوس)

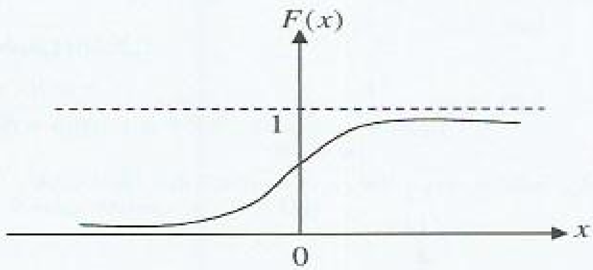
ويصل إلى نهايته العظمى عند النقطة $x=\mu$، ويتقارب بسرعة من محور السينات عندما يبتعد x عن μ (أي يتقارب طرفا منحنى التوزيع الطبيعي من الصفر عندما $x \to \infty$ ، $x \to -\infty$).

/الدالّة $f(x)$ لا تحدّد منحنياً واحداً بعينه وإنما تحدّد الشكل العام لعائلة من المنحنيات – لاحظ الشكل التالي حيث القيم الصغيرة لـ σ تعني قمة مرتفعة، أي توزيعاً أقل انتشاراً على جانبي μ/.



ويعطى تابع توزيعه بالعلاقة: $F(x)=\dfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx$

ويأخذ بيانه الشكل التالي:



* القيم المميّزة للتوزيع الطبيعي:

. التوقع الرياضي:

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} xe^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx$$

بفرض أنّ: $\frac{x-\mu}{\sigma}=t \Rightarrow dx=\sigma\, dt$

$$E(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma t+\mu)e^{-\frac{t^2}{2}}dt$$

$$=\frac{1}{\sqrt{2\pi}}[\sigma\underbrace{\int_{-\infty}^{\infty} t\, e^{-\frac{t^2}{2}}dt}_{*}+\mu\underbrace{\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt}_{**}]$$

ولكن:

$$*=\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}d(\frac{t^2}{2})=-e^{-\frac{t^2}{2}}]_{-\infty}^{+\infty}=0$$

$$**=\sqrt{2\pi}\quad ;\ \frac{t^2}{2}=y \Rightarrow t=\sqrt{2y}\Rightarrow dt=\frac{dy}{\sqrt{2y}}$$

$$\Rightarrow \sqrt{2}\int_{0}^{\infty} y^{-\frac{1}{2}}\, e^{-y}\, dy=\sqrt{2}\ \Gamma(\frac{1}{2})=\sqrt{2\pi}$$

ومنه: $E(X)=\frac{1}{\sqrt{2\pi}}[0+\mu\sqrt{2\pi}]=\mu$

وهوالوسيط العددي الأول في تابع الكثافة المذكور.

. التباين: نأخذ العلاقة: $\sigma^2(X)=E(X^2)-\{E(X)\}^2$:

نوجد $E(X^2)$:

$$E(X^2)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2\, e^{-\frac{1}{2}(\frac{x-\mu}{2})^2}dx$$

نضع:

$$\frac{x-\mu}{\sigma}=t \Rightarrow dx=\sigma\, dt\ ;\ x^2=\sigma^2t^2+\mu^2+2\mu\sigma t$$

$$E(X^2)=\frac{1}{\sqrt{2\pi}}[\ \sigma^2\int_{-\infty}^{\infty}t^2 e^{-\frac{t^2}{2}}dt+\mu^2\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt+2\mu\sigma\int_{-\infty}^{\infty}t\ e^{-\frac{t^2}{2}}dt\ ]$$

ولحساب التكاملات الثلاثة في العلاقة الأخيرة نلاحظ أنّ:

$$\int_{-\infty}^{\infty}t^2 e^{-\frac{t^2}{2}}dt=2\sqrt{2}\,\Gamma(\frac{3}{2})=\sqrt{2\pi}\quad ;\ \Gamma(\frac{3}{2})=\Gamma(\frac{3}{2}-1+1)$$

$$=(\frac{3}{2}-1)\Gamma(\frac{3}{2}-1)=\frac{1}{2}\Gamma(\frac{1}{2})=\frac{\sqrt{\pi}}{2}$$

$$\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}\quad ;\ \frac{t^2}{2}=y$$

$$\int_{-\infty}^{\infty}t\ e^{-\frac{t^2}{2}}dt=0$$

بالتبديل ينتج: $E(X^2)=\frac{1}{\sqrt{2\pi}}[\sigma^2\sqrt{2\pi}+\mu^2\sqrt{2\pi}+0]=\sigma^2+\mu^2$

ويكون التباين: $\sigma^2(X)=\sigma^2+\mu^2-\mu^2=\sigma^2$

وهو الوسيط العددي الثاني في تابع الكثافة، وهو موجب دوماً كما نعلم مما يجعل $f(x)$ غير سالب. واضح أنّ الانحراف المعياري هو: $\sigma(X)=\sigma$.

(اختصاراً نكتب: $X : N(\mu,\sigma^2)$ ونعني أنّ المتغيّر X يخضع لتوزيع طبيعي بمتوسط μ وتباين σ^2). علماً أنّه في بعض المراجع يكتب بدلالة المتوسط والانحراف المعياري بالشكل: $X : N(\mu,\sigma)$. وسنستخدم الاختصار الأول.

. الدالّة المولّدة للعزوم:

$$\Psi_X(t)=E(e^{tX})=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{tx}\,e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx$$

$$\frac{x-\mu}{\sigma} = y \Rightarrow dx = \sigma\, dy$$

$$\Rightarrow \Psi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma y+\mu)}\, e^{-\frac{y^2}{2}}\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\, e^{\mu t} \int_{-\infty}^{\infty} e^{t\sigma y - \frac{y^2}{2}}\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\, e^{\mu t} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(y-t\sigma)^2 + t^2\sigma^2 - t^2\sigma^2]}\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\, e^{\mu t}\, e^{\frac{t^2\sigma^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t\sigma)^2}\, dy$$

$$= \frac{1}{\sqrt{2\pi}}\, e^{\mu t + \frac{\sigma^2}{2}t^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y-t\sigma)^2}\, dy$$

وإذا فرضنا في العلاقة الأخيرة أنّ: $z = (y - t\sigma)$، فإنّ قيمة التكامل تساوي $\sqrt{2\pi}$،

ونحصل بالنتيجة على أنّ: $\Psi_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

وكل متغيّر له الدالّة المولّدة للعزوم هذه، نقول إنّه يخضع للتوزيع الطبيعي $N(\mu, \sigma^2)$.

التوزيع الطبيعي المعياري:

التوزيع الطبيعي المعياري هو التوزيع الطبيعي الذي لأجله يكون: ($\mu = 0$ و $\sigma = 1$).

وبالحقيقة إذا كان: $X : N(\mu, \sigma^2)$ وكان $Z = \dfrac{X-\mu}{\sigma}$ فإن: $Z : N(0,1)$

ونسمي عندئذٍ المتغيّر الجديد Z الذي حصلنا عليه بالتحويل $Z = \dfrac{X-\mu}{\sigma}$، نسميه بالمتغيّر المعياري، حيث يوافق كل قيمة من قيم X قيمة واحدة لـ Z وبالعكس. وتسمى قيم Z عندئذٍ بالقيم المعياريّة المقابلة لقيم X.

مثلاً: من أجل $X : N(50,36)$ يكون للمتغيّر العشوائي $Z = \frac{X-50}{6}$ توزيع طبيعي معياري، والقيمة المعياريّة المقابلة للقيمة $x_1 = 68$ (على سبيل المثال) هي:

$$Z_1 = \frac{68-50}{6} = \frac{18}{6} = 3$$

المساحات تحت التوزيع الطبيعي:

من الواضح أنّ المساحة تحت أي منحني كثافة طبيعي (وفوق محور السينات) تعتمد على μ و σ، ولما كانت أسرة المنحنيات الطبيعيّة كثيرة جداً (لا حصر لها) فإننا لا نستطيع وضع جداول لكل قيم μ و σ . وبما أنّه يمكن تحويل أي توزيع طبيعي $X : N(\mu, \sigma^2)$ إلى توزيع طبيعي معياري $Z = \frac{X-\mu}{\sigma}$، فإنّ جدول التوزيع الطبيعي المعياري يكون كافياً لحساب هذه المساحة.

مثال (1):

إذا كان $Z : N(0,1)$ أوجد:

أ ـ المساحة تحت منحني الكثافة الواقعة إلى اليسار من $Z = 1.24$

أي أوجد: $P(Z < 1.24)$

ب ـ $P(Z > 2)$

ج ـ $P(0 < Z < 1.25)$

الحل:

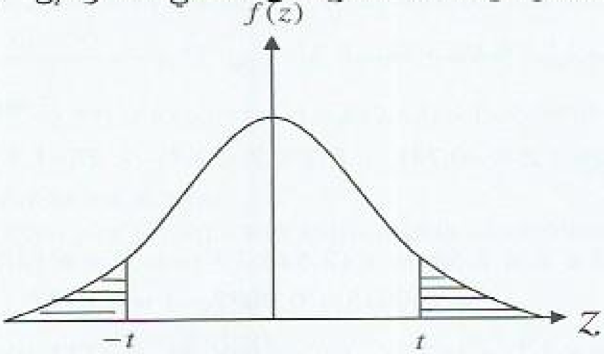
أ) لحساب $P(Z < 1.24)$ ندخل الجدول وفق السطر 1.2 والعمود 0.04 فنجد القيمة 0.8925 . أي العدد 0.8925 يمثّل المساحة المطلوبة بمعنى: $P(Z < 1.24) = 0.8925$.

ب) $P(Z > 2.0) = 1 - P(Z \leq 2.0) = 1 - 0.9772 = 0.0228$ (استخدمنا المتممة)

ج) $P(0 < Z < 1.25) = P(Z < 1.25) - P(Z < 0) = 0.8944 - 0.5000 = 0.3944$

ويشكل عام:

. من المعروف أنَّ: $P(Z \le t) = F(t)$ وهي المساحة تحت المنحني الطبيعي المعياري الواقعة إلى اليسار من النقطة t ، وبما أنَّ المنحني متناظر فإنَّ: $F(-t) = 1 - F(t)$



حيث: $F(-t)$ هي المنطقة المظللة على اليسار من $Z = -t$

$F(t)$ هي مجموع المنطقة المظللة على اليسار والمنطقة غير المظللة في الوسط.

بالتالي: $1 - F(t)$ هي المنطقة المظللة في أقصى اليمين.

وبحكم التناظر فإن المنطقتين المظللتين على الشكل متساويتين أي: $F(-t) = 1 - F(t)$ (سنبرهنها رياضيّاً)، وبالتالي لحساب $F(-t)$ يكفي حساب $F(t)$ من الجدول ثم نطرح القيمة الناتجة من 1 (هذا في حال الجدول لا يحوي قيماً سالبة لـ Zوهو الغالب).

. أيضاً من المعروف (باستخدام المتممة) أنَّ: $P(A) = 1 - P(\overline{A})$

بالتالي نجد مباشرة أنَّ: $P(Z > t) = 1 - P(Z \le t) = 1 - F(t)$

. ولو كان المطلوب هو حساب: $P(a < Z \le b)$

فإنه يمكن التعبير عن الحدث $(Z \le b)$ كاجتماع لحدثين منفصلين على الشكل:

$$(Z \le b) = (Z \le a) \cup (a < Z \le b)$$

بالتالي: $P(Z \le b) = P(Z \le a) + P(a < Z \le b)$

ومنه: $P(a < Z \le b) = P(a \le Z \le b) = P(a < Z < b) = F(b) - F(a)$

وهذا واضح من كون الاحتمال الموافق لنقطة في التوزيعات المستمرّة يساوي صفراً كما نعلم. وهذا تعبير واقعي عن استحالة توصل الإنسان إلى أجهزة قياس دقيقة بصورة مطلقة.

مثال (2):

إذا كان: $Z : N(0,1)$ أوجد:

$P(-1.3 \le Z \le 2.54)$ ، $P(2 \le Z \le 2.7)$ ، $P(-1.36 \le Z \le -0.74)$.

الحل:

$$P(-1.3 \le Z \le 2.54) = F(2.54) - F(-1.3) = F(2.54) - [1 - F(1.3)]$$
$$= 0.9945 + 0.9032 - 1 = 0.8977$$

$$P(2 \le Z \le 2.7) = F(2.7) - F(2) = 0.9972 - 0.9772 = 0.020$$

$$P(-1.36 \le Z \le -0.74) = F(-0.74) - F(-1.36) = [1 - F(0.74)] - [1 - F(1.36)]$$
$$= -F(0.74) + F(1.36) = 0.7704 + 0.9131 = 0.1427$$

واضح أنّ الحسابات السابقة تتعلق بالمتغيّر المعياري Z.

- إذا كان: $X : N(\mu, \sigma^2)$ فإنّه لحساب الاحتمالات المتعلقة بالمتغيّر X ، فإنّنا نحوّل جميع قيم X إلى قيم معيارية مقابلة لها ثم نستعمل جدول التوزيع الطبيعي المعياري. وهذا ما أكدناه أنّ علاقة المعايرة $Z = \dfrac{X - \mu}{\sigma}$ تستخدم بطريقة عكسيّة. فنحن نحدّد سلفاً Z أي القياس المطلوب على السلم المعياري، ونريد القياس المقابل له على السلم الأصلي (قبل المعايرة).

مثال (3):

إذا كان: $X : N(50,49)$ أوجد:

$P(X \ge 64)$ ، $P(X \le 32)$ ، $P(30 < X \le 65)$ ، $P(X \le 46)$

الحل:

واضح أنّ: $Z = \dfrac{X - 50}{7} : N(0,1)$ أي: يخضع لتوزيع طبيعي معياري.

بالتالي: القيمة المعياريَّة المقابلة للقيمة 64 هي: $Z = \frac{64-50}{7} = 2$

ويكون عندئذٍ:

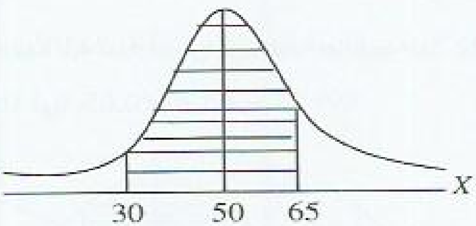
$$P(X \geq 64) = P(Z > 2) = 1 - P(Z \leq 2) = 1 - F(2) = 1 - 0.9772 = 0.0228$$

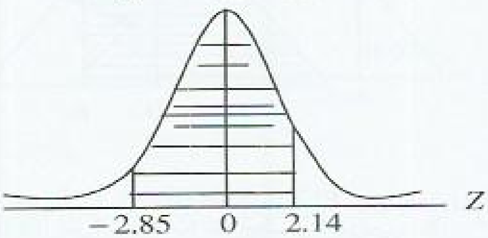
والقيمـة المعياريَّـة المقابلـة للقيمـة 32 هـي: $Z = \frac{32-50}{7} = -2.57$ وعندئـذٍ يكـون:

$$P(X \leq 32) = P(Z \leq -2.57) = F(-2.57) = 1 - F(2.57) = 1 - 0.9949 = 0.0051$$

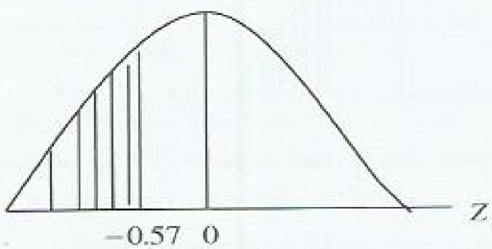
وبالمثل يكون:

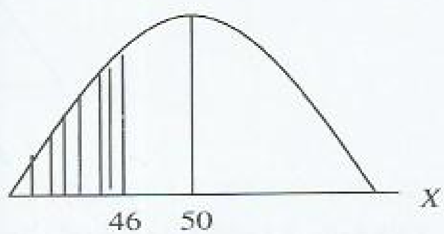
$$P(30 < X \leq 65) = P(-2.85 < Z \leq 2.14)$$
$$= F(2.14) - F(-2.85) = F(2.14) - [1 - F(2.85)] = 0.9838 - 1 + 0.9978 = 0.9816$$





وبالمثل نجد: $P(X \leq 46) = P(Z \leq -0.57) = 0.2843$ ويمكن تمثيلها أيضاً بالشكل:





وكما أنّ جدول التوزيع الطبيعي المعياري يحدّد المساحات تحت المنحني فإنّه أيضاً يعطي قيم Z التي تقابلها مساحات معيّنة إلى يسارها حيث نأخذ أقرب قيمة للمساحة المعطاة.

مثال (4):

إذا كان $Z : N(0,1)$ أوجد:

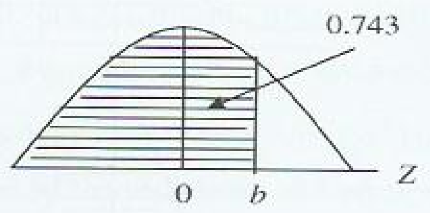
1. قيمة b بحيث يكون: $P(Z \le b) = 0.743$

2. قيمة b بحيث يكون: $P(Z \ge b) = 0.025$

الحل:

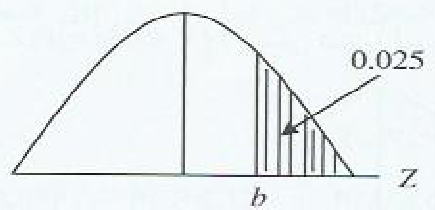
1) نبحث داخل الجدول عن القيمة 0.743 أو عن أقرب قيمة لها فنجد أنّها 0.7422

وقيمة Z المقابلة نجدها 0.65 أي: $b = 0.65$



2) $P(Z > b) = 0.025$ يعني أنّ : $P(Z \le b) = 0.975$ ومن الجدول مباشرة نجد:

$b = 1.96$



واضح هنا أيضاً أنّ الحسابات تتعلق بالمتغيّر المعياري Z.

- وإذا كان: $X : N(\mu, \sigma^2)$ فإنّه بمعرفة المساحة يمكن إيجاد قيمة X المقابلة، وذلك بحل المسألة في حالة التوزيع الطبيعي المعياري ثم نحوّل القيمة المعياريّة الناتجة إلى قيمة عادية فنحصل على قيمة X المطلوبة.

مثال (5):

إذا كان: $X : N(60,81)$

أوجد a بحيث يكون: $P(X > a) = 0.02$

الحل: نوجد أولاً b بحيث يكون: $P(Z > b) = 0.02$

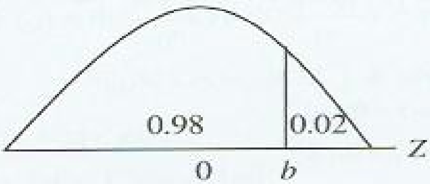
بالتالي: $P(Z \le b) = 1 - P(Z > b) = 1 - 0.02 = 0.98$

ومن الجدول الطبيعي المعياري نجد قيمتين قريبتين من 0.98 هما: 0.9798 تقابلها $Z = 2.05$ ، ثم القيمة 0.9803 وتقابلها $Z = 2.06$ ولذلك نأخذ القيمة المتوسطة بينهما أي نأخذ:

$$Z = \frac{2.05 + 2.06}{2} = 2.055$$

وهذا يعني أنّ: $Z = 2.055$ هي القيمة المعياريّة للقيمة a ، الأمر الذي يعني أنّ:

$$2.055 = \frac{a - 60}{9} \Rightarrow a = 9 \times 2.055 + 60 = 78.495$$



ملاحظات وأمثلة:

1) إذا كان: $X : N(0,1)$ ورمزنا لتابع كثافته بالرمز $\varphi(x)$ ، ولتابع توزيعه بالرمز $\Phi(x)$ يكون:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad ; -\infty < x < \infty$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

ومـن جـدول التوزيـع الطبيعـي المعيـاري يمكـن حسـاب $\Phi(x)$ وذلـك مـن أجـل $x>0$.

وبالاستفادة من التناظر حيث: $\varphi(-x)=\varphi(x)$ يمكن الحصول على القيم التي لأجلها:

$x<0$ علماً أنَّ: $\Phi(-x)=1-\Phi(x)$ ، حيث:

$$\Phi(-x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-x}e^{-\frac{t^2}{2}}dt$$

$$-t=u\Rightarrow\left\{\begin{array}{l}t\to-\infty\Rightarrow u\to\infty\\ t\to-x\Rightarrow u\to x\end{array}\right\}\Rightarrow$$

$$\Phi(-x)=\frac{-1}{\sqrt{2\pi}}\int_{\infty}^{x}e^{-\frac{u^2}{2}}du$$

$$=\frac{1}{\sqrt{2\pi}}\int_{x}^{\infty}e^{-\frac{u^2}{2}}du=1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du=1-P(X\leq x)=1-\Phi(x)$$

أي:

$$\Phi(-x)=1-\Phi(x)$$
$$or:\ \Phi(-\lambda)=1-\Phi(\lambda)$$

2) إذا كان: $X:N(\mu,\sigma^2)$ وكان : $Y=aX+b$ ، فإنَّ: $Y:N(a\mu+b,|a|\sigma^2)$

البرهان:

$$\Psi_Y(t)=\Psi_{aX+b}(t)=e^{bt}.\Psi_X(at)=e^{bt}.e^{\mu at+\frac{1}{2}\sigma^2(at)^2}$$

$$=e^{(a\mu+b)t+\frac{a^2}{2}\sigma^2t^2}\Rightarrow Y:N(a\mu+b,|a|\sigma^2)$$

3) العلاقة بين $F(x)$ و $\Phi(x)$:

لدينا:

$$X:N(\mu,\sigma^2)\ \Rightarrow\ F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}dx$$

من أجل: $\frac{x-\mu}{\sigma}=t \Rightarrow x=\sigma t+\mu ; dx=\sigma dt$

يكون:

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{1}{2}t^2}\sigma dt$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}t^2} dt=\Phi(\frac{x-\mu}{\sigma})$$

أي:

$$F(x)=\Phi(\frac{x-\mu}{\sigma})$$

عندئذٍ يكون:

$$P(a<X\le b)=F(b)-F(a)=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})$$

4) من أجل: $X:N(0,1)$ و λ عدد موجب يكون:

$$P[|X-\mu|>\lambda\sigma]=1-P[|X-\mu|<\lambda\sigma]$$
$$=1-P[-\lambda\sigma<X-\mu<\lambda\sigma]$$
$$=1-\{\Phi(\lambda)-\Phi(-\lambda)\}$$
$$=2\{1-\Phi(\lambda)\} \quad ; \Phi(-\lambda)=1-\Phi(\lambda)$$

أي: $P[|X-\mu|>\lambda\sigma]=2\{1-\Phi(\lambda)\}$

بالتالي من أجل:

$$\lambda=1 \Rightarrow P[|X-\mu|>\sigma]=2\{1-\Phi(1)\}=0.3174$$

بمعنى آخر: من بين 10000 قيمة لـ X ، فإنّه يمكن أن يقع خارج المجال $[\mu-\sigma,\mu+\sigma]$ 3174 قيمة ، وبالتالي ما يقع داخله يشكل 68.26%

فمثلاً: إذا كان $X:N(100,100)$ فإنّ:

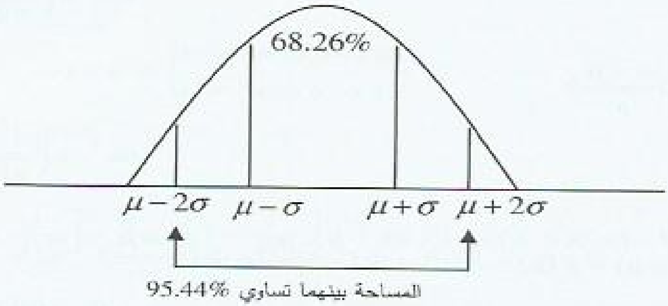
$$P(90<X<110)=\Phi(1)-\{1-\Phi(-1)\}=2\Phi(1)-1=0.6826$$

أي أنّ %68.26 من المساحة التي تحت منحني $\varphi(x)$، تقع بين $(\mu-\sigma,\mu+\sigma)$.

بمعنى: من بين 10000 قيمه لـ X يمكن أن يقع منها 6826 قيمة داخل المجال المذكور.

ومن أجل: $\lambda=2 \Rightarrow P[|X-\mu|>2\sigma]=2\{1-\Phi(2)\}=0.0456$

أي: من بين 10000 قيمة لـ X، فإنّه يمكن أن يقع خارج المجال $[\mu-2\sigma,\mu+2\sigma]$ فقط 456 قيمة.



والجداول الخاصة بالعلاقة:

$$P[|X-\mu|>\lambda_p\sigma]=\frac{p}{100} \quad ; 0\le p\le 100$$

تعطي قيمة p عند معرفة λ_p، وبالعكس.

5) إذا كانت: $X_i : N(\mu_i,\sigma_i^2),\ i=1,2,3,\dots,n$ مستقلة، فإنّ:

$$\sum_{i=1}^{n} X_i = X : N(\sum_{i=1}^{n}\mu_i,\sum_i \sigma_i^2)$$

البرهان:

لدينا: $\Psi_{X_i}(t)=e^{\mu_i+\frac{1}{2}\sigma_i^2 t^2}\ ,\ \forall i=\overline{1,n}$، وبالتالي:

$$\Psi_X(t) = E(e^{tX}) = E[e^{t(X_1+X_2+...+X_n)}] = E[e^{tX_1}.e^{tX_2}...e^{tX_n})$$
$$= E(e^{tX_1}).E(e^{tX_2}) \ldots E(e^{tX_n})$$
$$= \Psi_{X_1}(t).\Psi_{X_2}(t) \ldots \Psi_{X_n}(t)$$
$$= \prod_{i=1}^{n} \Psi_{X_i}(t) = \prod_{i=1}^{n} e^{\mu_i + \frac{1}{2}\sigma_i^2 t^2}$$
$$= e^{\sum_i \mu_i + \frac{1}{2}\sum_i \sigma_i^2 t^2}$$

أي أنَّ:

$$X : N(\sum_{i=1}^{n} \mu_i , \sum_i \sigma_i^2)$$

وبالتالي من أجل: $X_i : N(\mu, \sigma^2)$, $\forall i = 1,2,3,...,n$ فإنَّ المجموع $\sum_{i=1}^{n} X_i$ يخضع للتوزيع الطبيعي، أي: $X = \sum_{i=1}^{n} X_i : N(n\mu, n\sigma^2)$ وكذلك المتوسط الحسابي يخضع للتوزيع الطبيعي، أي: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i : N(\mu , \frac{\sigma^2}{n})$

ومنه نستنتج أنَّ: $\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} : N(0,1)$

(6-6) توزيع χ^2 *Distribution* (مربع كاي):

إذا كان تابع كثافة المتغيّر العشوائي X من الشكل:

$$f_n(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} e^{-\frac{x}{2}} . x^{\frac{n}{2}-1} \quad ; 0 \le X < \infty$$

عندها نقول إنَّ لـ X توزيعاً بشكل: χ^2 وبـ: n درجة حريَّة (*degrees of freedom*)، (n عدد صحيح موجب)، ونكتب اختصاراً: $X : \chi^2$.

وبالواقع أنَّ فكرة هذا التوزيع جاءت من التوزيع الطبيعي المعياري، بمعنى:

إذا كانت $X_1, X_2,..., X_n$ متغيّرات عشوائيّة مستقلة لكلٍ منها توزيع طبيعي معياري، فإنَّ مجموع مربعات هذه المتغيّرات يُرمَز له بالرمز: $\chi^2 = X_1^2 + X_2^2 + ... + X_n^2$، وتوزيع هذا المجموع هو ما يُعرَف في الإحصاء بتوزيع χ^2، وبالتالي عندما نقول إنَّ

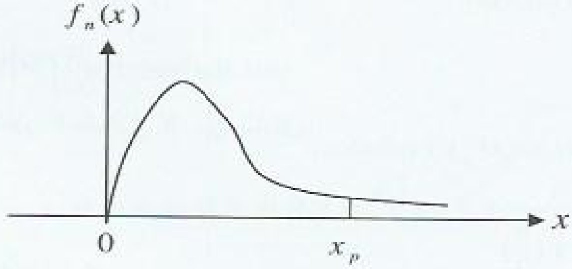
للمتغيِّر X توزيعاً بشكل χ_n^2، فهذا يعني أنَّ هذا المتغيِّر هو عبارة عن مجموع مربعات n من المتغيِّرات العشوائيَّة المستقلة والتي يخضع كل منها لتوزيع طبيعي معياري، أي أنَّ عدد درجات الحريَّة n يتبع عدد المتغيِّرات، ولهذا العدد n أهمية خاصة تظهر في أثناء استخدام جداول χ_n^2، وهذه الجداول تُعطي قيمة الاحتمال التالي:

$$P(X > x_0) = \int_{x_0}^{\infty} f_n(x)dx \; .$$

حيث x_0 قيمة محدَّدة ، وبالعكس، إذا كان المطلوب معرفة x التي لأجلها: $P(X > x) = \frac{p}{100}$ ، حيث p عدد موجب أصغر من 100، فإنَّه يتوجَّب علينا حل المعادلة: $P(X > x) = \int_{x}^{\infty} f_n(x)dx = \frac{p}{100}$

التي ستعطي جذراً تابعاً للكميَّة p مثل: $x = x_p$ ، حيث يُسمَّى هذا الجذر بالقيمة $\frac{p}{100}$ في توزيع χ_n^2.

ويكون لتابع كثافة χ_n^2 البيان التالي:



من الواضح أنَّ: $f_n(x) \geq 0$

كما أنَّ:

$$\int_{-\infty}^{\infty} f_n(x)dx = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^{\infty} e^{-\frac{x}{2}}.x^{\frac{n}{2}-1}d \qquad ; \frac{x}{2}=y \Rightarrow dx = 2dy$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^{\infty} e^{-y}.(2y)^{\frac{n}{2}-1}2dy$$

$$= \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2})} = 1$$

وبالتالي فإنّ $f_n(x)$ هو فعلاً تابع كثافة.

حالات خاصة:

1- من أجل: $n=2$ ، فإنّ تابع الكثافة السابق يأخذ الشكل:

$$f_2(x) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} & ,\forall x > 0 \\ 0 & ,\forall x \le 0 \end{cases}$$

وهو يمثل تابع كثافة التوزيع الأسي (لابلاس) (حيث : $c=\frac{1}{2}$).

2- وإذا بدّلنا في التابع $f_n(x)$ كل x بـ $2y$ يصبح:

$$f_n(2y) = \begin{cases} \frac{1}{\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-y} & ,\forall y > 0 \\ 0 & ,\forall y \le 0 \end{cases}$$

وهو يمثل تابع الكثافة لتوزيع Γ .

. الصفات المميّزة لتوزيع χ^2:

$$\Psi_X(t) = E(e^{tX}) = \int_0^\infty e^{tX} f_n(x)dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^\infty e^{tX} e^{-\frac{x}{2}} . x^{\frac{n}{2}-1} dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^\infty e^{-\frac{x}{2}(1-2t)} . x^{\frac{n}{2}-1} dx$$

$$; \frac{x}{2}(1-2t) = y \Rightarrow dx = \frac{2dy}{1-2t}$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^\infty e^{-y}[\frac{2y}{1-2t}]^{\frac{n}{2}-1}[\frac{2}{1-2t}]dy$$

$$= (1-2t)^{-\frac{n}{2}} \frac{1}{\Gamma(n)} \underbrace{\int_0^\infty e^{-y} y^{\frac{n}{2}-1}]dy}_{\Gamma(\frac{n}{2})} = (1-2t)^{-\frac{n}{2}}$$

وبتطبيق العلاقة: $\alpha_i = \Psi'_X(0)$ وبالحساب ينتج:

$$\alpha_1 = E(X) = n$$
$$\alpha_2 = n^2 + 2n$$
$$\alpha_3 = n^3 + 6n^2 + 8n$$
$$\sigma^2(X) = \alpha_2 - \alpha_1^2 = 2n$$
$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 = ... = 8n$$
$$\gamma = \frac{\mu_3}{\sigma^3(X)} = ... = \frac{4}{\sqrt{2n}}$$

واضح أنّه بزيادة n يقترب γ من الصفر، وعندما $n \to \infty$ يكون التوزيع متناظراً.

كان من الممكن حساب: $E(X)$ ، $\sigma^2(X)$ انطلاقاً من التعريف، علماً أنّه وبالاستفادة من التابع Γ يمكن البرهان أيضاً أنّ: $\alpha_i = 2^i \dfrac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2})}$ بالشكل التالي:

لدينا:

$$\alpha_i = E(X^i) = \int_0^\infty x^i f_n(x)\,dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^\infty e^{-\frac{x}{2}}.\, x^{\frac{n}{2}+i-1}\,dx$$

$$;\ \frac{x}{2} = y \Rightarrow dx = 2dy$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\int_0^\infty e^{-y}.\,(2y)^{\frac{n}{2}+i-1}\,2\,dy$$

$$= \frac{2^i}{\Gamma(\frac{n}{2})}\underbrace{\int_0^\infty e^{-y}.\,(2y)^{\frac{n}{2}+i-1}\,dy}_{\Gamma(\frac{n}{2}+i)}$$

$$= 2^i\,\frac{\Gamma(\frac{n}{2}+i)}{\Gamma(\frac{n}{2})}$$

ومنه:

$$\alpha_1 = E(X) = 2\frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})} = 2\,.\,\frac{\frac{n}{2}\,\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2})} = n$$

$$\alpha_2 = E(X^2) = 4\frac{\Gamma(\frac{n}{2}+1+1)}{\Gamma(\frac{n}{2})} = 4\,\frac{(\frac{n}{2}+1)\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})}$$

$$= 4\,\frac{(\frac{n}{2}+1)\,.\,\frac{n}{2}\,.\,\Gamma(\frac{n}{2})}{\Gamma(\frac{n}{2})} = 2n(\frac{n}{2}+1) = n^2+2n$$

ملاحظات وأمثلة:

-1 إذا كان: $X : N(0,1)$ فإنّ: $Y = X^2 : \chi_1^{\ 2}$.

من الواضح أنّ: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ بالتالي: $g(y) = 2f(x)\left|\frac{dx}{dy}\right|_{x\to y}$

(حيث $f(x)$ دالّة متماثلة).

ومن أجل: $y = x^2 \Rightarrow dy = 2xdx \Rightarrow \frac{dx}{dy} = \frac{1}{2x}$

يكون:

$$g(y) = \frac{2}{\sqrt{2}}\frac{1}{\sqrt{\pi}} e^{-\frac{y}{2}} . \frac{1}{2\sqrt{y}}$$
$$= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{\pi}} e^{-\frac{y}{2}} y^{-\frac{1}{2}} \equiv \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{y}{2}} y^{-\frac{1}{2}}$$

أي أنّ تابع كثافة X^2 يُكتب بالشكل:

$$g(x) = \frac{1}{2^{\frac{1}{2}}\ \Gamma(\frac{1}{2})} e^{-\frac{x}{2}}\ x^{-\frac{1}{2}}$$

وهو بالحقيقة يطابق تابع كثافة χ^2 بدرجة حريّة واحدة ، أي : $X^2 : \chi_1^2$.

