جامعة دمشق

المعهد العالي للتنمية الإدارية

سلسلة محاضرات في الطرق الكمية الإحصائية في الإدارة

البرنامج: ماجستير في الأعمال الدولية (بحثي)

إسم المقرر: الطرق الكمية و التطبيقات المعلوماتية في الإدارة

العام الدراسي: 2008 – 2009

أستاذ المقرر: د. معاذ الشرفاوي الجزائرلي

ملاحظة: هذا الملف هو عبارة عن مجموعة من شرائح "باور بوينت" تم نقلها إلى العربية من كتاب:
Groebner et. al. Business Statistics: A Decision-Making Approach. 6th ed.

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الأولى /القسم النظري/

جمع البيانات: من الألف إلى الياء!

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

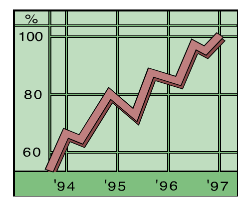
مخرجات المحاضرة الأولى:

تأكد بنهاية هذه المحاضرة أنك تعلمت:

- وصف الطرائق الأساسية للمعاينة.
- التمييز بين الأزواج التالية من الاصطلاحات الإحصائية:
  - العينة و المجتمع (الإحصائي).
  - البيانات الأولية و البيانات الثانوية.
  - البيانات الكمية و البيانات النوعية.
  - بيانات السلاسل الزمنية و البيانات المقطعية.
- تمييز الإحصاء الوصفي عن الإحصاء الاستدلالي.
- أنواع البيانات و طرائق جمعها.

أدواتنا كإحصائيين:

- **الإحصاء الوصفي:**
  - **جمع ⇦ عرض ⇦ وصف**



$$\frac{\sum x_i}{n}$$

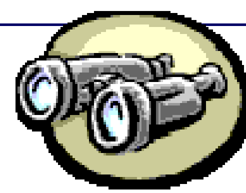
- **الإحصاء الاستدلالي:**
  - **معلومات من العينة ⇦ استنتاجات عن المجتمع.**

مصادر البيانات

جمع البيانات الأولية

- مسح

- مشاهدة

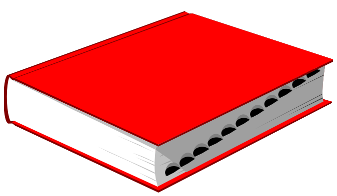


- تجارب

جمع البيانات الثانوية

بيانات منشورة:

مطبوعة / ألكترونية



خطوات تصميم الاستبيان

- حدد المشكلة بوضوح.
- حدد المجتمع المستهدف.
- قم بصياغة أسئلة الاستبيان.
- اختبر الاستبيان قبل تطبيقه.
- حدد حجم العينة و طريقة المعاينة.
- قم باختيار العينة و نفذ المسح ميدانياً.

أنواع الأسئلة:

- **الأسئلة المغلقة.**
  - الاختيار من قائمة من الخيارات المعرفة سلفاً.
- **الأسئلة المفتوحة.**
  - نمط حر في الإجابة تقبل في أية عبارة / قيمة / ...
- **الديموغرافيات:**
  - عن الخصائص الشخصية للمستقصى عنه.

المجتمعات و العينات

- **المجتمع** هو كافة العناصر أو الأفراد محل الاهتمام

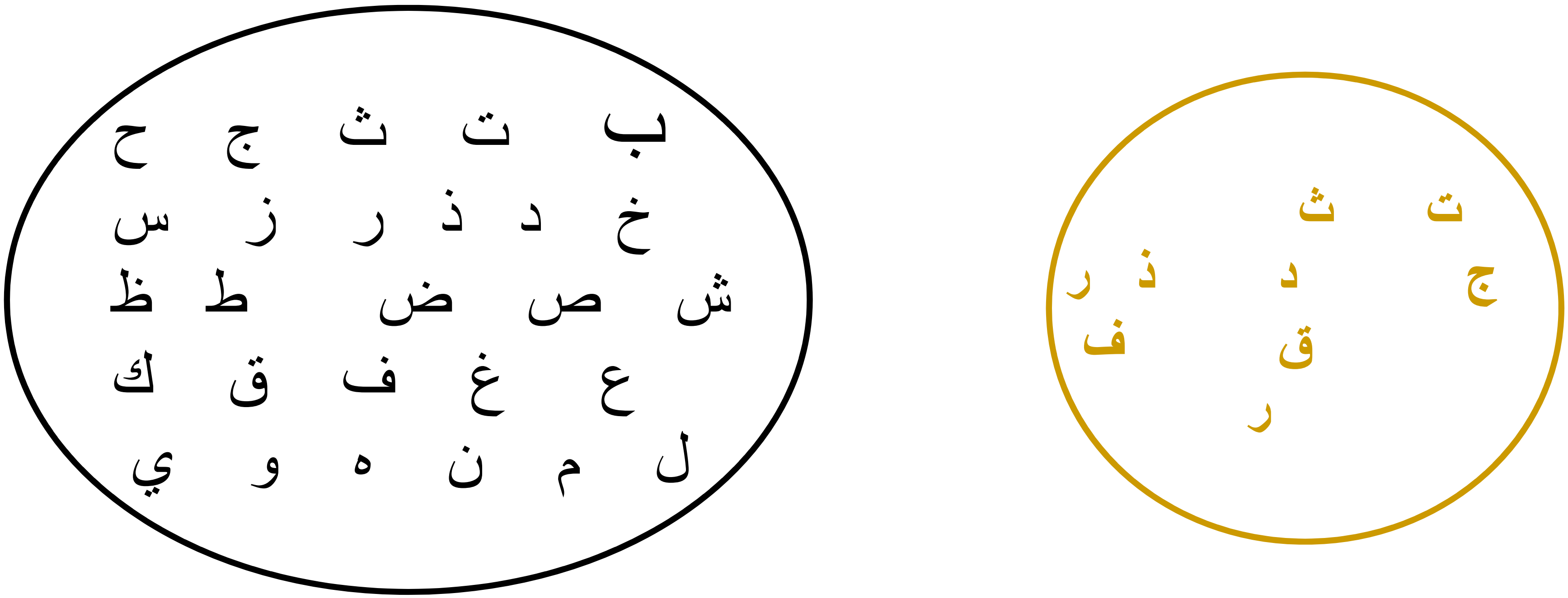
مثال :	كافة الناخبين المحتملين في الانتخابات القادمة. كافة القطع المنتجة في الشهر الماضي.

- **العينة** هي مجموعة جزئية من المجتمع.

مثال :	ألف ناخب تم اختيارهم عشوائياً لإجراء مقابلات لاستطلاع الرأي. أربعين قطعة من إنتاج الشهر الماضي تم سحبها لإجراء الاختبارات.

المجتمع و العينة

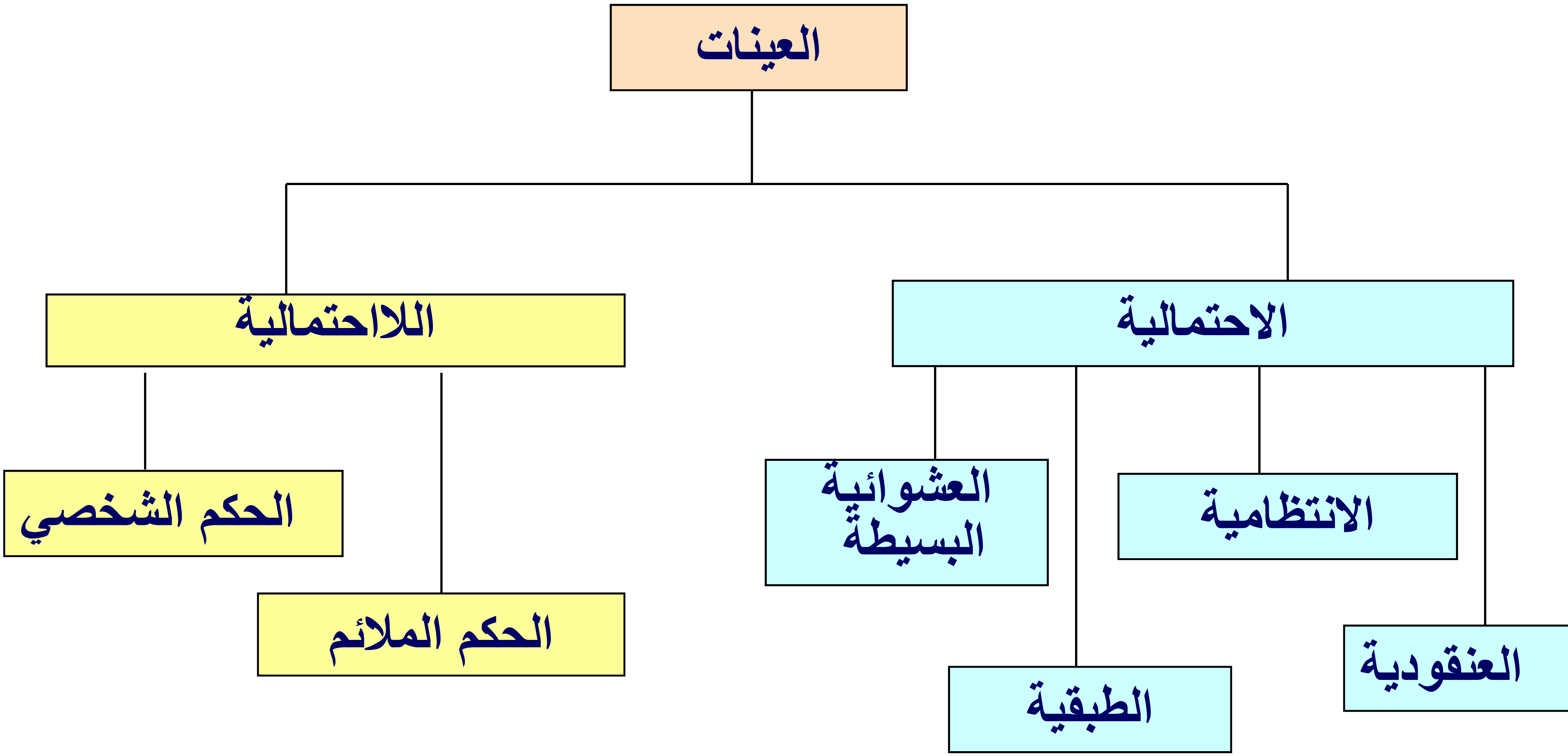
العينة **المجتمع**



لماذا نسحب عينة؟

- **الوقت:** هل لديك الوقت الكافي لقياس كافة مفردات المجتمع؟
- **التكلفة:** هل لديك الميزانية اللازمة لمسح المجتمع بأكمله؟
- **الدقة:** هل من داعٍ لمسح المجتمع إذا كانت العينة تعطي تقديرات مجتمعية دقيقة إلى حد مرضٍ؟