-2 إذا كان X_1 و X_2 متغيّرين عشوائيّين مستقلين بحيث:

$X_1 : \chi^2_{n_1}$ و: $X_2 : \chi^2_{n_2}$ عندئذٍ: $X = X_1 + X_2 : \chi^2_{n_1+n_2}$

البرهان: لدينا:

$$\Psi_X(t) = E(e^{tX}) = E(e^{t(X_1+X_2)}) = E(e^{tX_1}.e^{tX_2})$$
$$= E(e^{tX_1}).E(e^{tX_2}) = (1-2t)^{-\frac{n1}{2}}.(1-2t)^{-\frac{n2}{2}}$$
$$= (1-2t)^{-\frac{n_1+n_2}{2}} \Rightarrow X = X_1 + X_2 : \chi^2_{n_1 n_2}$$

وهذا يعني أنّ المتغيّر: $X = X_1 + X_2$ يخضع لتوزيع χ^2 وبـ: $(n_1 + n_2)$ درجة حريّة.

وبالحالة العامة:

إذا كانت: $X_1, X_2, \ldots, X_n$ مجموعة من المتغيّرات العشوائيّة المستقلة بالتبادل

بحيث: $X_i : \chi^2_{n_i} \ , \forall i = \overline{1,n}$ وكان: $X = \sum_{i=1}^{n} X_i$ فإنّ : $X : \chi^2_{\sum_{i=1}^{n} n_i}$

3- إذا كان: X_1, X_2 متغيّرين عشوائيّين مستقلين، بحيث: $X_1 : \chi^2_{n_1}$ ، و: $X = X_1 + X_2 : \chi^2_n$ حيث: $n > n_1$ ، عندئذٍ : $X_2 : \chi^2_{n_2 = n - n_1}$.

البرهان: لدينا:

$$(1-2t)^{-\frac{n}{2}} = \Psi_X(t) = \Psi_{X_1 + X_2}(t) = E(e^{t(X_1+X_2)})$$
$$= E(e^{tX_1})E(e^{tx_2}) = \Psi_{X_1}(t)\Psi_{X_2}(t)$$
$$= (1-2t)^{-\frac{n_1}{2}} \Psi_{X_2}(t) \Rightarrow \Psi_{X_2}(t) = (1-2t)^{-\frac{n-n_1}{2}}$$

أي أنّ لـ X_2 توزيع بشكل χ^2 وبـ $(n_2 = n - n_1)$ درجة حريّة .

مما تقدّم ينتج أنّه من أجل:

$$X_i : N(\mu, \sigma^2) \Rightarrow \frac{X_i - \mu}{\sigma} : N(0,1) \ , \ i = \overline{1,n}$$
$$\Rightarrow (\frac{X_i - \mu}{\sigma})^2 : \chi_1^2 \Rightarrow \sum_{i=1}^{n} (\frac{X_i - \mu}{\sigma})^2 : \chi_n^2$$

وهذا يؤكّد ما بيّناه في مقدمة توزيع χ^2 .

(7-6) توزيع ت: *T- Distribution*

إذا كانت: $X_1, X_2, \ldots, X_{n+1}$ متغيّرات عشوائيّة مستقلة بحيث: $X_i : N(0,1) \ ; \ i = \overline{1 \ , \ n+1}$ عندئذٍ يكون للمتغيّر : $U = \frac{X_1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X^2{}_{i+1}}}$ ، حيث: $(-\infty < U < \infty)$ تابع الكثافة التالي:

$$f(u)=\frac{1}{\sqrt{n}\,\beta(\frac{1}{2},\frac{n}{2})}\cdot\frac{1}{(1+\frac{u^2}{n})^{\frac{n+1}{2}}}$$

وهذا التابع هو ما يُعرَف بتوزيع T وبـ n درجة حريّة.

بمعنى آخر: نقول عن المتغيّر العشوائي X إنّه يخضع لتوزيع T و بـ n درجة حريّة إذا كان تابع كثافته من الشكل:

$$f(x)=\frac{1}{\sqrt{n}\;\beta(\frac{1}{2},\frac{n}{2})}\;\frac{1}{(1+\frac{x^2}{n})^{\frac{n+1}{2}}}$$

(لاحظ أنّ عدد درجات الحريّة أقل بواحد من عدد المتغيّرات)

بالحقيقة أنّ فكرة التوزيع اشتُقت من التوزيع الطبيعي، ولنبيّن ذلك:

إذا كان لدينا: $X_1:N(0,1)\Rightarrow f_1(x_1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x_1^2}{2}}\;,-\infty<X_1<\infty$

وبما أنّ: $X_i:N(0,1)\;;\;i=\overline{1,n+1}$ ، فإنّ : $Y=\sum_{i=1}^{n}X_{i+1}^{2}:\chi_n^{2}$

وبالتالي يكون تابع كثافة Y من الشكل:

$$f_2(y)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}e^{-\frac{y}{2}}.y^{\frac{n}{2}-1}\;;0<y<\infty$$

وبما أنّ: X_1 و Y مستقلان، فإنّ تابع الكثافة المشترك لهما يعطى بالعلاقة:

$$f(x_1,y)=f_1(x_1)f_2(y)=\frac{1}{\sqrt{\pi}\;2^{\frac{n+1}{2}}\Gamma(\frac{n}{2})}e^{-\frac{1}{2}(x_1^2+y)}.y^{\frac{n}{2}-1}$$

ولكن لدينا: $U=\frac{X_1}{\sqrt{Y/n}}$ ، وبوضع: $Z=\frac{Y}{n}$ يكون تابع الكثافة المشترك للشعاع (U,Z) (تبعاً لتوزيع شعاع عشوائي تابع لشعاع عشوائي آخر) بالشكل:

$$f(u,z)=f(x,y)\left\|\frac{\partial(x,y)}{\partial(u,v)}\right\|_{(x,y)\to(u,z)}$$

$$= \frac{1}{2^{\frac{n+1}{2}} . \Gamma(\frac{n}{2}) \sqrt{\pi}} e^{-\frac{1}{2}(u^2 z + nz)} . (nz)^{\frac{n}{2}-1} n z^{\frac{1}{2}}$$

$$= \underbrace{\frac{n^{\frac{n}{2}}}{2^{\frac{n+1}{2}} . \Gamma(\frac{n}{2}) \sqrt{\pi}}}_{C} e^{-\frac{1}{2}(u^2+n)z} . z^{\frac{n-1}{2}}$$

ومنه، فإنّ تابع كثافة U يكون: $f(u) = \int_0^{\infty} f(u,z) dz$

وإذا فرضنا: $\frac{1}{2}(u^2+n)z = t \Rightarrow dz = 2(u^2+n)^{-1} dt$

يكون:

$$f(u) = c . \int_0^{\infty} e^{-t} [2(u^2+n)^{-1} t]^{\frac{n-1}{2}} . 2(u^2+n)^{-1} dt$$

$$= c . 2^{\frac{n+1}{2}} (u^2+n)^{-(\frac{n+1}{2})} . \underbrace{\int_0^{\infty} t^{\frac{n-1}{2}} . e^{-t} dt}_{\Gamma(\frac{n+1}{2})}$$

$$f(u) = c . 2^{\frac{n+1}{2}} [n(1+\frac{u^2}{2})]^{-(\frac{n+1}{2})} . \Gamma(\frac{n+1}{2})$$

$$= \frac{n^{-\frac{1}{2}}}{\Gamma(\frac{n}{2}) \Gamma(\frac{1}{2})} \frac{\Gamma(\frac{n+1}{2})}{(1+\frac{u^2}{n})^{\frac{n+1}{2}}} \quad ; \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

$$= \frac{1}{\sqrt{n}} \underbrace{\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{n}{2})}}_{\frac{1}{\beta(\frac{1}{2},\frac{n}{2})}} \frac{1}{(1+\frac{u^2}{n})^{\frac{n+1}{2}}} = \frac{1}{\sqrt{n}\, \beta(\frac{1}{2},\frac{n}{2})} \frac{1}{(1+\frac{u^2}{n})^{\frac{n+1}{2}}}$$

أي: $U : t_n$ وهو المطلوب.

عزوم توزيع t:

واضح أنّه إذا كان $X : t_n$ فإنّ تابع كثافته $f(x)$ سيكون متناظراً بالنسبة لنقطة البدء أي $f(-x) = f(x)$ ، وبالتالي فإنّ العزوم الابتدائيّة من المراتب الفرديّة معدومة كلها، وبالحالة العامة لدينا:

$$\alpha_{2i} = E(U^{2i}) = E(\frac{X_1}{\sqrt{Y/n}})^{2i} = n^i E(X_1)^{2i} E(Y^{-1})$$

وبما أنّ: $X_1 : N(0,1)$ فإنّ:

$$E(X_1^{\ 2i}) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x_1^{\ 2i} . e^{-\frac{x_1^2}{2}} dx_1$$

فإذا فرضنا: $\frac{x_1^{\ 2}}{2} = t$ وبالاختصار يمكن أن نكتب:

$$E(X_1^{\ 2i}) = 2^i \frac{\Gamma(i+\frac{1}{2})}{\Gamma(\frac{1}{2})}$$

وكذلك بما أنّ : $Y : \chi_n^{\ 2}$ فإنّ:

$$E(Y^{-i}) = \int_0^{\infty} y^{-i} . k_n(y) dy = \frac{1}{2^{\frac{n}{2}} . \Gamma(\frac{n}{2})} \int_0^{\infty} y^{\frac{n}{2}-i-1} . e^{-\frac{y}{2}} . dy$$

وإذا فرضنا: $\frac{y}{2} = Z$ ، وبالاختصار يمكن أن نكتب:

$$E(Y^{-i}) = 2^{-i} \frac{\Gamma(\frac{n}{2}-i)}{\Gamma(\frac{n}{2})}$$

بالتالي:

$$\alpha_{2i} = n^i \frac{\Gamma(i+\frac{1}{2}).\Gamma(\frac{n}{2}-i)}{\Gamma(\frac{1}{2}).\Gamma(\frac{n}{2})}$$

ومنه:

$$\alpha_2 = n\frac{\Gamma(1+\frac{1}{2}).\Gamma(\frac{n}{2}-1)}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} = \frac{n}{n-2}$$

مع ملاحظة أنّ:

$$\Gamma(\frac{n}{2}) = \Gamma(\frac{n}{2}-1+1) = (\frac{n}{2}-1)\Gamma(\frac{n}{2}-1)$$

$$\Gamma(n+\frac{1}{2}) = \Gamma(n+\frac{1}{2}-1+1) = \Gamma(n-\frac{1}{2}+1) = (n-\frac{1}{2})\Gamma(n-\frac{1}{2})$$

$$\sigma^2 = \alpha_2 - \alpha_1{}^2 = \frac{n}{n-2} \; ; n > 2$$

أضف إلى أنّ: $\gamma = 0$.

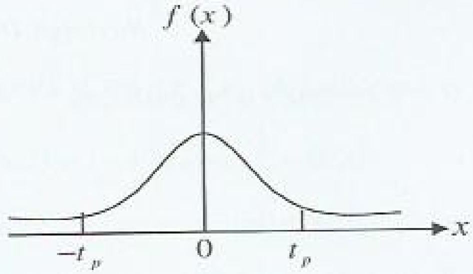
وأخيراً: إذا كان للمتغيّر T توزيع بشكل t_n فإنّ للعلاقة:

$$P[|T| > t_p] = \frac{p}{100} \; ; 0 \leq p \leq 100$$

جداول خاصة تُدعى جداول التوزيع t ، وهي تُفيد في معرفة p فيما لو عُرفت t وبالعكس.

وإذا كانت p معروفة فإنّ جذر المعادلة السابقة الذي نحصل عليه من الجداول والذي هو قيمة t نرمز له بالرمز $t_{n,p}$ ويسمى القيمة $p\%$ للتوزيع t_n ، ولكن وبما أنّ التوزيع متناظر بالنسبة للنقطة 0 فإنّ : $t_{n,p} = -t_{n,(1-p)}$

وبيان تابع الكثافة للتوزيع t_n كما في الشكل:



(6-8) بعض التوزيعات الأخرى المستمرّة (كوشي– فيشر – كاي– ريلاي– ويبل– باريتو) بشكل مختصر:

*** توزيع كوشي:** *Caushy Distribution:*

نقول عن المتغيّر العشوائي X إنه يتبع توزيع كوشي إذا كانت دالّة كثافته الاحتماليّة من الشكل:

$$f(x)=\frac{1}{\pi(1+x^2)} \;;\quad -\infty<x<\infty$$

*** توزيع فيشر:** *F – Distribution*

نقول عن المتغيّر العشوائي X إنّه يتبع توزيع فيشر بدرجات حريّة n ، m إذا كانت دالّة كثافته الاحتماليّة من الشكل:

$$f(x)=\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{n})\,\Gamma(\frac{n}{2})}\,(\frac{m}{n})^{\frac{m}{2}}\,\frac{x^{\frac{m}{2}-1}}{[1+\frac{m}{n}x]^{\frac{m+n}{2}}} \;;\; x\geq 0$$

*** توزيع كاي:** *χ - Distribution*

نقول عن المتغيّر العشوائي X إنّه يتبع توزيع كاي بالوسيط: ($\sigma>0$) وبـ: n درجة حريّة، إذا كانت دالة كثافته الاحتماليّة من الشكل:

$$f(x)=\frac{2(\frac{n}{2})^{\frac{n}{2}}}{\sigma^n\Gamma(\frac{n}{2})}\,x^{n-1}\,e^{-\frac{n}{2\sigma^2}x^2} \;;\quad x\geq 0\;,\; n=1,2,3,...$$

*** توزيع ريلاي:** *Rayleigh Distribution*

نقول عن المتغيّر العشوائي X إنّه يتبع توزيع ريلاي بالوسيط: ($\alpha>0$) إذا كانت دالّة كثافته الاحتماليّة من الشكل:

$$f(x)=\frac{1}{\alpha^2}\,x\,e^{-\frac{x^2}{2\alpha^2}} \;;\quad x>0$$

من الملاحظ أنّه من أجل: $\sigma=\alpha\sqrt{2}$ ، $n=2$ فإنّ توزيع ريلاي يؤول إلى توزيع كاي.

* **توزيع ويبل:** *Weibull Distribution*

نقول عن المتغيِّر العشوائي X إنَّه يتبع توزيع ويبل بالوسيطين : α , c إذا كانت دالَّة كثافته الاحتماليَّة من الشكل:

$$f(x) = c\alpha x^{\alpha-1}\, e^{-cx^{\alpha}} \;\; ; \;\; x \geq 0 \;,\; c\,,\alpha > 0$$

ومن الملاحظ أيضاً في هذا التوزيع أنَّه من أجل: $\alpha = 1$ فإنَّ توزيع ويبل يؤول إلى التوزيع الأسي.

* **توزيع باريتو:** *Pareto Distribution*

نقول عن المتغيِّر العشوائي X إنَّه يتبع توزيع باريتو إذا كانت دالَّة كثافته الاحتماليَّة من الشكل:

$$f(x) = \frac{a b^{a}}{x^{a+1}} \;\; ; \;\; x \geq 0 \;,\;\; a\,, b \geq 0$$

مثال:

إذا كان العمر بالسنين لنوع محدَّد من قطع الغيار الميكانيكيَّة يخضع لتوزيع ريلاي التالي:

$$f(x) = \frac{x}{4}\, e^{-\frac{x^2}{8}} \;\; ; \;\; x \geq 0$$

المطلوب: أوجد تابع التوزيع $F(x)$ ، ثم أوجد احتمال أن تعمل قطعة أكثر من سنتين ؟

الحل:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{4}\int_{0}^{x} x\, e^{-\frac{x^2}{8}}\, dx$$

$$= [-e^{-\frac{x^2}{8}}]_0^x = 1 - e^{-\frac{x^2}{8}}$$

وبالتالي: $P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = e^{-\frac{1}{2}} \approx 0.60$

....

(6-9) تمارين الفصل السادس:

1- إذا كان احتمال حصول طالب في كلية ما على معدّل تخرُّج بدرجة ممتاز هو 0.1، ولنفرض أنّه تخرّج في ذات الدورة 100 طالب، فما احتمال ألا يقل عدد أولئك ذوو المعدّل الممتاز عن 10طلاب ؟ وذلك في الحالات:

1) إذا فرضنا أنّ X يمثل عدد المتخرجين ويخضع لتوزيع ثنائي الحد .

2) إذا كان X يخضع لتوزيع بواسون.

3) إذا كان X يخضع للتوزيع الطبيعي.

2- إذا رمزنا بـ X لعدد مرات تكرار تجربة إلقاء قطعة نقود متجانسة حتى ظهور الصورة لأول مرّة. احسب : $p(X<5)$, $E(x)$, $\sigma(x)$.

3- إذا كان 8% من إنتاج مصنع ما لمادة ما غير صالح ، وسحبنا عيّنة عشوائيّة مقدارها 24 قطعة من نفس المادة . فما احتمال أن تحوي هذه العيّنة على أربع قطع غير صالحة في الحالتين:

1) باستخدام توزيع بواسون.

2) باستخدام توزيع ثنائي الحد.

4- إذا كان : $X : P(\lambda)$ ، وكان : $\lambda = 2$

احسب $P(X=0)$ ، $P(0<X<3)$

5- إذا كان: $X : N(0,\frac{1}{4})$ وكان لـ y توزيع منتظم على المجال [0,1]، وإذا فرضنا أنّ X و Y مستقلان وكان: $Z = X + Y$

أوجد: $f(x,y)$ ، $E(Z)$.

6- أوجد الاحتمال $P[1 < X < 2]$ إذا كان X يتوزع وفق:

1) التوزيع الطبيعي من النموذج (1,1)

2) توزيع ثنائي الحد حيث: $n = 10 , p = 0.1$

3) توزيع بواسون بالوسيط $\lambda = 1$

4) التوزيع فوق الهندسي (حيث: $N = 100 , p = 0.1 , n = 10$)

5) التوزيع المستطيلي على المجال $\left[\frac{1}{2}, \frac{3}{2}\right]$

7- إذا كان: $X : \chi_n^2$ ، أوجد تابع كثافة المتغيّر: $Z = 2X - 1$

8- إذا كان: $X : \tau_n$ ، أوجد توزيع المتغيّر: $Z = X^2$

9- إذا كان احتمال إصابة شخص بمرض معيَّن ضمن فترة زمنية محدَّدة هو 0.001، ووضعنا تحت المراقبة وفي نفس الفترة الزمنيّة 4000 شخص. فما احتمال إصابة 5 أشخاص منهم على الأكثر؟

10- بفرض: $f(x) = \frac{1}{2a} \quad ; \ -a \le X \le a$

المطلوب:

1) بيِّن أنَّ العزم من الرتبة r يمكن أن يُكتَب بالشكل:

$$\mu_r = \frac{1}{2a}\left[\frac{a^{r+1}}{r+1} - \frac{(-a)^{r+1}}{r+1}\right]$$

2) احسب γ (عامل التناظر)، ثم تحقَّق أنَّ: $\frac{\mu_4}{\sigma^4} = \frac{9}{5}$

11- بفرض X يخضع للتوزيع الأسي بتابع الكثافة التالي:

$f(x) = ke^{-kx} \ ; \ 0 \le X \le \infty$ حيث: k ثابت موجب.

بيِّن أنَّ العزم من الرتبة m يمكن أن يُكتَب بالشكل: $\mu_m = \frac{m!}{k^m}$

ثم احسب عامل التناظر γ .

12- إذا كان: $f(x)=\frac{1}{\Gamma(n)}e^{-x}.\,x^{n-1}\,;\,0\le x\le\infty$ (حيث n ثابت موجب)،

بيِّن أنَّ العزم من الرتبة r يُكتَب بالشكل: $\mu_r=\frac{\Gamma(n+r)}{\Gamma(n)}$، ثم احسب عامل التناظر γ.

13- صندوق يحوي 15 مصباحاً منها 5 غير صالة والباقية سليمة. سُحبَت عيِّنة من دون إعادة حجمها $(n = 6)$ مصابيح. فإذا رمزنا بـ X لعدد المصابيح السليمة التي تظهر في هذه العيِّنة، المطلوب:

1) أوجد $E(X)$, $\sigma(X)$

2) أوجد $P(X\le 2)$

....

الفصل السابع

التوزيع المشترك لجملة متغيّرات عشوائيّة

كما عرّفنا المتغيّر العشوائي X بأنّه التطبيق : $X:\Omega \to R$ ، فإنّنا وبطريقة مشابهة، نعرّف الجملة من المتغيّرات العشوائيّة $(X_1, X_2, ..., X_n)$ بالتطبيق:

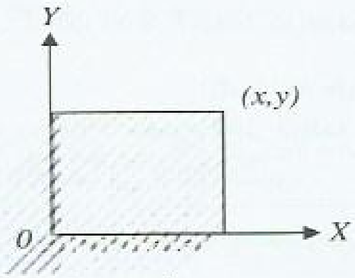
$(X_1, X_2, ..., X_n) : \Omega \to R^n$ ، حيث تمثّل قيمته بالنقطة: $(x_1, x_2, ..., x_n) \in R^n$.

من الأشعة العشوائيّة التي سندرسها، الشعاع العشوائي الثنائي كالشعاع (X,Y)، وبالتالي تُمثّل قيمة الشعاع (X,Y) بالنقطة $(x,y) \in R^2$.

وبالواقع أنّ دراسة التوزيع المشترك للشعاع (X,Y) مشابهة من حيث المبدأ إلى حدٍ كبيرٍ لدراسة المتغيّر العشوائي الوحيد الذي درسناه من قبل.

(7-1) تابع التوزيع المشترك:

نرمز لتابع توزيع الشعاع العشوائي (X,Y) بالرمز: $F(x,y)$ ويعرّف بالشكل: $F(x,y) = P(X \le x, Y \le y)$ وهو يمثّل تابع التوزيع المشترك للمتغيّرين العشوائيين X و Y .



منطقة غير منتهية

وبما أنّه تابع توزيع فهو يحقّق الشروط التالية:

$$0 \le F(x,y) \le 1$$

$$F(x,-\infty) = P(X \le x, Y \le -\infty) = 0$$

$$F(-\infty, y) = P(X \le -\infty, Y \le y) = 0$$

$$F(+\infty, +\infty) = 1$$

$$F(-\infty, -\infty) = 0$$

أضف إلى أنّه غير متناقص، بمعنى:

$$x_1 \le x_2 \Rightarrow F(x_1, y) \le F(x_2, y)$$
$$y_1 \le y_2 \Rightarrow F(x, y_1) \le F(x, y_2)$$

ولنبرهن الحالة الأولى منها (برهان الحالة الثانية مشابه):

واضح أنّ الأحداث ترتبط كما يلي:

$$X \le x_2 = (X \le x_1) \cup (x_1 < X \le x_2)$$

وبما أنّ الحدثين: $X \le x_1 , x_1 < X \le x_2$ متنافيان، فإنّ:

$$P(X \le x_2) = P(X \le x_1) + P(x_1 < X \le x_2)$$

ومن أجل قيمة y ثابتة يكون:

$$\underbrace{P(X \le x_2, Y \le y)}_{F(x_2, y)} = \underbrace{P(X \le x_1, Y \le y)}_{F(x_1, y)} + \underbrace{P(x_1 < X \le x_2, Y \le y)}$$

بالتـــالي: $F(x_2, y) - F(x_1, y) \ge 0$ لأن $P(x_1 < X \le x_2, Y \le y)$ تمثـــل احتمـــالاً، ومنه: $F(x_1, y) \le F(x_2, y)$.

في الصورة الميكانيكيّة: يمثّل التابع $F(x, y)$ مجموع الكتل الواقعة فوق المنطقة المحدّدة بـ$\{X \le x, Y \le y\}$ من المستوي XOY كما في الشكل السابق .

وإذا كـان (X, Y) مسـتمراً ، فـإنّ تـابع كثافـة الشـعاع العشـوائي (X, Y) تُحسَـب مـن العلاقة: $f(x, y) = \dfrac{\partial}{\partial x}\dfrac{\partial}{\partial y} F(x, y)$

أضف إلى أنّه يحقّق الشّرطين التاليين:

- $f(x, y) \ge 0$

- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$

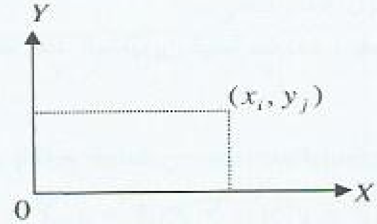
وهو يعني في الصورة الميكانيكيّة أنّ حجم الكتلة موزعة بشكل مستمر، وبكثافة مقدارها $f(x,y)$ في النقطة: $(x,y)\in X0Y$.

نرمز للتابع الاحتمالي للشعاع المنفصل (المتقطع) (X,Y) بالرمز:

$$P_{ij}=P[X=x_i,Y=y_j]\ ;\ i,j=1\,,2\,,\ldots$$

وهو يحقّق الشّرطين التاليين: $\sum_i\sum_j P_{ij}=1$ ، $P_{ij}\geq 0$, $\forall i,j$

حيث تمثّل P_{ij} (في الصورة الميكانيكيّة) الكتلة الموجودة في النقطة (x_i,y_j) من المستوي X0Y .



وكما نعلم فإنّه يوجد نوعين من الأشعة العشوائيّة (منفصلة ومستمرة)، فلذلك يوجد نوعان من التوزيعات منفصلة ومستمرة.

(7-2) التوزيع المشترك المنفصل:

تعريف: نقول عن الشعاع العشوائي (X,Y) إنّه من النوع المنفصل، إذا كانت القيم الممكنة لكلٍ من X و Y منفصلة من الشكل:

$X:x_1,x_2,x_3,\ldots,x_n$

$Y:y_1,y_2,y_3,\ldots,y_m$

حيث: $P_{ij}=P(X=x_i,Y=y_j)\ \ ;i,j=1,2,3,\ldots$

وهذا يكافئ القول: إنّ تابع توزيعه $F(x,y)$ هو من النوع المنفصل، ويُكتب بالشكل:

$$F(x,y)=P(X\leq x,Y\leq y)=\sum_{x_i\leq x}\ \sum_{y_j\leq y}P_{ij}$$

وأنّ جدول التوزيع المشترك للشعاع (X,Y) من الشكل:

$X \backslash Y$	y_1	y_2	...	y_j	...	y_m
x_1	p_{11}	p_{12}	...	p_{1j}	...	p_{1m}
x_2	p_{21}	p_{22}	...	p_{2j}	...	p_{2m}
x_3	p_{31}	p_{32}	...	p_{3j}	...	p_{3m}
...	...	...	...	...	...	...
x_i	p_{i1}	p_{i2}	...	p_{ij}	...	p_{im}
x_n	p_{n1}	p_{n2}	...	p_{nj}	...	p_{nm}

(7-2-1) التوزيعات الهامشيَّة المنفصلة:

والمقصود بالهامشيَّة هو أن يأخذ أحد المتغيِّرين قيمة محدَّدة ، بصرف النظر عن جميع القيم التي يأخذها المتغيِّر الآخر.

نرمز للتابع الاحتمالي الهامشي (التابع الهامشي– *Marginal Function*) لـ X بالرمز:

$$P_{i.} = P[\bigcup_j (X = x_i, Y = y_j)] = \sum_j \underbrace{P(X = x_i, Y = y_j)}_{P_{ij}} = \sum_j P_{ij}$$

أي: $P_{i.} = \sum_j P_{ij}$ ، وهو يمثِّل الاحتمال كي يأخذ X القيمة x_i بدون الاهتمام بالقيم التي يأخذها المتغيِّر Y . وبشكل مكافئ نكتب:

$P_{i.} = P(X = x_i, Y =$ أي قيمة من قيمه)

ويكون جدول التوزيع الهامشي لـ X عندئذٍ من الشكل:

X	x_1	x_2	...	x_i	...
$P_{i.}$	$P_{1.}$	$P_{2.}$	...	$P_{i.}$	...

كما نرمز لتابع التوزيع الهامشي لـ X بالرمز:

$$F_1(x) = P(X \leq x) = \sum_{x_i \leq x} P_{i.} = \sum_{x_i \leq x} \sum_j P_{ij}$$

وبطريقة مشابهة يكون التوزيع الهامشي لـ Y من الشكل : $P_{.j} = \sum_i P_{ij}$

كما يكون جدول توزيع Y بالشكل:

Y	y_1	y_2	...	y_j	...
$P_{.j}$	$P_{.1}$	$P_{.2}$	...	$P_{.j}$	...

ويكون تابع التوزيع الهامشي لـ Y:

$$F_2(y) = P(Y \leq y) = \sum_{y_j \leq y} P_{.j} = \sum_{y_j \leq y} \sum_i P_{ij}$$

(يُرمَز في بعض المراجع لـ $P_{i.}$ بالرمز $P_i(X)$، ولـ $P_{.j}$ بالرمز $P_j(Y)$)

ملاحظة:

يمكن تدوين ما تقدّم بالجدول العام التالي:

$X \backslash Y$	y_1	y_2	...	y_j	...	$P_{i.} = \sum_j P_{ij}$
x_1	p_{11}	p_{12}	...	p_{1j}	...	$P_{1.}$
x_2	p_{21}	p_{22}	...	p_{2j}	...	$P_{2.}$
x_3	p_{31}	p_{32}	...	p_{3j}	...	$P_{3.}$
...	...	...		...	...	...
x_i	p_{i1}	p_{i2}	...	p_{ij}	...	$P_{i.}$
...	...		...	...	...	...
$P_{.j} = \sum_i P_{ij}$	$P_{.1}$	$P_{.2}$	...	$P_{.j}$	...	1

لاحظ أنّ: العمود الأخير يمثّل جدول التوزيع الهامشي لـ X. في حين أنّ الصف الأخير يمثّل جدول التوزيع الهامشي لـ Y، لذلك سَمِّيت هذه التوزيعات بالهامشيّة.

(3-7) التوزيع المشترك المستمر:

وهنا تُحسَب الاحتمالات عن طريق المساحات تحت بيان الدالة المعطاة، وبما أنّه لا توجد مساحة تحت نقطة، فالاحتمال صفر، وهذا ما يميِّز التوزيعات المنفصلة عن التوزيعات المستمرة .

تعريف: نقول عن الشعاع العشوائي (X,Y) إنّه من النوع المستمر، إذا كانت القيم الممكنة لكلٍ من X و Y من النوع المستمر .

هذا، وكما بيّنا سابقاً: يُرمز لتابع توزيع الشعاع العشوائي (X,Y) بالرمز $F(x,y)$، ولتابع كثافته بالرمز $f(x,y)$ وهو يحقّق شرطين أساسيين كما نعلم، وهو يعني في الصورة الميكانيكيّة أنّ واحدة الكتل موزعة بشكل مستمر وبكثافة مقدارها $f(x,y)$ في النقطة: $(x,y)\in XOY$.

وقياساً بحالة متغيّر عشوائي واحد حيث كان:

$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(x)dx$$

نكتب:

$$F(x,y)=P(X\le x,Y\le y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy$$

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\,\partial y}$$

(7-3-1) التوزيعات الهامشيّة المستمرة:

بفرض أنّ: X و Y متغيّران عشوائيان لهما توزيع مشترك من النوع المستمر، وإذا رمزنا لتابع التوزيع الهامشي لـ X بالرمز $F_1(x)$، ولتابع كثافته الهامشي بـ $f_1(x)$، ورمزنا لتابع التوزيع الهامشي لـ Y بالرمز $F_2(y)$، ولتابع كثافته بـ $f_2(y)$، فإنّ:

$$F_1(x)=P[X\le x,Y=\text{أية قيمةٍ من قيمه}]=\int_{-\infty}^{x}[\int_{-\infty}^{\infty} f(x,y)dy]dx$$

$$f_1(x)=\frac{dF_1(x)}{dx}=\int_{-\infty}^{\infty} f(x,y)\,dy \Leftrightarrow F_1(x)=\int_{-\infty}^{x} f_1(x)dx$$

كذلك:

$$F_2(y) = P[X = \text{أية قيمةٍ من قيمه}\ , Y \leq y] = \int_{-\infty}^{y}[\int_{-\infty}^{\infty} f(x, y)dx]dy$$

$$f_2(y) = \frac{dF_2(y)}{dy} = \int_{-\infty}^{\infty} f(x, y)\,dx \Leftrightarrow F_2(y) = \int_{-\infty}^{y} f_2(y)dy$$

(4-7) التوزيعات الشَّرطيَّة:

نريد أن ندرس احتمال وقوع أحد حدثين، شريطة وقوع الحدث الآخر، وذلك في الحالتين التاليتين :

1- إذا كان الشعاع (X, Y) منفصلاً:

يكون:

$$P_{ij} = P(\underbrace{X = x_i}_{A}, \underbrace{Y = y_j}_{B}) = P(A \cap B) = P(A \setminus B)P(B)$$

$$= P(\underbrace{X = x_i \setminus Y = y_j}_{P_{i\setminus j}})\, P(\underbrace{Y = y_j}_{P_{.j}})$$

أي: $P_{ij} = P_{i\setminus j}\ P_{.j}$

ومنه:

$$P_{i\setminus j} = \frac{P_{ij}}{P_{.j}}\ ;\ P_{.j} \neq 0\ : \qquad (1)$$

وهو يمثِّل احتمال وقوع الحدث $X = x_i$ شريطة وقوع الحدث $Y = y_j$، وقد اعتبرنا هنا أنَّ: $x_i \equiv i$ وكذلك : $y_j \equiv j$.

وجدول التوزيع المشروط لـ X في هذه الحالة يكون من الشكل:

$X \setminus Y = y_j$	x_1	x_2	...	x_i	...
	$P_{1\setminus j}$	$P_{2\setminus j}$	...	$P_{i\setminus j}$	...

وبطريقة مشابهة يكون الاحتمال الشَّرطي (المشروط) لـ Y :

$$P_{j\backslash i} = \frac{P_{ij}}{P_{i.}} \; ; \; P_{i.} \neq 0 \quad : \quad (2)$$

وجدول توزيع Y يكون من الشكل :

$Y \backslash X = x_i$	y_1	y_2	...	y_j	...
	$P_{1\backslash i}$	$P_{2\backslash i}$	...	$P_{j\backslash i}$	...

وأياً كانت الاحتمالات، مشروطة أم غير ذلك، يبقى مجموع الاحتمالات مساوياً للواحد، أي:

$$\sum_i P_{i\backslash j} = \sum_i \frac{P_{ij}}{P_{.j}} = \frac{1}{P_{.j}} \underbrace{\sum_i P_{ij}}_{P_{.j}} = \frac{P_{.j}}{P_{.j}} = 1$$

وكذلك:

$$\sum_j P_{j\backslash i} = \sum_j \frac{P_{ij}}{P_{i.}} = \frac{1}{P_{i.}} \underbrace{\sum_j P_{ij}}_{P_{i.}} = \frac{P_{i.}}{P_{i.}} = 1$$

من (1) و (2) ينتج:

$$P_{ij} = P_{i\backslash j}\, P_{.j} = P_{j\backslash i}\, P_{i.} \quad : \quad (3)$$

2- إذا كان الشعاع (X,Y) مستمراً:

ورمزنا لتابع الكثافة الشّرطي لـ X بالرمز $f(x\backslash y)$ ، ولتابع الكثافة الشّرطي لـ Y بالرمز : $f(y\backslash x)$.