تقنيات المعاينة



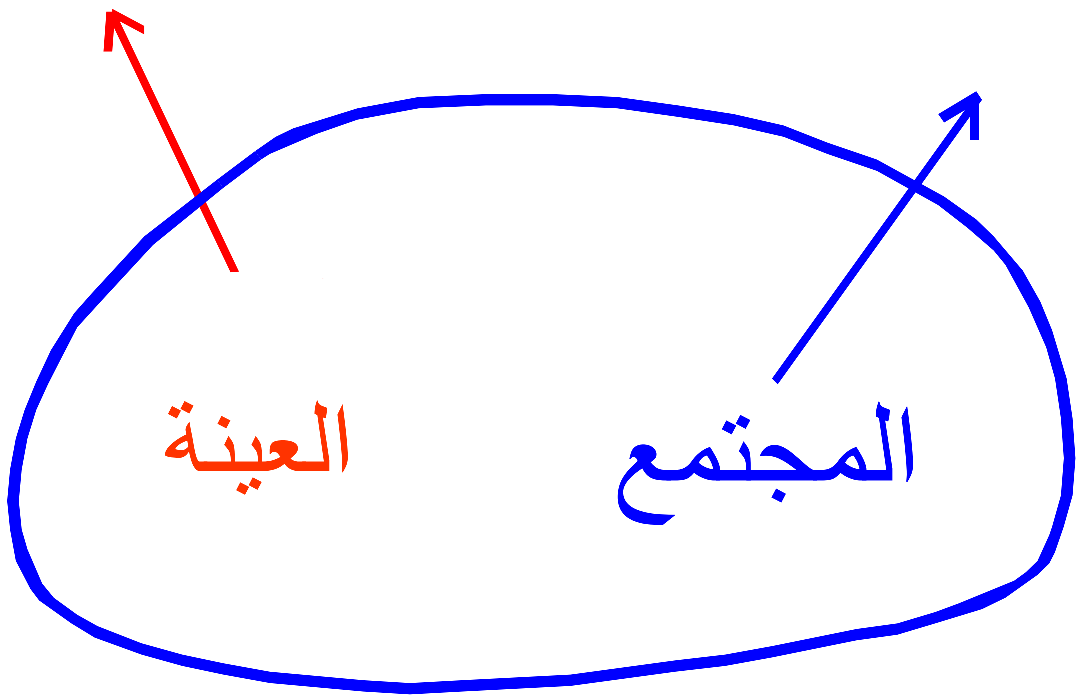
الإحصاء الاستدلالي

- فحص نتائج المعاينة بغرض الوصول لاستنتاجات حول المجتمع

بارامترات المجتمع ← إحصاءات العينة

(مجهولة و لكن بالإمكان تقديرها) (معلومة)

استدلال

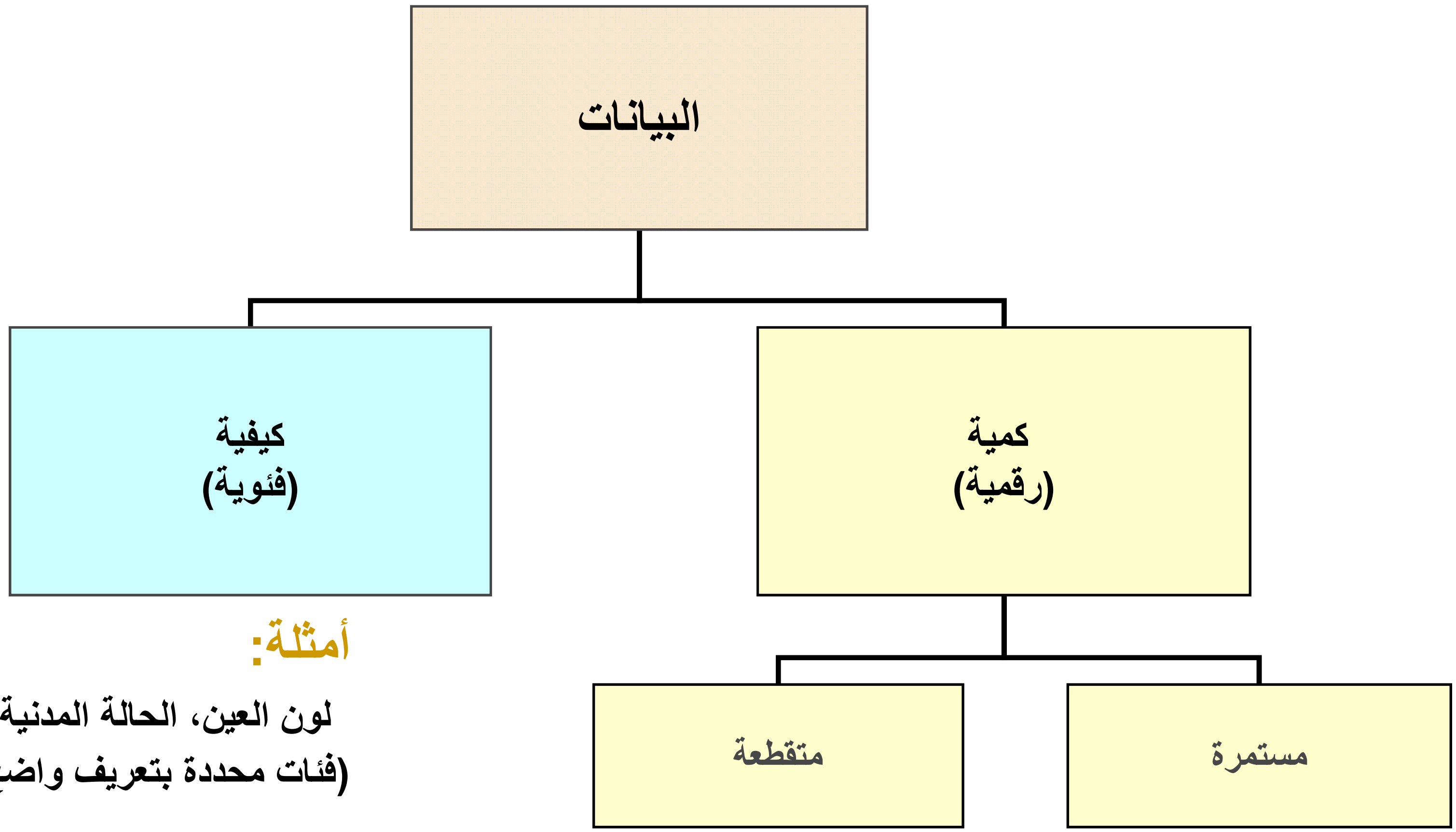


الإحصاء الاستدلالي

الوصول لاستنتاجات و / أو صنع قرارات بخصوص المجتمع بناءً على النتائج المستخلصة من العينة.

- **التقدير:**
  - تقدير وسطي وزن طلاب المرحلة الإعدادية من الذكور على مستوى محافظة دمشق بأكملها.
- **اختبار الفرضية:**
  - استخدام قرائن من العينة المسحوبة عشوائياً من طلاب الإعدادية (خمسمائة طالب من مدارس دمشق) لاختبار الزعم القائل بأن متوسط الوزن لا يزيد عن 55كغ.

أنواع البيانات: الكمية و الكيفية.



أمثلة:

الوزن، الفولت، الطول

(خصائص للقياس)

أمثلة:

عدد الأطفال، عدد الأعطال في اليوم

(أشياء للعد)

أمثلة:

لون العين، الحالة المدنية.

(فئات محددة بتعريف واضح)

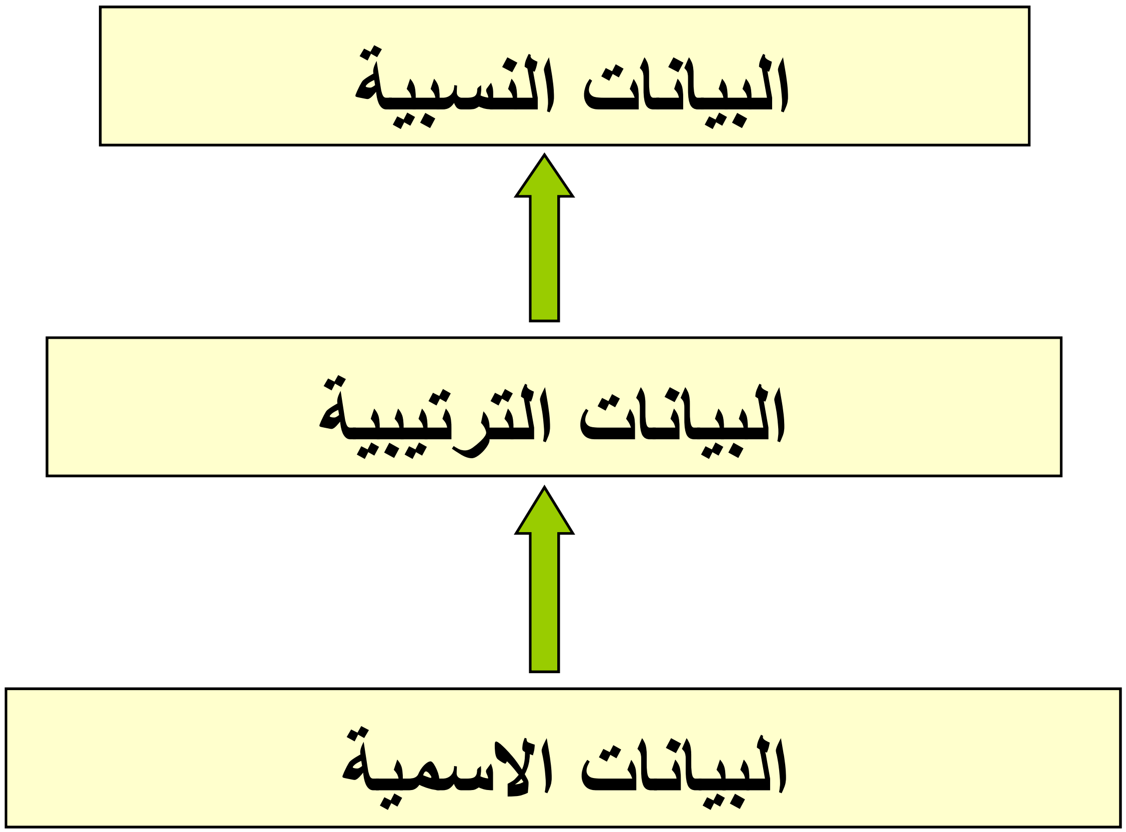
أنواع البيانات: الزمنية و المقطعية.

	حجم الإنتاج الموزع على المحافظات (ملايين الليرات السورية)			
	2004	2005	2006	2007
دمشق	490	475	460	435
حلب	395	375	345	320
المنطقة الوسطى	395	410	390	405
المنطقة الساحلية	280	285	270	260

بيانات سلسلة زمنية

بيانات مقطعية

مستويات قياس البيانات



ملخص المحاضرة

الطرائق الأساسية للمعاينة.

- أهم الاصطلاحات الإحصائية:
  - العينة و المجتمع (الإحصائي).
  - البيانات الأولية و البيانات الثانوية.
  - البيانات الكمية و البيانات النوعية.
  - بيانات السلاسل الزمنية و البيانات المقطعية.
- الإحصاء الوصفي و الإحصاء الاستدلالي.
- أنواع البيانات و طرائق جمع أكثرها شيوعاً.

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الثانية /القسم النظري/

وصف البيانات بالمخططات و الجداول

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

مخرجات المحاضرة الثانية:

تأكد بنهاية هذه المحاضرة أنك تستطيع:

- بناء التوزيعات التكرارية يدوياً و بمساعد الحاسب.
- بناء و تفسير هستوغرام يدوياً و بمساعد الحاسب.
- إنشاء و تفسير مخططات الكعكة.
- إنشاء و تفسير مخططات الأعمدة بأنواعها المختلفة.
- إنشاء و تفسير مخططات الانتشار و التبعثر.
- تفسير مخططات أخرى (باريتو، الوريقة، إلخ)
- اختيار المخطط المناسب للبيانات و الظاهرة محل الاهتمام.

التوزيعات التكرارية

ما هي؟

هي عبارة عن **قائمة/جدول** تحتوي **قيم متغير** أو مجموعة من الفئات التي تقع فيها البيانات، مع **التكرارات الموافقة** لقيم المتغير أو للفئات التي تقع داخلها البيانات.

- لم نستخدمها؟
  - تلخص البيانات.
  - تكثف البيانات الخام في صيغ أكثر فائدة.
  - تساعد في الوصول إلى تفسير مرئي يوفر الوقت و الجهد.

التوزيعات التكراري: البيانات المتقطعة

- هل تذكر ما هي البيانات المتقطعة؟ (المحاضرة الأولى!)

مثال:

قامت جريدة يومية تصدر في هونولولو بسؤال مائتي مواطن عن عدد المرات التي يقوم القارئ فيها بشراء الجريدة في الأسبوع، و قد تم تلخيص نتائج الاستبيان في الجدول إلى اليسار.

عدد الأيام	التكرارات
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
الإجمالي	**200**

التكرار النسبي

تابع المثال السابق:

$$\frac{44}{200} = .22$$

التكرار النسبي	التكرار	عدد الأيام
.22	44	0
.12	24	1
.09	18	2
.08	16	3
.10	20	4
.11	22	5
.13	26	6
.15	30	7
1.00	**200**	**Total**

22% من القراء المشمولين بالعينة أفادوا بأنهم لا يقرؤون الجريدة و لا حتى مرة واحدة في الأسبوع.

التوزيعات التكراري: البيانات المستمرة

هل تذكر ما هي البيانات المستمرة؟ (المحاضرة الأولى!)

هل تعد قياسات درجات الحرارة من البيانات المستمرة؟

مثال:

قامت شركة تستورد المكيفات بقياس درجة الحرارة في عشرين يوماً صيفياً فحصلت على :

24, 35, 27, 21, 24, 37, 26, 43, 44, 30,
32, 23, 22, 38, 41, 43, 44, 27, 43, 27

تجميع البيانات في فئات

رتب بيانات الحرارة ترتيباً تصاعدياً:

22, 22, 23, 27, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 42, 43, 43, 44, 44

- أوجد المدى: 44 – 12 =32
- اختر عدد الفئات: خمسة**!** (عادة من 5 إلى عشرين فئة)
- أحسب عرض الفئة: 44 / 5 = 10 (بعد التقريب نحو الأعلى)
- حدد حدود الفئات: 20، 30، 40، 50.
- أحسب منتصف كل فئة: 25، 35، 45
- قم بعد المشاهدات و تسجيلها من أجل كل فئة.

مثال على التوزيعات التكرارية

رتب البيانات تصاعدياً

22, 22, 23, 27, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 42, 43, 43, 44, 44

التوزيع التكراري

الفئة	التكرار	التكرار النسبي
تساوي/أكبر من 20 و أصغر من 30	9	.45
تساوي/أكبر من 30 و أصغر من 40	5	.25
تساوي/أكبر من 40 و أصغر من 50	6	.30
الإجمالي	20	1.00

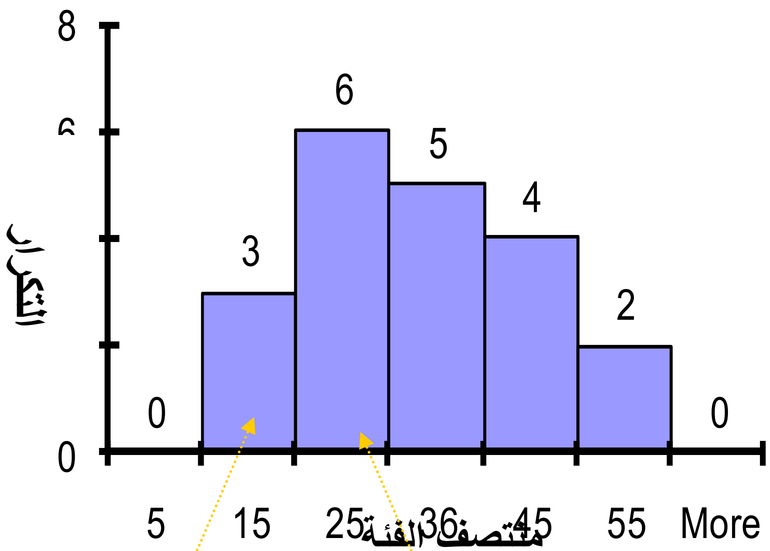
الـهيستوغرام

إذا كان لديك مخططاً بالمواصفات التالية:

- الفئات على المحور الأفقي.
- **التكرارات على المحور العمودي**
- أطوال الأعمدة تمثل عدد المشاهدات في الفئات.

فالمخطط الذي لديك هو **"هيستوغرام"**

هيستوغرام بعلامات عشرين طالباً في مادة الإحصاء!



مثال: علامات الطلاب في مادة الإحصاء!

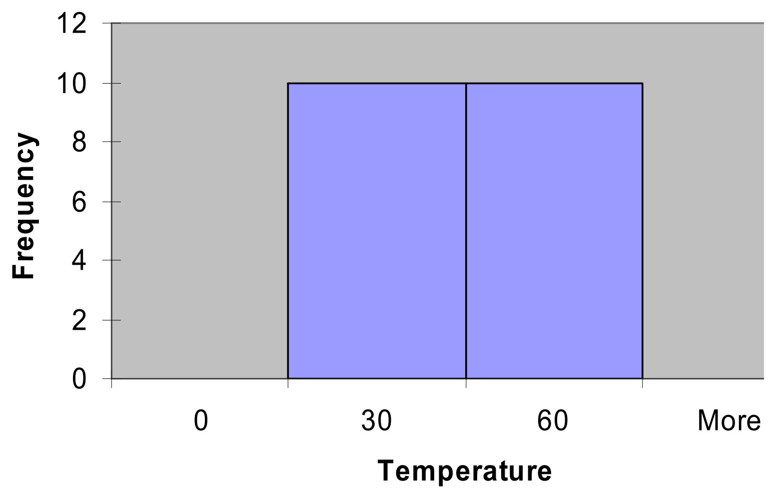
رتب علامات الطلاب تصاعديا ثم قسمهم إلى فئات

22, 22, 23, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

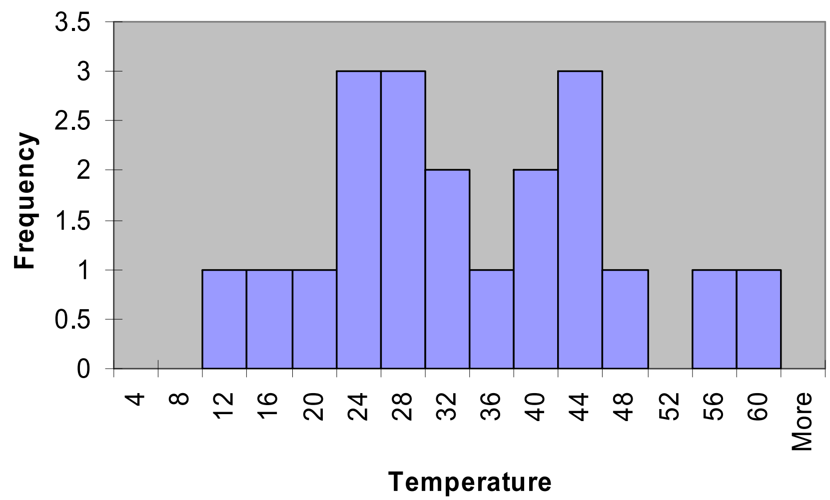
أسئلة عن تجميع البيانات في فئات:

1. كم هو عرض الفئة؟ كيف نحدده؟
   1. (بالتالي، كم هو عدد الفئات الواجب إنشاؤها)
2. كيف نحدد النقاط الحدية للفئة؟