عندئذٍ ، وقياساً بالعلاقات الثلاث السابقة، نحصل على العلاقات المكافئة التالية:

$$f(x\backslash y) = \frac{f(x,y)}{f_2(y)} \qquad (1)'$$

$$f(y\backslash x) = \frac{f(x,y)}{f_1(x)} \qquad (2)'$$

$$f(x,y) = f(x\backslash y) f_2(y) = f(y\backslash x) f_1(x) \qquad (3)'$$

لاحظ أنّ:

$$\int_{-\infty}^{\infty} f(x \setminus y)dx = \int_{-\infty}^{\infty} \frac{f(x,y)}{f_2(y)}dx = \frac{1}{f_2(y)} \underbrace{\int_{-\infty}^{\infty} f(x,y)dx}_{f_2(y)} = \frac{f_2(y)}{f_2(y)} = 1$$

أيضاً سيكون: $\int_{-\infty}^{\infty} f(y \setminus x)dy = 1$.

(5-7) الاستقلال في التوزيعات المشتركة:

نقول عن المتغيّرين العشوائييّن X و Y إنّهما مستقلان بالتبادل إذا كان:

1- الحالة المنفصلة: $P_{i\setminus j} = P_{i.} \; or: \; P_{ij} = P_{i.}P_{.j}$

2- الحالة المستمرة: $\left.\begin{array}{l} f(x \setminus y) = f_1(x) \\ f(y \setminus x) = f_2(y) \end{array}\right\} \quad or : \quad f(x,y) = f_1(x) f_2(y)$

مثال:

بفرض للشعاع العشوائي (X,Y) توزيع مشترك منفصل من الشكل:

$$P_{ij} = k(2x_i + y_j) \; ; \begin{cases} i = 0, 1, 2 \\ j = 0, 1, 2, 3 \end{cases}$$

المطلوب:

1- أوجد قيمة الثابت k ثم احسب : $P(X = 1, Y = 3)$ ، $P(X \geq 1, Y \leq 2)$.

2- أوجد جدول التوزيع المشترك للشعاع (X,Y) ثم الهامشي لكلٍ من X و Y .

3- أوجد جدول التوزيع المشروط للمتغيّر: $Y \setminus X = 1$.

4- بيّن ما إذا كان X و Y مستقلين أم لا ؟

5- أوجد: $E(X)$ ، $E(Y)$ ، $E(XY)$ ، $\sigma^2(X)$ ، $\sigma^2(Y)$.

6- أوجد: $\Psi_X(t)$ ، $\Psi_Y(t)$.

الحل:

1– لدينا الشرط:

$$1 = \sum_{i,j} P_{ij} = k\sum_{j=0}^{3}\{j + 2+j + 4+j\}$$

$$= k\sum_{j=0}^{3}(3j+6) = k(6+9+12+15) = 42k \Rightarrow k = \frac{1}{42}$$

واضح أنّ:

$$P(X=1, Y=3) = k(2+3) = 5k = \frac{5}{42}$$

$$P(X \geq 1, Y \leq 2) = \sum_{i\geq 1}\sum_{j\leq 2} P_{ij}$$

$$= \sum_{i\geq 1}(P_{i0} + P_{i1} + P_{i2})$$

$$= P_{10} + P_{11} + P_{12} + P_{20} + P_{21} + P_{22}$$

$$= \frac{1}{42}(2+3+4 + 4+5+6) = \frac{4}{7}$$

2– جدول التوزيع المشترك والهامشي:

$X \backslash Y$	0	1	2	3	$P_{i.}$
0	0	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{3}{42}$	$\frac{6}{42}$
1	$\frac{2}{42}$	$\frac{3}{42}$	$\frac{4}{42}$	$\frac{5}{42}$	$\frac{14}{42}$
2	$\frac{4}{42}$	$\frac{5}{42}$	$\frac{6}{42}$	$\frac{7}{42}$	$\frac{22}{42}$
$P_{.j}$	$\frac{6}{42}$	$\frac{9}{42}$	$\frac{12}{42}$	$\frac{15}{42}$	1

3–جدول التوزيع المشروط للمتغيّر: $Y \backslash X = 1$

$Y \backslash X = 1$	0	1	2	3
	$\frac{2}{14}$	$\frac{3}{14}$	$\frac{4}{14}$	$\frac{5}{14}$

4–شرط الاستقلال: $P_{ij} = P_{i.}\ P_{.j}$

لدينا:

$$\left.\begin{aligned} P_{.0} &= \frac{6}{42} \\ P_{0.} &= \frac{6}{42} \\ P_{00} &= 0 \end{aligned}\right\} \Rightarrow P_{00} \neq P_{0.}\ P_{.0}$$

والمتغيّران غير مستقلين.

5–لدينا:

$$E(X) = \sum_{i=0}^{2} i\ .\ P_{i.} = 0 + 1.\frac{14}{42} + 2.\frac{22}{42} = \frac{58}{42}$$

وبالمثل:

$$E(Y) = \sum_{j=0}^{3} j\ .\ P_{.j} = \ldots = \frac{78}{42}$$

$$E(XY) = \sum_{i,j} i\,j\ P_{ij} = \sum_{i} (0 + i\,P_{i1} + 2i\ P_{i2} + 3i\ P_{i3})$$

$$= 0 + 1.P_{11} + 2.P_{12} + 3.P_{13} + 2.P_{21} + 4.P_{22} + 6.P_{23}$$

$$= \frac{102}{42}$$

$$\sigma^2(X) = E(X^2) - [E(X)]^2$$

$$= \sum_{i=0}^{2} i^2\ .\ P_{i.} - \{\sum_{i=0}^{2} i.\ P_{i.}\}^2 = \frac{102}{42} - [\frac{58}{42}]^2 \approx 0.521$$

6– لدينا:

$$\Psi_X(t) = E(e^{tX}) = \sum_{i=0}^{2} P_{i.}\ e^{it} = P_{0.} + P_{1.}\ e^{t} + P_{2.}\ e^{2t}$$

$$= \frac{6}{42} + \frac{14}{42}.e^{t} + \frac{22}{42}.e^{2t}$$

وتحسب $\Psi_Y(t)$ بنفس الطريقة، حيث:

$$\Psi_Y(t) = E(e^{tY}) = \sum_{j=0}^{3} P_{.j}\, e^{jt} = \ldots$$

(6–7) توزيع شعاع عشوائي تابع لشعاع عشوائي آخر:

بفرض: $Z = (X,Y)$ شعاع عشوائي معلوم التوزيع و $W = (U,V)$ شعاع عشوائي آخر مجهول التوزيع .

فإذا كان: $Z = \Psi(W)\ ;\ W = \varphi(Z)$، يُطلَب إيجاد توزيع $W = (U,V)$ في الحالتين:

1– $Z = (X,Y)$ منفصل:

عندئذٍ يكون أيضاً للشعاع $W = (U,V)$ توزيع منفصل على النقاط:

$$(u_i, v_j) = [\varphi_1(x_i, y_j), \varphi_2(x_i, y_j)]$$

بحيث: $P[U = u_i, V = v_j] = P[X = \Psi_1(u_i, v_j), Y = \Psi_2(u_i, v_j)]$

لأنَّ:

$$\begin{aligned} P[U = u_i, V = v_j] &= P[W = (u_i, v_j)] = P[\varphi(Z) = (u_i, v_j)] \\ &= P[Z = \Psi(u_i, v_j)] = P[X = \Psi_1(u_i, v_j), Y = \Psi_2(u_i, v_j)] \end{aligned}$$

2– $Z = (X,Y)$ مستمر:

عندئذٍ يكون أيضاً للشعاع $W = (U,V)$ توزيعاً مستمراً، وتابع كثافته يعطى إما بالعلاقة:

$$g(u,v) = |J|\, f(\Psi_1(u,v), \Psi_2(u,v))$$

حيث:

$$J = \left|\frac{\partial(x,\ y)}{\partial(u,v)}\right| = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

ويسمّى الجاكوبيان (*Jacobia*) أو معامل التحويل (وهو عبارة عن محدّد من الرتبة الثانية للتفاضلات الجزئيّة بين المتغيّرات) كما هو واضح، أو يعطى بالعلاقة المكافئة:

$$g(u,v) = f(x,y)\left\|\begin{matrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{matrix}\right\|_{(x,y)\to(u,v)}$$

حيث تُكتب العلاقة الأخيرة (بالحالة العامة) بالشكل التالي:

$$g(y_1, y_2,..., y_n) = f(x_1, x_2,..., x_n)\left\|\frac{\partial(x_1, x_2,..., x_n)}{\partial(y_1, y_2,..., y_n)}\right\|_{(x_1, x_2,..., x_n)\to(y_1, y_2,..., y_n)}$$

وبمعنى آخر: إذا كان $X = [X_1, X_2,..., X_n]$ شعاع عشوائي، تابع كثافته المشترك $f(x_1, x_2,..., x_n)$ معلوم، و $Y = [Y_1\ ,\ Y_2 ,..., Y_n]$ شعاع عشوائي آخر مجهول التوزيع يرتبط مع X بالعلاقة: $X = \Psi(Y)$; $Y = \varphi(X)$، عندئذٍ: تابع الكثافة المشترك لـ Y يُعطى بالعلاقة:

$$g(y_1, y_2,..., y_n) = f(x_1, x_2,..., x_n)\left\|\frac{\partial(x_1, x_2,..., x_n)}{\partial(y_1, y_2,..., y_n)}\right\|_{(x_1, x_2,..., x_n)\to(y_1, y_2,..., y_n)}$$

مثال (1):

إذا كان للشعاع العشوائي (X,Y) توزيع مشترك من النوع المنفصل بالشكل:

$$\left.\begin{matrix} X : 1\ ,2 \\ Y\ : 3\ ,\ 5 \end{matrix}\right\}\ ;\ P_{ij} = P(X = i, Y = j) = c(2i + j)$$

المطلوب: أوجد قيمة الثابت c، ثم احسب الاحتمال: $P(\frac{X}{Y} < 1)$

الحل:

يمكن حساب الثابت c من جدول التوزيع المشترك للشعاع (X,Y) بالشكل:

X \ Y	3	5	sum
1	$5c$	$7c$	$12c$
2	$7c$	$9c$	$16c$
sum	$12c$	$16c$	$28c$

ومنه: $28c = 1 \Rightarrow c = \frac{1}{28}$

نضع:
$$\left.\begin{array}{l} U=\dfrac{X}{Y}=\varphi_1(X,Y) \\ V=Y\ =\varphi_2(X,Y) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} X=UV=\Psi_1(u,v) \\ Y\ =V\quad =\Psi_2(u,v) \end{array}\right.$$

ويلزمنا جدول التوزيع المشترك للشعاع (U,V)، حيث القيم الممكنة لكلٍ من U و V بالشكل:

$$U\ :\ \frac{1}{5}\ ,\ \frac{1}{3}\ ,\ \frac{2}{5}\ ,\ \frac{2}{3}$$
$$V\ :\ 3\ ,\ 5$$

بالتالي:

$$P(U=\frac{1}{5},\ V=3)=P[X=\Psi_1(u,v)=UV\ \ ,\ Y=\Psi_2(u,v)=V]$$
$$=P(X=\frac{3}{5},Y=3)=0$$

$$P(U=\frac{1}{5},V=5)=P(X=1,Y=5)=7c=\frac{7}{28}$$

$$P(U=\frac{1}{3},V=3)=P(X=1,Y=3)=5c=\frac{5}{28}$$

$$P(U=\frac{1}{3},V=5)=P(X=\frac{5}{3},Y=5)=5c=\frac{5}{28}=0$$

$$P(U=\frac{2}{5},V=3)=P(X=\frac{6}{5},Y=3)=0$$

$$P(U=\frac{2}{5},V=5)=P(X=2,Y=5)=9c=\frac{9}{28}$$

$$P(U=\frac{2}{3},V=3)=P(X=2,Y=3)=7c=\frac{7}{28}$$

$$P(U=\frac{2}{3},V=5)=P(X=\frac{10}{3},Y=5)=0$$

(مع ملاحظة أنّ X لا يأخذ قيماً كسريّة).

ومما تقدّم، ينتج أنّ جدول التوزيع المشترك للشعاع (U,V) يكون من الشكل:

$U \backslash V$	3	5	sum
$\frac{1}{5}$	0	$\frac{7}{28}$	$\frac{7}{28}$
$\frac{1}{3}$	$\frac{5}{28}$	0	$\frac{5}{28}$
$\frac{2}{5}$	0	$\frac{9}{28}$	$\frac{9}{28}$
$\frac{2}{3}$	$\frac{7}{28}$	0	$\frac{7}{28}$
Sum	$\frac{12}{28}$	$\frac{16}{28}$	1

ويكون جدول التوزيع الهامشي لكلٍ من U و V من الشكل:

U	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{2}{3}$
	$\frac{7}{28}$	$\frac{5}{28}$	$\frac{9}{28}$	$\frac{7}{28}$

V	3	5
	$\frac{12}{28}$	$\frac{16}{28}$

ومنه ينتج مباشرةً أنّ: $P[\frac{X}{Y}=U<1]=\frac{7}{28}+\frac{5}{28}+\frac{9}{28}+\frac{7}{28}=1$ وهو المطلوب.

مثال (2):

بفرض أنّ للشعاع العشوائي (X,Y) توزيعاً مستمراً موصوفاً بتابع الكثافة التالي:

$$f(x,y)=\frac{1}{4}e^{-\frac{x+y}{2}} \quad ; x,y>0$$

المطلوب:

أوجد تابع التوزيع الهامشي $g_1(u)$ للمتغيّر: $U = \frac{X}{Y}$، ثم احسب الاحتمال: $P(U < 1)$

الحل:

لدينا: $U = \frac{X}{Y} = \varphi_1(x, y) \Rightarrow X = UV = \Psi_1(u, v)$

ولنضع: $V = Y = \varphi_2(x, y) \Rightarrow Y = V = \Psi_2(u, v)$

عندئذٍ فإنّ تابع الكثافة المشترك للشعاع (U, V) يكون من الشكل:

$$\begin{aligned} g(u,v) &= |\,j\,|\; f(\Psi_1(u,v), \Psi_2(u,v)) \\ &= |\,j\,|\; f(uv, v) \end{aligned}$$

حيث:

$$|\,j\,| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

بالتالي:

$$g(u,v) = \frac{1}{4}\, v \, . \, e^{-\frac{uv+v}{2}} \quad ; \; uv \, , \, v > 0$$

ومنه:

$$g_1(u) = \int_{-\infty}^{\infty} g(u,v)\, dv = \frac{1}{4}\int_{0}^{\infty} v \, . e^{-\frac{u+1}{2} v}\, dv$$

نفرض:

$$\frac{u+1}{2} v = t \Rightarrow v = \frac{2t}{u+1} \Rightarrow dv = \frac{2\,dt}{u+1}$$

بالتالي:

$$g_1(u) = \frac{1}{4}\int_0^{\infty} \frac{2t}{u+1} . e^{-\frac{u+1}{2}\cdot\frac{2t}{u+1}} . \frac{2dt}{u+1}$$

$$= \frac{1}{(u+1)^2} \underbrace{\int_0^{\infty} t . e^{-t}\, dt}_{\Gamma(2)=1} = \frac{1}{(u+1)^2} \;, u > 0$$

ومنه:

$$P(U < 1) = \int_0^1 g_1(u)\, du = \int_0^1 \frac{du}{(1+u)^2} = [arc\ Tg\ x]_0^1 = \frac{\pi}{4}$$

(7-7) بعض القيم المميّزة في التوزيعات المشتركة:

(7-7-1) التوقُّع الرياضي:

- توقُّع دالة بمتغيّرين:

بفرض: (X,Y) شعاع عشوائي معلوم التوزيع ، و $Z = \varphi(X,Y)$ شعاع عشوائي آخر مجهول التوزيع، عندئذٍ فإنّ توقُّع المتغيّر Z يُحسَب من العلاقة:

$E(Z) = E[\varphi(X,Y)] = \sum_i \sum_j \varphi(x_i, y_j) P_{ij}$ إذا كانت المتغيّرات منفصلة .

$E(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x,y) f(x,y) dxdy$ إذا كانت المتغيّرات مستمرة .

فمثلاً:

. من أجل: $Z = X\,Y$ يكون للمتغيّرات المنفصلة والمستمرة على الترتيب:

$$E(Z) = \sum_i \sum_j x_i\, y_j P_{ij}$$

$$E(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y) dxdy$$

. ومن أجل: $Z = X + Y$ يكون للمتغيّرات المنفصلة والمستمرة على الترتيب:

$$E(Z) = \sum_i \sum_j (x_i + y_j) P_{ij}$$

$$E(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f(x,y) dxdy$$

وهكذا....

. توقُّع مجموع متغيِّرين:

$$E(X \mp Y) = E(X) \mp E(Y)$$

البرهان:

$$E(X+Y) = \sum_i \sum_j (x_i + y_j) P_{ij} = \sum_i x_i \underbrace{(\sum_j P_{ij})}_{P_{i.}} + \sum_j y_j \underbrace{(\sum_i P_{ij})}_{P_{.j}}$$

$$= \sum_i x_i P_{i.} + \sum_j y_j P_{.j} = E(X) + E(Y)$$

أو :

$$E(X+Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f(x,y) dx dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y) dx dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y) dx dy$$

$$= \int_{-\infty}^{\infty} x [\int_{-\infty}^{\infty} f(x,y) dy] dx + \int_{-\infty}^{\infty} y [\int_{-\infty}^{\infty} f(x,y) dx] dy$$

$$= \int_{-\infty}^{\infty} x f_1(x) dx + \int_{-\infty}^{\infty} y f_2(y) dy = E(X) + E(Y)$$

وتبقى العلاقة صحيحة، سواءً كان الشعاع (X,Y) منفصلاً أم مستمراً .

وبصفة عامة: $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$ شريطة وجود التوقُّع الرياضي لكلٍ منها .

. توقُّع حاصل ضرب متغيِّرين عشوائيين مستقلين:

إذا كان X و Y مستقلين فإنّ: $E(XY) = E(X)E(Y)$

البرهان:

$$E(XY) = \sum_i \sum_j x_i y_j P_{ij} = \underbrace{\sum_i x_i P_{i.}}_{E(X)} \underbrace{\sum_j y_j P_{.j}}_{E(Y)}$$

$$= E(X)E(Y) \quad ; P_{ij} = P_{i.} P_{.j}$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f(x,y) dx dy = \underbrace{\int_{-\infty}^{\infty} x f_1(x) dx}_{E(X)} \underbrace{\int_{-\infty}^{\infty} y f_2(y) dy}_{E(Y)}$$

$$= E(X)E(Y) \quad ; \ f(x,y) = f_1(x) f_2(y)$$

وفي الحالة العامة يكون لدينا: $E[\prod_{i=1}^{n} X_i] = \prod_{i=1}^{n} E[X_i]$

شريطة استقلال المتغيِّرات $(i=\overline{1,n})$ ، X_i .

- التوقُّع الشّرطي:

$$E(X \setminus Y) = \sum_i x_i\, P_{i\setminus j} = \frac{1}{P_{.j}} \sum_i P_{i\,j}$$

$$E(X \setminus Y) = \int_{-\infty}^{\infty} x\, f(x \setminus y)\, dx = \frac{1}{f_2(y)} \int_{-\infty}^{\infty} x\, f(x, y) dx$$

أيضاً:

$$E(Y \setminus X) = \sum_j y_j\, P_{j\setminus i} = \frac{1}{P_{i.}} \sum_j y_j P_{i\,j}$$

$$E(Y \setminus X) = \int_{-\infty}^{\infty} y\, f(y \setminus x)\, dy = \frac{1}{f_1(x)} \int_{-\infty}^{\infty} y\, f(x, y)\, dy$$

- توقُّع حاصل الضرب بالحالة العامة:

$$E(XY) = E[X.E(Y \setminus X)] = E(X).E(Y \setminus X)$$
$$E(XY) = E[Y.E(X \setminus Y)] = E(Y).E(X \setminus Y)$$

البرهان:

لنبرهن الأولى منها (باعتبار X و Y متغيِّرين عشوائيين غير مستقلين):

$$\begin{aligned} E(X)E(Y \setminus X) &= \sum_i x_i P_{i.} \sum_j y_j P_{j\setminus i} \\ &= \sum_i \sum_j x_i\, y_j P_{i.}\, P_{j\setminus i} \\ &= \sum_i \sum_j x_i\, y_j P_{ij} \;=\; E(X\,Y) \end{aligned}$$

وذلك باعتبار الشعاع (X,Y) من النوع المنفصل.

$$E(X)E(Y\setminus X) = \int_{-\infty}^{\infty} x f_1(x)dx \int_{-\infty}^{\infty} y f(y\setminus x)dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \underbrace{f_1(x) f(y\setminus x)}_{f(x,y)} dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dx\,dy = E(XY)$$

وهذا في حالة الاستمرار.

(7-7-2) العزوم في التوزيعات المشتركة:

تعريف: بفرض أنّ : X و Y متغيّران عشوائيان لهما توزيع مشترك. عندئذٍ:

. العزم الابتدائي من المرتبة $r+s$: يُرمّز ويُكتَب بالشكل:

$\alpha_{rs} = E(X^r Y^s)$; r, s أعداد صحيحة غير سالبة

. العزم المركزي من المرتبة $r+s$: : يُرمّز ويُكتَب بالشكل:

$$\mu_{rs} = E[(X-m_X)^r (Y-m_Y)^s] \;\; ; m_X \equiv E(X) \;,\; m_Y \equiv E(Y)$$

حيث: s , r أعداد صحيحة غير سالبة أيضاً.

نتائج:

/التوقُّع الرياضي لـ X / : $\alpha_{10} = E(X)$

/التوقُّع الرياضي لـ Y / : $\alpha_{01} = E(Y)$

/ مستقلان X , Y / ; $\alpha_{11} = E(XY) = E(X)E(Y)$

/غير مستقلين X,Y / ; $= E(X)E(Y\setminus X)$

$$\mu_{00} = 1 \,,\; \mu_{10} = 0 \,,\; \mu_{01} = 0$$
$$\mu_{20} = \sigma^2(X) = E(X^2) - \{E(X)\}^2 = \alpha_{20} - \alpha_{10}^{\;2}$$
$$\mu_{02} = \sigma^2(Y) = E(Y^2) - \{E(Y)\}^2 = \alpha_{02} - \alpha_{01}^{\;2}$$

أضـف إلـى أنّ: $\mu_{11} = E[(X - m_X)(Y - m_Y)] \equiv E(XY) - E(X)E(Y)$ وهـو يمثّـل العزم المركزي من المرتبة الثانية ويسمّى:

(3–7–7) تمام تباين متغيّرين ـ $Cov(X,Y)$:

وجدنا سابقاً أنّه إذا كان X و Y مستقلين وكان: $Z = X + Y$، فإنّ:

$$\sigma^2(Z) = \sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$$

وإذا كان X و Y مرتبطين وكان: $Z = X + Y$، فإنّ:

$$\sigma^2(Z) = \sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y) + 2\underbrace{E[(X - m_X)(Y - m_Y)]}_{\mu_{11}}$$

$$= \sigma^2(X) + \sigma^2(Y) + 2\mu_{11}$$

أي أنّ: μ_{11} يرتبط بتبـاين المتغيّرين X و Y ويسـمّى بتمام التبـاين (*Covariance*) أو التغاير، وُيرمز له أحياناً بالرمز: $\text{Cov}\ (X,Y)$، وله مجموعة خواص.

. خواص $Cov\,(X,Y)$:

1– إذا كان X و Y متغيّرين عشوائيين مستقلين فإنّ: $Cov(X,Y) = 0$

2– $\text{Cov}\ (X,Y) = \text{Cov}(Y,X)$ أي: التغاير بين X و Y هو ذاته التغاير بين Y و X

3– $\text{Cov}(X,X) = \sigma^2(X)$ أي أن التغاير بين المتغيّر ونفسه يمثّل تباين المتغيّر

4– $\text{Cov}(X,a) = 0$ (حيث: a ثابت) لأنّ:

$$Cov(X,a) = E(Xa) - E(X).E(a)$$
$$= aE(X) - aE(X) = 0$$

5– $Cov(aX,bY) = ab.Cov(X,Y)$

لأنّ:

$$Cov(aX,bY) = E(aX.bY) - E(aX).E(bY)$$
$$= ab.E(XY) - aE(X).bE(Y)$$
$$= ab[E(XY) - E(X)E(Y)] = ab.Cov(X,Y)$$

أي: $Cov(aX,bY)=ab.Cov(X,Y)$ (حيث: a و b ثوابت)

6- $Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$

حيث:

$$\begin{aligned}
Cov(X+Y,Z) &= E[(X+Y)Z]-E(X+Y).E(Z) \\
&= E(XZ+YZ)-[E(X+E(Y)]E(Z) \\
&= E(XZ)+E(YZ)-E(X)E(Z)-E(Y)E(Z) \\
&= [E(XZ)-E(X)E(Z)]+[E(YZ)-E(Y)E(Z)] \\
&= Cov(X,Z)+Cov(Y,Z)
\end{aligned}$$

وبالحالة العامة، تُكتَب العلاقة السابقة بالشكل:

$$Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

فمثلاً:

$$\begin{aligned}
Cov(X_1+X_2+X_3, Y_1+Y_2) &= Cov(\sum_{i=1}^{3} X_i, \sum_{j=1}^{2} Y_j) = \sum_{i=1}^{3}\sum_{j=1}^{2} Cov(X_i, Y_j) \\
&= Cov(X_1,Y_1)+Cov(X_1,Y_2) \\
&+ Cov(X_2,Y_1 + Cov(X_2,Y_2) \\
&+ Cov(X_3,Y_1)+Cov(X_3,Y_2)
\end{aligned}$$

7- $Cov(X+Y)=\frac{1}{2}[\sigma^2(X+Y)-\sigma^2(X)-\sigma^2(Y)] \; : \; \forall X,Y$

ويمكن برهان هذه العلاقة بالشكل:

لدينا:

$$\begin{aligned}
\sigma^2(X+Y) &= E(X+Y)^2-[E(X+Y)]^2 \\
&= E(X^2+Y^2+2XY)-[E(X)+E(Y)]^2 \\
&= E(X^2)+E(Y^2)+2E(XY)-[E(X)+E(Y)]^2 \\
&= E(X^2)+E(Y^2)+2E(XY)-[E(X)]^2 \\
&\qquad -[E(Y)]^2-2E(X).E(Y) \\
&= E(X^2)-[E(X)]^2+E(Y^2)-[E(Y)]^2 \\
&\qquad +2[E(XY)-E(X).E(Y)] \\
&= \sigma^2(X)+\sigma^2(Y)+2Cov(X,Y)
\end{aligned}$$

ومنه:
$$Cov(X+Y)=\frac{1}{2}[\sigma^2(X+Y)-\sigma^2(X)-\sigma^2(Y)]$$

أيضاً وانطلاقاً من العلاقة:
$$\sigma^2(X-Y)=\sigma^2(X)+\sigma^2(Y)-2Cov(X,Y)$$

يمكن أن نكتب:
$$Cov(X,Y)=\frac{1}{2}[\sigma^2(X)+\sigma^2(Y)-\sigma^2(X-Y)]\ ,\ \forall X,Y$$

(7-8) عامل(معامل) الارتباط :

نرمز لعامل(معامل) الارتباط بين X و Y بالرمز $\rho(X,Y)$ ويعرّف بالعلاقة التالية:

$$\rho(X,Y)=\frac{E[(X-m_X)(Y-m_Y)]}{\sqrt{E[(X-m_X)^2].E[(Y-m_Y)^2]}}$$
$$=\frac{\mu_{11}}{\sigma(X)\sigma(Y)}=\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$

علماً أنّه يمكن كتابة μ_{11} بالشكل: $\mu_{11}=E(XY)-E(X)E(Y)$

لأنّ:

$$\begin{aligned}\mu_{11}&=E[(X-m_X)(Y-m_Y)]\\&=E[XY-X\,m_Y-Y\,m_X+m_X\,m_Y]\\&=E(XY)-m_X\,m_Y-m_X\,m_Y+m_X\,m_Y\\&=E(XY)-m_X\,m_Y\\&=E(XY)-E(X)E(Y)\end{aligned}$$

بالتالي يمكن حساب معامل الارتباط بالعلاقة المكافئة التالية:

$$\rho(X,Y)=\frac{E(XY)-E(X)E(Y)}{\sigma(X)\sigma(Y)}$$

- بعض خواص معامل الارتباط:

1- $\rho(X,Y)=\rho(Y,X)$

2- $\rho(X\mp a,Y\mp b)=\rho(X,Y)$ حيث : a و b ثوابت .

3- $\rho(aX,bY)=\rho(X,Y)$

4- $\rho(X,X)=1$

5- أياً كان X مرتبطاً بـ Y فإنَّ: $-1 \le \rho(X,Y) \le 1$

6- إذا كان X و Y مستقلان، فإنَّ: $\rho(X,Y)=0$ والعكس غير صحيح، أي : إذا كان معامل الارتباط صفراً، فهذا لا يُعتبر شرطاً كافياً للاستقلال، لأنَّ تعريف الاستقلال يختلف عن ذلك.

مثال (1) :

إذا كان جدول التوزيع المشترك للشعاع العشوائي (X,Y) من الشكل:

$X \backslash Y$	1	2	sum
1	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{6}{12}$
2	$\frac{1}{12}$	$\frac{4}{12}$	$\frac{5}{12}$
3	$\frac{1}{12}$	0	$\frac{1}{12}$
Sum	$\frac{4}{12}$	$\frac{8}{12}$	1

أوجد: $Cov(X,Y)$ ، $Cov(3X,2Y)$ ، $\rho(X,Y)$.

الحل:

واضح أنَّ جدول توزيع كلٍ من X و Y بالشكل:

X	1	2	3
	$\frac{6}{12}$	$\frac{5}{12}$	$\frac{1}{12}$

Y	1	2
	$\frac{4}{12}$	$\frac{8}{12}$

وجدول توزيع XY بالشكل:

XY	1	2	3	4	6
	$\frac{2}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	$\frac{4}{12}$	0

ولدينا:

$$Cov(X,Y) = E(XY) - E(X).E(Y)$$

بالتالي:

$$E(X) = \sum_i x_i P_i = ... = \frac{19}{12}$$

$$E(Y) = \sum_j y_j P_j = ... = \frac{20}{12}$$

$$E(XY) = \sum_i \sum_j x_i y_j . P_{ij} = ... = \frac{31}{12}$$

ومنه:

$$Cov(X,Y) = \frac{31}{12} - \frac{380}{144} = -0.0556$$

$$Cov(3X, 2Y) = 6\,Cov(X,Y) = -0.3336$$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$

$$\sigma^2(X) = E(X^2) - [E(X)]^2$$
$$= \frac{35}{12} - (\frac{19}{12})^2$$
$$= 0.4097$$

وبطريقة مشابهة يكون: $\sigma^2(Y) = 0.2222$

بالتالي: $\rho(X,Y) = \dfrac{-0.0556}{(0.64)(0.471)} = -0.184$

مثال (2):

بفرض أنْ:

$$f(x,y) = k(6 + x - y) \quad ;\begin{cases} 0 \le x \le 3 \\ 0 \le y \le 2 \end{cases}$$

أوجد k بحيث يكون $f(x,y)$ تابع كثافة مشترك للشعاع العشوائي (X,Y)، ثم أوجد كلاً من $\rho(X,Y)$ ، $Cov(X,Y)$

الحل:

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = k\int_{0}^{3}\int_{0}^{2}(6+x-y)\,dx\,dy = \ldots = 39k \Rightarrow k=\frac{1}{39}$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)\ ;$$

$$E(X)=\int_{0}^{3} x\, f_1(x)\,dx\ ;$$

$$f_1(x)=\int_{0}^{2} f(x,y)\,dy$$

$$=\frac{1}{39}\int_{0}^{2}(6+x-y)\,dy=\ldots=\frac{1}{39}(10+2x)$$

ومنه:

$$E(X)=\frac{1}{39}\int_{0}^{3} x(6+x-y)\,dx=\ldots=\frac{63}{39}$$

وبطريقة مشابهة يكون:

$$E(Y)=\ldots=\frac{74}{78}$$

$$E(XY)=\int_{0}^{3}\int_{0}^{2} xy.\ f(x,y)dx\,dy$$

$$=\frac{1}{39}\int_{0}^{3}\int_{0}^{2} xy(6+x-y)dxdy=\ldots=\frac{60}{39}$$

بالتالي:

$$Cov(X,Y)=\frac{60}{39}-(\frac{63}{39})(\frac{74}{78})=0.00631$$

لدينا:

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$

حيث:

$$\sigma^2(X) = E(X^2) - [E(X)]^2$$

$$= \int_0^3 x^2 f_1(x)dx - \{\int_0^3 x f_1(x)dx\}^2 = \ldots = \frac{261}{78} - \frac{63}{39} = \frac{135}{78}$$

$$\Rightarrow \sigma(X) = \sqrt{\frac{135}{78}} = 0.859$$

وبطريقة مشابهة ينتج: $\sigma(Y) = 0.575$ (تحقَّق من ذلك)

وبالتالي:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{0.00631}{(0.859)(0.575)} = 0.01277$$

....

(7–9) تمارين الفصل السابع:

1– بفرض X و Y متغيِّران عشوائيان مستقلان، لكلٍ منهما جدول توزيع كما في الشكل:

X	-1	0	1
	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

Y	-1	0	1
	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

المطلوب:

1) أوجد جدول توزيع المتغيِّر : $Z = X + Y$

2) أوجد: $\sigma^2(Z)$, $E(XY)$

2– إذا كان جدول التوزيع المشترك للشعاع العشوائي (X,Y) من الشكل:

$X \backslash Y$	-1	0	1
-1	$\frac{1}{8}$	0	$\frac{1}{8}$
1	$\frac{4}{8}$	$\frac{2}{8}$	0

المطلوب:

1) بيِّن ما إذا كان X و Y مستقلين أم لا ؟

2) أوجد: $P(X=-1 \backslash Y=1)$, $P(X=-1, Y \le 0)$

3) أوجد: $\rho(X,Y)$

3– بفرض تابع الكثافة المشترك للشعاع العشوائي (X,Y) من الشكل:

$$f(x,y)=\begin{cases} c(-2xy+x+y) & ; \begin{cases} x\in[0,1] \\ y\in[0,1]\end{cases} \\ 0 & ; \begin{cases} x\notin[0,1] \\ y\notin[0,1]\end{cases}\end{cases}$$

المطلوب:

1) أوجد قيمة الثابت c

2) برهن أنَّ X و Y غير مستقلين

3) أوجد: $P(X \le 2, Y \le 1)$, $P(X=1, Y=1)$

4) أوجد: $E(Y \setminus X=x)$, $F_1(x)$, $\rho(X,Y)$

4- بفرض:

فيما عدا ذلك: $f(x,y)=\begin{cases} 2\,;\ 0<x<y<1 \\ 0\,;\ \ otherwise\end{cases}$

أوجد : $f_1(x)$, $f_2(y)$, $F(x,y)$

5- بفرض X و Y متغيّران عشوائيان منفصلان. إذا كان:

$$P_{ij}=\frac{i+j}{21}\,;\begin{cases} i=1,2,3 \\ j=1,2\end{cases}$$

أوجد: $P_{i.}$, $P_{.j}$, $P(X=2)$, $P(Y=1)$

6- بفرض: $\left\{ \begin{array}{l} X:\ 0\ ,\ 1\ ,\ 2 \\ Y\ :\ 0\ ,\ 1\ ,\ 2 \end{array} \right.$ ، حيث X و Y متغيّرين عشوائيين مستقلين.