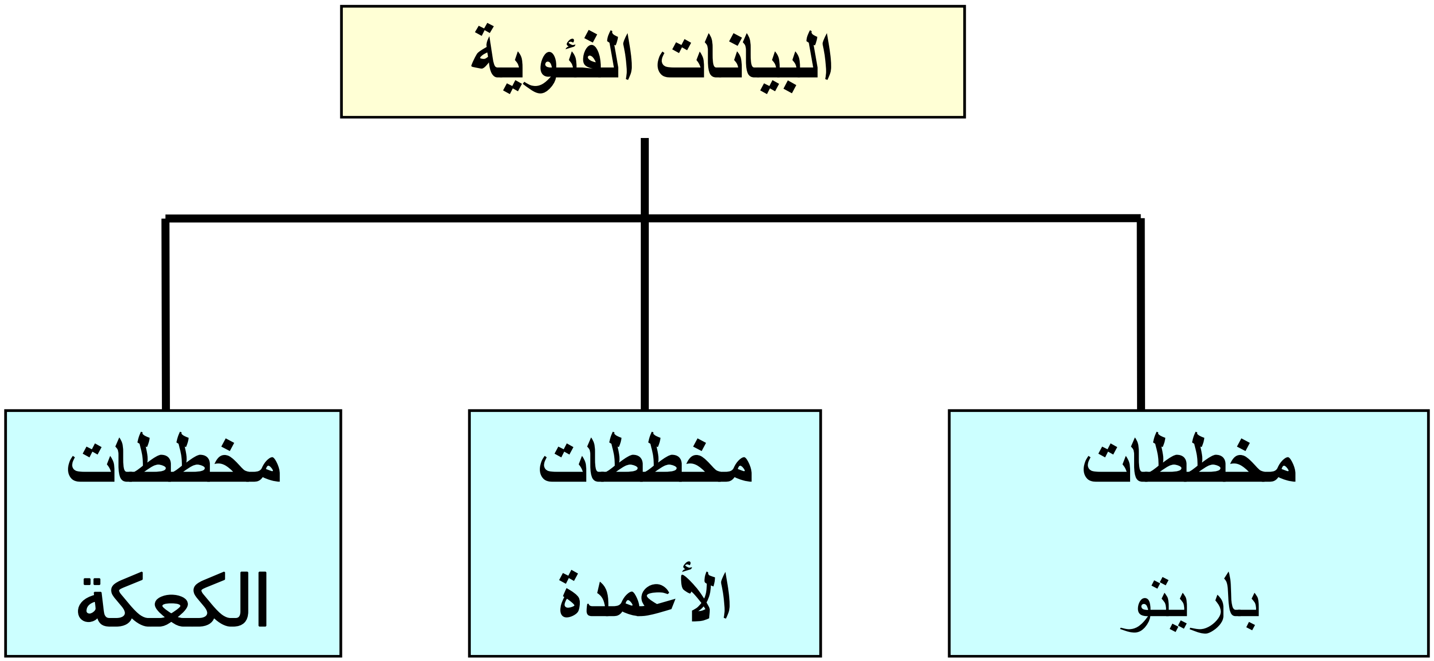
أم هذه؟



هذه؟

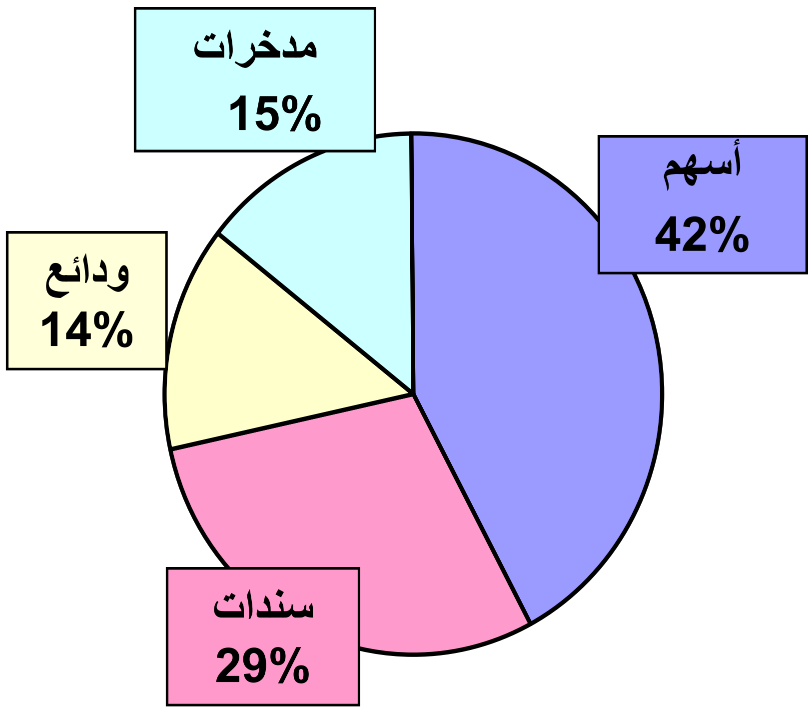


التمثيل البياني للبيانات الفئوية

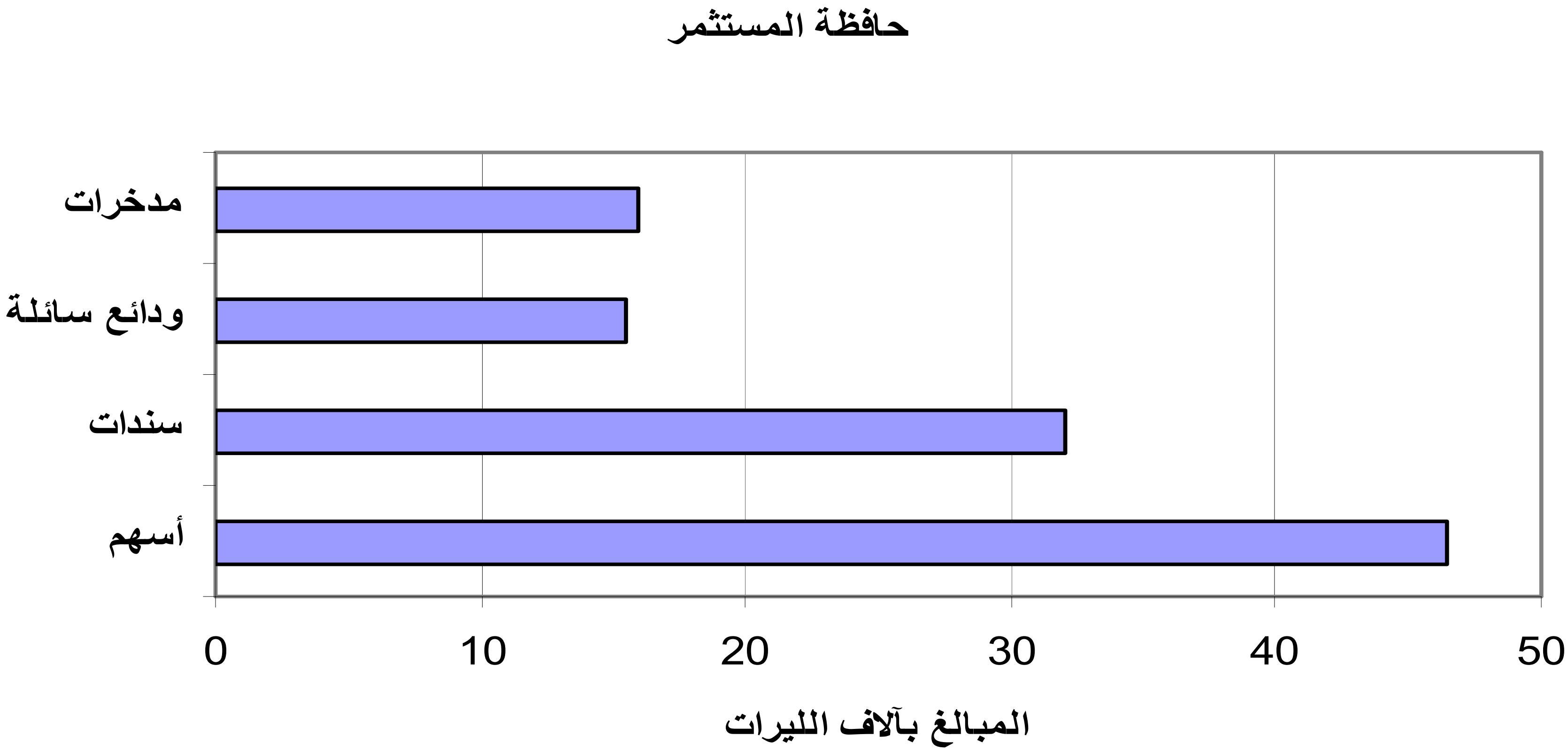


مثال: مخطط كعكة

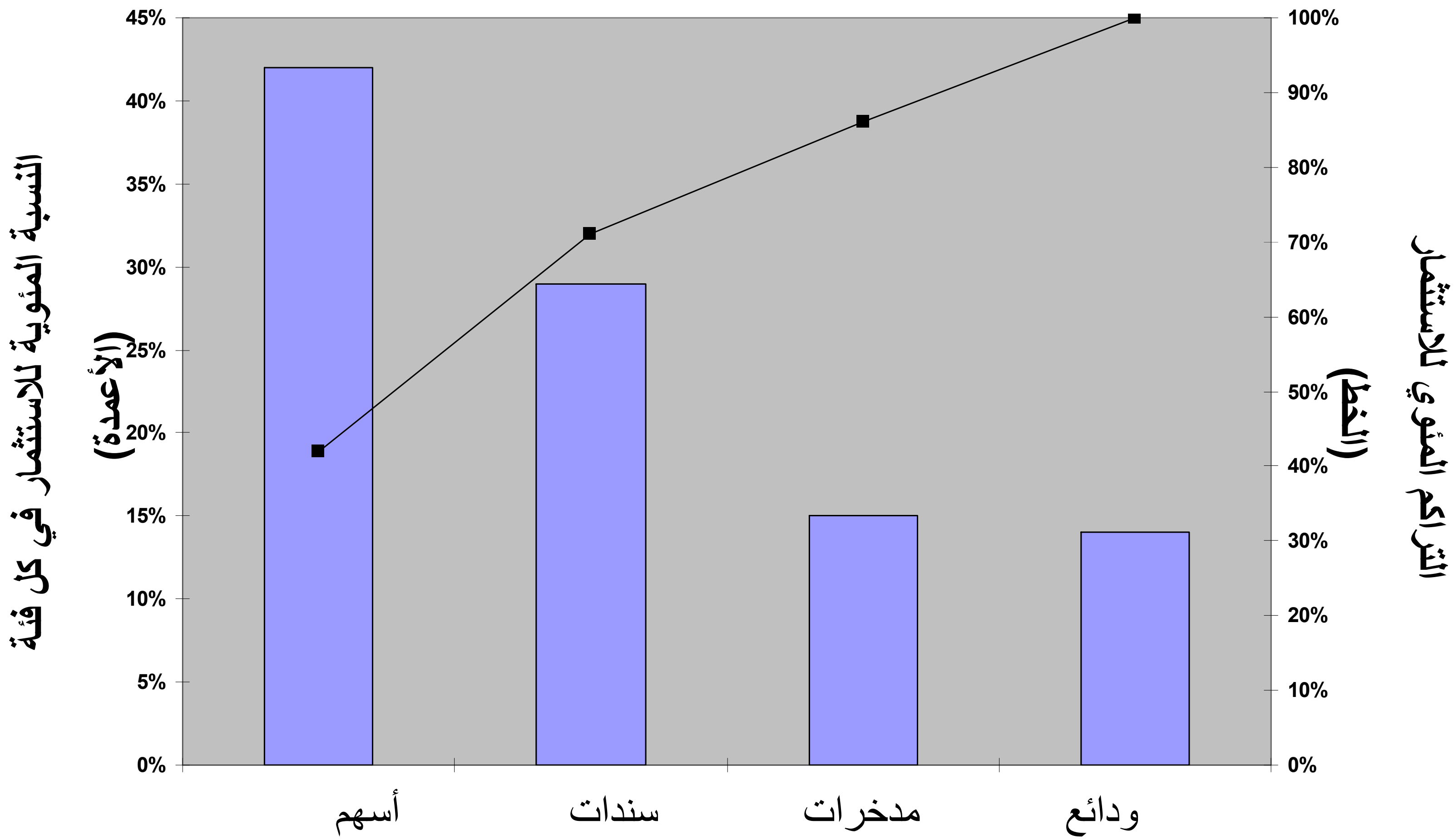
مثال: الحافظة الاستثمارية الحالية



مثال: مخطط أعمدة

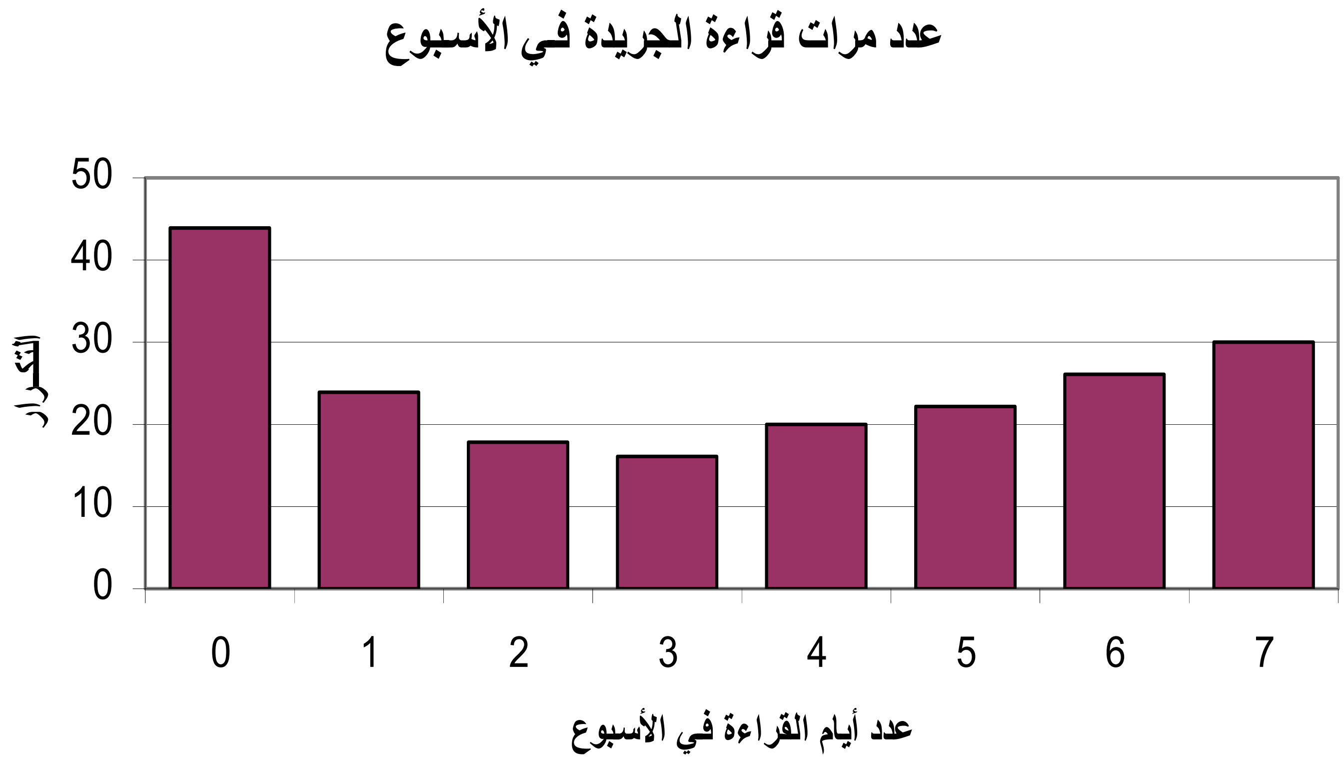


مثال: مخطط باريتو



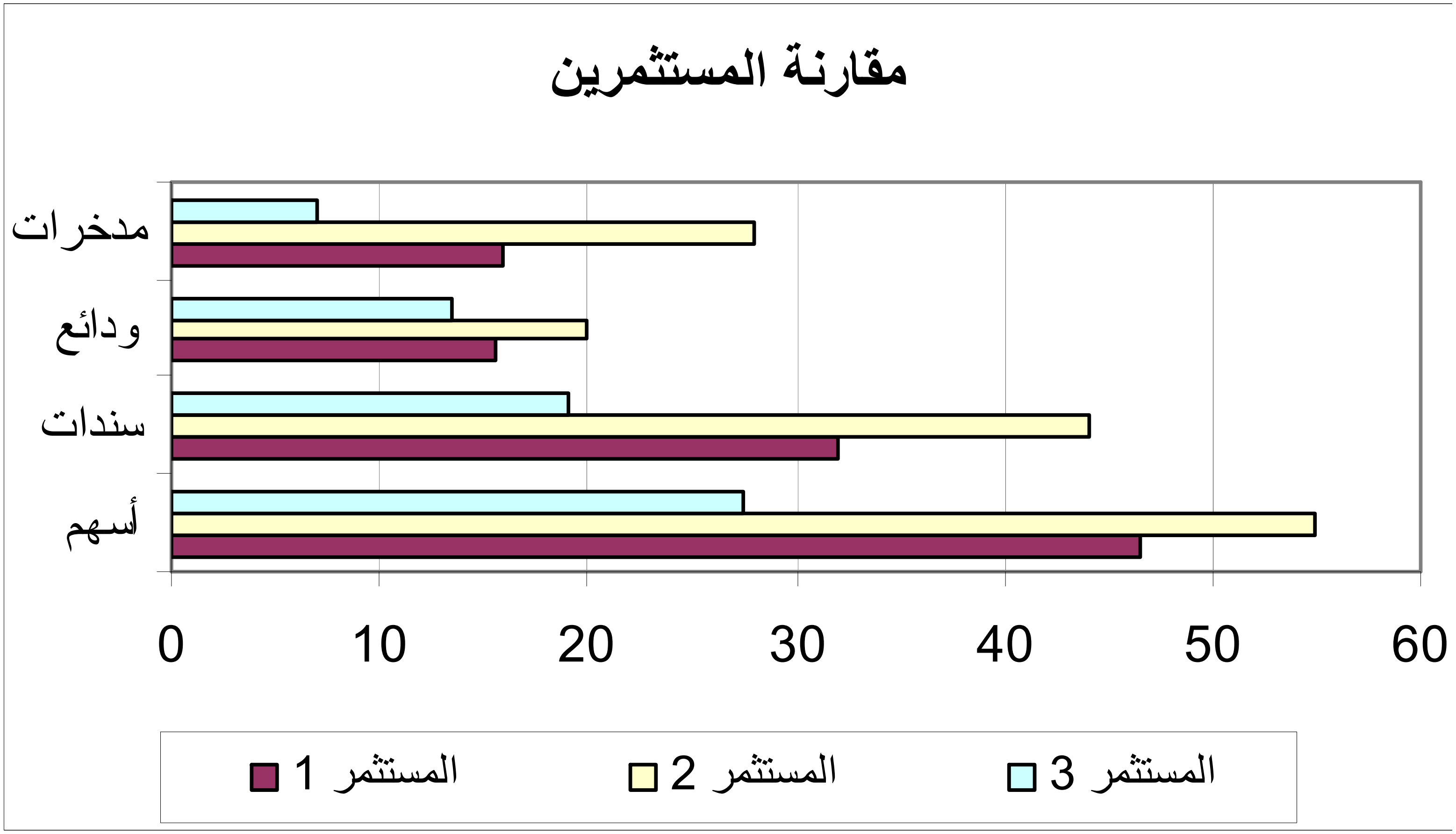
مثال: مخطط أعمدة بسيط

عدد الأيام	التكرارات
0	44
1	24
2	18
3	16
4	20
5	22
6	26
7	30
الإجمالي	**200**

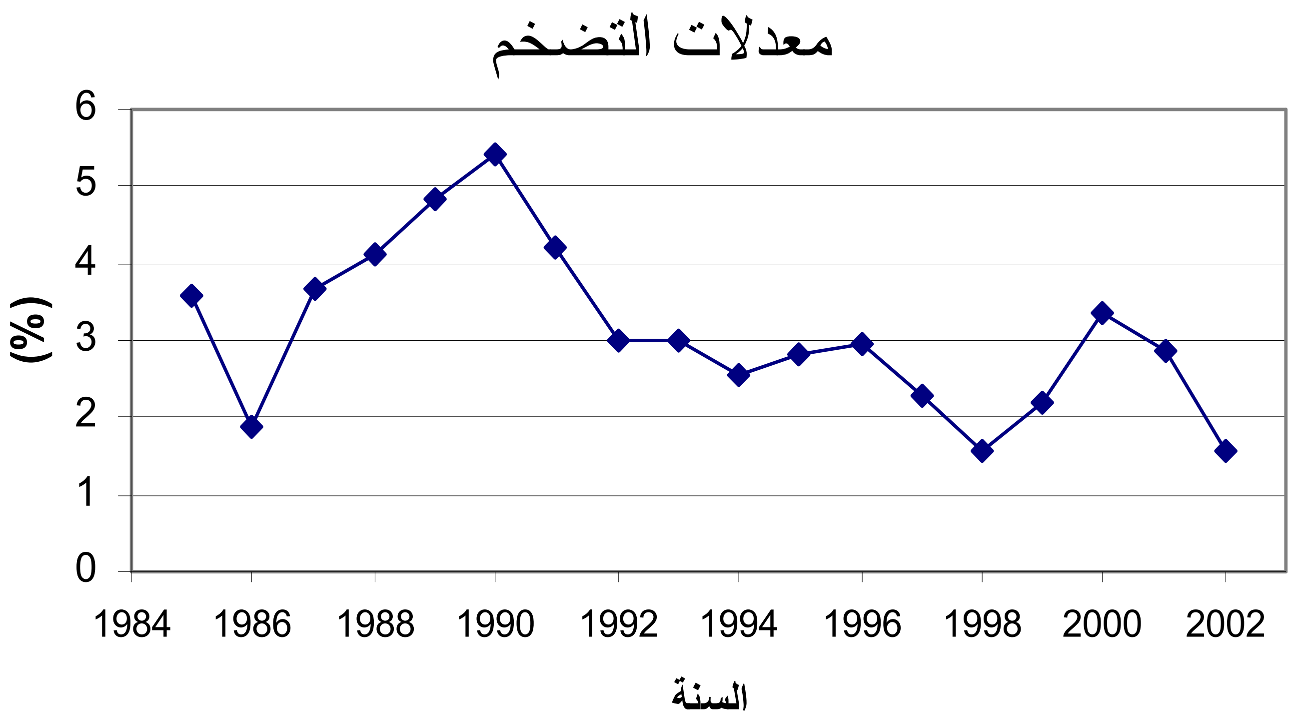


مثال: مخطط أعمدة متعدد

- "ساید باي ساید تشارت"



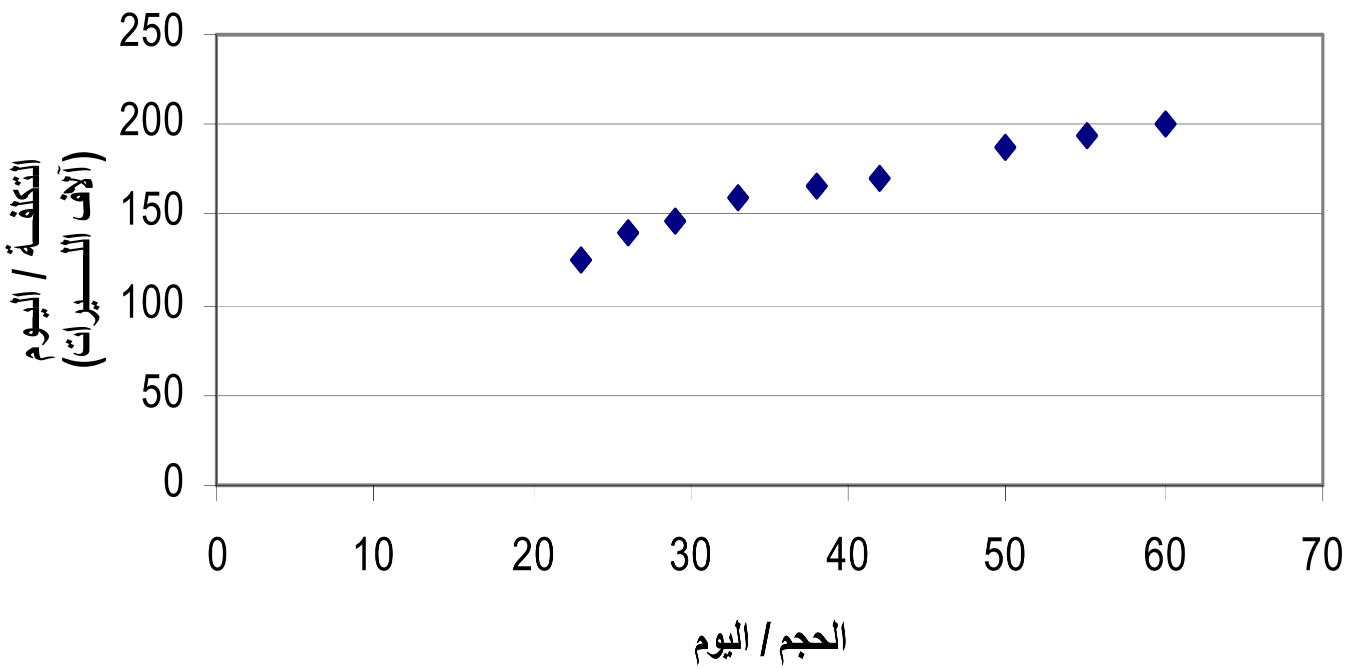
مثال: مخطط خطي



السنة	معدل التضخم
1985	3.56
1986	1.86
1987	3.65
1988	4.14
1989	4.82
1990	5.40
1991	4.21
1992	3.01
1993	2.99
1994	2.56
1995	2.83
1996	2.95
1997	2.29
1998	1.56
1999	2.21
2000	3.36
2001	2.85
2002	1.58