إذا كان: $\begin{cases} P(X=0)=\dfrac{1}{2} \\ P(X=2)=\dfrac{1}{6} \end{cases}$ ، وكان : $\begin{cases} P(Y=0)=\dfrac{1}{6} \\ P(Y=1)=\dfrac{1}{3} \end{cases}$

المطلوب:

1ة) أوجد: $\Psi_X(t)$, $F(x)$, $E(X)$.

2ة) أوجد: $\sigma^2(Y)$, $E(Y)$.

3ة) إذا كان: $Z=X+Y$ أوجد : $P(Z=3)$.

7- بفرض X و Y متغيّرين عشوائيين مستمرين، تابع كثافتهما المشترك من الشكل:

$$f(x,y)=\frac{1}{32}(8-x-y)\ ;\ \begin{cases} x\in(0,4) \\ y\in(1,3) \end{cases}$$

المطلوب:

1ة) أوجد: $f_2(y)$, $f_1(x)$.

2ة) احسب: $P(Y<2)$, $P(X\le 3)$.

3ة) أوجد: $\rho(X,Y)$.

8- بفرض للشعاع العشوائي المستمر (X,Y) تابع الكثافة التالي:

$$f(x,y)=e^{-x^2y}\ \ ;\ x\ge 1\ ,\ y\ge 0$$

بيّن أنّ: $P(X^2Y)>\dfrac{1}{e}$

9- إذا كان X و Y متغيّرين عشوائيين مستقلين بحيث:

$$P(X=i)=pq^i \quad , i=0,1,2,...$$
$$P(Y=j)=pq^j \ , \ j=0,1,2,...$$

أوجد جدول توزيع المتغيّر: $Z=X+Y$

10- إذا كان: $f(y \setminus x)=\dfrac{8-x-y}{12-2x} \ ; 0 \le x \le 4 \ , \ 1 \le y \le 3$

وكان: $F_1(x)=\dfrac{1}{32}(12x-x^2) \ ; 0 \le x \le 4$

المطلوب:

1) أوجد: $f(x,y)$

2) أوجد: $f_2(y)$

3) احسب الاحتمال: $P(Y \le 2)$

10- أكمل الجدول التالي: (علماً أنّ X و Y متغيّرين عشوائيين مستقلين)

$X \setminus Y$	0	1	2
3	?	?	0.12
4	?	0.20	0.16
5	?	0.15	0.12
	0.10	?	?

11- بفرض للشعاع العشوائي المنفصل (X,Y) جدول التوزيع التالي:

$X \setminus Y$	1	2
3	$\frac{5}{28}$	$\frac{7}{28}$
5	$\frac{7}{28}$	$\frac{9}{28}$

المطلوب:

1) من أجل: $Z = X\,Y$ أوجد جدول توزيع Z

2) إذا كان تابع التوزيع المشترك للشعاع المذكور من الشكل:

$$P_{ij} = \frac{i+2j}{28}\ ,\ \begin{cases} i = 3,5 \\ j = 1,2 \end{cases}$$

3) أوجد: $P_{i.}$ بدلالة i ، ثم استنتج $P(X=3)$

4) أوجد: $\Psi_Y(t)$، ثم احسب الاحتمال: $P(Y=1 \setminus X=5)$

5) أوجد: $Cov(2X, 3Y)$.

....

الفصل الثامن

توزيع العيّنة والتقدير (النقطي – المجالي)

يهدف الإحصاء باعتباره فرعاً من فروع الرياضيات التطبيقيّة إلى دراسة خصائص عدديّة للمجتمعات الإحصائيّة، والمقصود بالمجتمع الإحصائي هو جملة الأشياء أو العناصر التي تشترك بصفة أو أكثر والتي تشكّل هدف الدراسة.

بفرض أنّ A مجتمع إحصائي، وأنّ X هو المتغيّر العشوائي الذي يمثّل الصفة أو الظاهرة التي نريد التعرُّف عليها في هذا المجتمع (تقديرها)، علماً أنّ عبارة مجتمع إحصائي قد تعني: مجموعة بشريّة، مجموعة آراء ، مجموعة نتائج تجربة عشوائيّة، مجموعة أزهار في حديقة...الخ.

وقد يكون المجتمع واقعيّاً (عدد الطلبة في كليّة ما) أو افتراضيّاً (نتائج قذف قطعة نقود عندما تُكرَّر التجربة عدد لانهائي من المرات).

فإذا كان لـ X توزيع طبيعي، نقول عن A إنّه مجتمع طبيعي، وإذا كان لـ X توزيع بواسوني نقول عن A إنّه مجتمع بواسوني، وهكذا صفة المجتمع تتبع توزيع المتغيّر X .

وفي كل الأحوال ننظر لـ X كمتغيّر عشوائي له توزيع احتمالي يعتمد على وسطاء من المجتمع (معالم) تحدّد ذلك التوزيع (مثلاً: في التوزيع الطبيعي يوجد وسيطان هما μ و σ)، علماً أنّ X قد يكون منفصلاً أو مستمراً، وبالتالي نحصل على مجتمع منفصل أو مستمر.

نريد دراسة المجتمع/ صفة أو ظاهرة ما، وضع اجتماعي أو اقتصادي.../ من خلال دراسة عيّنة أو عدة عيّنات عشوائيّة مسحوبة من هذا المجتمع وذلك تبعاً لعدد عناصر المجتمع الذي نرمز له بالرمز N (قد يكون N منتهياً أو غير منتهٍ)، ونسميه بحجم المجتمع، وتتم الدراسة عادةً من خلال العيّنات وذلك لتعذُّر دراسة كل فرد فيه على حده ، ولو كان حجم المجتمع صغيراً بقدرٍ كافٍ لتمكنّا من ذلك، وبالتالي لن يكون هناك عندئذٍ من داعٍ لدراسة العيّنات.

إذاً المسألة المطروحة بالشكل:

لنفترض أنّه لدينا مجتمع إحصائي حجمه N، ونريد أن نسحب عيّنة عشوائيّة من هذا المجتمع حجمها n ($n < N$) وقد يستلزم الأمر سحب عدة عيّنات كما قلنا، أضف إلى أنّ حجم العيّنة يجب أن يكون محدوداً أياً كان حجم المجتمع منتهياً أم غير منتهي، ومن خلال دراسة هذه العيّنة أو تلك العيّنات، نعمّم الدراسة على كل المجتمع، وهذه الدراسة، أي الانتقال من الخاص إلى العام تسمى بالاستدلال أو الاستقراء، وهو يُبنى على الطرق التجريبيّة حتماً، ويلعب دوراً هاماً في توسيع المعارف البشريّة، في حين يسمّى الانتقال من العام إلى الخاص بالاستنتاج، وهذا الأخير يعتمد العلوم غير التجريبيّة وفي مقدمتها الرياضيات البحتة، بينما تسمّى الظاهرة أو الصفة التي نودُّ التعرّف إليها بالمادة الإحصائيّة.

فإذا تتبعنا الوضع الاجتماعي أو الاقتصادي لمجتمع إحصائي ما خلال عشر سنوات خلت طبقاً لمعايير معيّنة، فإنّه وتبعاً لذات المعايير يمكن استقراء حالة المجتمع بعد سنة أو عدة سنوات أخرى، ولكن ليس بشكل مؤكّد، في حين عندما نقول: إنّ B مثلّث متساوي الأضلاع، فإنّه ببساطة نستنتج أنّ كل زاوية فيه هي $(60)^{\circ}$، وهذا الاستنتاج مؤكّد.

بمعنى آخر: إنّ نتائج الاستدلال أو الإستقراء هي نتائج احتماليّة، في حين نتائج الاستنتاج هي نتائج مؤكّدة وليست احتماليّة.

إنّ عملية اتخاذ قرار (قرار إحصائي) تعتمد على جمع كميّة من المعلومات، ومن ثمّ تصنيف وتحليل هذه المعلومات، وكلما كانت المعلومات المقدّمة صحيحة ودقيقة، كلما كان القرار المُتّخذ أمثليّاً، وبغير ذلك نكون أبعد عن الصواب.

وقد توجد مؤثّرات عشوائيّة ومعلومات تتحكم فيها الصدفة، وبالتالي سيتأثّر القرار المُتّخذ عندئذٍ بهذه المعلومات الاحتماليّة ولن يعكس بدقة تامة حالة كل فرد من أفراد المجتمع، ويبقى الهدف دائماً هو محاولة الحصول على قرار أمثل قدر الإمكان بنتائجه وانعكاساته الإيجابيّة على المجتمع الإحصائي المدروس.

(8-1) توزيع العيّنة:

بفرض A مجتمع إحصائي حجمه N، موصوف بالمتغيّر X، والسؤال المطروح: كيف يمكن معرفة توزيع X على المجتمع من خلال معرفة توزيعه على عيّنة عشوائيّة حجمها n من ذات المجتمع ؟ بمعنى آخر: إذا فرضنا أنّ: $F^*(x)$ تابع توزيع X على العيّنة معلوم، و $F(x)$ تابع توزيع X على المجتمع الإحصائي مجهول:

نريد معرفة $F(x)$ من خلال معرفة $F^*(x)$ أياً كان حجم المجتمع A. ونؤكّد مسبقاً أنّ $F(x)$ الناتج سيكون تقريبيّاً طالما أنّ معلوماتنا مستقاة من عيّنة عشوائيّة، ولن يعكس تماماً حالة كل فرد في المجتمع الإحصائي المدروس.

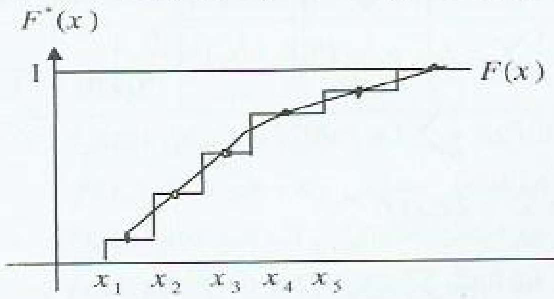
فإذا كانت قيم X وتردُّداتها (تكراراتها)على العيّنة العشوائيّة من الشكل:

$$X \ : \ x_1 \ , x_2 \ , ... \ x_n$$
$$f_1 \ , f_2 \ , ... f_n$$

عندئذٍ يكون: $P_i = P(X = x_i) = \dfrac{f_i}{n}$ حيث: f_i هو تردد النقطة x_i.

بالتالي: $F^*(x) = P(X \le x) = \sum\limits_{x_i \le x} P_i = \dfrac{1}{n} \sum\limits_{x_i \le x} f_i = \dfrac{f}{n}$ حيث: f مجموع الترددات الواقعة في النقطة x وما قبلها.

ومن بيان $F^*(x)$ التالي (وهو مدرّج الشكل كأي تابع توزيع):



نستنتج شكل التابع $F(x)$ المجهول وذلك بتقريب الخط الواصل بين منتصفات النقاط الموضّحة على الشكل إلى منحني، هذا المنحني، يُقدَّر على أنّه $F(x)$ (تقريبي بالطبع).

بالحقيقة إنَّ حساب $F(x)$ (وبالتالي حساب تابع كثافته $f(x)$) في المجتمع الإحصائي هو من المسائل المهمة جداً في الإحصاء، أضف إلى أنّه مهما تكن دقة القياس (وهي محدودة بالطبع لأنَّ ذلك يرتبط بحواس الإنسان)، ومهما كانت المعلومات مستمرة، فإنَّ الإنسان لا يمكن أن يستوعبها إلا في شكلها المتقطّع، بمعنى:

$$\forall F^{*}(x)\ (\text{مستمر أم منقطع}) \Rightarrow F(x)\ \text{منقطع}$$

مما تقدّم نلاحظ أنَّ توزيع X في العيّنة يختلف بنسبة ما عن توزيعه في المجتمع، وبالتالي، وبذات النسبة سيختلف $F(x)$ عن $F^{*}(x)$، وأياً كان X ، فإنَّ لـ X توزيعين: توزيعاً على العيّنة، وتوزيعاً على المجتمع، وفي هذه الحالة نستنتج الصفات القياسيّة المميّزة لـ X على المجتمع المدروس من خلال صفاته القياسيّة المميّزة على العيّنة المأخوذة من ذات المجتمع، والتي سنرمّزها كما في الجدول التالي:

القيمة الوسيطية المميّزة	رمزها في توزيع المجتمع	رمزها في توزيع العيّنة
التوقع الرياضي	μ	$\bar{x}$
التباين	σ^2	s^2
العزم الابتدائي من الرتبة i	α_i	a_i
العزم المركزي من الرتبة i	μ_i	m_i
القيمة التصفيّة	$\Im_2$	z_2

والمسألة الآن هي حساب القيم المميّزة ($\bar{x}$ ، s^2 ، a_i ، m_i ، z_2) في توزيع العيّنة التي حجمها n والمأخوذة من ذات المجتمع الإحصائي الذي حجمه N حيث:

$$\bar{x} = \sum_{i=1}^{n} x_i P_i = \frac{1}{n}\sum_i x_i f_i \ ;\ p_i = P(X = x_i) = \frac{f_i}{n}$$

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 P_i = \frac{1}{n}\sum_i (x_i - \bar{x})^2 f_i$$

$$= \frac{1}{n}\sum_i (x_i^2 + \bar{x}^2 - 2x_i\bar{x}) f_i$$

$$= \frac{1}{n}\sum_i x_i^2 f_i + \bar{x}^2 \underbrace{\left(\frac{1}{n}\sum_i f_i\right)}_{1} - 2\bar{x}\underbrace{\left(\frac{1}{n}\sum_i x_i f_i\right)}_{\bar{x}}$$

$$= \frac{1}{n}\sum_i x_i^2 f_i + \bar{x}^2 - 2\bar{x}^2 = \frac{1}{n}\sum_i x_i^2 f_i - \bar{x}^2$$

ولكن: $a_i = E(X^i) \Rightarrow a_2 = E(X^2) = \sum_i x_i^2 P_i = \frac{1}{n}\sum_i x_i^2 f_i$

بالتالي يمكن كتابة s^2 بالشكل: $s^2 = a_2 - \bar{x}^2$ ، كذلك نجد:

$$m_j = E(X-\bar{x})^j = \sum_{i=1}^{n}(x_i-\bar{x})^j P_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^j f_i$$

. أخيراً: القيمة النصفية z_2 هي بالتعريف تلك القيمة التي لأجلها يكون: $F^*(z_2) = \frac{1}{2}$

(2-8) المتغيّرات الإحصائيّة:

ليكن المجتمع الإحصائي $\{A, N, X, f(x), F(x)\}$ أي: (اسم المجتمع وحجمه ومتغيّره العشوائي وتابع كثافته وتوزيعه)، ولنفرض أنّ N كبير بقدرٍ كافٍ، بحيث يتطلّب الأمر سحب عدة عيّنات لها نفس الحجم n، وهذا ما يُفترض أن يكون بالحالة العامة.

فإذا رمزنا لقيم X الملاحظة في العيّنات المسحوبة بالرموز التالية كما يلي:

العيّنة الأولى بالرمز	x_n	.....	x_3	x_2	x_1
العيّنة الثانية بالرمز	x_n'	.....	x_3'	x_2'	x_1'
العيّنة الثالثة بالرمز	x_n''	.....	x_3''	x_2''	x_1''
وهكذا	...	...	...	...	...
:	↓	↓	↓	↓	↓
:	X_n	.....	X_3	X_2	X_1
:	↓	↓	↓	↓	↓
:	$f(x_n)$	.....	$f(x_3)$	$f(x_2)$	$f(x_1)$

وبترتيب قيم هذه العيّنات كما في الشكل الوارد، فإنّه يمكن اعتبار:

قيم العمود الأول هي قيمٌ لمتغيّر عشوائي X_1 ، تابع كثافته $f(x_1)$

قيم العمود الثاني هي قيمٌ لمتغيّر عشوائي X_2 ، تابع كثافته $f(x_2)$

قيم العمود الثالث هي قيمٌ لمتغيّر عشوائي X_3 ، تابع كثافته $f(x_3)$

....

وقيم العمود ذي الرقم i هي قيم لمتغيّر عشوائي X_i ، تابع كثافته $f(x_i)$.

وهكذا نحصل على عدد من المتغيّرات العشوائيّة: $X_1, X_2, X_3, ..., X_n$ والتي يكون عددها n حيث: $X_i : \quad x_i, x_i', x_i'', x_i''', ...$

نسمي هذه المتغيّرات بمتغيّرات القيم الملاحظة لعيّنة عشوائيّة (عيّنة إحصائيّة) حجمها n ،وباعتبار {الحالة العامة: $n << N$ أصغر بكثير }، فإنّه يكون لهذه المتغيّرات الخاصتان المهمتان التاليتان:

1- $X_i, i = \overline{1,n}$ مستقلة بالتبادل مثنى مثنى.

2- لكلٍ من : $X_i, i = \overline{1,n}$ توزيع مماثل لتوزيع X في المجتمع الإحصائي A .

وهذا يكافئ القول: إنّ كلاً من $f(x_i)$, $i = \overline{1,n}$ مماثل لـ $f(x)$ في المجتمع A ، وبالتالي، إذا رمزنا بـ: $g(x_1, x_2, ..., x_n)$ لتابع الكثافة المشترك للمتغيّرات $X_1, X_2, X_3, ..., X_n$ ، فإنّه وانطلاقاً من الاستقلال يمكن أن نكتب:

$$g(x_1, x_2, ..., x_n) = f(x_1).f(x_2)...f(x_n)$$

سندعو (من الآن فصاعداً) المتغيّرات $X_1, X_2, X_3, ..., X_n$:عيّنة عشوائيّة حجمها n ، ونسمي أية علاقة أو كمية نحصل عليها من هذه العيّنة بهدف تقدير قيمة وسيطية ما من القيم المميّزة للمجتمع الإحصائي، نسميها متغيّراً إحصائياً أو إحصاء فقط .

إذاً: الإحصاء أو المتغيّر الإحصائي هو: كل تابع لعناصر العيّنة العشوائيّة لا يعتمد على الوسطاء المجهولة.

فمثلاً:

إذا كانت: $X_1, X_2, X_3, ..., X_n$ متغيّرات القيم الملاحظة لعيّنة عشوائيّة حجمها n وأخذنا: $m_s = \frac{1}{n}\sum_i X_i^{\ s}$ (وهو يمثّل عزم عيّنة عشوائيّة من الرتبة s). عندئذٍ :

$m_1 = \frac{1}{n}\sum_i X_i$ ، $m_2 = \frac{1}{n}\sum_i X_i^{\ 2}$ ، $m_3 = \frac{1}{n}\sum_i X_i^{\ 3}$ ، كلها متغيّرات إحصائيّة، وكذلك التعابير من الشكل: { $\sum_{i=1}^{n} X_i$. $\prod_{i=1}^{n} X_i$ } تمثل إحصاء أو متغيّرات إحصائيّة،

في حين التعابير من الشكل: $\mu - \frac{X_i + X_j}{\sigma}$ ، $\sigma(X_i + X_j) + \mu$ وما شابه ليست بإحصاء ، طالما بقيت σ , μ مجهولة.

* **المتغيّر الإحصائي $\overline{X}$:(متوسط العيّنة):**

وهو يُعرّف بالشكل: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ حيث: $(i = \overline{1,n})$, X_i هي متغيّرات القيم الملاحظة في العيّنة العشوائيّة المسحوبة من المجتمع الإحصائي المدروس، حيث المجتمع الإحصائي هنا، هو مجتمع طبيعي حصراً، وبناءً على استقلاليتها، ومماثلة توزيع كلٍ منها لتوزيع المجتمع، فإنّه يمكن دراسة صفاتها المميّزة:

. الصفات المميّزة للمتغيّر الإحصائي $\overline{X}$:

1- التوقع الرياضي:

$$E(\overline{X}) = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i} E(X_i)$$
$$= \frac{1}{n}(n.E(X)) = E(X)$$

أي: $E(\overline{X}) = E(X)$ ⇐ توقع متوسط العيّنة = توقع X في المجتمع.

وهذه النتيجة تبيّن الدور الهام للمتغيّر الإحصائي $\overline{X}$ في دراسة المجتمعات الإحصائيّة.

2- التباين:

$$\sigma^2(\overline{X}) = \sigma^2(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2(X_i)$$
$$= \frac{1}{n^2}(n\,\sigma^2(X)) = \frac{\sigma^2(X)}{n}$$

أي: $$\sigma^2(\overline{X}) = \frac{\sigma^2(X)}{n} \Rightarrow \sigma(\overline{X}) = \frac{\sigma(X)}{\sqrt{n}}$$

3- الدالّة المولّدة للعزوم:

$$\Psi_{\overline{X}}(t) = E(e^{t\overline{X}}) = E(e^{\frac{t}{n}\sum_i X_i})$$

$$= E(e^{\frac{t}{n}X_1} e^{\frac{t}{n}X_2} \dots e^{\frac{t}{n}X_n})$$

$$= E(e^{\frac{t}{n}X_1}).E(e^{\frac{t}{n}X_2}) \dots E(e^{\frac{t}{n}X_n})$$

$$= \Psi_{X_1}(\frac{t}{n}).\Psi_{X_2}(\frac{t}{n}) \dots \Psi_{X_n}(\frac{t}{n}) = [\Psi_X(\frac{t}{n})]^n$$

أي: $\Psi_{\overline{X}}(t) = [\Psi_X(\frac{t}{n})]^n$

وبالتالي من أجل: $X : N(\mu, \sigma^2)$ فإنَّ : $\Psi_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

وعندئذٍ يكون: $\Psi_{\overline{X}}(t) = e^{\mu t + \frac{1}{2}\frac{\sigma^2}{n}t^2}$ أي: $\overline{X} : N(\mu, \frac{\sigma^2}{n})$

الأمر الذي يعني أنَّ: $\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} : N(0,1)$

وجدنا ممّا سبق أنّ: $E(\overline{X}) = E(X)$ ، ولكن على صعيد الواقع، وخاصةً عندما يكون حجم المجتمع كبيراً أو لانهائيّاً، هل يكون الأمر كذلك تماماً ؟

بالحقيقة، ولكوننا نستقي معلوماتنا من عيّنة عشوائيّة محدّدة من المجتمع الإحصائي، فلا بدَّ من وجود ارتياب في حساب $E(X)$ ، ويبقى السؤال المهم التالي مطروحاً: ما هي درجة الثّقة في حساب الكمية $E(X)$ ؟ أو بمعنى آخر: هل يمكن حساب $E(X)$ بأقل خطأ ممكن؟

والجواب بالإيجاب حسب النظريّة المهمة التالية:

(3-8) قانون الأعداد الكبيرة: *Law of large numbers*

(يسمى أحياناً نظريّة تشيبتشف)

ليكن المجتمع الإحصائي $\{ A, X, f(x) \}$، ولنفرض أنّ: $\{ E(X) = \mu \ , \sigma^2(X) = \sigma^2 \}$

عندئذٍ:

$$\forall \varepsilon, \delta > 0 , \exists n_0 \in \mathbb{N} \; ; n > n_0 \Rightarrow$$
$$P\{|\overline{X} - \mu| < \varepsilon\} \geq 1 - \delta$$

أي أنّه باحتمال كبير (احتمال قريب من الواحد) يمكن أن نجعل الكمية $|\overline{X} - \mu|$ صغيرة بالقدر الذي نريد، وذلك عندما تكون n كبيرة بشكلٍ كافٍ و: ε ، δ صغيرين جداً.

البرهان:

لدينا حسب متراجحة تشيبتشف: $P[|X - \mu| < k\sigma] \geq 1 - \frac{1}{k^2} \; ; k \in R^+$

وفي توزيع العيّنة حيث: $\sigma = \frac{\sigma}{\sqrt{n}}$ ، $X = \overline{X}$ ، تُكتب هذه المتراجحة بالشكل:

$$P[|\overline{X} - \mu| < k\frac{\sigma}{\sqrt{n}}] \geq 1 - \frac{1}{k^2}$$

وبالتالي إذا اعتبرنا أنّ: $\varepsilon = k\frac{\sigma}{\sqrt{n}}$ ، $\delta = \frac{1}{k^2}$ يكون قد تمّ المطلوب .

(4-8) نظريّة النهاية المركزيّة: *Central limit theorem*

إذا كانت: $X_i, i = \overline{1,n}$ عيّنة عشوائيّة (أي مستقلة ومتماثلة التوزيع)

بحيث: $\forall i = \overline{1,n}$ ، $\sigma(X_i) = \sigma$ ، $E(X_i) = \mu$ وكان: $Y = \sum_{i=1}^{n} X_i$ عندئذٍ:

$Z = \frac{Y - E(Y)}{\sigma(Y)} : N(0,1)$ وذلك عندما $n \to \infty$ ، أي: n كبيرة جداً.

(إنّ أوّل من عرض هذه النظريّة هو لابلاس 1812م، وبرهنها ليابونوف 1901م)، ولهذه النظريّة أهميّة بالغة في الإحصاء كما سنجد.

إنّ أهمية هذه النظريّة تتأتّى من كونها اعتمدت متغيّرات عشوائيّة مختلفة، بصرف النظر عن طبيعة المجتمع الإحصائي الذي جاءت منه (طبيعياً كان أم غير ذلك)، إذ يكفي أن يكون لهذه المتغيّرات العشوائيّة توقع رياضي وانحراف معياري أياً كان شكل توزيعاتها.

البرهان:

يكفي أن نبرهن أنّ: $\lim\limits_{n\to\infty} \Psi_Z(t) = e^{\frac{t^2}{2}}$

نعلم أنّه من أجل:

$$X : N(\mu,\sigma) \Rightarrow \Psi_X(t) = e^{\mu t+\frac{1}{2}\sigma^2 t^2}$$

$$X : N(0,1) \quad \Rightarrow \Psi_X(t) = e^{\frac{t^2}{2}}$$

لدينا: $E(Y) = E(\sum_i X_i) = n\,E(X_i) = n\mu$

كذلك : $Y = n\overline{X} \Rightarrow \sigma^2(Y) = n^2\sigma^2(\overline{X}) = n^2 \dfrac{\sigma^2(X)}{n} = n\sigma^2$

بالتالي: $Z = \dfrac{n\overline{X} - n\mu}{\sigma\sqrt{n}} = \dfrac{\sqrt{n}}{\sigma}\overline{X} - \dfrac{\sqrt{n}}{\sigma}\mu = a\overline{X} + b \quad ; a = \dfrac{\sqrt{n}}{\sigma} \ , b = \dfrac{-\sqrt{n}}{\sigma}\mu$

ومنه:

$$\begin{aligned}\Psi_Z(t) &= \Psi_{a\overline{X}+b}(t) = E[e^{(a\overline{X}+b)t}] \\ &= e^{bt}E(e^{a\overline{X}t}) = e^{bt}\Psi_{\overline{X}}(at)\end{aligned}$$

وبالاستفادة من العلاقة: $\Psi_{\overline{X}}(t) = [\Psi_X(\frac{t}{n})]^n$ يمكن أن نكتب:

$$e^{bt}\Psi_{\overline{X}}(at) = e^{bt}[\Psi_X(\frac{at}{n})]^n$$

ويصبح عندئذٍ:

$$\Psi_Z(t) = e^{bt}[\Psi_X(\frac{at}{n})]^n = [e^{\frac{b}{n}t}\Psi_X(\frac{at}{n})]^n =$$

$$= [\Psi_{\frac{a}{n}X+\frac{b}{n}}(t)]^n = [\Psi_{\frac{1}{\sqrt{n}}\underbrace{(\frac{X-\mu}{\sigma})}_{U}}(t)]^n = [\Psi_{\frac{U}{\sqrt{n}}}(t)]^n$$

حيث استبدلت كلٍ من a و b بقيمهما، وفرضنا أنّ:

$$U = \frac{X-\mu}{\sigma} \Rightarrow \begin{cases} E(U) = \alpha_1 = 0 \\ \sigma^2(U) = 1 \end{cases}$$

بالتالي:

$$\Psi_Z(t)=\left\{E(e^{\frac{tU}{\sqrt{n}}})\right\}^n=\left\{\int_{-\infty}^{\infty}e^{\frac{tu}{\sqrt{n}}}f(u)du\right\}^n$$

حيث: $f(u)$ تابع كثافة المتغيّر المنتظم U، وحسب نشر (ماك – لوران):

$$e^{\frac{tu}{\sqrt{n}}}=1+\frac{tu}{\sqrt{n}}+\frac{t^2u^2}{2n}+\frac{t^3u^3}{6n^{\frac{3}{2}}}+\dots$$

وبالتالي:

$$\Psi_Z(t)=\left\{\int_{-\infty}^{\infty}(1+\frac{tu}{\sqrt{n}}+\frac{t^2u^2}{2n}+\dots)f(u)du\right\}^n$$

$$=\left\{\int_{-\infty}^{\infty}f(u)du+\frac{t}{\sqrt{n}}\int_{-\infty}^{\infty}uf(u)du+\frac{t^2}{2n}\int_{-\infty}^{\infty}u^2f(u)du+\dots\right\}^n$$

$$=\left\{1+\frac{t}{\sqrt{n}}\alpha_1+\frac{t^2}{2n}\alpha_2+\frac{t^3}{6n^{\frac{3}{2}}}\alpha_3+\dots\right\}^n$$

$$=\left\{1+0+\frac{t^2}{2n}.1+0+\dots\right\}^n=\left\{1+\frac{t^2}{2n}\right\}^n$$

$$\Rightarrow\underbrace{\lim}_{n\to\infty}\Psi_Z(t)=\underbrace{\lim}_{n\to\infty}\left\{1+\frac{t^2}{2n}\right\}^n=e^{\frac{t^2}{2}}$$

وهو يمثّل الدالّة المولّدة للعزوم للتوزيع الطبيعي المعياري، الأمر الذي يعني أنّه عندما: $n\to\infty$ فإنّ: $Z=\frac{Y-E(Y)}{\sigma(Y)}:N(0,1)$، وبالتالي: $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}:N(0,1)$.

وهذه النتيجة تبرز الأهمية الكبيرة لنظريّة النهاية المركزيّة في دراسة الإحصاء الرياضي.

ملاحظة: ينتج من النظريّة السابقة أنّه إذا كان A مجتمعاً إحصائياً موصوفاً بمتغيّر عشوائي X مجهول التوزيع بحيث: $E(X)=\mu$ و: $\sigma(X)=\sigma$، وإذا كانت: $X_1, X_2, \dots, X_n$ متغيّرات القيم الملاحظة لعيّنة عشوائيّة حجمها n ، كبير إلى الحد الذي يسمح بالاستفادة من نظريّة النهاية المركزيّة، وإذا وضعنا:

$Y=n\overline{X}=X_1+X_2+\dots+X_n$ عندئذٍ يكون: $Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}:N(0,1)$

ومنه ينتج أنَّ: $\overline{X} : N(\mu, \frac{\sigma^2}{n})$ وهذا ما وجدناه أثناء دراسة المتغيّر الإحصائي $\overline{X}$.

*** المتغيّر الإحصائي S^2 (تباين العيّنة):**

وهو يُعرّف بالشكل:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$E(S^2) = \frac{1}{n}E[\sum_{i=1}^{n}(X_i - \overline{X})^2] = \frac{1}{n}E[\sum_{i}(X_i - \mu + \mu - \overline{X})^2]$$

$$= \frac{1}{n}E\left\{\sum_{i}[(X_i - \mu)^2 + (\mu - \overline{X})^2 + 2(X_i - \mu)(\mu - \overline{X})]\right\}$$

$$= \frac{1}{n}E[\sum_{i}(X_i - \mu)^2 + n(\mu - \overline{X})^2 + 2\sum_{i}(X_i - \mu)(\mu - \overline{X})]$$

ولكن:

$$\sum_{i}(X_i - \mu) = X_1 - \mu + X_2 - \mu + ... + X_n - \mu$$

$$= X_1 + X_2 + ... + X_n - n\mu = n\overline{X} - n\mu = -n(\mu - \overline{X})$$

بالتبديل يكون:

$$E(S^2) = \frac{1}{n}E[\sum_{i}(X_i - \mu)^2 - n(\mu - \overline{X})^2]$$

$$= \frac{1}{n}[\sum_{i}E(X_i - \mu)^2 - nE(\mu - \overline{X})^2]$$

ولكن:

$$\sum_{i}E(X_i - \mu)^2 = \sigma_1^{\ 2} + \sigma_2^{\ 2} + ... + \sigma_n^{\ 2} = n\sigma^2 \quad ; E(X_i) = E(X)$$

$$E(\mu - \overline{X})^2 = E(\overline{X} - \mu)^2 = \sigma^2(\overline{X}) = \frac{\sigma^2}{n}$$

بالتالي: $E(S^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$ ، وهذا يعني أنّ S^2 هو مقدّر جيّد لـ σ^2 رغم أنّ: $\frac{n-1}{n}\sigma^2 = E(S^2) \neq \sigma^2$ ، ولكن عندما تكون n كبيرة جداً $(n \to \infty)$ فإنّ: $E(S^2) = \sigma^2(X)$ مع ارتكاب خطأ يتناسب عكساً مع n .

فـي التطبيقـات العمليّـة: نعتبـر n كبيـرة عنـدما $(n \geq 30)$، وبشـكل عـام، يُعرَّف متغيِّـر إحصائي آخر يسمّى: متغيِّر تباين العيِّنة المعدَّل $S^{\wedge 2}$.

. تباين العيِّنة المعدَّل $S^{\wedge 2}$:

ويعرَّف بالعلاقة: $$S^{\wedge 2} = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

ويُستخدَم في غالب الأحيان بدلاً من S^2 في الدراسات الإحصائيَّة لأنَّ:

$$E(S^{\wedge 2}) = \frac{n}{n-1} E(S^2) = \frac{n}{n-1} \frac{n-1}{n} \sigma^2 = \sigma^2$$

أي: $E(S^{\wedge 2}) = \sigma^2$ وهو مقدِّر أجود من S^2 كمـا هو واضـح الآن ، وكمـا سنجده فـي أثناء دراسة التقدير .

(8-5) تقريب التوزيع الثنائي بالتوزيع الطبيعي:

فـي أثنـاء دراسـة توزيـع ثنـائي الحد وفـي حالـة n كبيرة ، كـان هنـاك صـعوبات كبيـرة فـي الحسابات، لذلك اعتمدنا تقريب بواسون ، وبالحقيقة إنَّ نظريَّة النهاية المركزيَّة تُقدِّم حلاً لهذه المشكلة: حيث فـي توزيـع ثنـائي الحد اقتضـى الأمـر حسـاب احتمـال أن يأخـذ X (وهو عدد النجاحات) من بين n تكرار، قيمة معيَّنة، أو قيمة تقع ضمن مجال معيَّن، ولم تكن عندها n كبيرة بسبب صعوبة الحسابات، ولكن حسب نظريَّة النهاية المركزيَّة، يمكن النظر إلى عدد النجاحات X كمجموع يحقِّق شروط نظريَّة النهاية المركزيَّة، بحيث إذا قرنّا النجاح بالعدد 1، والفشل بالعدد 0 ، فعندئذٍ تكون نتائج التكرارات المستقلة والتي عددها n ، عبارة عن متتالية من المتغيِّرات المستقلة $X_1, X_2, X_3, ..., X_n$ حيث يأخذ كل X_i إما القيمة 1 وإما القيمة 0 ،

وعندها يكون عدد النجاحات هو بالضبط: عدد مرات ورود الـ 1 في تلك المتتالية أو مجموعها، أي: $X = \sum_{i=1}^{n} X_i$. وبما أنَّ كل من X_i يتوزع وفق توزيع ثنائي الحد، عندئذٍ تصبح نتـائج التكرارات المسـتقلة الـ n وهـي: $X_1, X_2, X_3, ..., X_n$ ، تصـبح أو يمكـن اعتبارها عيِّنة عشوائيَّة من مجتمع ثنائي الحد، ويصبح X مجموع هذه العيِّنة.