مثال: مخطط تبعثر

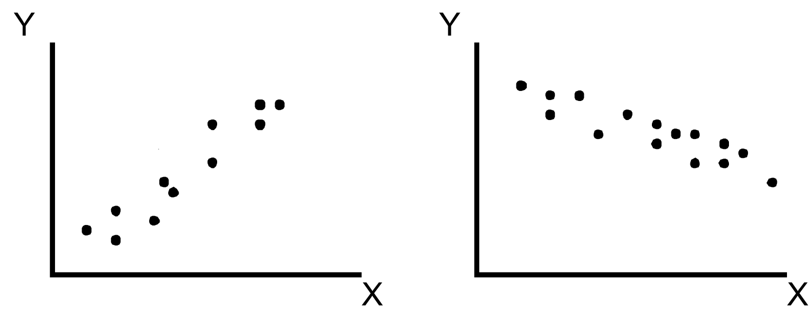
الحجم اليومي للانتاج إزاء التكلفة اليومية المرافقة



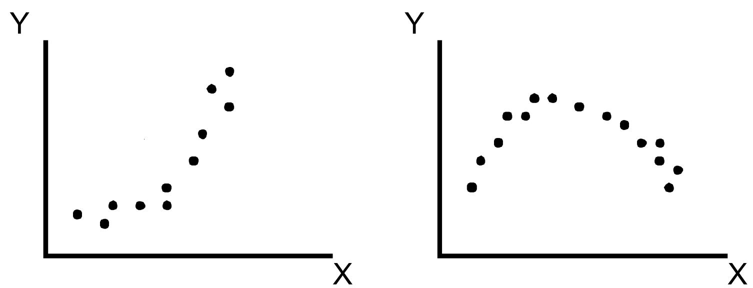
التكلفة اليومية	الحجم اليومي للانتاج
125	23
140	26
146	29
160	33
167	38
170	42
188	50
195	55
200	60

أنواع العلاقات:

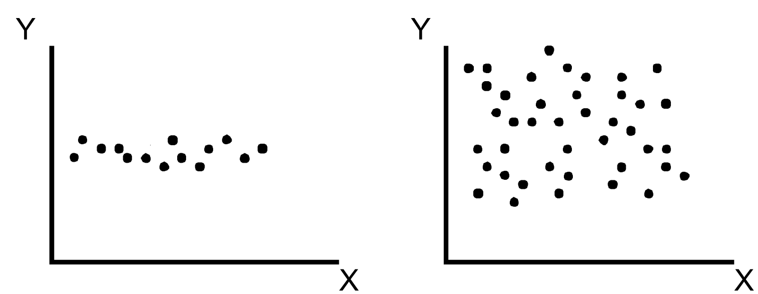
العلاقات الخطية



العلاقات "الانحناخطية"



انعدام العلاقة



ملخص المحاضرة الثانية

- بناء و تفسير التوزيعات التكرارية.
- بناء و تفسير هستوغرام.
- إنشاء و تفسير مخططات الكعكة.
- إنشاء و تفسير مخططات الأعمدة بأنواعها المختلفة.
- إنشاء و تفسير مخططات التبعثر و تمييز أنواع العلاقات.
- تفسير مخططات أخرى (باريتو، الوريقة، إلخ)
- اختيار المخطط المناسب للبيانات و الظاهرة محل الاهتمام.

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الثالثة /القسم النظري/

وصف البيانات باستخدام المقاييس الرقمية

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

مخرجات المحاضرة الثالثة:

تأكد بنهاية هذه المحاضرة أنك تستطيع:

- حساب **الوسط** و **الوسيط** و **المنوال** لمجموعة من البيانات و تفسيره.
- حساب **المدى** و **التشتت** و **الانحراف المعياري** لمجموعة من البيانات و تفسير ما تعنيه هذه المقاييس بدقة.
- حساب و تفسير **معامل التباين** تفسيره و فهم الفائدة من استخدامه بوضوح.
- حساب و استخدام الـ **"زي-سكور"**
- تفسير (و إن أمكن إنشاء) مخططات **بوكس أند فيسكر**.
- استخدام الحسابات الرقمية مع المخططات الإيضاحية و الجداول الرقمية بغرض وصف و عرض البيانات.

المقاييس التلخيصية.

- وصف البيانات رقمياً
  - المركز و الموضع
    - الوسط
    - الوسيط
    - المنوال
    - الوسط المثقل
  - مقاييس موضعية أخرى
    - المئويات
    - الربيعيات
  - التباين
    - المدى
    - المدى الربيعي
    - التشتت
    - الانحراف المعياري
    - معامل التباين

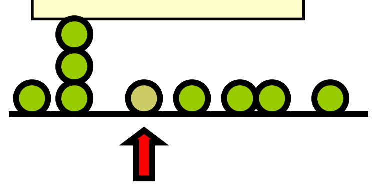
Center and Location مقاييس المركز و الموقع

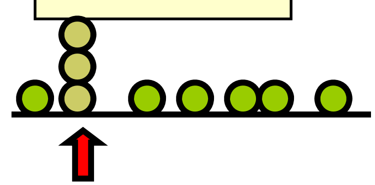
المركز و الموقع

الوسط	الوسيط	المنوال	الوسط المثقل
Mean	**Median**	**Mode**	**Weighed Mean**

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

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$





$$\bar{X}_W = \frac{\sum w_i x_i}{\sum w_i}$$

$$\mu_W = \frac{\sum w_i x_i}{\sum w_i}$$

الوسط (الوسط الحسابي) 1

- **الوسط** هو ببساطة الوسط الحسابي لمجموعة من البيانات!
  - وسط العينة

n = حجم العينة

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

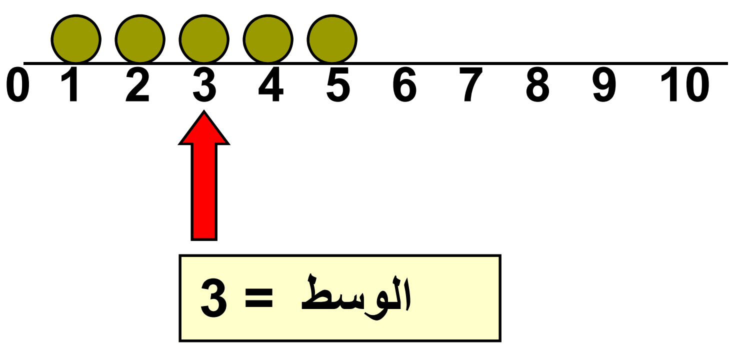
  - وسط المجتمع

N = حجم المجتمع

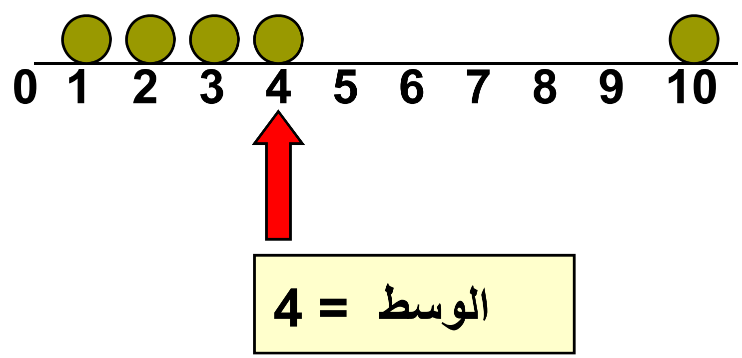
$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

الوسط (الوسط الحسابي) 2

- الأكثر شيوعاً بين مقاييس النزعة المركزية.
- يساوي مجموع القيم مقسوماً على عددها.
- يتأثر بالقيم الشاذة.



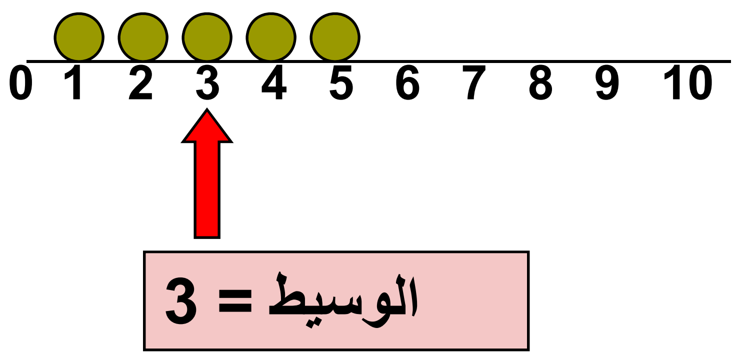
$$\frac{1+2+3+4+5}{5}=\frac{15}{5}=3$$

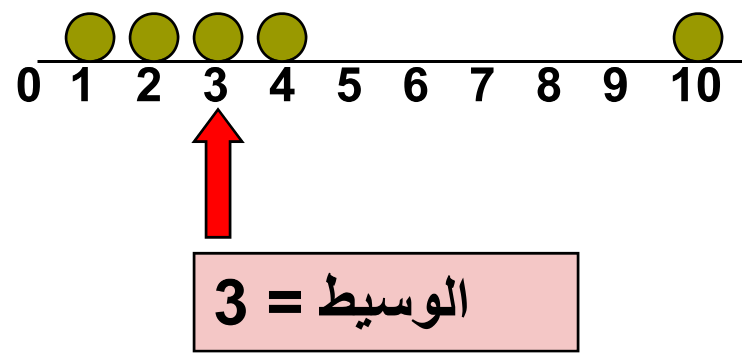


$$\frac{1+2+3+4+10}{5}=\frac{20}{5}=4$$

Median الوسيط

- لا يتأثر بالقيم الشاذة

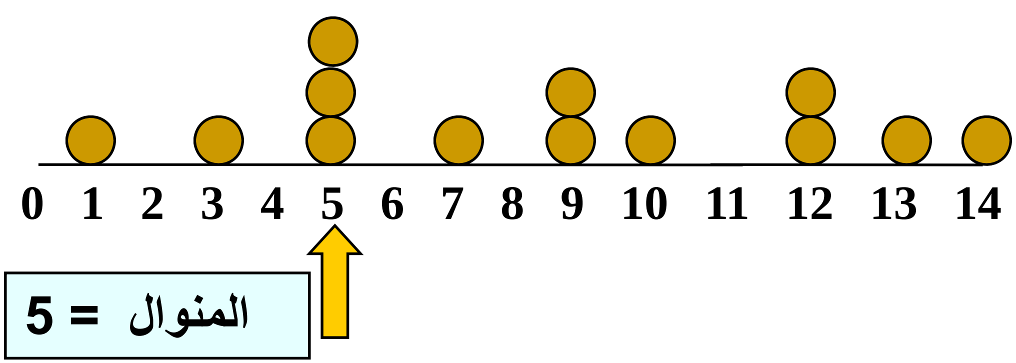


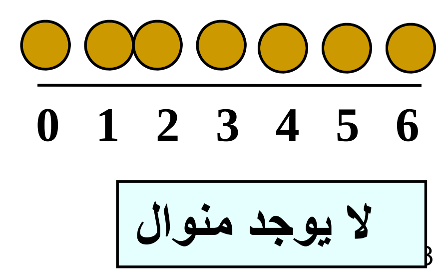


- في شعاع مرتب من البيانات يمثل الوسيط الرقم الأوسط.
  - هو العدد الأوسط إذا كان حجم البيانات فردياً.
  - أما إذا كان حجمها زوجياً فالوسيط هو وسطي العددين الأوسطين.

المنوال Mode

- هو مقياس للنزعة المركزية.
- يمثل القيمة **الأكثر ظهوراً** (حدوثاً أو مشاهدةً)
- لا يتأثر بالقيم الشاذة
- يمكن استخدامه للبيانات الرقمية و الفئوية.
- لا يشترط وجوده بالضرورة.
- و قد يوجد من أكثر من واحد.





Weighed Mean الوسط المُثَقَّل

- يستخدم عند تجميع القيم بحسب التكرار أو الأهمية النسبية

مثال: عينة من

26 مشروع إصلاح

عدد الأيام اللازمة للانجاز	التكرار
5	4
6	12
7	8
8	2

الوسطي المثقل للأيام اللازمة للانجاز

$$\bar{X}_W = \frac{\sum w_i x_i}{\sum w_i} = \frac{(4\times 5)+(12\times 6)+(8\times 7)+(2\times 8)}{4+12+8+2}$$

$$= \frac{164}{26} = 6.31 \text{ days}$$

مثال للمراجعة

- خمس بيوت على شاطئ البحر!

أسعار البيوت:

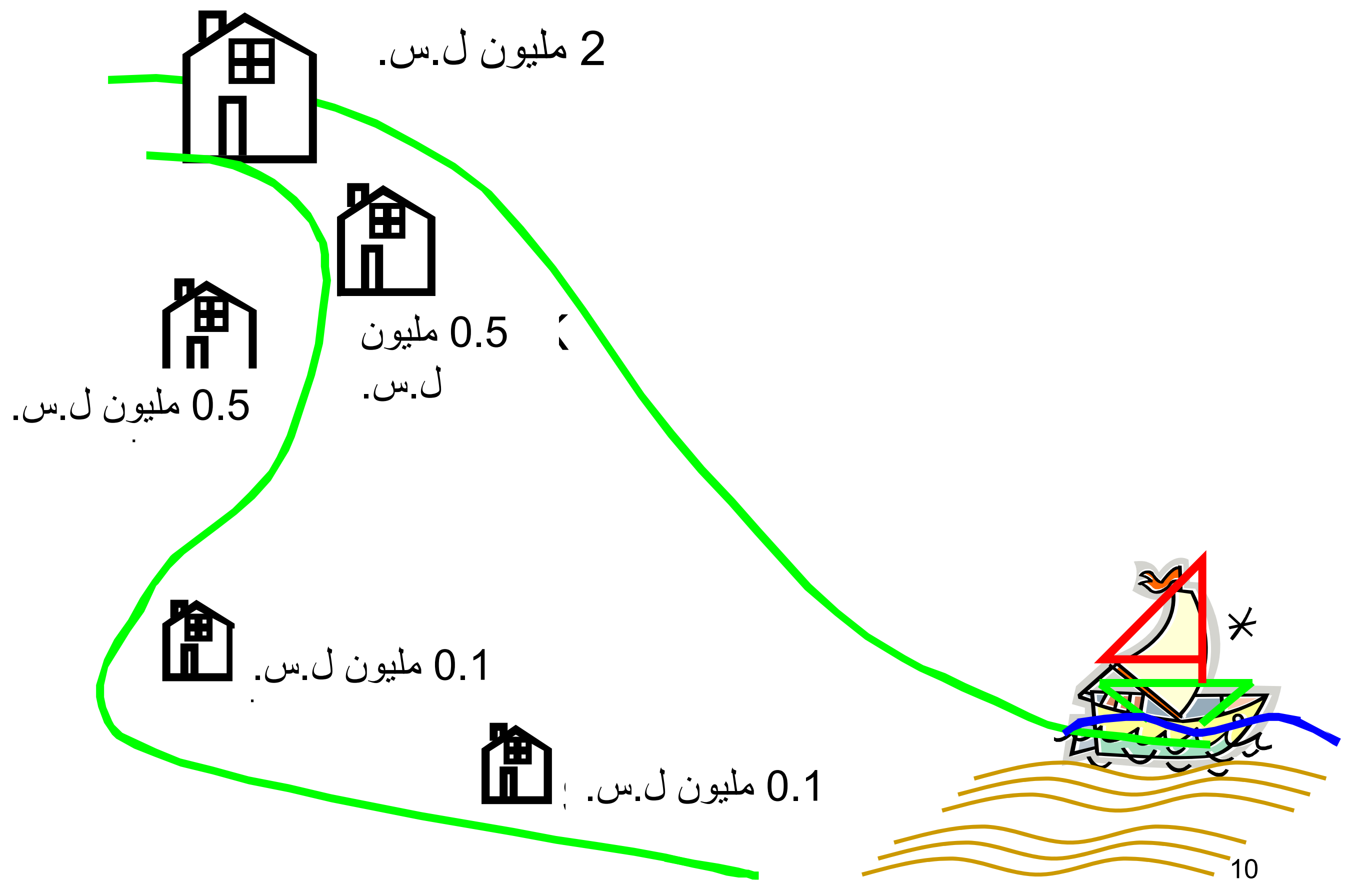
2,000,000

500,000

300,000

100,000

100,000



Summary stat. الإحصاءات التلخيصية

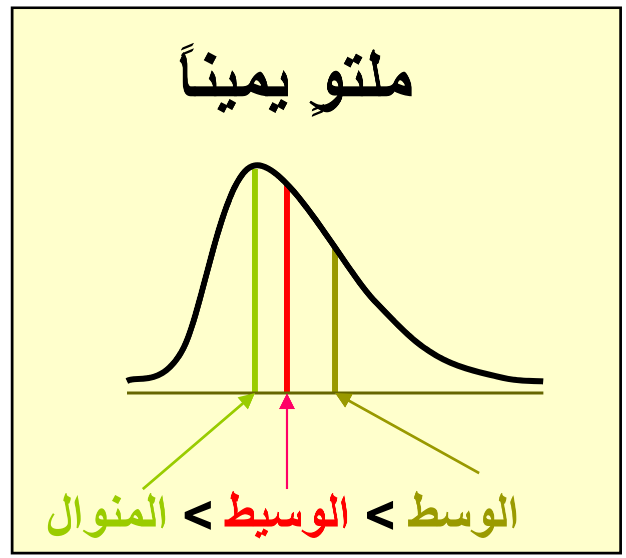
أسعار البيوت:

	2,000,000
	500,000
	300,000
	100,000
+	100,000
	3,000,000

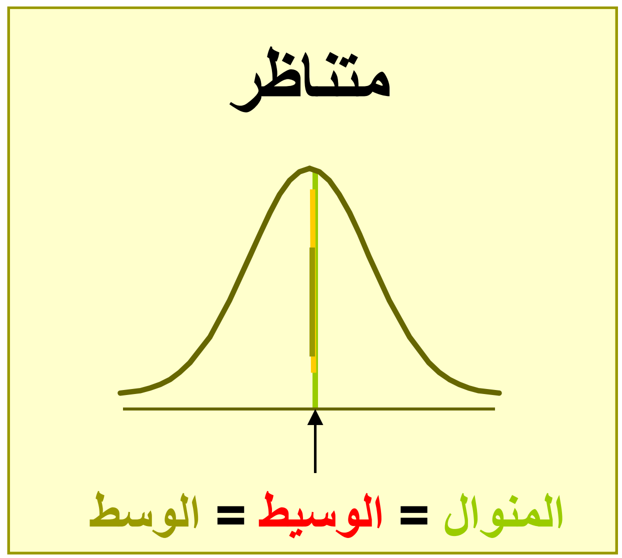
- الوسط: (3,000,000/5)
  600,000 =
- الوسيط: القيمة الوسطى ضمن بيانات مرتبة
  300,000 =
- المنوال: القيمة الأكثر تكراراً.
  100,000 =

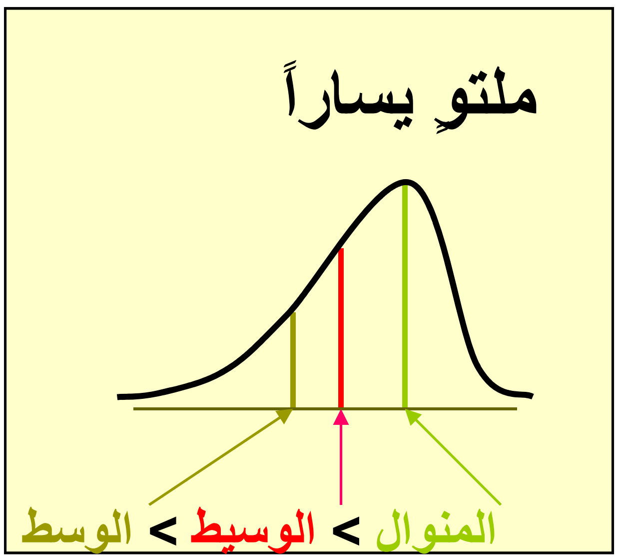
Distribution Shape شكل التوزيع

- يصف توزيع البيانات
  - **متناظر** أو **ملتوٍ**



(الذيل الأطول يمتد يميناً)





(الذيل الأطول يمتد يساراً)

مقاييس موضعية أخرى

مقاييس موضعية أخرى

الكسيرات المئوية

percentiles

الربيعيات

quartiles

ر1 = 1^{st} quartile = 25^{th} percentile = الكسيرة المئوية الخامسة و العشرين

ر2 = 2^{nd} quartile = 50^{th} percentile = الكسيرة المئوية الخمسين

= الوسيط median

ر3 = 3^{rd} quartile = 75^{th} percentile = الكسيرة المئوية الخامسة و السبعين

Percentiles الكسيرات المئوية

- الكسيرة المئوية ذات الترتيب p في شعاع مرتب مكون من n قيمة هي القيمة المتموضعة في الموضع i من الترتيب حيث:

$$i = \frac{p}{100}(n+1)$$

- **مثال:** الكسيرة المئوية الستين في شعاع بيانات مرتب مكون من 19 قيمة هي القيمة المتموضعة في الموضع الثاني عشر من الترتيب.

$$i = \frac{p}{100}(n+1) = \frac{60}{100}(19+1) = 12$$

Quartiles الربيعيات

- تقسّم الربيعيات البيانات المرتبة إلى أربعة أقسام متساوية.

25%	25%	25%	25%
↑ Q1 ر1	↑ Q2 ر2	↑ Q3 ر3	

مثال: أوجد الربيع الأول

بيانات العينة في شعاع مرتب :	22	21	18	17	16	16	13	12	11

$\frac{25}{100}$

(n = 9)

ر1 = Q1 = 25th percentile

$\frac{25}{100}(9+1) = 2.5$ position

Q1 = 12.5

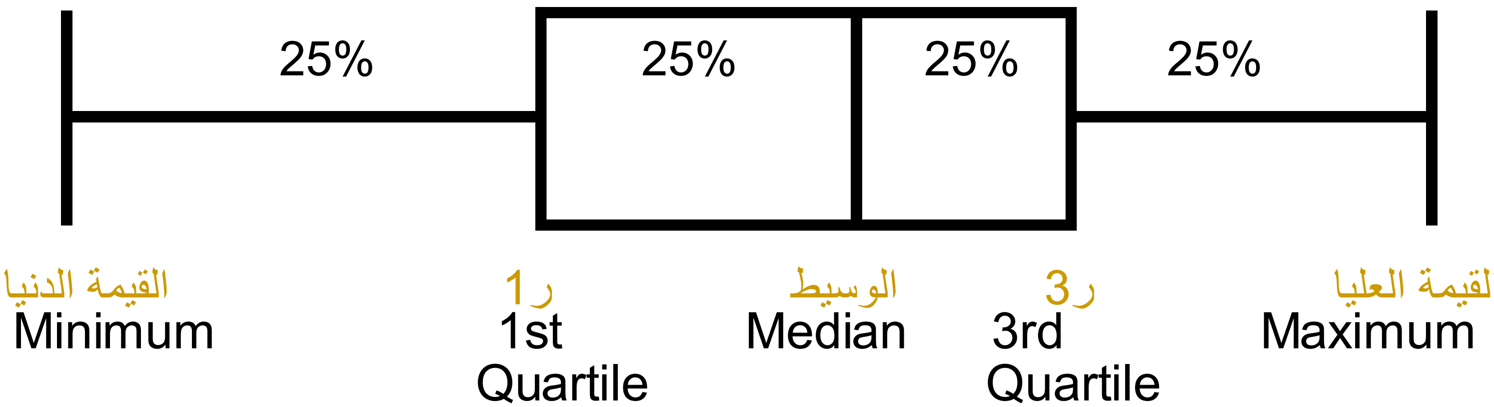
Box-and-Whisker Plot مخطط بوكس و فيسكر

- عرض بياني باستخدام خمس أرقام!