وفقاً لنظريّة النهاية المركزيّة يكون التوزيع التقريبي لـ X ، في حالة n كبيرة بشكلٍ كافٍ، هو التوزيع الطبيعي بتوقُّع رياضي يساوي $n\,p$ وتباين يساوي $n\,p\,q$.

وبالتالي يمكن استخدام جداول التوزيع الطبيعي المعياري لحساب احتمالات تتعلق بالمتغيّر X ، ولكن بصورة تقريبيّة.

كحالة خاصة: في المجتمعات الطبيعية والتي هي من أكثر المجتمعات الإحصائيّة بروزاً في الواقع العملي، لدينا النظريّة التالية والتي نقبلها دون برهان:

نظريّة: في المجتمع الطبيعي يكون المتغيّران الإحصائيّان $\overline{X}$ و S^2 مستقلين حيث:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad ، \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ملاحظات وأمثلة:

1 . إذا فرضنا أنّ: $(A\,;X : N(\mu_1,\sigma_1^2)$ و: $B\,;Y : N(\mu_2,\sigma_2^2)$ مجتمعان طبيعيان مستقلان . عندئذٍ يكون للمتغيّر الإحصائي $\overline{X}-\overline{Y}$ توزيع طبيعي من الأنموذج:

$$\overline{X}-\overline{Y} : N(\mu_1-\mu_2\,,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}) \quad \text{أي:} \quad (\mu_1-\mu_2\,,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$

حيث: $\overline{X}$ يقابل متوسط العيّنة : $X_1, X_2, X_3, \ldots, X_n$ من المجتمع الإحصائي $A\,;X : N(\mu_1,\sigma_1^2)$.

$\overline{Y}$ يقابل متوسط العيّنة: $Y_1, Y_2, Y_3, \ldots, Y_n$ من المجتمع الإحصائي $B\,;Y : N(\mu_2,\sigma_2^2)$.

البرهان:

واضح أنّ: $\overline{X} : N(\mu_1,\frac{\sigma_1^2}{n_1})$ و : $\overline{Y} : N(\mu_2,\frac{\sigma_2^2}{n_2})$

بالتالي: $\Psi_{\overline{X}-\overline{Y}}(t) = E(e^{(\overline{X}-\overline{Y})t}) = E(e^{\overline{X}t}).E(e^{-\overline{Y}t})$

$$= \Psi_{\overline{X}}(t) \,.\, \Psi_{\overline{Y}}(-t)$$

$$= e^{\mu_1 t + \frac{1}{2}\frac{\sigma_1^2 t^2}{n_1}} \,.\, e^{-\mu_2 t + \frac{1}{2}\frac{\sigma_2^2 t^2}{n_2}} = e^{(\mu_1-\mu_2)+\frac{1}{2}t^2(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})}$$

الأمر الذي يعني أنّ: $\overline{X}-\overline{Y} : N(\mu_1-\mu_2 \,, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$

ومنه (حسب نظريّة النهاية المركزيّة) يكون:

$$V = \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} : N(0,1)$$

2- لدينا: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

بالتالي (بعد إضافة وطرح μ) يكون:

$$nS^2 = \sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2$$

ومنه:

$$\underbrace{\frac{nS^2}{\sigma^2}}_{A} = \underbrace{\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2}_{B} - \underbrace{(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}})^2}_{C}$$

وبما أنّ: $\overline{X}$ و S^2 مستقلان، فإن A و C مستقلان أيضاً (نقبلها دون برهان)

بالتالي: $\Psi_B(t) = \Psi_{(A+C)}(t) = \Psi_A(t).\Psi_C(t)$

وبما أنّ: $C = (\frac{\overline{X}-\mu}{\sigma/\sqrt{n}})^2 : \chi_1^2$: و $B = \sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 : \chi_n^2$

يكون: $\Psi_A(t) = \frac{\Psi_B(t)}{\Psi_C(t)} = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-(\frac{n-1}{2})}$

أي: $\Psi_{\frac{nS^2}{\sigma^2}}(t) = (1-2t)^{-(\frac{n-1}{2})}$. الأمر الذي يعني أنّ: $\frac{nS^2}{\sigma^2} : \chi_{n-1}^2$

ومنه يمكن أيضاً استنتاج الصفات القياسيّة للمتغيّر الإحصائي S^2 حيث:

$$E(\frac{nS^2}{\sigma^2}) = n-1 \Rightarrow \frac{n}{\sigma^2}E(S^2) = n-1$$

ومنه: $E(S^2) = \frac{n-1}{n}\sigma^2$. وقد وجدنا ذلك سابقاً .

أضف إلى أنّ: $\sigma^2(\frac{nS^2}{\sigma^2}) = 2(n-1) \Rightarrow \frac{n^2}{\sigma^4}\sigma^2(S^2) = 2(n-1)$

ومنه: $\sigma^2(S^2) = \frac{2(n-1)}{n^2}\sigma^4$.

وبناءً عليه يمكن حساب $\sigma^2(S^{\wedge 2})$ بالشكل:

$$\sigma^2(S^{\wedge 2}) = \sigma^2(\frac{n}{n-1}S^2) = \frac{n^2}{(n-1)^2}\sigma^2(S^2)$$

$$= \frac{n^2}{(n-1)^2}\frac{2(n-1)}{n^2}\sigma^4 = \frac{2}{n-1}\sigma^4$$

3- وجدنا في توزيع t أنّه إذا كان: $X : N(0,1)$، و $Y : \chi^2_n$، وكان X و Y مستقلين، فإنّ: $U = \frac{X}{\sqrt{Y/n}} : t_n$

ونتيجة لذلك يمكن القول: بما أنّ : $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} ; N(0,1)$ و $\frac{nS^2}{\sigma^2} : \chi^2_{n-1}$

(وقد وجدنا ذلك سابقاً ، وهما مستقلان)، فإنّ:

$$T = \frac{\frac{\overline{X}-\mu}{\sigma\sqrt{n}}}{\sqrt{\frac{nS^2}{\sigma^2}\Big/(n-1)}} = \sqrt{n-1}\,\frac{\overline{X}-\mu}{S} = \frac{\overline{X}-\mu}{S^{\wedge}/\sqrt{n}}$$

بالتالي: $T = \frac{\overline{X}-\mu}{S^{\wedge}/\sqrt{n}} : t_{n-1}$

أي: المتغيّر T يخضع لتوزيع ستودينت و بـ $(n-1)$ درجة حريّة.

(6-8) المقدّر والتقدير الإحصائي النقطي:

التقدير النقطي *(Point Estimation)*:

نعلم أنّ دراسة توزيع مجتمع إحصائي ما (أو دراسة تابع كثافته) تتم بشكل عام من خلال دراسة عيّنة أو عدة عيّنات عشوائيّة مسحوبة من ذات المجتمع، وبالتالي القرار

(القرار الإحصائي) المتوجب قبوله أو رفضه، سيكون تبعاً لنتائج دراسة هذه أو تلك العيّنات.

لنفرض أنّ القرار الإحصائي المنتظر اتخاذه يتوقف على وسيط مجهول θ ، أو عدة وسطاء مجهولة $\theta_1, \theta_2, ..., \theta_n$ نريد تقديرها واستقرائها بشكل أمثل لكي يكون القرار الإحصائي المتوجب اتخاذه بناءً على أفضلها، أمثليّاً . وبالواقع إنّ التقدير يجب ألاّ يكون مُجرّد تخمين شخصي مبني على مقدرة الإنسان الذاتية في التوقُّع والاستنتاج... إنما يجب أن يتم بطرق واستقراءات علميّة سليمة، بحيث نتمكن بدرجة عالية من الثّقة القول بأنّ قرارنا أمثليّاً. (المقصود بالحل أو القرار الأمثل – هو ذاك الحل أو القرار الذي لا يوجد أفضل منه في حينه).

فإذا كان $f(x,\theta)$ تابع كثافة مجتمع إحصائي ما، حيث θ وسيط مجهول، فإنّ المسألة الأولى في الإحصاء هي إيجاد تقدير أمثل لهذا الوسيط ، علماً أنّ θ قد يأخذ قيمة محدّدة وحيدة (وهذا ما نسمّيه بالتقدير النقطي) أو قد يأخذ قيمته ضمن فترة أو مجال $[a,b]$ (ويسمى في هذه الحالة بالتقدير المجالي، وهذا ما سندرسه في الفقرات القادمة بعون الله)، والسؤال المهم المطروح في التقديرين هو: كيف نوجد هذه المقدّرات ؟ وما مدى ثقتنا بها إن وُجِدَت ؟

بفرض $f(x,\theta)$ تابع كثافة المجتمع الإحصائي A ، والمهمة الأولى كما قلنا هي إيجاد مقدّر (نرمز له بـ T) للوسيط θ بحيث يكون هذا المقدّر أمثليّاً.

فمثلاً: إذا كان الوسيط المجهول في المجتمع θ عبارة عن التوقع الرياضي لـ X ، فإنّ الدراسة الإحصائيّة توحي بأنّ المقدّر T يمكن أن يكون التوقع الرياضي للعيّنة. أي:

$$T = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

وبالتالي، فالمقدّر ـ هو الصيغة أو الدالّة التي تستخدم في التقدير، وهو متغيّر عشوائي تختلف قيمته من عيّنة إلى أخرى، في حين التقدير ـ هو قيمة المقدّر بعد تعويض بيانات العيّنة بالصيغة أو بالدالّة المقترحة، أي هو إحدى قيم هذا المتغيّر، وفي كل الأحوال، فإنّه لإيجاد المقدّر T لوسيط θ، لا بدّ من الخطوات الأساسيّة التالية:

1- تحديد مجموعة كل القيم الممكنة للوسيط θ في العيِّنة العشوائيَّة المسحوبة من ذات المجتمع، وتسمى هذه المجموعة عادةً بفراغ الوسيط θ ويرمز له بالرمز Ω.

وبالتالي إذا كان θ عبارة عن توقع رياضي، فإنَّ: $\Omega = R$، وإذا كان θ عبارة عن تباين فإنَّ: $\Omega =]0, \to[$، وأعتقد جازماً بأنَّ المسألة ستكون أكثر تعقيداً بكثير فيما لو كان الوسيط θ عبارة عن سلوك أو أخلاقيات في المجتمع.

2- اختيار المقدَّرات الملائمة (حيث يوجد بالحالة العامة أكثر من مقدَّر)، ومن ثم انتخاب هذه المقدَّرات لاختيار أمثلها. وفي هذه الحالة ستكون النتائج السلبيَّة المترتِّبة، إن وُجِدَت، أقل ما يمكن، وعندها يُتَّخذ الحل أو القرار الإحصائي بناءً على المقدَّر النهائي المُختار للوسيط θ. وبشكل عام: عند اختيار المقدَّر الأمثل، ستكون هناك مجازفة أو مخاطرة، علماً أن كلمة مجازفة قد تأخذ معنى الربح أو الخسارة (الناتج أو الناتجة) عن اختيارنا لـ T كمقدِّر للوسيط θ. ولفهم النتائج السلبيَّة المترتِّبة عند اختيار مقدَّر ما، ندرس ما يسمى:

(7-8) تابع المجازفة:

من الطبيعي أن تكون المجازفة أكبر أو أصغر حسب ابتعاد أو اقتراب المقدَّر T عن القيمة الحقيقيَّة للوسيط θ.

فإذا رمزنا بـ $f(t)$ لتابع كثافة T، (حيث ينظر لـ T كمتغيِّر)، و بـ $r(T,\theta)$ لكمية الغرامة المترتِّبة أو الخسارة الناجمة عن هذا الاختيار، فإنَّ تابع المجازفة الذي سنرمز له بالرمز R يُعرَّف بالشكل:

$$R = E[\ r\ (T,\theta\ )\ ] = \int_{-\infty}^{\infty} r(t,\theta) f(t) dt$$

وبالتالي، فمن المنطقي أن يكون المقدَّر الأفضل هو ذاك المقدَّر الذي لأجله يكون R في نهايته الصغرى، هذا وقد حدَّد الإحصائيون تابع الخسارة $r(T,\theta)$ بالعلاقة التالية: $r(T,\theta) = (T-\theta)^2$ وعندئذٍ يصبح تابع المجازفة بالشكل: $R = \int_{-\infty}^{\infty} (t-\theta)^2 f(t) dt$

مثال (1): أوجد تابع المجازفة الناتج عن اختيار المقدّر $T = \overline{X}$ في المجتمع الإحصائي الموصوف بالمتغيّر $X = N(\mu,1)$. (واضح ما نريده هو تقدير μ)

الحل:

لدينا في هذه الحالة: $r(T,\theta) = r(\overline{X},\mu) = (\overline{X}-\mu)^2$

بالتالي: $R = \int_{-\infty}^{\infty} (\bar{x}-\mu)^2 f(\bar{x}) d\bar{x} = \sigma^2(\overline{X}) = \frac{1}{n}$

حيث n حجم العيّنة العشوائيّة المسحوبة من المجتمع الإحصائي المذكور، ومنه فإنّ تابع المجازفة أقل كلما كبرت n، وبالعكس.

مثال (2):

أوجد تابع المجازفة الناتج عن اختيار المتغيّر الإحصائي $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ كمقدّر إحصائيّ للوسيط σ^2 في المجتمع الإحصائي الموصوف بالمتغيّر $X : N(0,\sigma^2)$

الحل:

واضح أنّ تابع الخسارة من الشكل: $r(T,\theta) = r(S^2,\sigma^2) = (\overline{S}-\sigma^2)^2$، بالتالي تابع المجازفة يكون:

$$
\begin{aligned}
R &= E[r(T,\theta) = E(S^2-\sigma^2)^2 \\
&= E[(S^2 - M + M - \sigma^2)^2] \quad ; \; M = E(S^2) \\
&= E[(S^2-M)^2 + (M-\sigma^2)^2 + 2(S^2-M)(M-\sigma^2)] \\
&= E(S^2-M)^2 + (M-\sigma^2)^2 + 0 \\
&= \sigma^2(S^2) + (M-\sigma^2)^2
\end{aligned}
$$

ولكن:

$$M = E(S^2) = \frac{n-1}{n}\sigma^2$$

$$\sigma^2(S^2) = \frac{2n-1}{n^2}\sigma^4$$

بالتالي: $R = \frac{2n-1}{n^2}\sigma^2$ ، وهو تابع المجازفة المطلوب.

من الواضح أنَّ $R \to 0$ عندما $n \to \infty$ ، في حين تكبر R كلما صَغُرَت n .

(8-8) خواص المقدّرات:

قلنا إنَّ القرار الأمثل هو ذاك القرار الذي لا يوجد أفضل منه في حينه، وقد لا يعجبك!، وبما أنَّ قيمة المقدّر تُحسَب من عيّنة عشوائيّة مأخوذة من المجتمع، فهي بشكل عام لن تتطابق مع القيمة الحقيقيّة للوسيط θ إلا بالصدفة، ذلك لأنَّ المجتمع الإحصائي أصلاً غير متجانس، ورغم ذلك فإنَّ المقدّر يحتفظ بفائدته طالما كانت قيمته الرقميّة قريبة من الوسيط المدروس بدرجة معقولة، هذا وللمفاضلة بين مقدّرين أو أكثر سنتعرف إلى بعض المعايير والخصائص التي تسمح لنا باختيار مقدّر دون آخر، ولذلك يلزم دراسة الخواص التالية:

(8-8-1) خاصة عدم التحيُّز (الإنصاف):

تعريف: نقول عن المقدّر T إنّه مقدّر غير متحيّز للوسيط θ في المجتمع الإحصائي الموصوف بتابع الكثافة $f(x,\theta)$ إذا كان: $E(T) = \theta$.

أما إذا كان: $E(T) = \theta + \delta$ ، فإنّنا نسمي T عندئذٍ بالمتغيّر المتحيّز بالمقدار δ .

مثال (1):

إذا كان: $X : B(n,p)$ ، عندئذٍ النسبة الملاحظة (المشاهدة) للنجاح $\frac{X}{n}$ ، هي مقدّر غير متحيّز، ذلك لأنَّ: $E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}np = p$

مثال (2):

إذا كان: $X : N(\mu,\sigma)$ ، عندئذٍ من أجل المتغيّر الإحصائي $T = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

يكون: $E(\overline{X}) = \mu$ ، وبالتالي فإنّ $\overline{X}$ هو مقدّر غير متحيّز لـ μ.

مثال (3):

إذا كان: $X : N(\mu,\sigma)$ ، فإنّه من أجل المتغيّر الإحصائي $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

يكون: $E(S^2) = \frac{n-1}{n}\sigma^2 = \sigma^2 - \frac{1}{n}\sigma^2 = \sigma^2 - \delta$

أي أنّ المقدّر $T = S^2$ هو مقدّر متحيّز للوسيط σ^2، ومقدار التحيّز هو:

$\delta = -\frac{1}{n}\sigma^2$ في حين نلاحظ أنّ المقدّر : $\hat{S}^2 = \frac{n}{n-1}S^2$، هو مقدّر غير متحيّز لـ σ^2 لأنّ: $E(\hat{S}^2) = \frac{n}{n-1}E(S^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$

(لاحظ أنّنا حصلنا على مقدّر غير متحيّز من مقدّر متحيّز وذلك بعد ضربه بمعامل مناسب ...).

مثال (4):

في المجتمع الطبيعي $X : N(\mu,\sigma^2)$ وجدنا أنّ: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n}X_i$ هو مقدّر غير متحيّز لـ μ، وكذلك كل من المتغيّرات : $X_i \ ; \ i = \overline{1,n}$ هو مقدّر غير متحيّز لـ μ لأنّ: $E(X_i) = \mu \ , \ \forall \ i = \overline{1 \ , n}$. ومن أجل المقدّر: $T = \frac{X_1 + X_2}{2}$ يكون أيضاً: $E(T) = \frac{1}{2}[E(X_1) + E(X_2)] = \mu$ أي أنّ T هو مقدّر غير متحيّز أيضاً.

إذاً كل المقدّرات $\frac{X_1 + X_2}{2}$ ، $\overline{X}$ ، $X_i \ ; \ i = \overline{1 \ , n}$ هي مقدّرات غير متحيّزة للوسيط μ. والسؤال الذي يطرح نفسه هو: أي من هذه المقدّرات نختار؟ ألا توجد صفات أخرى لهذه المقدّرات بحيث نتمكن من مقارنتها وبالتالي انتخاب أفضلها ؟ والجواب بالإيجاب كما سنجد ...

نتيجة: إذا كان T مقدّراً غير متحيّز للوسيط θ فإنّ:

$$R = E[\ r\ (T,\theta\ )\ ] = E(T-\theta)^2$$
$$= E(T - E(T))^2 = \sigma^2(T)$$

أي: المقدّر غير المتحيّز هو المقدّر الذي يكون لأجله تابع المجازفة مساوياً لتباين ذات المقدّر.

(8-8-2) الفعاليّة (الكفاءة):

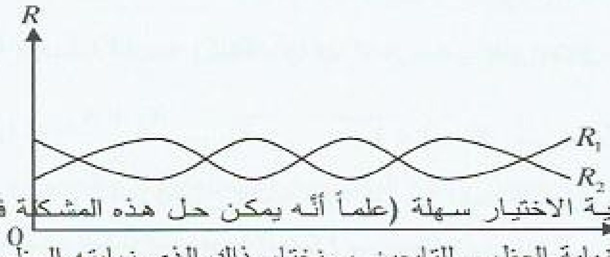
ينتج مما سبق أنّه إذا وجد مقدّران T_1 و T_2 لوسيط ما θ، فإنّه يتوجب علينا اختيار ذاك المقدّر الذي تباينه أصغر ما يمكن (حيث تابع الخسارة أقل).

تعريف:

نقول عن المقدّر T إنّه مقدّر فعّال للوسيط θ إذا كان: $\sigma^2(T) \le \sigma^2(T_i)\ ,\ \forall\ i$

ويُعدُّ هذا المعيار مقياساً لتفضيل المقدّرات المختلفة أو الممكنة لوسيط θ، وعندها نقول إنَّ T أكفأ المقدّرات غير المتحيّزة، وقد تكون المقارنة بين مقدّرين غير متحيّزين لوسيط θ، سهلة، إذا كان لهذين المقدّرين التوزيع نفسه. فمثلاً في الشكل التالي:

حيث يتضح من الشكل أن: $\sigma^2(T_2) < \sigma^2(T_1)$ ، لذلك نختار المقدّر T_2 وفيما عدا ذلك، أي إذا لم يكن لهما نفس التوزيع، كأن يتقاطع التابعان كما في الشكل:



فعندئذٍ لن تكون عملية الاختيار سهلة (علماً أنّه يمكن حل هذه المشكلة في حالات خاصة، حيث نوجد النهاية العظمى للتابعين ، ونختار ذاك الذي نهايته العظمى أصغر) وفي هذه الحالة نوجد قيمة النسبة: $\dfrac{1/\sigma^2(T_1)}{1/\sigma^2(T_2)} = \dfrac{\sigma^2(T_2)}{\sigma^2(T_1)}$ والتي تسمى بالإنتاجيّة النسبيّة للمقدّر T_1 بالنسبة للمقدّر T_2، أو فعالية المقدّر T_1 بالنسبة للمقدّر T_2. فإذا كانت هذه الإنتاجية أكبر من الواحد، عندئذٍ $\sigma^2(T_1) < \sigma^2(T_2)$ وعندها نختار T_1 رغم

أنَّ كليهما مقدِّر غير متحيِّز. والسؤال المطروح الآن: ألا يوجد مقدِّر ثالث مثل T_3 بحيث يكون $\sigma^2(T_3) < \sigma^2(T_1)$؟ وعندها سنختار T_3. ولنفرض أنه وجد مثل هذا المقدِّر يبقى السؤال قائماً بالنسبة لمقدِّرات أخرى، وبالتالي ما الكلمة الفصل في هذه المسألة؟ بمعنى آخر ألا يوجد حد أصغري للتباينات في صف المقدّرات غير المتحيِّزة للوسيط θ؟ بالواقع الجواب بالإيجاب ، ويكمن في النظريَّة المهمة التالية:

(3-8-8) نظريَّة كرامر . راو (أفضل المقدّرات):

إذا كان: $T(X_1, X_2, ..., X_n)$ مقدّراً غير متحيِّزاً للوسيط θ، وكان $f(x,\theta)$ تابع كثافة المجتمع الإحصائي المدروس فإنَّ:

$$\sigma^2(T) \geq \frac{1}{n\ E\{\ [\frac{\partial}{\partial\theta}\log f(x,\theta)]^2\ \}}$$

حيث نسمي الكميَّة: $\frac{1}{n\ E\{[\frac{\partial}{\partial\theta}\log f(x,\theta)]^2\}}$ حد كرامر المتعلق بالوسيط θ.

البرهان:

بما أنَّ T مقدّر غير متحيِّز لـ θ فإنَّ: $E(T) = \theta$

بالتالي إذا كان تابع الكثافة المشترك لمتغيِّرات العيِّنة $(X_1, X_2, ..., X_n)$ هو: $L = \prod_{i=1}^{n} f(x_i,\theta)$ فإنّه سيحقّق كأي تابع كثافة العلاقة التالية:

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} L\ dx_1 \ldots dx_n = 1$$

وباشتقاق الطرفين بالنسبة لـ θ يكون:

$$(1) \quad : \quad \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{\partial L}{\partial\theta}\ dx_1 \ldots dx_n = 0$$

بما أنَّ: $\frac{\partial}{\partial\theta}\log f(x,\theta) = \frac{\frac{\partial}{\partial\theta}f(x,\theta)}{f(x,\theta)}$ أي : $\frac{\partial}{\partial\theta}\log L = \frac{1}{L}\frac{\partial L}{\partial\theta}$

ومنه: $$\frac{\partial L}{\partial \theta} = L\,\frac{\partial \log L}{\partial \theta} = L \sum_{i=1}^{n} \frac{\partial \log f(x_i,\theta)}{\partial \theta}$$

وبالتعويض في (1) يكون:

$$(2) \quad : \quad \int_{-\infty}^{\infty}\!\!\dots\!\int_{-\infty}^{\infty} L\ Z\ dx_1 \dots dx_n = 0 \equiv E(Z)$$

حيث فرضنا أنّ: $$Z = \sum_{i=1}^{n} \frac{\partial \log f(x_i,\theta)}{\partial \theta}$$

وبما أنّ T مقدّر غير متحيّز لـ θ يكون: $\int_{-\infty}^{\infty}\!\dots\!\int_{-\infty}^{\infty} T.L\ \ dx_1 \dots dx_n = \theta$

وباشتقاق الطرفين بالنسبة لـ θ :

$$(3) \quad : \quad \int_{-\infty}^{\infty}\!\!\dots\!\int_{-\infty}^{\infty} T\,\frac{\partial L}{\partial \theta}\ dx_1 \dots dx_n = \int_{-\infty}^{\infty}\!\!\dots\!\int_{-\infty}^{\infty} T\ L\ Z\ \ dx_1 \dots dx_n = 1 \equiv E(T.Z)$$

هذا من ناحية، ومن ناحية أخرى نعلم أنّ معامل الارتباط بين T و Z يمكن أن يُكتَب بالشكل:

$$\rho_{TZ} = \frac{E(T.Z) - E(T)\,.\,E(Z)}{\sigma^2{}_T\,.\,\sigma^2{}_Z}$$

وبالاستفادة من (2) و (3) يكون: $$\rho_{TZ} = \frac{1}{\sigma^2{}_T\,.\,\sigma^2{}_Z}$$

وحسب خواص معامل الارتباط لدينا: $$\rho_{TZ} \le 1 \Rightarrow \frac{1}{\sigma^2{}_T\,.\,\sigma^2{}_Z} \le 1$$

الأمر الذي يعني أنّ:

$$(4) \quad : \quad \sigma^2{}_T \ge \frac{1}{\sigma^2{}_Z}$$

لنحسب الآن $\sigma^2{}_Z$: (مع ملاحظة أنّ الحدود في العلاقة : $Z = \sum_{i=1}^{n} \frac{\partial \log f(x_i,\theta)}{\partial \theta}$ مستقلة لأنّها ترتبط بالمتغيّرات المستقلة X_i):

لدينا:

$$\sigma^2_Z = E(Z^2) - \{E(Z)\}^2 = E(Z^2) \quad ; \; E(Z) = 0$$

$$= E\{ [\sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta}]^2 \} = n . E\{ [\frac{\partial \log f(x_i, \theta)}{\partial \theta}]^2 \}$$

وبالتبديل في (4) يكون:

$$\sigma^2(T) \geq \frac{1}{n \, . \, E\{ \, [\, \frac{\partial \log f(x_i, \theta)}{\partial \theta} \,]^2 \,\}}$$

وبما أنَّ توزيع $X_i \; ; \; i = \overline{1 \, , \, n}$ يماثل توزيع X فإنَّ:

$$\sigma^2(T) \geq \frac{1}{n \, . \, E\{ \, [\frac{\partial}{\partial \theta} \log f(x, \theta)]^2 \,\}}$$

وهو المطلوب.

(تنطبق هذه النظريَّة على المتغيِّرات المنفصلة أيضاً).

نتيجــة: ينتج من النظريَّـة السـابقة أنَّـه إذا كـان T مقدِّراً غير متحيِّزاً للوسيط θ بحيث تباينه يساوي حد كرامر أي: $\sigma^2(T) = \frac{1}{n \, . \, E\{ \, [\frac{\partial}{\partial \theta} \log f(x, \theta)]^2 \,\}}$

عندئذٍ لن يوجد أي مقدِّر آخر غير متحيِّز لـ θ يكون تباينه أصغر من تباين T، بالتالي (وفي هذه الحالة) يكون المقدِّر T هو أفضل المقدِّرات غير المتحيِّزة للوسيط θ، وبمعنى آخر: المقدِّر الذي يصل إلى حد كرامر هو أفضل المقدِّرات، ويقال عنه في هذه الحالة أنَّـه مقدِّر كفء *(Efficient)*. وإذا كـان T مقدِّراً كفئاً، فإنَّ: $\frac{1}{\sigma^2(T)}$ تمثل كميَّة المعلومات التي يحويها T من المعلومات التي تقدِّمها العيِّنة.

مثال (1):

بفرض المجتمع الإحصائي $X : N(\mu, \sigma^2)$. ولنقدِّر σ^2 على اعتبار أنَّ μ معلومة (أي لنوجد حد كرامر المتعلق بالوسيط σ^2).

الحل:

لدينا: $f(x,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

$$\log f(x,\sigma^2)=\log\frac{1}{\sigma\sqrt{2\pi}}-\frac{1}{2}(\frac{x-\mu}{\sigma})^2$$

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}\log f(x,\sigma^2)&=\frac{\partial}{\partial\sigma^2}[\log\frac{1}{\sigma\sqrt{2\pi}}-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]\\&=\frac{\partial}{\partial\sigma^2}[-\log\sqrt{2\pi}-\frac{1}{2}\log\sigma^2-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]\\&=-\frac{1}{2\sigma^2}+\frac{(x-\mu)^2}{2\sigma^4}\end{aligned}$$

ومنه:

$$\begin{aligned}[\frac{\partial}{\partial\sigma^2}\log f(x,\sigma^2)]^2&=[-\frac{1}{2\sigma^2}+\frac{(x-\mu)^2}{2\sigma^4}]^2\\&=\frac{1}{4\sigma^4}-\frac{(x-\mu)^2}{2\sigma^6}+\frac{(x-\mu)^4}{4\sigma^8}\end{aligned}$$

$$E\{[\frac{\partial}{\partial\sigma^2}\log f(x,\sigma^2)]^2\}^2=\frac{1}{4\sigma^4}-\frac{1}{2\sigma^6}\sigma^2+\frac{1}{4\sigma^8}\mu_4$$

$$;\ E(x-\mu)^2=\sigma^2\ ,\ (E(x-\mu)^4=\mu_4$$

$$=-\frac{1}{4\sigma^4}+\frac{\mu_4}{4\sigma^8}=\frac{1}{2\sigma^4}\quad;\mu_4=3\sigma^4$$

وحد كرامر المتعلق بالوسيط σ^2 يكون:

$$\frac{1}{n.E\{[\frac{\partial}{\partial\sigma^2}\log f(x,\sigma^2)]^2\}}=\frac{2\,\sigma^4}{n}$$

بالتالي أفضل المقدّرات لـ T هو ذاك المقدّر الذي لأجله يكون: $\sigma^2(T)=\frac{2\sigma^4}{n}$

وإذا اعتبرنا T هو المتغيّر الإحصائي: $T=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$

يكون: $E(T)=\frac{1}{n}\ n\sigma^2=\sigma^2$

لنحسب الآن $\sigma^2(T)$:

لـدينا: $T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \Rightarrow \frac{nT}{\sigma^2} = \sum_{i=1}^{n}(\frac{X_i - \mu}{\sigma})^2$ وبالتـالي فـإنّ المقـدّر T هـو مقدّر غير متحيّز للوسيط σ^2.

ولكن: $X_i : N(\mu,\sigma) \Rightarrow \frac{X_i - \mu}{\sigma} : N(0,1)$

وبالتالي يكون: $\sum_{i=1}^{n}(\frac{X_i - \mu}{\sigma}) : \chi_n^{\ 2}$ أي:

$$\frac{nT}{\sigma^2} : \chi_n^{\ 2} \Rightarrow \begin{cases} E(\frac{nT}{\sigma^2}) = \quad n \Rightarrow E(T) = \sigma^2 \\ \sigma^2(\frac{nT}{\sigma^2}) = 2n \Rightarrow \sigma^2(T) = \frac{2\sigma^4}{n} \end{cases}$$

واضـح ممـا سبق أنّ: T مقدّر غير متحيّز لـ σ^2، وتباينـه يصـل إلـى حد كرامر، وتبعاً لهذين الشرطين فإنّ T هو أفضل مقدّر لـ σ^2.

مثال (2):

بفرض A مجتمـع إحصـائي موصـوف بـالمتغيّر $X : P(\lambda)$ (أي يخضـع لتوزيـع بواسـون) ولنبيّن أنّ المقدّر الإحصائي $T = \frac{1}{n}\sum_{i=1}^{n} X_i$ هو أفضل مقدّر للوسيط λ.

الحل:

$T = \overline{X} \Rightarrow E(T) = E(\overline{X}) = \lambda$. وهو مقدّر غير متحيّز.

لنوجد الآن حد كرامر المتعلق بالوسيط λ:

لدينا:

$$f(x,\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \quad ; \lambda = 0,1,2,...$$

$$\log f(x,\lambda) = -\lambda + x\log\lambda - \log x!$$

$$\frac{\partial}{\partial\lambda}\log f(x,\lambda) = -1 + \frac{x}{\lambda}$$

بالتالي:

$$E\{[\frac{\partial}{\partial\lambda}\log f(x,\lambda)]^2\} = E\,(1+\frac{x^2}{\lambda^2}-\frac{2x}{\lambda})$$
$$=1+\frac{1}{\lambda^2}E(x^2)-\frac{2}{\lambda}E(x)$$
$$=\frac{1}{\lambda}$$

ومن توزيع بواسون نعلم أنَّ: $E(x)=\lambda$ ، $E(x^2)=\lambda^2+\lambda$

وبالتالي حد كرامر المتعلق بالوسيط λ يكون:

$$\frac{1}{n.E\{[\frac{\partial}{\partial\lambda}\ \log f(x,\lambda)]^2\}}=\frac{\lambda}{n}$$

ومن جهة أخرى، لدينا: $\sigma^2(T)=\sigma^2(\overline{X})=\frac{\sigma^2}{n}=\frac{\lambda}{n}$

مما تقدَّم نلاحظ: T مقدَّر غير متحيِّز للوسيط λ، وتباينه يصل إلى حد كرامر، وبالتالي يعتبر T أفضل مقدَّر للوسيط λ.

مثال (3):

أوجد حد كرامر المتعلق بالوسيط μ في المجتمع الإحصائي الموصوف بالمتغيِّر $X : N(\mu,\sigma^2)$

الحل:

لدينا:

$$f(x,\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$\log f(x,\mu)=\log\frac{1}{\sigma\sqrt{2\pi}}-\frac{1}{2}(\frac{x-\mu}{\sigma})^2$$

$$\frac{\partial}{\partial \mu} \log f(x,\mu)=\frac{x-\mu}{\sigma^2}$$

$$E\,[\frac{\partial}{\partial \mu} \log f(x,\mu)]^2 = E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^4} E\,(x-\mu)^2$$

$$=\frac{\sigma^2}{\sigma^4}=\frac{1}{\sigma^2}$$

ومنه يكون حد كرامر هو : $\frac{\sigma^2}{n}$

وبالتالي فإنّ أي مقدّر غير متحيّز مثل U سوف يحقّق : $\sigma^2(U) \geq \frac{\sigma^2}{n}$. وبما أنّ: $E(\overline{X})=\mu$ و $\sigma^2(\overline{X})=\frac{\sigma^2}{n}$، فإنّ أفضل مقدّر للوسيط μ في حالة التوزيع الطبيعي هو $\overline{X}$.

(4-8-8) الاتساق *(Cosistency)* خاصية الإمكانيّة:

بفرض A مجتمع إحصائي موصوف بالمتغيّر X الذي تابع كثافته $f(x,\theta)$ ، وليكن T مقدّراً إحصائياً للوسيط θ .

لنسحب من هذا المجتمع العيّنة العشوائيّة $(X_1, X_2,...., X_n)$ ذات الحجم n .

من الواضح أنّ المقدّر T لوسيط θ يكون أفضل كلما كان حجم العيّنة أكبر ، وبالتالي إذا اعتبرنا أنّ: $T = T_n$ عندما يكون T معيّناً من خلال عيّنة حجمها n ، فإنّه من الطبيعي في هذه الحالة أن تكون المعلومات التي يقدّمها T_2 عن θ أفضل من تلك التي يقدّمها T_1 عن θ، والمعلومات التي يقدّمها T_3عن θ ستكون أفضل من تلك التي يقدمها T_2 ... وهكذا تكون المعلومات التي يقدّمها T_n عن θ أفضل من تلك التي يقدّمها T_{n-1} . أي أنّ ما نسميه الاتساق أو خاصة الإمكانيّة ترتبط بالسلوك النهائي للمقدّر عندما $n \to \infty$.

بمعنى آخر : إنّ متتالية المقدّرات (T_n) تنتهي احتمالياً لـ θ عندما $n \to \infty$ أو بتعبير أدق: نقول عن المقدّر T إنّه مقدّر متّسق (أي يحقّق خاصية الإمكانيّة) للوسيط θ عندما يتناقص احتمال اختلافه عن القيمة الحقيقيّة للوسيط θ مع زيادة حجم العيّنة n، ونعبّر عن ذلك رياضياً بالتعريف التالي:

تعريف:

نقول عن المقدّر T إنّه مقدّر متّسق (معقول) للوسيط θ إذا تحقّق (لكل ثابت موجب δ) الشرط التالي: $\lim_{n\to\infty} P[|T_n-\theta|\geq\delta]=0$

أو بصورة أخرى: $\lim_{n\to\infty} P[|T_n-\theta|<\delta]=1$

وإذا جعلنا $\delta\to 0$ فهذا يعني أنّ انطباق T_n على θ هو حدث مؤكّد عندما $n\to\infty$.
ومن المفيد ذكره الآن أنّه يمكن الاستفادة من متراجحة تشيبتشيف لإثبات ما إذا كان مقدّراً ما متّسقاً أم لا ؟ كما في الأمثلة التالية:

مثال (1):

بيّن أنّ المتغيّر الاحصائي $\overline{X}$ هو مقدّر متّسق للوسيط μ :

نُذكّر بمتراجحة تشيبتشيف التي تنص على أنّه إذا كان X متغيّراً عشوائيّاً توقعه الرياضي μ وتباينه σ^2، فإنّه لكل ثابت موجب c يكون:

$$P[|X-\mu|\geq c]\leq\frac{\sigma^2(X)}{c^2}$$

فإذا وضعنا $X=\overline{X}$ يكون: $P[|\overline{X}-\mu|\geq c]\leq\frac{\sigma^2(\overline{X})}{c^2}$

ولكن: $\sigma^2(\overline{X})=\frac{\sigma^2}{n}$ بالتالي: $P[|\overline{X}-\mu|\geq c]\leq\frac{\sigma^2}{nc^2}$

وبأخذ نهاية الطرفين عندما تؤول n إلى ما لانهاية نحصل على:

$$\lim_{n\to\infty} P[|\overline{X}-\mu|\geq c]=0$$

وهذا يثبت (حسب تعريف الاتساق) أنّ $\overline{X}$ مقدّر متسق للوسيط μ .

مثال (2):

في توزيع ثنائي الحد $X:B(n,p)$. اثبت أنّ النسبة المشاهدة للنجاح $\frac{X}{n}$ هي مقدّر متّسق للوسيط p .

الحل:

نستخدم متراجحة تشيبتشيف: $P[|X-\mu| \geq c] \leq \frac{\sigma^2}{c^2}$

ومن توزيع ثنائي الحد نعلم أنّ: $E(X) = n\, p \equiv \mu$ و $\sigma^2(X) = n\,p\,q$

وإذا بدّلنا X بـ $\frac{X}{n}$ يكون: $E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}\, n\, p = p$

$$\sigma^2(\frac{X}{n}) = \frac{1}{n^2}\sigma^2(X) = \frac{1}{n^2} n\,p\,q = \frac{p\,q}{n}$$

وبالتعويض في متراجحة تشيبتشيف يكون: $P\{\ \left|\frac{X}{n} - p\right| \geq c\ \} \leq \frac{p^2 q^2}{n^2 c^2}$

ومنه: $\lim_{n\to\infty} P\{\ \left|\frac{X}{n} - p\right| \geq c\} = 0$ أي : $\frac{X}{n}$هو مقدّر متّسق للوسيط p .

نتيجة:

يكون المقدّر T تقديراً متّسقاً للوسيط θ إذا تحقّق الشرطان:

1- T غير متحيّز لـ θ

2- $\sigma^2(T) \to 0$ عندما $n \to \infty$.

مثال (3):

من أجل: $\hat{S}^2 = \frac{n}{n-1} S^2$ وجدنا أنّ:

$$E(\hat{S}^2) = \frac{n}{n-1} E(S^2) = \sigma^2 \quad ، \quad \sigma^2(\hat{S}^2) = \frac{n}{n-1}\sigma^2(S^2) = \frac{2\sigma^4}{n-1}$$

كما نلاحظ أنّ: $\lim_{n\to\infty} \sigma^2(\hat{S}^2) = \lim_{n\to\infty} \frac{2\sigma^4}{n-1} = 0$

بالتالي: $\hat{S}^2$ هو مقدّر غير متحيّز لـ σ^2 وتباينه يؤول إلى الصفر عندما $n \to \infty$، أي حقق الشرطين السابقين ، وبالتالي فهو مقدّر متّسق للوسيط σ^2.

نظريّة:

الشرط الكافي كي يكون المقدّر T متّسقاً للوسيط θ هو :

$\lim_{n\to\infty} E[(T_n - \theta)^2] = 0$ حيث $T_n = T$ عندما يعين T من خلال عيّنة حجمها n.

البرهان:

بفرض δ أي عدد موجب، بالتالي عندما يأخذ المقدّر T القيمة $t \notin [\theta-\delta\ ,\ \theta+\delta]$

فإنّ: $|t-\theta| > \delta$ ، ومنه يكون: $(t-\theta)^2 > \delta^2$

وبناءً عليه يمكن أن نكتب:

$$:(*)\quad \begin{cases} \int_{-\infty}^{\theta-\delta}(t-\theta) f(t)dt \ > \delta^2 \int_{-\infty}^{\theta-\delta} f(t)dt \\ \int_{\theta-\delta}^{\theta+\delta}(t-\theta)^2 f(t)dt \geq 0 \\ \int_{\theta+\delta}^{\infty} (t-\theta)^2 f(t)dt > \delta^2 \int_{\theta+\delta}^{\infty} f(t)dt \end{cases}$$

هذا من ناحية، ومن ناحية أخرى لدينا:

$$E[(T_n-\theta)^2] = E[(T-\theta)^2] = \int_{-\infty}^{\infty}(t-\theta)^2 f(t)dt$$

$$= \int_{-\infty}^{\theta-\delta}(t-\theta)^2 f(t)dt + \int_{\theta-\delta}^{\theta+\delta}(t-\theta)^2 f(t)dt + \int_{\theta+\delta}^{\infty}(t-\theta)^2 f(t)dt +$$

و حسب $(*)$:

$$\geq \delta^2 [\int_{-\infty}^{\theta-\delta} f(t)dt + \int_{\theta+\delta}^{\infty} f(t)dt] = \delta^2[1-\int_{\theta-\delta}^{\theta+\delta} f(t)dt]$$

$$= \delta^2[1-P(\theta-\delta < T_n < \theta+\delta)] = \delta^2[1-P(-\delta < T-\theta_n < \delta)$$

$$= \delta^2[1-P(|T_n-\theta|) < \delta)] \qquad = \delta^2[P(|T_n-\theta|) \geq \delta)]$$

وبالتالي: $$\frac{E[(T_n-\theta)^2]}{\delta^2} \geq [P(|T_n-\theta|) \geq \delta] \geq 0$$

وبملاحظة أنّ الطرف الأيسر يؤول إلى الصفر عندما $n \to \infty$

يكون: $\lim_{n\to\infty} P[|T_n - \theta| \geq \delta] = 0$ وهو المطلوب .

نتائج:

1 . نعلم أنّه إذا كان T مقدّراً غير متحيّزاً لـ θ فإنّ: $E(T) = \theta$

وبالتالي: $$E[(T-\theta)^2] = E[(T-E(T))^2] = \sigma^2(T)$$

وحسب النظريّة السابقة:

$$\lim_{n\to\infty} E[(T-\theta)^2] = 0 \Rightarrow \lim_{n\to\infty} \sigma^2(T) = 0 \quad ; \ (T = T_n)$$

أي: كل مقدّر ينتهي تباينه إلى الصفر عندما $n \to \infty$ هو مقدّر متّسق أيضاً.

2 . إذا كانت $a_1 , a_2 ,..., a_n$ مجموعة من الأعداد الثابتة بحيث: $\sum_{i=1}^{n} a_i = 1$

وكان: $T = \sum_{i=1}^{n} a_i \ T_i$ (حيث $T_i \ ; \ i = \overline{1,n}$ مقدّرات غير متحيّزة لـ θ) فإنّ T يكون مقدّراً غير متحيّزاً لـ θ ، حيث واضح أنّ:

$$E(T) = E[\sum_{i=1}^{n} a_i T_i] = \sum_{i=1}^{n} a_i \ E(T_i) = 1 \ . \ E(T_i) = 1 \ . \ \theta = \theta$$

(8-8-5) الكفاية:

تعريف:

نقول عن المقدّر T إنّه مقدّر كافٍ للوسيط θ إذا لم يستطع أي مقدّر آخر مأخوذ من نفس العيّنة إعطاء معلومات إضافيّة عن T.

وبصورة أخرى: يكون $T = T(X_1, X_2,...., X_n)$ مقدّراً كافياً للوسيط θ إذا تحقّق الشرط التالي: من أجل أي مقدّر $U = U(X_1, X_2,...., X_n)$ لـ θ يكون تابع الكثافة الشّرطي $f(u/t)$ مستقلاً عن θ. والمقدّر الذي يتّصف بخاصية الكفاية هذه يعني أنّه لخّص واحتوى كل ما تحويه العيّنة من معلومات عن الوسيط المعطى.

(8-9) نظريّة التحليل المعاملي:

يكون المقدّر $T = T(X_1, X_2,...., X_n)$ مقدّراً كافياً للوسيط θ إذا وفقط إذا أمكن تحليل

تابع الكثافة المشترك للعيّنة إلى حاصل جداء تابعين موجبين g و h بالشكل:

$$f(x_1, x_2, \dots, x_n, \theta) = g(T, \theta) \,.\, h(x_1, x_2, \dots, x_n)$$

بحيث تعتمد g على T و θ ، في حين لا تعتمد h على θ مهما كانت قيمة المقدّر T.

مثال (1):

برهن أنّ $T = \overline{X}$ مقدّر كافٍ للوسيط μ في المجتمع الطبيعي $X : N(\mu, 1)$

لدينا تابع الكثافة: $$f(x, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}$$

وتابع الكثافة المشترك لمتغيرات العيّنة:

$$f(x_i, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} \quad ; \quad f(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

حيث يمكن تحليل المقدار $\sum_{i=1}^{n}(x_i - \mu)^2$ كما يلي:

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}[(x_i - \overline{x}) + (\overline{x} - \mu)]^2$$

$$= \sum_{i=1}^{n}(x_i - \overline{x})^2 + 2(\overline{x} - \mu)\sum_{i=1}^{n}(x_i - \overline{x}) + n(\overline{x} - \mu)^2$$

ولكن: $$\sum_{i=1}^{n}(x_i - \overline{x}) = n\overline{x} - n\overline{x} = 0$$

بالتالي: $$f(x_1, x_2, \dots, x_n, \theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2 + \frac{n}{2}(\overline{x}-\mu)^2\right]}$$

والتي يمكن كتابتها كحاصل جداء دالّتين: $f_1(\overline{x}, \mu) \,.\, f_2(x_1, x_2, \dots, x_n)$

حيث:

$$f_1(\bar{x},\mu) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{n}{2}(\bar{x}-\mu)^2} \equiv g(T,\theta)$$

$$f_2(x_1, x_2, ..., x_n) = e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

واضح أنَّ f_1 تعتمد على T و θ ، في حين f_2 خالية من μ ولا تعتمد عليها.

مثال (2):

بيّن أنَّ المتغيّر الإحصائي $\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ هو مقدّر غير كافٍ للوسيط σ^2 في المجتمع الطبيعي $X : N(\mu,\sigma)$ حيث μ مجهولة.

الحل:

تابع الكثافة المشترك:

$$f(x_1, x_2, ..., x_n, \sigma^2) = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

ولا يمكن تحليل هذه الدالّة إلى دالّتين، إحداهما في المقدّر والوسيط ، والثانية خالية من الوسيط ، وبالتالي المتغيّر $\hat{S}^2$ هو مقدّر غير كافٍ لـ σ^2 .

مثال (3):

في مجتمع طبيعي إذا فرضنا أنَّ μ معلومة. اثبت أنَّ المقدّر $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$ هو مقدّر كافٍ للوسيط σ^2.

الحل:

تابع الكثافة المشترك:

$$f(x_1, x_2, ..., x_n, \sigma^2) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2} = (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2}$$

وبما أنّ: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \Rightarrow \sum_{i=1}^{n}(x_i-\mu)^2 = nS^2$

وبالتبديل أعلاه يكون:

$$= (\frac{1}{\sigma})^n \; e^{-\frac{nS^2}{2\sigma^2}} \, . \, (\frac{1}{\sqrt{2\pi}})^n \; f(x_1, x_2, ..., x_n, \sigma^2)$$

$$= f_1(s^2, \sigma^2) \, . \, f_2$$

حيث: $f_1(s^2, \sigma^2) = (\frac{1}{\sigma})^n \; e^{-\frac{nS^2}{2\sigma^2}}$ ، $f_2 = (\frac{1}{\sqrt{2\pi}})^n$

مثال (4):

بيّن أنّ المقدّر $T = \overline{X}$ هو مقدّر كافٍ للوسيط λ في توزيع بواسون.

الحل:

تابع الكثافة المشترك للعيّنة:

$$f(x_1, x_2, ..., x_n, \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}}{x!} \lambda^x \quad ; x = 1, 2, ...$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda}}{x!} \lambda^{x_1} \, . \, \frac{e^{-\lambda}}{x!} \lambda^{x_2} \; ... \; \frac{e^{-\lambda}}{x!} \lambda^{x_i}$$

$$= \frac{e^{-n\lambda} \; \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x!}$$

وبما أنّ: $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \Rightarrow \sum_{i=1}^{n} X_i = n \overline{X}$

بالتالي: $f(x_1, x_2, ..., x_n, \lambda) = e^{-n\lambda} \; \lambda^{n\bar{x}} \; \frac{1}{\prod_{i=1}^{n} x!} \equiv f_1(\bar{x}, \lambda) \, . \, f_2(x_1, x_2, ..., x_n)$

وهذا يعني أنّ $\overline{X}$ مقدّر كافٍ للوسيط λ في التوزيع المذكور.

(8–10) دالّة المعلومات: *The Information Function* :

(8–10–1) دالّة المعلومات لمشاهدة (لملاحظة) واحدة:

وهي تقيس المعلومات المقدّمة عن الوسيط المجهول، وتعرّف رياضياً بالشكل:

$I(\theta)=\sigma^2(\frac{\partial}{\partial\theta}\log f(x,\theta))$ ، (من الواضح أنَّ $I(\theta)\geq 0$ لأنّها تمثّل تباين)

ملاحظة:

- يمكن إثبات أنّ: $I(\theta)=E(-\frac{\partial^2}{\partial\theta^2}\log f(x,\theta))$ حيث الصيغة الأخيرة قد تكون أبسط في الحسابات.

. كما يمكن إثبات أنّ: $E\{\frac{\partial \log f(x,\theta)}{\partial\theta}\}^2=E\{-\frac{\partial^2\log f(x,\theta)}{\partial\theta^2}\}$

(2-10-8) دالّة المعلومات للعيّنة:

وهي تمثّل كمية المعلومات المقدّمة من عيّنة (مجموعة من المشاهدات) حول الوسيط، وتعرّف رياضيّاً بالشكل:

$$I_{SA}(\theta)=\sigma^2[\frac{\partial}{\partial\theta}\log f(x_1,x_2,\dots,x_n,\theta)]$$
$$\equiv E[-\frac{\partial^2}{\partial\theta^2}\log f(x_1,x_2,\dots,x_n,\theta)]$$

حيث الصيغة الأخيرة أبسط في الحسابات.

ملاحظة:

يمكن إثبات أنّ: $I_{SA}(\theta)=n.I(\theta)$ كما في المثال التالي:

مثال (1):

a) أوجد دالّة المعلومات $I(\mu)$ في المجتمع الطبيعي $X:N(\mu,\sigma^2)$.

b) إذا كانت: $X_1,X_2,\dots,X_n$ عيّنة عشوائيّة من الملاحظات المسحوبة من المجتمع الطبيعي $X:N(\mu,\sigma^2)$ أوجد: $I_{SA}(\mu)$ ، ثم تحقّق أنّ: $I_{SA}(\mu)=n.I(\mu)$.

الحل:

لدينا: $I(\mu)=E(-\frac{\partial^2}{\partial\mu^2}\log f(x,\mu))$

لحساب $I(\mu)$ نتبع الخطوات التالية:

1 . نكتب دالّة الكثافة: $f(x,\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

2 . نأخذ لوغاريتم الدالّة: $\log f(x,\mu)=\log(\frac{1}{\sigma\sqrt{2\pi}})-\frac{1}{2}(\frac{x-\mu}{\sigma})^2$

3 . نفاضل لوغاريتم الدالّة بالنسبة لـ μ:

$$\frac{\partial}{\partial\mu}\log f(x,\mu)=\frac{1}{\sigma^2}(x-\mu)=\frac{x}{\sigma^2}-\frac{\mu}{\sigma^2}$$

4 . نحسب المشتق الثاني: $\frac{\partial^2}{\partial\mu^2}\log f(x,\mu)=-\frac{1}{\sigma^2}$

5 . نبدّل: $I(\mu)=E(-\frac{\partial^2}{\partial\mu^2}\log f(x,\mu))=E(-\frac{-1}{\sigma^2})=E(\frac{1}{\sigma^2})=\frac{1}{\sigma^2}$

وهي دالّة المعلومات للمشاهدة الواحدة.

لحساب $I_{SA}(\mu)$ نتبع الخطوات التالية:

1- نوجد الدالّة المشتركة:

2- $$f(x_1,x_2,\dots,x_n,\mu)=(\frac{1}{\sigma\sqrt{2\pi}})^n\ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

3- $$\log f(x_1,x_2,\dots,x_n,\mu)=\log(\frac{1}{\sigma\sqrt{2\pi}})^n-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

4- $$\frac{\partial}{\partial\mu}\log f(x_1,x_2,\dots,x_n,\mu)=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=\frac{\sum_{i=1}^{n}(x_i)}{\sigma^2}-\frac{n\mu}{\sigma^2}$$

$$\frac{\partial^2}{\partial\mu^2}\log f(x_1,x_2,\dots,x_n,\mu)=-\frac{n}{\sigma^2}$$

5- $$I_{SA}(\mu)=E[-\frac{\partial^2}{\partial\mu^2}\log f(x_1,x_2,\dots,x_n,\mu)]=E(\frac{n}{\sigma^2})=\frac{n}{\sigma^2}$$

أي أنّ: $I_{SA}(\mu)=\frac{n}{\sigma^2}=n.I(\mu)$

(3-10-8) دالّة معلومات المقدّر:

وهي تقيس المعلومات التي يقدمها المقدّر T عن الوسيط θ، وتعرّف رياضياً بالشكل:

$$I_T(\theta) = \sigma^2 (\frac{\partial}{\partial\theta}\log\ g(t,\theta))$$
$$\equiv E[-\frac{\partial^2}{\partial\theta^2}\log\ g(t,\theta)]$$

حيث: $g(t,\theta)$ دالّة كثافة المقدّر T، (الصيغة الأخيرة أسهل في الحسابات)

مثال (2):

أوجد دالّة المعلومات التي يقدّمها المقدّر $\overline{X}$ عن الوسيط μ في المجتمع الطبيعي $X : N(\mu,\sigma^2)$.

الحل:

1- دالّة الكثافة للمقدّر: $g(\overline{x},\mu) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}e^{-\frac{n}{2\sigma^2}(\overline{x}-\mu)^2}$

2- $\log\ g(\overline{x},\mu) = \log(\frac{\sqrt{n}}{\sigma\sqrt{2\pi}}) - \frac{n}{2\sigma^2}(\overline{x}-\mu)^2$

3- $\frac{\partial}{\partial\mu}\log\ g(\overline{x},\mu) = \frac{n}{\sigma^2}(\overline{x}-\mu) = \frac{n\overline{x}}{\sigma^2} - \frac{n\mu}{\sigma^2}$

4- $\frac{\partial^2}{\partial\mu^2}\log\ g(\overline{x},\mu) = -\frac{n}{\sigma^2}$

وبالتالي دالّة المعلومات التي يقدّمها المقدّر $\overline{X}$ عن الوسيط μ تكون:

$$I_{\overline{X}}(\mu) = E[-\frac{\partial^2}{\partial\mu^2}\log\ g(\overline{x},\mu)] = E(\frac{n}{\sigma^2}) = \frac{n}{\sigma^2}$$

(8-11) العلاقة بين دالّة معلومات المقدّر ودالّة معلومات العيّنة:

من الواضح أنّ معلومات المقدّر تعتمد على بيانات العيّنة، بالتالي يمكن أن نكتب: $I_T(\theta) \le I_{SA}(\theta)$.

- فإذا كان: $I_T(\theta) = I_{SA}(\theta)$ عندئذٍ يقال عن المقدّر T أنّه مقدّر كافٍ (أي يحتوي على كل المعلومات التي تقدّمها العيّنة عن الوسيط θ) ./ لاحظ، وجدنا في المجتمع الطبيعي $X : N(\mu,\sigma^2)$ أنّ: $I_{SA}(\mu) = I_{\overline{X}}(\mu) = \dfrac{n}{\sigma^2}$ وبالتالي $\overline{X}$ مقدّر كافٍ ويحتوي على كل المعلومات التي تقدمها العيّنة عن الوسيط μ / .

- وإذا كان: $I_T(\theta) < I_{SA}(\theta)$ فسيكون هناك خسارة في المعلومات، ولن يكون المقدّر كافياً في هذه الحالة، والخسارة في المعلومات تقاس عندئذٍ بمقدار الفرق بين الدالتين أي: { الخسارة في المعلومات } $= I_{SA}(\theta) - I_T(\theta)$.

ملاحظة:

باستخدام الرموز السابقة يمكن كتابة متراجحة كرامر بالشكل التالي: $\sigma^2(T) \geq \dfrac{1}{I_{SA}(\theta)}$ حيث T مقدّر غير متحيّز لـ θ، ولكي يكون له أقل تباين يجب أن يكون: $\sigma^2(T) = \dfrac{1}{I_{SA}(\theta)}$، ويمكن إثبات أنّ الشرط اللازم لتحقيق المعادلة الأخيرة هو:

$$\frac{\partial}{\partial\theta}\log f(x_1, x_2, \ldots, x_n, \theta) = I_{SA}(\theta)[T-\theta]$$

أي: كي نثبت أنّ المقدّر غير المتحيّز له أقل تباين ممكن ، يجب أن نثبت أنّ المشتقة التفاضلية الأولى للوغاريتم الدالّة المشتركة يساوي جداء مقدارين: أولهما دالّة معلومات العيّنة، وثانيهما الفرق بين المقدّر والوسيط . والمثال التالي يوضّح ذلك:

مثال (3):

في المجتمع الطبيعي $X : N(\mu,\sigma^2)$ اثبت أنّ المقدّر $T = \overline{X}$ له أقل تباين كمقدّر للوسيط μ .

الحل:

سنثبت الشرط التالي:

$$\frac{\partial}{\partial \mu} \log f(x_1, x_2, ..., x_n, \mu) = I_{SA}(\mu)[\overline{X} - \mu]$$

وجدنا في الأمثلة السابقة أنَّ: $I_{SA}(\mu) = \dfrac{n}{\sigma^2}$

لدينا:

$$f(x_1, x_2, ..., x_n, \mu) = (\frac{1}{\sigma\sqrt{2\pi}})^n \; e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\log f(x_1, x_2, ..., x_n, \mu) = \log(\frac{1}{\sigma\sqrt{2\pi}})^n - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial}{\partial \mu} \log f(x_1, x_2, ..., x_n, \mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu)$$

$$= \frac{1}{\sigma^2}(n\overline{X} - n\mu) = \frac{n}{\sigma^2}(\overline{X} - \mu)$$

الأمر الذي يعني أنَّ: $\dfrac{\partial}{\partial \mu} \log f(x_1, x_2, ..., x_n, \mu) = I_{SA}(\mu)[\overline{X} - \mu]$

وهذه العلاقة هي حاصل جداء مقدارين أولهما: هو دالّة معلومات العيِّنة $I_{SA}(\mu)$، وثانيهما: هو الفرق بين المقدّر $\overline{X}$ والوسيط μ، أي أنَّ $\overline{X}$ هو مقدّر غير متحيِّز وتباينه أقل ما يمكن، وبالتالي فهو أفضل مقدّر.

(8-12) بعض طرق التقدير النقطي:

بعد أن تعرَّفنا إلى خواص المقدّرات، ننتقل الآن لاستعراض أهم طريقتين مستخدمتين في إيجاد المقدّرات، وهما: طريقة الإمكانيّة العظمى (تسمى أحياناً طريقة الإمكان الأكبر)، وطريقة العزوم.

(8-12-1) طريقة الإمكانيّة العظمى :(*Method of maximum Likelihood*)

تعتبر هذه الطريقة من أهم طرق التقدير النقطي وخاصة في حالة العيّنات الكبيرة، وتُعزى هذه الطريقة لرونالد فيشر (*R.Fisher*)، وهي تعطي مقدّرات كافية إن وجدت، إلا أنّها قد تكون متحيّزة أحياناً، ولكن في الحالة، عندما تؤول n (حيث n حجم العيّنة)

إلى ما لانهاية، فإنّها تعطي مقدّرات غير متحيّزة، ولها أقل تباين، ومتّسقة دوماً، كما يؤول توزيعها إلى التوزيع الطبيعي مع زيادة حجم العيّنة.

بفرض المجتمع الإحصائي $\{A\ ,\ X\ ,\ f(x,\theta)\}$ حيث θ الوسيط المراد تقديره . ولنفرض أنّ متغيّرات القيم الملاحظة في العيّنة من الشكل: $X_1, X_2, \ldots, X_n$

(وهي كما نعلم مستقلة ومتماثلة التوزيع) ، عندئذٍ تابع الكثافة المشترك للعيّنة

(ولنرمز له بالرمز $L(\theta)$) هو :

$$L(\theta) = f(x_1,\theta).f(x_2,\theta)\ldots f(x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$$

وفي طريقة الإمكانية العظمى يتم اختيار قيمة θ التي تجعل تابع الكثافة المشترك (والذي يسمى بتابع الإمكانيّة العظمى أو تابع الإمكان) أكبر ما يمكن. ويسمى المقدّر في هذه الحالة بالمقدّر المبني على التوقع الأعظمي أو مقدّر الإمكان الأكبر.

فإذا كانت $T = T(x_1, x_2, \ldots, x_n)$ هي قيمة θ التي تعظّم الدالّة $L(\theta) = \prod_{i=1}^{n} f(x_i,\theta)$

أي إذا كان: $L(T) = \max_{\theta} L(\theta)$ عندئذٍ نقول عن T إنّه مقدّر مبني على التوقع الأعظمي للوسيط θ، وبما أنّ T يجب أن يحقّق النهاية العظمى لتابع الإمكانيّة العظمى، فإنّه يمكن الحصول على هذه النهاية عادةً عن طريق الاشتقاق بالنسبة للوسيط θ، ثم مساواة الناتج بالصفر.

بمعنى آخر: إذا كان $L(\theta)$ تابعاً قابلاً للمفاضلة بالنسبة لـ θ، فإنّ الشرط اللازم لكي يبلغ $L(\theta)$ نهايته العظمى هو: $\frac{\partial L}{\partial \theta} = 0$. ولما كان $L(\theta)$ هو إما جداء توابع احتمال، أو جداء توابع كثافة، فإنّه موجب دوماً، وبالتالي يكون $\log L(\theta)$ معرّفاً دوماً، وهو تابع متزايد باضطراد. وبالحالة العامة يكون الوصول إلى أعظميّة $\log L(\theta)$ أسهل من الوصول إلى أعظميّة $L(\theta)$، لأنّه بأخذ اللوغاريتم تُحوّل الدالّة إلى حاصل جمع، وهذا يُسهّل العمليات الرياضيّة، في حين $L(\theta)$ عبارة عن جداء دوال.

وبشكل مكافئ: إنّ تغيُّرات التابع: $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$ تماثل تغيُّرات التابع: $\log L(\theta)=\log\prod_{i=1}^{n} f(x_i,\theta)$، وقيمة θ التي تجعل هذا التابع في نهايته العظمى تعطى من حل المعادلة:

$$\frac{\partial}{\partial\theta}[\sum_{i=1}^{n}\log f(x_i\,,\,\theta)]=0$$

والمقدّر T المحسوب بهذه الطرقة يسمى: المقدّر المبني على التوقع الأعظمي.

مثال (1):

في المجتمع الطبيعي $X : N(\mu,\sigma^2)$ وبفرض μ الوسيط المجهول.

أوجد المقدّر المبني على التوقع الأعظمي للوسيط μ .

الحل: الخطوات كما يلي:

1) دالّة الكثافة من الشكل: $f(x,\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

وبالتالي: $f(x_i,\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2}$

2) $\log f(x_i,\mu)=\log(\frac{1}{\sigma\sqrt{2\pi}})-\frac{1}{2\sigma^2}(x_i-\mu)^2$

3) نأخذ المجموع:

$$\sum_{i=1}^{n}\log f(x_i,\mu)=\sum_{i=1}^{n}\log(\frac{1}{\sigma\sqrt{2\pi}})-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$
$$=n\log\frac{1}{\sigma\sqrt{2\pi}}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

4) نشتق بالنسبة للوسيط μ :

$$\frac{\partial}{\partial\mu}\sum_{i=1}^{n}\log f(x_i,\mu)=0+\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

5) نساوي المشتق بالصفر ونحل المعادلة الناتجة:

$$\frac{\partial}{\partial \mu}\sum_{i=1}^{n}\log f(x_i,\mu)=0 \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0$$

$$\Rightarrow \sum_{i=1}^{n}x_i - n\mu = 0$$

$$\Rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n}x_i$$

بالتالي المقدَّر المبني على التوقع الأعظمي للوسيط μ في المجتمع الطبيعي المذكور هو: $T=\frac{1}{n}\sum_{i=1}^{n}X_i \equiv \overline{X}$ مع ملاحظة أنّه كان من الممكن أن نكتب مباشرةً:

$$L(\mu)=\prod_{i=1}^{n}f(x_i,\mu)=(\frac{1}{\sigma\sqrt{2\pi}})^n.e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\log L(\mu)=n(\frac{1}{\sigma\sqrt{2\pi}})-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial}{\partial \mu}\log L(\mu)=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

وبمساواتها بالصفر يكون:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0 \Rightarrow \sum_{i=1}^{n}x_i-n\mu=0 \Rightarrow \mu=\frac{1}{n}\sum_{i=1}^{n}x_i$$

إذاً: $T=\frac{1}{n}\sum_{i=1}^{n}X_i \equiv \overline{X}$ وهو المقدَّر المبني على التوقع الأعظمي للوسيط μ في المجتمع المذكور.

بالحالة العامة: إذا كان التوزيع يحتوي على أكثر من وسيط مجهول، فإنّه بطريقة مشابهة يمكن الحصول على قيم الوسطاء التي تعظّم تابع الإمكانيّة العظمى بالشكل:

إذا كان تابع الكثافة من الشكل: $f(x,\theta_1,\theta_2,....,\theta_n)$

فإنّ تابع الإمكانيّة العظمى يكون: $L(\theta_1,\theta_2,....,\theta_n)$

نأخذ لوغاريتم تابع الإمكانيّة العظمى: $\log L(\theta_1,\theta_2,...,\theta_n)$ ،

ثم نشتق جزئياً بالنسبة لكل وسيط من الوسطاء ونساوي بالصفر، وعندها سنحصل على k من المعادلات التي بحلها نحصل على مقدّرات التوقع الأعظمي $T_1, T_2, ..., T_k$ والتي تحقّق المعادلات:

$$\frac{\partial \log\ L(\theta_1, \theta_2, ..., \theta_n)}{\partial \theta_i} = 0 \quad ; i = \overline{1, k}$$

مثال (2):

في المجتمع الطبيعي $X : N(\mu, \sigma^2)$. أوجد المقدّرين المبنيين على التوقع الأعظمي للوسيطين μ و σ^2.

الحل: حسب الخطوات السابقة نكتب:

تابع الإمكانيّة العظمى:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = (\frac{1}{\sqrt{2\pi}})^n \ (\frac{1}{\sigma})^n \ . \ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= (\frac{1}{\sqrt{2\pi}})^n \ (\sigma^2)^{-\frac{n}{2}} \ . \ e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

بالتالي:

$$\log\ L(\mu, \sigma^2) = C - \frac{n}{2} \ \log\ \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \quad ; C = \log(\frac{1}{\sqrt{2\pi}})^n$$

وبالاشتقاق الجزئي : مرة بالنسبة لـ μ، وأخرى بالنسبة لـ σ^2 يكون:

$$\frac{\partial}{\partial \mu} \log\ L(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \qquad (*)$$

$$\frac{\partial}{\partial \sigma^2} \log\ L(\mu, \sigma^2) = \frac{-n1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad (**)$$

وبمساواة كلٍ من (*) و (**) بالصفر، نحصل على:

$T_1 = \frac{1}{n}\sum_i x_i = \overline{X}$ وهو المقدّر المبني على التوقع الأعظمي للوسيط μ.

$T_2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2$ وهـو المقـدّر المبنـي علـى التوقـع الأعظمـي للوسـيط σ^2، وهـو مقدّر متحيّز، وهذا ما أوردناه سابقاً عندما قلنا بأنّ المقدّرات المبنيّة على التوقع الأعظمي قد تكون متحيّزة.

(2-12-8) طريقة العزوم *(Method of moments)*

في المجتمع الإحصائي $\{A\ ,\ X\ ,\ f(x,\theta)\}$ نعلم أنّ العزوم من الرتبة i حول الصفر تعطى بالشكل:

$$\alpha_i = E(X^i) = \sum_i x_i p_i$$

$$or = \int_{-\infty}^{\infty} x^i f(x)dx$$

والعزوم حول التوقع الرياضي ومن ذات الرتبة هي:

$$\mu_i = E[(X-m)^i] = \sum_i (x_i - m)^i p_i \quad ; m = E(X)$$

$$or = \int_{-\infty}^{\infty} (x-m)^i f(x)dx$$

تعريف:

بفرض: $X_i\ ,\ i = \overline{1,n}$ متغيُّرات القيم الملاحظة لعيّنة عشوائيّة حجمها n. عندئذٍ نعرّف العزم الابتدائي حول الصفر، ومن الرتبة s، لهذه العيّنة بالشكل:

$$\hat{\alpha_s} = \frac{1}{n}\sum_{i=1}^{n} X_i^s \quad ; s = \overline{1,k}$$

ومنه ينتج: $\hat{\alpha_1} = \frac{1}{n}\sum_{i=1}^{n} X_1^s\ ,\ \hat{\alpha_2} = \frac{1}{n}\sum_{i=1}^{n} X_2^s\ ,\ \hat{\alpha_3} = \frac{1}{n}\sum_{i=1}^{n} X_3^s$ الخ ...

وكل منها يعتبر متغيّراً إحصائيّاً (أشرنا إلى ذلك في أثناء دراسة المتغيُّرات الإحصائيّة)، وبالواقع يعتبر كل منها أيضاً مقدّرا غير متحيّز ومعقول للوسيط θ كما تؤكّده النظريّة التالية:

نظريّة:

عـزوم العيّنـة : $\hat{\alpha_1},\hat{\alpha_2},....,\hat{\alpha_S}$ هـي مقـدّرات غيـر متحيّـزة ومعقولـة لعـزوم المجتمـع: $\alpha_1,\alpha_2,...,\alpha_S$ على الترتيب ، وذلك من أجل قيمة محدّدة لـ S وبشرط وجود α_S .

البرهان:

$$E(\hat{\alpha_s})=E(\frac{1}{n}\sum_i X_i^s)=\frac{1}{n}\sum_i E(X_i^s)$$
$$=\frac{1}{n}.n\,E(X^s)=E(X^s)=\alpha_s$$

وهذا يؤكّد صفة عدم التحيُّز أو الإنصاف.

ومن جهة أخرى، لدينا:

$$\sigma^2(\hat{\alpha_s})=\sigma^2(\frac{1}{n}\sum_i X_i^s)=\frac{1}{n^2}\sum_i \sigma^2(X_i^s)$$
$$=\frac{1}{n^2}.n\,\sigma^2(X^s)=\frac{\sigma^2(X^s)}{n}\xrightarrow[n\to\infty]{}0$$

وهذا بدوره يؤكّد خاصة الاتّساق أو الإمكانيّة، وهو المطلوب.

وبـالواقع تعتبـر طريقـة العـزوم لإيجـاد المقـدّرات مـن أقـدم طـرق التقـدير، وهـي تـتلخّص بالخطوات التالية:

a) نوجد العزوم الخاصـة بالمجتمع الإحصائي (قد تكون حول الصفر، أو حول التوقع الرياضي، أو قد نختار العزوم الـ k الأولى أو غير ذلك) .

b) نوجد العزوم المناظرة لها في العيّنة.

c) نعادل العزوم المتنـاظرة فـي كلٍ مـن المجتمـع والعيّنـة (أي نسـاوي بينهما)، ثم نحل المعادلـة (أو المعـادلات)الناتجـة حسـب مـا يكـون المتغيّر X تابعـاً لوسيط واحد أم لعدة وسطاء .

مثال (1):

بفرض: $(X_1, X_2, ..., X_n)$ عيّنة عشوائيّة من التوزيع الطبيعي $X : N(\mu, \sigma^2)$. ولنقدّر الوسيط المجهول μ اعتماداً على طريقة العزوم.

الحل:

بما أنَّ μ هو التوقع الرياضي في التوزيع الطبيعي، فإنّه من المناسب اختيار العزم الأول حول الصفر، وكنّا قد وجدنا سابقاً أنّ العزم الأول حول الصفر (في المجتمع الطبيعي) هو: $\alpha_1 = E(X) = \mu$ ، وفي العيّنة هو: $\hat{\alpha_1} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

وبوضع: $\alpha_1 = \hat{\alpha_1}$ ينتج: $\mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$ أي: المقدّر المطلوب هو: $T = \overline{X}$ وهو ذات المقدّر الذي حصلنا عليه بطريقة الإمكانيّة العظمى، وهو مقدّر متّسق وغير متحيّز.

مثال (2):

ليكن المطلوب هو تقدير σ^2 كوسيط مجهول في المثال(1)، وذلك بطريقة العزوم.

الحل:

بما أنّ σ^2 يمثّل التباين في التوزيع الطبيعي، فإنّه من المناسب اختيار العزم الثاني حول التوقع الرياضي. أي:

في المجتمع: $\mu_2 = E(X-\mu)^2 = \sigma^2$

وفي العيّنة: $\hat{\mu_2} = \frac{1}{n}\sum_{i}(x_i - \overline{x})^2$

بالتالي: $\mu_2 = \hat{\mu_2} \Rightarrow \sigma^2 = \frac{1}{n}\sum_{i}(x_i - \overline{x})^2$

أي أنّ: $T = \frac{1}{n}\sum_{i}(x_i - \overline{x})^2$ هو مقدّر التباين في التوزيع الطبيعي بطريقة العزوم، وهو مقدّر متّسق، لكنه متحيّز.

مثال (3): استخدم طريقة العزوم لتقدير الوسيط θ في المجتمع الإحصائي الذي تابع كثافته من الشكل: $f(x,\theta) = \theta . x^{\theta-1} \;\; ; \; (\; 0 < x < 1 \;\; , \;\; \theta > 0)$

الحل: العزم الأول حول الصفر (للمجتمع):

$$\alpha_1 = E(X) = \int_0^1 x.f(x,\theta)\,dx = \int_0^1 x.\theta x^{\theta-1}\,dx$$

$$= \theta \int_0^1 x^{\theta}\,dx = \frac{\theta}{\theta+1}[x^{\theta+1}]_0^1 = \frac{\theta}{\theta+1}$$

العزم الأول حول الصفر (للعيّنة): $\hat{\alpha_1} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

بالتالي: $\alpha_1 = \hat{\alpha_1} \Rightarrow \frac{\theta}{\theta+1} = \overline{X} \Rightarrow \theta = \frac{\overline{X}}{1-\overline{X}}$

أي أنّ المقدّر المطلوب هو : $T = \frac{\overline{X}}{1-\overline{X}}$

التقدير المجالي *(Interval Estimation)*:

(8-13) مجال الثّقة للوسيط μ للتوزيع الطبيعي $X : N(\mu,\sigma_0)$:

لقد تركّز اهتمامنا في التقدير النقطي لوسيط θ على حساب قيمة واحدة من بيانات العيّنة

وتهدف هذه الفقرة إلى إيجاد قيمتين بحيث تُحدّد درجة (أو معيار) ثقة بوقوع الوسيط بينهما، بمعنى آخر: الغاية من التقدير المجالي هو إعطاء المقدّر مجالاً يتحرك فيه بدرجة ثقة معيّنة مع ملاحظة أنّه في التقدير النقطي لم نستطع حساب احتمال صحة تقدير T لـ θ.

بفرض لدينا المجتمع الإحصائي $\{ A , X , f(x,\theta) \}$. فإذا كانت $\theta \in [T_1,T_2]$، فإنّه يُطلَبُ معرفة الاحتمال $P[T_1 \le \theta \le T_2]$، حيث نسمي: T_1 حد الثّقة الأدنى و T_2 حد الثّقة الأعلى.

وإذا كان $P[T_1 \le \theta \le T_2] = 1-\varepsilon$ حيث ε عدد حقيقي يحقّق $(0 \le \varepsilon \le 1)$، فإنّنا نسمي الكمية $1-\varepsilon$ بعامل (أو معيار) الثّقة ويرمز له عادةً بالرمز $1-\frac{p}{100}$ حيث $(0 \le p \le 100)$، في حين نسمي المجال $[T_1,T_2]$ بمجال الثّقة للوسيط θ على

المستوى $p\%$ ، وبالتالي يصبح المطلوب هو معرفة المجال $[T_1,T_2]$ لكي يكون:

$P[T_1 \leq \theta \leq T_2] = 1-\frac{p}{100}$ حيث الاحتمال $1-\frac{p}{100}$ معطى .

وإذا كانت القيَم المحسوبة لـ T_1 و T_2 من العيِّنة هي t_1 و t_2 على الترتيب، فإنَّنا نسمي المجال $[t_1,t_2]$ بمجال الثِّقة للوسيط θ وهو مجال محدَّد (غير عشوائي)، في حين المجال $[T_1,T_2]$ هو مجال عشوائي، وبالتالي فإنَّ الاحتمال $P[t_1 \leq \theta \leq t_2]$ هو إما صفر وإما واحد لأنَّ كل من t_1 و t_2 محدَّدة و θ إما أن تقع بينهما وإما لا تقع .

إذاً المسألة المطروحة الآن هي: كيف يمكن إيجاد مجال الثِّقة (يسمى في بعض المراجع فترة ثقة) إذا كان عامل الثِّقة معروفاً ؟ وبالحقيقة ، فإنَّه لمعرفة مثل هذا المجال، فإنَّنا نحتاج لما يسمى: الكميَّة المحوريَّة (*Pivotal Quantity*)، وهذه الكميَّة عبارة عن دالَّة تابعة للمقدِّر وللوسيط ، وتوزيعها لا يعتمد على الوسيط.

بمعنى آخر: يمكن إيجاد مجال ثقة لوسيط مثل θ إذا استطعنا إيجاد متغيِّر إحصائي T يتمتَّع بالصفتين التاليتين:

أ . T يحتوي على الوسيط θ ولا يحتوي على أي وسيط إحصائي آخر .

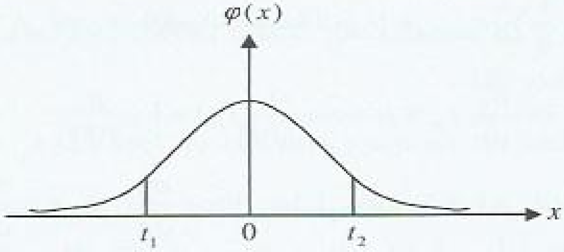
ب ـ توزيع T معروف ولا يعتمد على الوسيط θ ، ولا على أي وسيط آخر (مستقل عن θ) عندئذٍ ، وفي هذه الحالة فإنَّه عند معرفة p سنتمكن من حساب t_1 و t_2 من العلاقة: $P[t_1 \leq T \leq t_2] = F(t_2)-F(t_1) = 1-\frac{p}{100}$ (حيث F هو تابع توزيع T)، والذي يؤدي بدوره لمعرفة T_1 و T_2 في العلاقة: $P[T_1 \leq \theta \leq T_2] = 1-\frac{p}{100}$ ، حيث T_1 و T_2 تابعان فقط لـ t_1 و t_2 على الترتيب وللقيم الوسيطية المميَّزة في العيِّنة، والتي يمكن حساب كلٍ منها كما وجدنا في الفصل السابق . إذاً الصعوبة الآن تكمن بمعرفة المتغيِّر الإحصائي T المحقِّق

للشرطين أو للصفتين السابقتين ، وخاصةً في المجتمعات غير الطبيعية ، أما في المجتمعات الطبيعيّة فالمسألة أقل صعوبة كما سنجد .

من المعلوم أنّ $\overline{X}$ المحسوب من عيّنة عشوائيّة كمقدّر للوسيط μ في المجتمع المذكور له خصائص جيّدة مثل (عدم التحيُّز والكفاءة والكفاية) لذلك يكون من المنطقي استخدامه لإيجاد مجال ثقة لـ μ ، وبالواقع نجد أنّ المتغيّر الإحصائي(الكميّة المحوريّة) : $T = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ هي دالّة في المقدّر $\overline{X}$ والوسيط μ أي: يحتوي على الوسيط μ ولا يحتوي على أي وسيط آخر، هذا من ناحية . تحقق الشرط الأول . ومن ناحية ثانية: نعلم أنّ لهذا المتغيّر توزيعاً طبيعيّاً معيارياً . تحقق الشرط الثاني . لذلك أصبح من الممكن إيجاد العددين t_1 و t_2 بحيث يكون:

$$P[t_1 \leq T \leq t_2] \equiv P[\overline{X} - \frac{\sigma}{\sqrt{n}} t_2 \leq \mu \leq \overline{X} - \frac{\sigma}{\sqrt{n}} t_1] = 1 - \frac{p}{100} \quad : (*)$$

ومن دراسة التوزيع الطبيعي المعياري نعلم أنّ: $P[t_1 \leq T \leq t_2] = \Phi(t_2) - \Phi(t_1)$



حيث الطرف الأيمن هو عبارة عن المساحة المحصورة تحت المنحني الطبيعي المعياري $\varphi(x)$ وبين النقطتين t_1 و t_2، وبالتالي أصبح المطلوب هو إيجاد العددين t_1 و t_2 لكي تتحقّق العلاقة: $: (**) \quad \Phi(t_2) - \Phi(t_1) = 1 - \dfrac{p}{100}$.

أي: إيجاد العددين t_1 و t_2 بحيث تكون المساحة المذكورة مساوية لعامل الثّقة $1 - \dfrac{p}{100}$، وهذا يعني بدوره أنّه يوجد أكثر من مجال ثقة واحد ، من أجل عامل ثقة معيّن، ذلك لأنّ أي انزياح لليمين أو لليسار سيحدّد مجال ثقة مساوياً لعامل الثّقة المذكور، ونحن نريد

بالطبع إيجاد أقصر مجال ثقة ممكن، وهذا يتحقّق عندما نختار t_1 و t_2 أقرب ما يمكن لبعضهما (وبحيث لا تتغيّر المساحة) علماً أنّ طول مجال الثّقة كما يظهر من العلاقة (*) هو: $\frac{\sigma}{\sqrt{n}}(t_2 - t_1)$.

ومن أجل مساحة معيّنة فإنّ المسافة بين t_1 و t_2 تكون أقل ما يمكن عندما يكون $t_2 = -t_1$ وذلك لأنّ $\varphi(x)$ متناظر بالنسبة للنقطة 0 الممثّلة لنهايته العظمى.

وبالتعويض في (**) مع الأخذ بالاعتبار خواص Φ (حيث: $\Phi(-t_2) = 1 - \Phi(t_2)$)

يكون: $\Phi(t_2) - \Phi(-t_2) = 1 - \frac{p}{100} \Rightarrow \Phi(t_2) = 1 - \frac{1}{2}\frac{p}{100}$

ومن جداول التوزيع الطبيعي نوجد t_2 بعد معرفة p.

إذاً مجال الثّقة للوسيط μ على المستوى $p\ \%$ من أجل العيّنة: $X_1, X_2, ..., X_n$ وكما يظهر من العلاقة (*) هو: $(t_2 = -t_1 = t_p)$; $[\bar{x} - \frac{\sigma}{\sqrt{n}} t_p\ ,\ \bar{x} - \frac{\sigma}{\sqrt{n}} t_p]$

حيث نوجد قيمة t_p من جداول التوزيع الطبيعي بالاعتماد على العلاقة:

$\Phi(t_p) = 1 - \frac{1}{2}\frac{p}{100}$ في حين $\bar{x}$ (القيم الملاحظة للعيّنة) هي: $\bar{x} = \frac{1}{n}\sum_i x_i$

وبالتالي يكون: $\sigma_0 \equiv \sigma$; $P[\bar{x} - \frac{\sigma_0}{\sqrt{n}} t_p \le \mu \le \bar{x} + \frac{\sigma_0}{\sqrt{n}} t_p] = 1 - \frac{p}{100}$

أي أنّ المجال $]\bar{x} - \frac{\sigma_0}{\sqrt{n}} t_p\ ,\ \bar{x} + \frac{\sigma_0}{\sqrt{n}} t_p[$ يغطي القيمة الحقيقيّة للوسيط μ باحتمال مقداره $1 - \frac{p}{100}$ وهذا يكافئ القول: إنّه إذا كرّرنا عملية سحب العيّنة العشوائيّة ذات الحجم n ومن ذات المجتمع الطبيعي المذكور، وحسبنا في كل مرة مجال الثّقة أعلاه فإنّ $(100 - p)\%$ من المجالات التي نحصل عليها سوف تغطي القيمة الحقيقيّة للوسيط μ، ومما تقدّم يتّضح أنّ:

حد الثّقة الأدنى هو: $t_1 = \bar{x} - \frac{\sigma_0}{\sqrt{n}} t_p$، وحد الثّقة الأعلى هو: $t_2 = \bar{x} + \frac{\sigma_0}{\sqrt{n}} t_p$.

مثال:

أخذت العيّنة: 1.5 , 2.4 , 0.8 , 1.2 , 1.3 من المجتمع الطبيعي $X : N(\mu,1)$. اعتمد هذه العيّنة لتوجد مجال ثقة لـ μ على المستوى 99% ، ثم على المستوى 90 % .

الحل:

نعلم أنّ مجال الثّقة لـ μ بمعرفة σ هو: $[\ \bar{x}-\frac{\sigma_0}{\sqrt{n}}t_p\ ,\ \bar{x}+\frac{\sigma_0}{\sqrt{n}}t_p\ ]$

حيث: $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{5}\sum_{i=1}^{5}x_i=\frac{1.3+1.2+0.8+2.4+1.5}{5}=1.44$

وبالفرض لدينا: $p=99$ لذلك نوجد t_p من العلاقة:

$\Phi(t_p)=1-\frac{1}{2}\frac{p}{100}=1-\frac{99}{200}=0.505$ ومن جداول التوزيع الطبيعي ينتج أنّ: $t_p=0.01$. إذاً: مجال الثّقة المطلوب هو:

$$[\ 1.44-\frac{1}{\sqrt{5}}0.01\ ,\ 1.44+\frac{1}{\sqrt{5}}0.01]\approx[1.435,1.444]$$

وعلى المستوى 90 % سنحصل على المجال: [1.381 , 1.498]. لاحظ أنّ مجال الثّقة يتّسع بزيادة معامل الثّقة .

أي: تصغير العدد p $\Leftrightarrow$ توسيع مجال الثّقة $\Leftrightarrow$ زيادة الدقّة في التقدير.

ولكن، من المفضّل عادةً هو الحصول على تقدير دقيق ضمن مجال ثقة صغير، لذلك يحاول الإحصائيون أن يوفّقوا في كل مسألة من المسائل الإحصائيّة بين هذين الأمرين المتناقضين بقدر ما تسمح به طبيعة المسألة .

(14-8) مجال الثّقة للوسيط μ للتوزيع الطبيعي $X : N(\mu,\sigma^2)$

حيث: (σ مجهول وحجم العيّنة صغير):

نعلم أنّ لـ $\overline{X}$ المحسوب من العيّنة خصائص جيّدة كمقدّر لـ μ

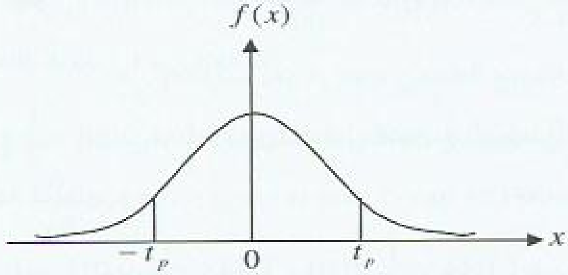
وكذلك لـ $\hat{S}^2$ المحسوب من العيّنة خصائص جيّدة كمقدّر لـ σ^2

وبالتالي إذا أخذنا المتغيّر الإحصائي (الكميّة المحوريّة) $T = \dfrac{\overline{X} - \mu}{\hat{S}/\sqrt{n}}$ فإنّنا نلاحظ :

أ . أنّه يحتوي على الوسيط μ ولا يحتوي على أي وسيط آخر (تحقق الشرط الأول).

ب . $T : t_{n-1}$ أي أن: T يتبع توزيع ستودينت وبـ $(n-1)$ درجة حريّة (تحقق الشرط الثاني).

وبما أنّ بيان تابع كثافة ستودينت يشبه إلى حد كبير شكل تابع كثافة التوزيع الطبيعي المعياري كما في الشكل:



لذلك أصبح من الممكن إيجاد العددين t_1 و t_2 بحيث يكون:

$$P[t_1 \le T \le t_2] \equiv P[\overline{X} - \frac{\hat{S}}{\sqrt{n}} t_2 \le \mu \le \overline{X} - \frac{\hat{S}}{\sqrt{n}} t_1] = 1 - \frac{p}{100} \quad : (*)$$

وأقصر مجال ثقة ممكن هو باختيار $t_2 = -t_1$ (حيث $f(x)$ متناظر)، وطول مجال الثّقة هو: $\dfrac{\hat{S}^2}{\sqrt{n}}(t_2 - t_1)$، وإذا فرضنا أنّ: $t_2 = -t_1 = t_p$ يصبح المطلوب هو إيجاد t_p الذي يحقّق العلاقة: $P[-t_p \le T \le t_p] = 1 - \dfrac{p}{100}$

أو العلاقة: $P[|T| \le t_p] = 1 - \dfrac{p}{100} \Rightarrow 1 - P[|T| > t_p] = 1 - \dfrac{p}{100}$

أي: $P[|T| > t_p] = \dfrac{p}{100}$ ولهذه العلاقة الأخيرة جداول خاصة تعطي قيمة t_p عند معرفة p وبالعكس. (حيث ننظر في الجدول إلى العمود الذي كُتِبَ في أعلاه الرقم p

وإلى السطر الذي كُتِبَ في أوله الرقم $n-1$ ، فيكون t_p عندئذٍ هو العدد الواقع في تقاطعهما) .

ومما تقدّم يكون مجال الثّقة للوسيط μ هو: $[\ \bar{x}-\frac{\hat{S}}{\sqrt{n}}\,t_p\ ,\ \bar{x}+\frac{\hat{S}}{\sqrt{n}}\,t_p\ ]$

حيث: $\bar{x}=\frac{1}{n}\sum_i x_i \quad , \quad \hat{S}^2=\frac{1}{n-1}\sum_i (x_i-\bar{x})^2$

ملاحظة:

إذا كانت n كبيرة، وكانت σ^2 مجهولة ، فإنّه يمكن استخدام نفس مجال الثّقة المذكور أعلاه، وذلك بعد تبديل σ بـ S ، أي أنّه يمكن استخدام التوزيع الطبيعي بدلاً من توزيع t .

مثال:

اعتمد العيّنة: 1.5 , 2.4 , 0.8 , 1.2 , 1.3 المسحوبة من المجتمع $X : N(\mu,\sigma^2)$ (حيث σ مجهولة) لتوجد مجال ثقة للوسيط μ على المستوى 90 % .

الحل:

مجال الثّقة المطلوب في هذه الحالة هو: $[\ \bar{x}-\frac{\hat{S}}{\sqrt{n}}\,t_p\ ,\ \bar{x}+\frac{\hat{S}}{\sqrt{n}}\,t_p\ ]$

حيث:

$$\bar{x}=\frac{1}{n}\sum_i x_i=1.44$$

$$\hat{S}^2=\frac{1}{n-1}\sum_i (x_i-\bar{x})^2=\frac{1}{4}(x_i-\bar{x})^2$$

$$=\frac{1}{4}(1.3-1.44)^2+\ \ldots\ +(1.5-1.44)^2=0.535$$

ومن جداول ستودينت (حيث: p = 90 ، n = 5) يكون: $t_p=0.132$ وبالتالي فإنّ مجال الثّقة المطلوب هو:

$$[\ 1.44-\frac{0.731}{\sqrt{5}}\,0.132 , 1.44+\frac{0.731}{\sqrt{5}}\,0.132\ ]\approx[1.396\ \ ,\ 1.483]$$

(15-8) مجال الثّقة للوسيط σ^2 في المجتمع الطبيعي $X : N(\mu,\sigma^2)$:

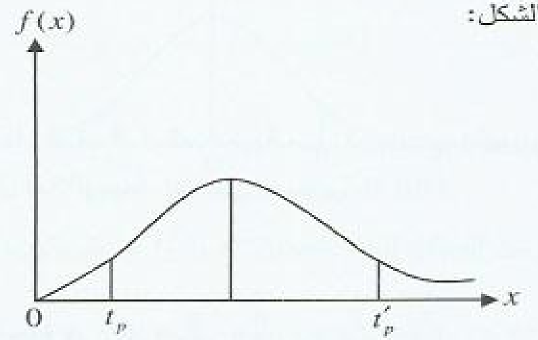
في المجتمع الإحصائي $X : N(\mu,\sigma^2)$ يُطلَبُ إيجاد مجال الثّقة للوسيط σ^2 على المستوى $p\%$. نعلم أنّ لـ S^2 المحسوب من العيّنة خصائص جيّدة كمقدّر لـ σ^2 ، وبالتالي إذا أخذنا المتغيّر الإحصائي (الكميّة المحوريّة) : $T = \dfrac{nS^2}{\sigma^2}$ فإنّنا نلاحظ :

أ . أنّه يحتوي على الوسيط σ^2 ولا يحتوي على وسيط آخر . تحقق الشرط الأول.

ب . $T : \chi^2_{n-1}$ أي يتبع توزيع χ^2 بـ $n-1$ درجة حريّة – تحقق الشرط الثاني.

علماً أنّ تابع كثافة χ^2_n هو: $f_n(x) = \dfrac{1}{2^{n/2}\ \Gamma(n/2)} e^{-\frac{x}{2}} \ .\ x^{\frac{n}{2}-1} \quad ; 0 \le x \le \infty$

وبيانه من الشكل:



لذلك أصبح من الممكن إيجاد العددين t_p و t'_p بحيث يكون:

$$P[t_p \le T \le t'_p] = P[\frac{nS^2}{t'_p} \le \sigma^2 \le \frac{nS^2}{t_p}] = 1 - \frac{p}{100} \qquad :(*)$$

كما يلي:

نعلم أنّ: $P[t_p \le T \le t'_p] = 1 - P[T < t_p] - P[T > t'_p]$

وبالتالي أصبح المطلوب هو إيجاد العددين t_p و t'_p بحيث يكون:

$$1 - P\,[T < t_p] - P[T > t'_p] = 1 - \frac{p}{100}$$

أو : $(**)$: $P\,[T < t_p] + P[T > t'_p] = \dfrac{p}{100}$

وبالنظر إلى بيان التابع $f(x)$ نلاحظ أنّ الكميّة: $P\,[T < t_p] + P[T > t'_p]$ تمثّل المساحة تحت البيان وخارج النقطتين t_p و t'_p . وبالتالي أصبح المطلوب هو إيجاد t_p و t'_p بحيث تكون هذه المساحة مساوية $\dfrac{p}{100}$. وبالواقع إنّ أي عددين يتركان خارجهما وتحت بيان التابع $f(x)$ مساحة قدرها $\dfrac{p}{100}$ يحقّقان المطلوب .

فإذا اخترنا في العلاقة $(**)$: $P\,[T > t'_p] = \dfrac{\frac{1}{2}p}{100}$, $P\,[T < t_p] = \dfrac{\frac{1}{2}p}{100}$

لتحقّق طلبنا، هذا من ناحية ، ومن ناحية أخرى ، إنّ هذا الاختيار هو الذي يجعل طول المجال $nS^2(\dfrac{1}{t_p} - \dfrac{1}{t'_p})$ حسب $(*)$، أقصر ما يمكن ، وهو ما نفضّله ، أضف إلى أنّ عملية حسابهما تكون سهلة وتتم كما يلي:

في توزيع χ^2 وجدنا العلاقة: $P(X > x_p) = \dfrac{p}{100}$ ، وبالتالي لو عرفنا p ، وعرفنا عدد درجات الحريّة لأمكن إيجاد x_p من جداول χ^2 . وبناءً عليه ، فإنّه يمكن حساب t'_p من العلاقة : $P\,[T > t'_p] = \dfrac{\frac{1}{2}p}{100}$ ، حيث: t'_p تساوي العدد الموجود في جداول χ^2 والواقع في تقاطع العمود الذي كُتِبَ في أعلاه الرقم $\frac{1}{2}p$ مع السطر الذي كُتِبَ في أوّله العدد $n-1$ (أي ندخل الجدول وفق العمود $\frac{1}{2}p$ والسطر $n-1$) .

ولحساب t_p : نلاحظ أنّ العلاقة: $P\,[T < t_p] = \dfrac{\frac{1}{2}p}{100}$ تكافئ العلاقة:

$$P\,[T > t_p] = 1 - \frac{\frac{1}{2}p}{100} = \frac{100 - \frac{1}{2}p}{100}$$

لذلك فإنَّ t_p تساوي إلى العدد الموجود في جداول χ^2 والواقع في تقاطع العمود الذي كُتِبَ في أعلاه العدد : $p\frac{1}{2}-100$ مع السطر الذي كُتِبَ في أوّله العدد : $n-1$ (أي: ندخل الجدول وفق العمود $p\frac{1}{2}-100$ والسطر $n-1$).

وبهذا الشكل (أي بعد حساب t_p و t'_p) نكون قد تمكنّا من إيجاد مجال الثّقة للوسيط σ^2 وهو : $[\frac{nS^2}{t_p},\frac{nS^2}{t'_p}]$ ، حيث: $S^2=\frac{1}{n}\sum_i(x_i-\bar{x})^2$ علماً أنَّ : $\bar{x}=\frac{1}{n}\sum_i x_i$

وبمعنى آخر: إنّ العبارة الاحتماليّة: $P[\frac{nS^2}{t_p}\leq\sigma^2\leq\frac{nS^2}{t'_p}]$ ، تعني أنّ المجال: $[\frac{nS^2}{t_p},\frac{nS^2}{t'_p}]$ يغطي القيمة الحقيقيّة لـ σ^2 باحتمال مقداره $1-\frac{p}{100}$.

وبما أنّ هذا المجال نتج عن عيّنة عشوائيّة حجمها n ، فإنّ العبارة الاحتماليّة السابقة تعني أيضاً أنّه لو كرّرنا عملية سحب العيّنة $x_1,x_2,....,x_n$ (ذات الحجم n) من المجتمع $X:N(\mu,\sigma^2)$ وحسبنا في كل مرة المجال $[\frac{nS^2}{t_p},\frac{nS^2}{t'_p}]$، فإنّ $(100-p)\%$ من المجالات التي نحصل عليها سوف تغطي القيمة الحقيقيّة للوسيط σ^2.

ملحوظة: المجال $[\frac{nS^2}{t_p},\frac{nS^2}{t'_p}]$ يكافئ المجال $[\frac{(n-1)\hat{S}^2}{t_p},\frac{(n-1)\hat{S}^2}{t'_p}]$، حيث: $\hat{S}^2$ هو متغيّر تباين العيّنة المعدّل .

مثال (1): أُخِذت عيّنة عشوائيّة حجمها $n=12$ من المجتمع $X:N(\mu,\sigma^2)$ ووجد أنّ: $S^2=18.3$. أوجد مجال ثقة لـ σ^2 على المستوى 10%.

الحل:

نعلم أنّ مجال الثّقة للوسيط σ^2 هو : $[\frac{nS^2}{t_p},\frac{nS^2}{t'_p}]$

ولإيجاد t'_p من جداول ${\chi_n}^2$ (ندخل الجدول وفق السطر $n-1$ والعمود 5 = $\frac{1}{2}p$) فنجد:

$t'_p = 19.675$، ثم لإيجاد t_p من جداول ${\chi_n}^2$ (ندخل الجدول وفق السطر $n-1$ والعمود $100-\frac{1}{2}p = 95$) فنجد: $t_p = 4.575$.

ويكون مجال الثقة المطلوب هو: $[\frac{12(18.3)}{19.675}, \frac{12(18.3)}{4.675}] = [11.16\ ,\ 48]$

مثال (2):

مجتمع طبيعي موصوف بالمتغيّر $X : N(0,\sigma^2)$. أُخذت منه عيّنة حجمها $n = 10$ فكانت القيم الملاحظة لهذه العيّنة هي: $-4\ ,3\ ,1\ ,0\ ,5\ ,-2\ ,-3\ ,-2\ ,1\ ,1$

اعتمد هذه العيّنة في إيجاد مجال الثقة للوسيط σ^2 على المستوى 10%.

الحل:

مجال الثقة للوسيط σ^2 هو: $[\frac{nS^2}{t'_p}, \frac{nS^2}{t_p}]$، حيث: $nS^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$

وبالحساب يكون:

$$\bar{x} = \frac{1}{n}\sum_i x_i = \frac{1}{10}(-4+3+1+0+5-2-3-2+1+1) = 0$$

وبالتالي: $nS^2 = 16+9+1+0+25+4+9+4+1+1 = 70$

ولحساب t'_p (ندخل الجدول وفق السطر $n-1 = 9$ والعمود 5 = $\frac{1}{2}p$) نجد: $t'_p = 16.919$. ومجال الثقة المطلوب يكون: $[\frac{70}{16.919}, \frac{70}{3.325}] \approx [4.13\ ,\ 21.05]$.

(8-16) مجال الثقة للفرق بين متوسطي مجتمعين طبيعيين $(\mu_1 - \mu_2)$.

بفرض: $A\ ;\ X : N(\mu_1, {\sigma_1}^2)$ و $B\ ;\ Y : N(\mu_2, {\sigma_2}^2)$: مجتمعين طبيعيين مستقلين، وليكن S_1^2 و S_2^2 متغيّري تباين العيّنة في المجتمعين A و B بالترتيب.

وجدنا سابقاً أنّ المتغيّر الإحصائي: $\bar{X} - \bar{Y} : N(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2})$

حيث: $\overline{X}$ يقابل العيِّنة $X_1, X_2,...., X_n$ من المجتمع A

$\overline{Y}$ يقابل العيِّنة $Y_1 , Y_2,...., Y_n$ من المجتمع B

ومن أجل: $\sigma_1 = \sigma_2 = \sigma$ يكون: $\overline{X} - \overline{Y} : N(\mu_1 - \mu_2 , \sigma^2 (\frac{1}{n_1} + \frac{1}{n_2}))$

وبالتالي فإنَّ: $V = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : N(0 , 1)$

ونعلم أيضاً أنَّ: $\dfrac{n_1 S_1^{\ 2}}{\sigma_1^{\ 2}} : \chi^2_{n_1 - 1}$ و : $\dfrac{n_2 S_2^{\ 2}}{\sigma_2^{\ 2}} : \chi^2_{n_2 - 1}$

بالتالي، ومن خاصة التجميع لتوزيع χ^2 يكون:

$$U = \frac{n_1 S_1^{\ 2}}{\sigma_1^{\ 2}} + \frac{n_2 S_2^{\ 2}}{\sigma_2^{\ 2}} = \frac{1}{\sigma^2} (n_1 S_1^2 + n_2 S_2^2) : \chi^2_{n_1 + n_2 - 2}$$

حيث المتغيِّران X و Y مستقلان .

ونتيجة لذلك يكون: $$T = \frac{V}{\sqrt{U / (n_1 + n_2 - 2)}} : t_{n_1 + n_2 - 2}$$

وبتبديل U و V بقيمهما يكون : $$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} : t_{n_1 + n_2 - 2}$$

وهو متغيِّر إحصائي (كميَّة محوريَّة) يحقِّق الشرطين:

أ . يحتوي على $(\mu_1 - \mu_2)$ ولا يحتوي على وسيط آخر .

ب . توزيعه معروف حيث يخضع لتوزيع ستوديَنت بـ $(n_1 + n_2 - 2)$ درجة حريَّة.

لذلك أصبح من الممكن إيجاد العددين t_1 و t_2 بحيث تتحقَّق العلاقة:

$$P[t_1 \leq T \leq t_2] = P[t_1 \leq \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq t_2 =] = 1 - \frac{p}{100}$$

والتي يمكن كتابتها على الشكل:

$$P\{t_1\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}\le(\overline{X}-\overline{Y})-(\mu_1-\mu_2)\le t_2\,\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}\}$$

$$=1-\frac{p}{100}$$

أو على الشكل:

$$P\{(\overline{X}-\overline{Y})-t_2\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}\le(\mu_1-\mu_2)\le(\overline{X}-\overline{Y})-t_1\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}\}$$

$$=1-\frac{p}{100}$$

وبنفس الطريقة السابقة إذا اعتبرنا أنّ: $t_2=-t_1=t_p$

فإنّ مجال الثّقة المطلوب على المستوى $p\,\%$ من أجل العيّنات المفروضة يكون:

$$[\overline{X}-\overline{Y}-t_p\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}\;,\;\overline{X}-\overline{Y}+t_p\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}})]$$

حيث: $\overline{y}=\frac{1}{n_2}\sum_{i=1}^{n_2}y_i \quad , \quad \overline{x}=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i$

$$S_2^2=\frac{1}{n_2}\sum_{i=1}^{n_2}(y_i-\overline{y})^2 \quad , \quad S_1^2=\frac{1}{n_1}\sum_{i=1}^{n_1}(x_i-\overline{x})^2$$

ونوجد t_p من الجدول وفق العمود p والسطر (n_1+n_2-2). نلاحظ مما تقدّم أنّ:

حد الثّقة الأدنى هو: $\overline{X}-\overline{Y}-t_p\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}$

وحد الثّقة الأعلى: $\overline{X}-\overline{Y}+t_p\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}$

ومنه ينتج:

. إذا كان التباين معلوماً لكلٍ من المجتمعين ، فإنّ حد الثّقة الأدنى يصبح:

$$\overline{X}-\overline{Y}-t_p\sqrt{\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1\,n_2}}$$

والأعلى: $\overline{X}-\overline{Y}+t_p\sqrt{\frac{n_1\sigma_1^2+n_2\sigma_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1 n_2}}$

. وفي الحالة الخاصة ، إذا كان: $\sigma_1=\sigma_2=\sigma$ وكان: $n_1=n_2=n$ فإنّ:

حد الثّقة الأدنى يصبح: $\overline{X}-\overline{Y}-t_p\ .\ \sigma\sqrt{\frac{2}{n-1}}$

وحد الثّقة الأعلى: $\overline{X}-\overline{Y}+t_p\ .\ \sigma\sqrt{\frac{2}{n-1}}$

ملاحظة:

إذا كان n_1 و n_2 كبيرين بدرجة كافية، فإنّ الدراسة السابقة تبقى صحيحة في مجتمعات غير طبيعية، سواءً كان σ_1 و σ_2 معروفين أم لا.

تطبيق:

أُخذت عيّنتان عشوائيّتان من توزيعين طبيعيين لهما التباين نفسه، حيث وجد:

للعيّنة الأولى أنّ: $n_1=10\ ,\ \overline{x}=60\ ,\ s_1^2=72$

وللعيّنة الثانية أنّ: $n_2=8\ ,\ \overline{y}=54\ ,\ s_2^2=78.75$

أوجد مجالات الثّقة: % 30 , % 5 للفرق بين الوسيطين : $\mu_1-\mu_2$.

الحل:

من أجل % 5 تكون : $t_p=2.120$ وبالتعويض في المجال:

$$[\overline{X}-\overline{Y}-t_p\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1 n_2}}\ ,\ \overline{X}-\overline{Y}+t_p\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}\cdot\frac{n_1+n_2}{n_1 n_2}})]$$

نحصل على مجال الثّقة المطلوب وهو : $[-74.6\ ,\ 86.6]$

ومن أجل: %30 تكون: $t_p=1.071$ ويكون مجال الثّقة المطلوب عندئذٍ هو : $[-80.3\ ,\ 92.3]$، ونذكّر بأنّ الفرضية $\mu=\mu_0$ تكون مقبولة على مستوى ثقة $p\%$ إذا كانت μ_0 واقعة ضمن مجال الثّقة الموافق .

(8-17) مجال الثّقة لتوقع مجتمع غير طبيعي في حالة عيّنات كبيرة الحجم:

بفـرض A مجتمعـاً إحصـائياً مـا، موصـوفاً بمتغيّـر عشـوائي X توقعـه الرياضـي μ وتباينه σ^2، وحيث لا نفترض هنا أنّ المجتمع A هو مجتمع طبيعي، ولكنّنا نفترض أنّ حجم العيّنـة n كبيـر $(n \geq 30)$ إلـى الحد الـذي يسـمح بالاسـتفادة مـن نظريّـة النهايـة المركزيّة واعتبار توزيع $\overline{X}$، مطابقاً تقريباً للتوزيع الطبيعي.

إذاً: بما أنّ n كبيرة، فإنّه حسب نظريّة النهاية المركزيّة يكون للمتغيّر: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ توزيعـاً مطابقـاً تقريبـاً لتوزيـع طبيعـي مـن الأنمـوذج $(\mu, \frac{\sigma^2}{n})$، ولكنّنـا نعلـم أنّـه مـن أجل $X : N(\mu, \sigma^2)$، فإنّ للعلاقة $P[\left|\overline{X} - \mu\right| > \lambda_p\ \sigma] = \frac{p}{100}$ جداول خاصـة تفيد في معرفة λ_p عند معرفة p وبالعكس، وينتج عن هذا أنّنا نستطيع إيجاد λ_p الموجودة في العلاقة: $P[\left|\overline{X} - \mu\right| > \lambda_p \frac{\sigma}{\sqrt{n}}] = \frac{p}{100}$ والتي يمكن كتابتها على الشكل التالي:

$$P[\overline{X} - \lambda_p \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \lambda_p \frac{\sigma}{\sqrt{n}}] = 1 - \frac{p}{100}$$

ويكون مجال الثّقة لـ μ على المستوى $p\%$ هو المجال :

$]\overline{X} - \lambda_p \frac{\sigma}{\sqrt{n}} , \overline{X} + \lambda_p \frac{\sigma}{\sqrt{n}}[$ ، وإذا لم يكن σ معروفاً ، فإننا نعوض عن σ بـ S ويكون مجال الثّقة في هذه الحالة هو المجال: $]\overline{X} - \lambda_p \frac{S}{\sqrt{n}} , \overline{X} + \lambda_p \frac{S}{\sqrt{n}}[$ ، حيث S هو الانحراف المعياري للعيّنة .

وقد جرت العادة على تسمية الحد $\frac{S}{\sqrt{n}}$ بالخطأ المعياري للقيمة $\overline{X}$ في العيّنات الكبيرة، علمـاً أنّ $\overline{X}$ (متوسـط العيّنـة) هـو تقـدير لــ μ (متوسـط المجتمـع). ويسـمى عـادة المقدار $e = \lambda_p \frac{S}{\sqrt{n}}$ بالحد الأعلى التقريبي لخطأ التقدير، علماً أنّه يمكن استبدال S بـ σ إذا كان σ معروفاً.

مثال (1):

عند دراسة عيّنة حجمها $n = 64$ من مجتمع إحصائي ما ، وجد أنّ:

$\overline{X} = 18.28$ ، $S = 4$. اعتمد هذه العيّنة لتوجد مجال ثقة لتَوقُّع هذا المجتمع على المستوى 1% . (أي: بمعامل ثقة مقداره 99%) .

الحل:

مجال الثّقة المطلوب حسابه هو: $[\overline{X} - t_p \frac{S}{\sqrt{n}} , \overline{X} + t_p \frac{S}{\sqrt{n}}]$

ولحساب t_p نلاحظ أنّ:

$$\Phi(t_p) = 1 - \frac{1}{2}\frac{p}{100} = 1 - \frac{1}{200} = 0.995 \Rightarrow t_p = 2.58$$

وبالتبديل يكون مجال الثّقة المطلوب حسابه هو: [16.99 , 19.57] والفرضيّة مقبولة.

مثال (2):

لنفرض أنّنا نرغب في تقدير متوسط الإنتاج اليومي في شركة للصناعات الكيميائية، حيث سُجّل الإنتاج اليومي لفترة يوم، وكان المتوسط والانحراف المعياري بالأطنان بالشكل: $\overline{X} = 941$ و $S = 23$. اعتمد هذه العيّنة لتوجد مجال ثقة لتوقُّع (لمتوسط الإنتاج اليومي في هذه الشركة على المستوى 5%.

الحل:

أيضاً مجال الثّقة من الشكل: $\overline{X} \mp t_p \frac{S}{\sqrt{n}}$ حيث تُحسَب t_p من العلاقة:

$$\Phi(t_p) = 1 - \frac{1}{2}\frac{p}{100} = 1 - \frac{5}{200} = 0.975 \Rightarrow t_p = 1.96$$

وبالتبديل يكون مجال الثّقة المطلوب هو: $941 \mp 1.96\frac{23}{\sqrt{60}} = 941 \mp 5.81$

وهكذا نقول بمعامل ثقة 95% إنّ التقدير 941 هو في حدود 5.81 طناً زيادة أو نقصاناً من القيمة الحقيقيّة لمتوسط الإنتاج.

(8-18) مجال الثّقة لنسبة (للعيّنات الكبيرة):

لنفرض أنّ صنفاً معيناً δ يوجد في مجتمع إحصائي كبير بنسبة تساوي π . إذا سحبنا عيّنة عشوائيّة حجمها $(n \geq 30)$ من هذا المجتمع الإحصائي، وعرّفنا النجاح بأنّه الحصول على عنصر من الصنف δ ، فإنّ احتمال النجاح عند كل سحب هو، عملياً

π، ونسبة النجاح في العيّنة هي عدد عناصر الصنف δ ولنرمز لها بـ $\frac{X}{n}$ (أي: عدد النجاحات مقسوماً على حجم العيّنة).

وفي أثناء دراستنا لتقريب التوزيع الثنائي بالتوزيع الطبيعي، وجدنا أنّه يمكن اعتبار عدد النجاحات هو مجموع عيّنة، وبالتالي تكون النسبة $\frac{X}{n}$ هي متوسط العيّنة. وكما أنّ تطبيق نظريّة النهاية المركزيّة على X يسمح لنا بالقول إنّ X يتوزع تقريباً وفق التوزيع الطبيعي، أي: $X : N(n\pi\,,\, n^2\pi^2(1-\pi)^2)$، فإنّ تطبيق ذات النظريّة على المتوسط $\frac{X}{n}$ يسمح لنا بالقول إنّ للنسبة $\frac{X}{n}$ توزيعاً مطابقاً تقريباً للتوزيع الطبيعي، أي:

$$\frac{X}{n} : N(\pi\,,\, \frac{\pi(1-\pi)}{n})$$

حيث:

$$E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{n\pi}{n} = \pi$$

$$\sigma^2(\frac{X}{n}) = \frac{1}{n^2}\sigma^2(X) = \frac{n\pi(1-\pi)}{n^2} = \frac{\pi(1-\pi)}{n}$$

وهذا محقّق دوماً شريطة أن يكون n كبيراً (مثلاً أكبر من 30 في حالة قيمة لـ π ليست قريبة من الصفر أو قريبة من الواحد) .

ومن دراسة التوزيع الطبيعي نعلم أنّه إذا كان: $X : N(\mu, \sigma^2)$ فإنّ للعلاقة: $P[\left|\overline{X}-\mu\right| > \lambda_p \frac{\sigma}{\sqrt{n}}] = \frac{p}{100}$ جداول خاصة تفيد في معرفة λ_p عند معرفة p وبالعكس .

ومنه ينتج أنّه من أجل: $\frac{X}{n} : N(\pi\,,\, \frac{\pi^2(1-\pi)^2}{n^2})$، فإنّه يمكن إيجاد قيمة λ_p الموجودة في العلاقة: $P[\left|\frac{X}{n}-\pi\right| > \lambda_p \sqrt{\frac{\pi(1-\pi)}{n}}] = \frac{p}{100}$

والتي يمكن كتابتها بالشكل: $P[\left|\frac{X}{n}-\pi\right| \le \lambda_p \sqrt{\frac{\pi(1-\pi)}{n}}] = 1-\frac{p}{100}$

أو على الشكل: $P[\frac{X}{n}-\lambda_p\sqrt{\frac{\pi(1-\pi)}{n}}\le \pi \le \frac{X}{n}+\lambda_p\sqrt{\frac{\pi(1-\pi)}{n}}]=1-\frac{p}{100}$

وإذا وضعنا $R=\frac{X}{n}$ ، فإنّ مجال الثّقة للنسبة $\frac{X}{n}$ ، على المستوى $p\%$ يكون:

$$[\ R-\lambda_p\sqrt{\frac{\pi(1-\pi)}{n}}\ \ ,\ \ R+\lambda_p\sqrt{\frac{\pi(1-\pi)}{n}}\ ]$$

ولكن وجود π المجهولة في هذه العبارة يمنع من تطبيقها، ويمكن الاستعاضة عن π، نسبة النجاح في المجتمع، بتقدير لها هو R، نسبة النجاح في العيّنة (تماماً كما عوضنا عن σ بـ S سابقاً) ويصبح بالتالي مجال الثّقة للنسبة π بالشكل:

$$[\ R-\lambda_p\sqrt{\frac{R(1-R)}{n}}\ \ ,\ \ R+\lambda_p\sqrt{\frac{R(1-R)}{n}}\ ]$$

ملاحظة: نفترض في تقدير النسب أن المجتمع له دوماً توزيع ثنائي .

مثال:

اختيرت 300 أسرة من بلدة كبيرة فوجد منها 103 أسَر بين أفرادها أُناس كرماء. اعتمد هذه العيّنة لتوجد مجال ثقة لنسبة الأُسَر التي فيها أُناس كُرماء في مجمل البلدة وذلك على مستوى دلالة 5% (أي بعامل ثقة 95%).

الحل:

لدينا: $n=300\ \ ,\ \ R=\frac{X}{n}=\frac{103}{300}=0.41$

والمجال المطلوب هو: $R\pm\lambda_p\sqrt{\frac{R(1-R)}{n}}$

حيث: $\Phi(\lambda_p)=1-\frac{1}{2}\frac{p}{100}=0.975\Rightarrow \lambda_p=1.96$

وبالتبديل يكون مجال الثّقة: $0.41\pm1.96\sqrt{\frac{0.41\ .\ 0.59}{300}}=0.41\pm0.0556$

أي إنَّ R واقعة بين 0.353 و 0.467، أو بمعنى آخر: أنَّ ما بين %35.3 و %46.7 من الأُسَر تضم أناس كُرماء.

....

(8-19) تمارين الفصل الثامن:

1) سُحبت عيّنة حجمها n من المجتمع الطبيعي $X : N(30,\sigma^2)$. فإذا كان:

$E(X^2) = 924$, $\sigma^2(\overline{X}) = 4$. أوجد حجم العيّنة n .

2) إذا كان تابع كثافة المجتمع الإحصائي A من الشكل:

$(x \ , \ \theta > 0) \ ; \ f(x,\theta) = \frac{1}{\theta}.e^{-\frac{x}{\theta}}$. اثبت أنّ $\overline{X}$ المسحوب من عيّنة عشوائيّة حجمها n ، هو مقدّر غير متحيّز للوسيط θ .

3) بفرض لـ X توزيع منتظم في المجال $(0 \ , \ \theta)$ ، أي:

$$f(x) = \begin{cases} \frac{1}{\theta} & ; \ x \in (0,\theta) \\ 0 & ; \ x \notin (0,\theta) \end{cases}$$

. المطلوب : تقدير θ بطريقة العزوم.

4) في توزيع برنولي حيث: $x = 0 \ , 1 \ ; \ f(x,p) = p^x(1-p)^{1-x}$

اثبت أنّ $\overline{X}$ المسحوب من عيّنة عشوائيّة حجمها n ، هو مقدّر مبني على التوقع الأعظمي للوسيط p .

5) أوجد المقدّر المبني على التوقع الأعظمي للوسيط θ في المجتمع الإحصائي الذي تابع كثافته من الشكل: $(0 < x < 1 \ , \ \theta > 0) \ ; \ f(x,\theta) = \theta.x^{\theta-1}$ ، ثم بيّن ما إذا كان المقدّر كافياً أم لا ؟

6) إذا كانت: $(X_1, X_2, ..., X_n)$ هي القيم الملاحظة المسحوبة من توزيع بواسون،

اثبت أنّ: $I_{SA}(\lambda) = \frac{n}{\lambda}$

7) إذا كانت: $(X_1, X_2, ..., X_n)$ هي القيم الملاحظة من مجتمع إحصائي يخضع للتوزيع الأسي بالوسيط $c = \frac{1}{\lambda}$ ، أي أنّ: $x \ , \ \lambda > 0 \ ; \ f(x,\lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$

أوجد المقدّر المبني على التوقع الأعظمي للوسيط λ .

8) أخذت عيّنتان عشوائيّتان حجماهما: $n_1 = 12$ و $n_2 = 15$ من مجتمعين طبيعيين: $N(\mu_1 , \sigma_1{}^2)$ و $N(\mu_2 , \sigma_2{}^2)$ على الترتيب. ووجد أنّ: $\hat{S}_1^2 = 20$ و $\hat{S}_2^2 = 25$ والمطلوب:

a. أوجد مجال ثقة لتباين المجتمع الأول على المستوى 5%.

b. أوجد مجال ثقة للفرق بين الوسيطين $\mu_1 - \mu_2$ على المستوى 5%.

c. أوجد ثقة 10% للنسبة σ_2^2/σ_1^2 بفرض أنّ التوزيعين مستقلان.

9) في دراسة من عيّنة عشوائيّة حجمها 100 طالب من طلاب إحدى الجامعات وجد أنّ عدد الطلاب الذين يستعملون يدهم اليسرى هو 15 . أوجد فترة ثقة لنسبة الطلاب في الجامعة الذين يستعملون يدهم اليسرى.

10) أخذت عيّنة عشوائيّة حجمها 25 عاملاً من أحد المصانع وكان الوسط الحسابي لأعمار العمال بالعيّنة 34.5 سنة بانحراف معياري 4.8 سنة.

المطلوب:

a. أوجد مجال ثقة 1% للوسط الحسابي للأعمار بالمصنع.

b. أوجد مجال ثقة 1% لتباين الأعمار.

c. إذا كان عدد العمال المدخنين هو 10عمال ، أوجد مجال ثقة 1% لنسبة المدخنين في المصنع.

11) أخذت عيّنتان عشوائيّتان من توزيعين طبيعيين لهما التباين نفسه ، ووجد أنّه:

للعيّنة الأولى: $n_1 = 10 \;,\; \overline{x_1} = 60 \;,\; \hat{S}_1^2 = 80$

وللعيّنة الثانية: $n_2 = 8 \;,\; \overline{x_2} = 54 \;,\; \hat{S}_2^2 = 90$

أوجد مجال ثقة للفرق بين الوسيطين $\mu_1 - \mu_2$ على المستوى 5%.

....

جداول إحصائيَّة

n=2						p							
x	**0.01**	**0.05**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**0.95**	**0.99**	**x**
0	0.980	0.902	0.810	0.640	0.490	0.360	0.250	0.160	0.090	0.040	0.010	0.002	0.000	0
1	1.000	0.998	0.990	0.960	0.910	0.840	0.750	0.640	0.510	0.360	0.190	0.098	0.020	1
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2

جدول (1) : التوزيع الثنائي: يعطي $P(x \leq k)$

n=3						p							
x	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	x
0	0.970	0.857	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	0.000	0.000	0
1	1.000	0.993	0.972	0.896	0.784	0.648	0.500	0.352	0.216	0.104	0.028	0.007	0.000	1
2	1.000	1.000	0.999	0.992	0.973	0.936	0.875	0.784	0.657	0.488	0.271	0.143	0.030	2
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3

n=4						p							
x	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	x
0	0.961	0.815	0.656	0.410	0.240	0.130	0.062	0.026	0.008	0.002	0.000	0.000	0.000	0
1	0.999	0.986	0.948	0.819	0.652	0.475	0.312	0.179	0.084	0.027	0.004	0.000	0.000	1
2	1.000	1.000	0.996	0.973	0.916	0.821	0.688	0.525	0.348	0.181	0.052	0.014	0.001	2
3	1.000	1.000	1.000	0.998	0.992	0.974	0.938	0.870	0.760	0.590	0.344	0.185	0.039	3
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	4

تابع جدول (1) :

n=5						P							
x	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	x
0	0.951	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000	0.000	0
1	0.999	0.977	0.919	0.737	0.528	0.337	0.188	0.087	0.031	0.007	0.000	0.000	0.000	1
2	1.000	0.999	0.991	0.942	0.837	0.683	0.500	0.317	0.163	0.058	0.009	0.001	0.000	2
3	1.000	1.000	1.000	0.993	0.969	0.913	0.812	0.663	0.472	0.263	0.081	0.023	0.001	3
4	1.000	1.000	1.000	1.000	0.998	0.990	0.969	0.922	0.832	0.672	0.410	0.226	0.049	4
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5

n=6 P

x	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	x
0	0.941	0.735	0.531	0.262	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000	0.000	0
1	0.999	0.967	0.886	0.655	0.42	0.233	0.109	0.041	0.011	0.002	0.000	0.000	0.000	1
2	1	0.998	0.984	0.901	0.744	0.544	0.344	0.179	0.07	0.017	0.001	0.000	0.000	2
3	1	1	0.999	0.983	0.93	0.821	0.656	0.456	0.256	0.099	0.016	0.002	0.000	3
4	1	1	1	0.998	0.989	0.959	0.891	0.767	0.58	0.345	0.114	0.033	0.001	4
5	1	1	1	1	0.999	0.996	0.984	0.953	0.882	0.738	0.469	0.265	0.059	5
6	1	1	1	1	1	1	1	1	1	1	1	1	1	6

n=7

P

x	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	x
0	0.932	0.698	0.478	0.21	0.082	0.008	0.002	0	0	0	0	0	0	0
1	0.998	0.956	0.85	0.577	0.329	0.159	0.062	0.019	0.004	0	0	0	0	1
2	1	0.996	0.974	0.852	0.647	0.42	0.227	0.096	0.029	0.005	0	0	0	2
3	1	1	0.997	0.967	0.874	0.71	0.5	0.29	0.126	0.033	0.003	0	0	3
4	1	1	1	0.995	0.971	0.904	0.773	0.58	0.353	0.148	0.026	0.004	0	4
5	1	1	1	1	0.996	0.981	0.938	0.841	0.671	0.423	0.15	0.044	0	5
6	1	1	1	1	1	0.998	0.992	0.972	0.918	0.79	0.522	0.302	0.07	6
7	1	1	1	1	1	1	1	1	1	1	1	1	1	7

جدول (2): توزيع بواسون

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.5
0	0.905	0.819	0.741	0.67	0.607	0.549	0.497	0.449	0.407	0.368	0.223
1	0.995	0.982	0.963	0.938	0.91	0.878	0.844	0.809	0.772	0.736	0.558
2	1	0.999	0.996	0.992	0.986	0.977	0.966	0.953	0.937	0.92	0.809
3		1	1	0.999	0.998	0.997	0.994	0.991	0.987	0.981	0.934
4				1	1	1	0.999	0.999	998	0.996	0.981
5							1	1	1	0.999	0.996
6										1	0.999
7											1

x	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7
0	0.135	0.082	0.055	0.033	0.018	0.011	0.007	0.004	0.003	0.002	0.001
1	0.406	0.287	0.199	0.136	0.092	0.061	0.04	0.027	0.017	0.011	0.007
2	0.677	0.544	0.423	0.321	0.238	0.174	0.125	0.088	0.062	0.043	0.03
3	0.857	0.758	0.647	0.537	0.433	0.342	0.265	0.202	0.151	0.112	0.082
4	0.947	0.891	0.815	0.725	0.629	0.532	0.44	0.358	0.285	0.224	0.173
5	0.983	0.958	0.916	0.858	0.785	0.703	0.616	0.529	0.446	0.369	0.301
6	0.995	0.986	0.966	0.935	0.889	0.831	0.762	0.686	0.606	0.563	0.45
7	0.999	0.996	0.988	0.973	0.949	0.913	0.867	0.809	0.744	0.673	0.599
8	1	0.999	0.996	0.99	0.979	0.96	0.932	0.894	0.847	0.792	0.729
9		1	0.999	0.997	0.992	0.983	0.968	0.946	0.916	0.877	0.83
10			1	0.999	0.997	0.993	0.986	0.975	0.957	0.933	0.901
11				1	0.999	0.998	0.995	0.989	0.98	0.966	0.947

		0.99	0.99	0.99	0.99	0.98	0.97
12	1	9	8	6	1	4	3
			0.99	0.99	0.99	0.99	0.98
13		1	9	8	6	3	7
				0.99	0.99	0.99	0.99
14			1	9	9	7	4
					0.99	0.99	0.99
15				1	9	9	8
							0.99
16					1	1	9
17							1

تابع توزیع بواسون:

x	7.5	8	8.5	9	9.5	10	12	15	20
0	0.001	0	0	0	0	0	0	0	0
1	0.005	0.003	0.002	0.001	0.001	0	0	0	0
2	0.02	0.014	0.009	0.006	0.004	0.003	0.001	0	0
3	0.059	0.042	0.03	0.021	0.015	0.01	0.002	0	0
4	0.132	0.01	0.074	0.055	0.04	0.029	0.008	0.001	0
5	0.241	0.191	0.15	0.116	0.089	0.067	0.02	0.003	0
6	0.378	0.313	0.256	0.207	0.165	0.13	0.046	0.008	0
7	0.525	0.453	0.386	0.324	0.269	0.22	0.09	0.018	0.001
8	0.662	0.593	0.523	0.456	0.392	0.333	0.155	0.037	0.002
9	0.776	0.717	0.653	0.587	0.522	0.458	0.242	0.07	0.005
10	0.862	0.816	0.763	0.706	0.645	0.583	0.347	0.118	0.011
11	0.921	0.888	0.849	0.803	0.752	0.697	0.462	0.185	0.021
12	0.957	0.936	0.909	0.876	0.836	0.792	0.576	0.268	0.039
13	0.978	0.966	0.949	0.926	0.898	0.864	0.682	0.363	0.066
14	0.99	0.983	0.973	0.959	0.94	0.917	0.772	0.466	0.105
15	0.995	0.992	0.986	0.978	0.967	0.951	0.844	0.568	0.157
16	0.998	0.996	0.993	0.989	0.982	0.973	0.899	0.664	0.221
17	0.999	0.998	0.997	0.995	0.991	0.986	0.937	0.749	0.297
18	1	0.999	0.999	0.998	0.996	0.993	0.963	0.819	0.381
19		1	0.999	0.999	0.998	0.997	0.979	0.875	0.47
20			1	1	0.999	0.998	0.988	0.917	0.559
21					1	0.999	0.994	0.947	0.644

22	1	0.997	0.967	0.721
23		0.999	0.981	0.787
24		0.999	0.989	0.843
25		1	0.994	0.888
26			0.997	0.922
27			0.998	0.948
28			0.999	0.966
29			0.999	0.978
30			1	0.987
31				0.992
32				0.995
33				0.997
34				0.999
35				0.999
36				1

جدول (3) : التوزيع الطبيعي المعياري $P(X \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706

1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

تابع : التوزيع الطبيعي المعياري :

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0059	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0078	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.001	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.001	0.001
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.002	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.003	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.004	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.006	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.008	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.011
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.015	0.0146	0.0143
-2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0427	0.0268	0.0262	0.0256	0.025	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0274	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455

-1.5	0.0668	0.0655	0.0643	0.063	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0722	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.102	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.123	0.121	0.1190	0.117
-1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1772	0.1772	0.1711	0.1685	0.166	0.1635	0.1611
-0.8	0.2119	0.209	0.2061	0.2033	0.2033	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.242	0.2389	0.2358	0.2327	0.2327	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2643	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.305	0.3015	0.2981	0.2981	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3336	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3707	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.409	0.4090	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4483	0.4404	0.4364	0.4325	0.4286	0.4247
0	0.5	0.496	0.492	0.488	0.4880	0.4801	0.4761	0.4721	0.4681	0.4641

دليل المصطلحات العلميّة

انكليزي – عربي

A

Addition law	قانون الجمع
Approximation	تقريب
Arithmetic mean	متوسط حسابي
Average	المتوسط
Axioms of probability	مسلمات الاحتمال
Alternative hypothesis	الفرض البديل

B

Beta distribution	توزيع بيتا
Best unbiased estimator	أفضل مقدّر غير متحيّز
Binomial distribution	توزيع ذي الحدين (ثنائي)
Bayes rule	دستور بايز
Bernoulli distribution	توزيع برونولي

C

Central limit theorem	نظريّة النهاية المركزية
Certain theorem	حدث مؤكد
Chi_ Square distribution	توزيع كاي مربع
Conditional expectation	توقع شرطي
Conditional probability	احتمال شرطي
Continues random variable	متغير عشوائي مستمر (متصل)
Chebyshev s inequality	متراجحة تشيبتشيف
Correspondence	تناظر
Covariance	تغاير
Coefficient	معامل
Characteristic function	الدالة المميزة
Consistency	الاتساق

Constant	ثابت
Complement	متمم (مكمل)
Combination	متوافقات (توافق)
Confidence coefficient	معامل الثّقة
interval	مجال الثّقة
level	مستوى (درجة) الثّقة
Correlation	ارتباط
Correlation Coefficient	معامل ارتباط
Counting principle	مبدأ العد
Curve	منحني

D

Deviation mean	متوسط الانحراف
Data	بيانات
Deciles	عشيرات
Density function	دالة الكثافة
Description statistics	إحصاء وصفي
Data	بيانات
Density function	تابع الكثافة
Derivation	اشتقاق
Discrete random variable	متغير عشوائي منفصل
Distribution function	تابع التوزيع
Degrees of freedom	درجات الحرية
Disjoint events	حوادث منفصلة
Decision	قرار

E

Event	حدث (حادثة)
Element	عنصر
Expectation	توقع
Efficiency	الكفاءة
Experiment of chance	تجربة عشوائية

Elementary event	حدث أولي
Events mutually exclusive	أحداث متنافية مثنى- مثنى
Expectation mathematical	توقع رياضي
Experiment	تجربة
Estimate	تقدير
Exponential distribution	التوزيع الأسي
Experiment	تجربة

F

Function	تابع (دالة)
Formula	صيغة رياضية
Factorial n	n عاملي (مضروب)
Factorization theorem	نظريّة التحليل المعاملي
Finite sample space	فراغ عيّنه منته
Frequency	تكرار
curve	منحني التكرار
distribution	توزيع التكرار
histogram	مدرج التكرار
polygon	مضلع التكرار
table	جدول التكرار

G

Gamma distribution	توزيع جاما
Generating	مولدة
Geometric distribution	توزيع هندسي
Graphic	بياني
Graphic presentation	تمثيل بياني

H

Histogram	مدرج
Hypergeometrice distribution	التوزيع فوق الهندسي

I

Intersection	تقاطع
Iterative methods	

	طرق التكرار (التجريب)
Interval estimation	التقدير بمجال
Impossible event	حدث مستحيل
Independent	مستقل
Infinite population	مجتمع لانهائي
Inequality	متباينة (متراجحة)
Information function	دالة المعلومات

J

Joint moments	عزوم مشتركة
Joint probability function	دالة الاحتمال المشتركة

L

Laplace distribution	توزيع لابلاس
Law of large numbers	قانون الأعداد الكبيرة
Likelihood function	دالة الإمكان

M

Marginal	هامشي
Marginal probfun	دالة الاحتمال الهامشية
Mathematical expectation	توقع رياضي
Mean square error	متوسط مربعات الخطأ
Measures of central tendency	مقاييس النزعة المركزية
Median	الوسيط
Mode	منوال
Model	انموذج
Moment generating function	الدالة المولدة للعزوم
Multiplication law	قانون الضرب

N

Negative binomial distribution	توزيع ذي الحدين السالب
Normal distribution	توزيع طبيعي

O

Odd - function	تابع فردي

One-to-one	أحادي
Observation	مشاهده، ملاحظة، قياس
Operation	عملية
Original	أصلي
Outcome	نتيجة
Ordinal data	بيانات ترتيبية

P

Poisson distribution	توزيع بواسون
Pascal distribution	توزيع باسكال
Population	المجتمع
Probability density function	دالة كثافة الاحتمال
Permutation	المتبادلات
Point estimation	تقدير نقطي
Properties	خواص
Proportion	نسبة
Parameter	وسيط (معلمة)
Partition	تجزئة
Percentage	نسبة مئوية
Percentiles	المئينات
Presentation of data	عرض البيانات

R

Random	عشوائي
Random experiment	تجربة عشوائية
Relative frequency	تكرار نسبي
Relation	علاقة
Random sample	عينة عشوائية
Ratio	نسبة
Real numbers	أعداد حقيقية

S

Sample mean	متوسط العينة

Sample space	فراغ العيّنة
Sampling	معاينة
Statistical distribution	توزيع إحصائي
Statistical inference	استدلال إحصائي
Set	مجموعة
Subset	مجموعة جزئية
Scatter diagram	مخطط الانتشار
Sampling with replacement	سحب عيّنة مع إعادة
Sampling without replacement	سحب عيّنة بدون إعادة
Some probability distribution	بعض التوزيعات الاحتماليّة
Sufficiency	الكفاية
Standard deviation	انحراف معياري
Standard normal distribution	التوزيع الطبيعي المعياري

T

t- distribution	توزيع ت
Transformation	تحويل

U

Uniformally distributed random variables	متغيرات عشوائية موزعة بانتظام
Uniform distribution	التوزيع المنتظم
Upper limit	الحد الأعلى

V

Variance	تباين
Variable(variate)	متغير
Venn-diagram	مخطط فن

المراجع العربية والأجنبية:

المراجع العلميّة

العربية :

1- د. أنيس كنجو : الإحصاء الرياضي- جامعة دمشق، 1978-1979 م.

2- د. أنيس كنجو : الإحصاء والاحتمال - جامعة الملك سعود، 1993م.

3- د. محمد خير أحمد: مبادئ الإحصاء النظري- جامعة تشرين، 1410هـ-1990م.

4- د.محمد جنيد العمر : الإحصاء النظري- جامعة حلب، 1411هـ-1990م.

5- د./الحسن + فرح + العمر/ نظريّة الاحتمالات- جامعة حلب، 1412هـ-1991م.

6- د. أحمد عوده: الاستدلال الإحصائي - جامعة الملك سعود، 1991م.

الأجنبية:

1- A.A Sveshenkov - Exercises theory of probability and mathematical statistics -Moscow 1970.
2- A.A Borovkov - theory of probability - Moscow 1976.
3- K.E.Evchenko - Mathematical statistical - Moscow 1984.
4- V.C.Karayk - theory of probability and mathematical statistics Keev 1978.

SYRIAN ARAB REPUBLIC
High Education Ministry
Tishreen university
Faculty of science

Principles in Probability and Statistics

Dr. Mubarak Dib

1429 - 1430
2008 - 2009

سعر المبيع للطالب (260) ل.س