Minimum -- Q1 -- Median -- Q3 – Maximum

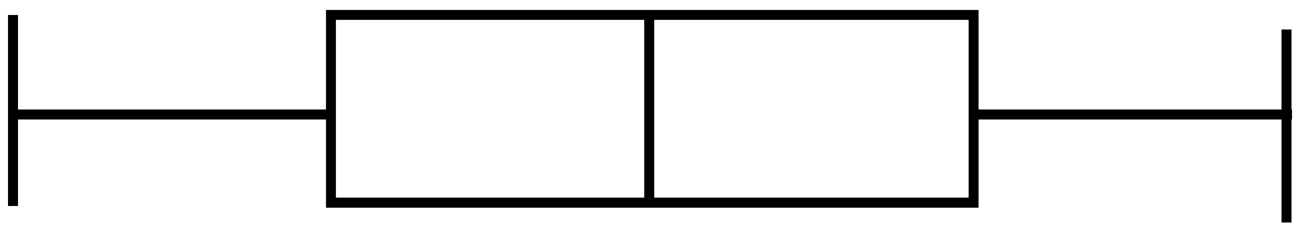
القيمة الدنيا ر1 الوسيط ر3 القيمة العليا

مثال:



شكل مخطط بوكس و فيسكر

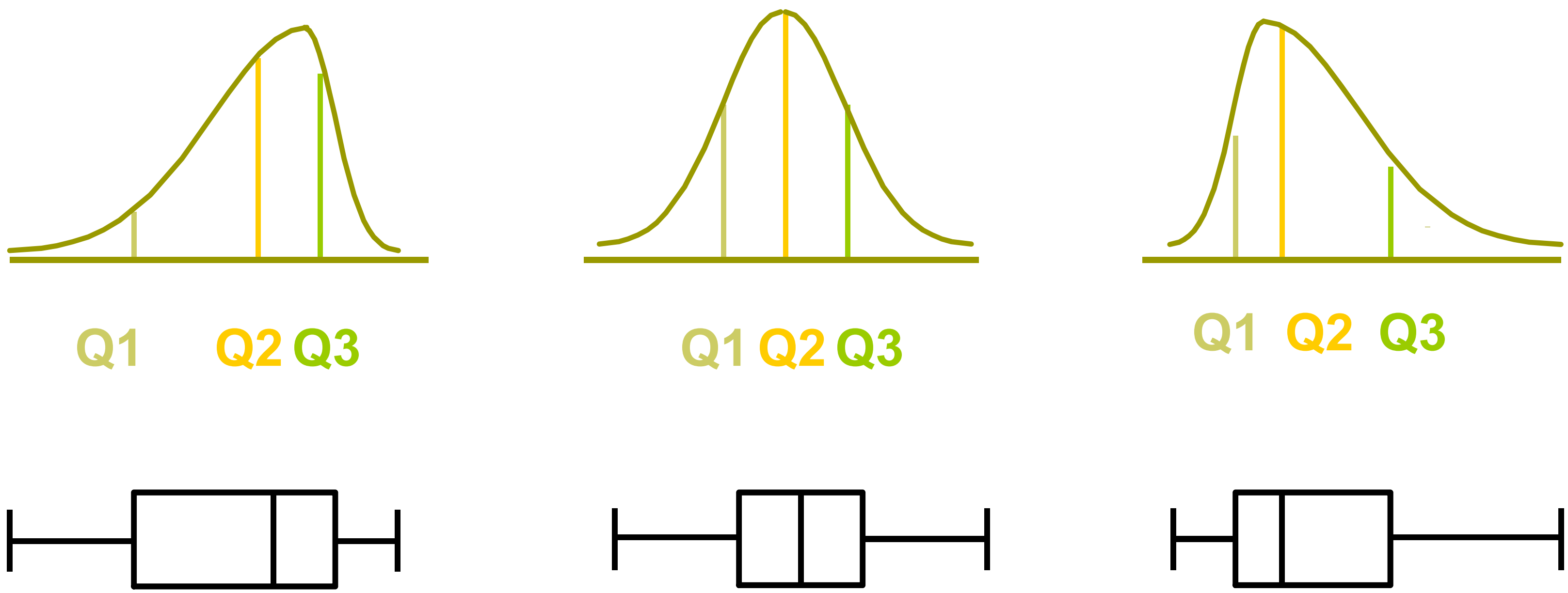
- الصندوق و الخط المركزي يتموضعان في منتصف المسافة بين النهايتين عندما تكون البيانات متناظرة حول الوسط.



- لا فرق في أن يتم رسم المخطط أفقياً أو عمودياً.

شكل التوزيع وشكل مخطط بوكس و فيكسر

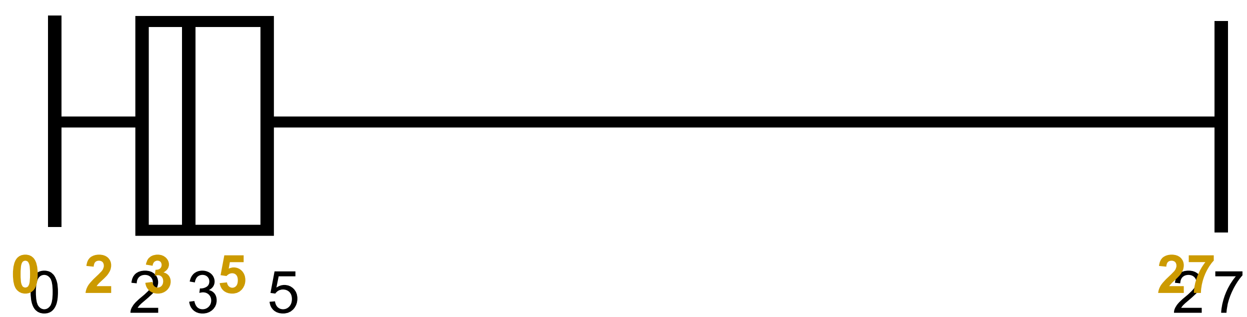
ملتوٍ يميناً | **متناظر** | **ملتوٍ يساراً**



Box-and-Whisker Plot مثال على

- مثال: قم ببناء مخطط بوكس-فيكسر من أجل البيانات التالية:
0,27,2,2,3,2,4,3,5,10,5

Min		Q1			Q2			Q3		Max
0	2	2	2	3	3	4	5	5	10	27



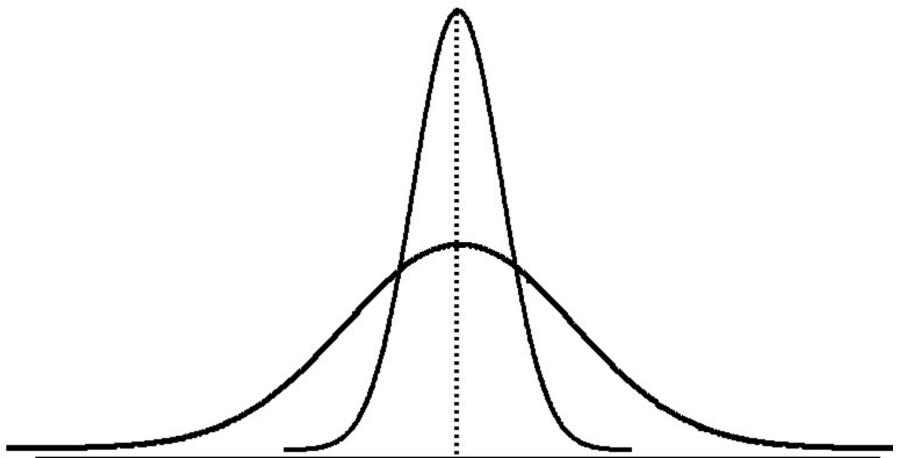
- لاحظ مقدار الالتواء نحو اليمين في هذه البيانات.

Variation Measures مقاييس التباين و الانحراف

- التباين و الانحراف **Variation**
  - المدى **Range**
    - المدى الربيعي **Interquartile Range**
  - التباين **Variance**
    - تباين المجتمع الإحصائي
    - تباين العينة
  - الانحراف المعياري **Standard Deviation**
    - الانحراف المعياري للمجتمع الإحصائي
    - الانحراف المعياري للعينة
  - معامل التباين و الانحراف **Coefficient of variation**

Variance التباين

- يعطي معلومات عن انتشار و تباين البيانات



المركز هو نفسه

أما التباين فهو مختلف

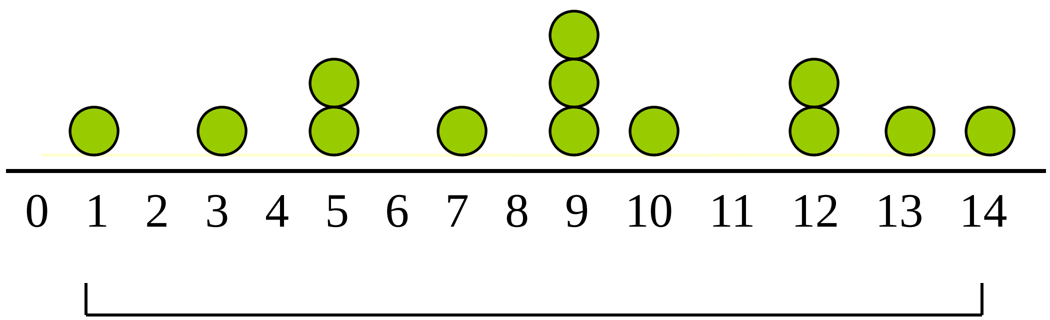
Range المدى

- ابسط مقاييس الانتشار و التباين و يعرف بأنه الفرق بين المشاهدة ذات القيمة العليا و المشاهدة ذات القيمة الدنيا

$$\text{Range} = x_{maximum} - x_{minimum}$$

المدى = القيمة العليا – القيمة الدنيا

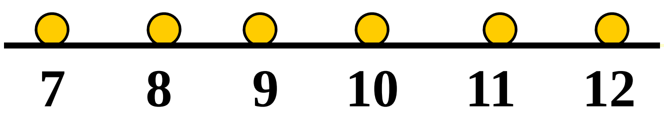
مثال:



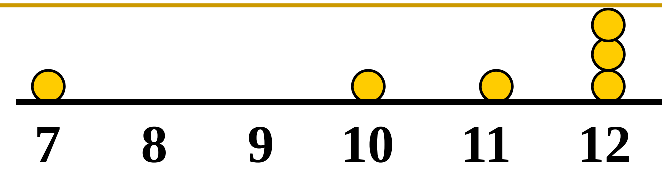
Range = 14 - 1 = 13

مثالب مقياس المدى

- يتجاهل توزيع البيانات:



Range = 12 - 7 = 5



Range = 12 - 7 = 5

- حسّاس للقيم المتطرفة:

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,5

Range = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,120

Range = 120 - 1 = 119

Interquartile Range المدى الرُبَيْعي

- يمكن استخدامه لتخفيف مساوئ المدى.
- يتخلص من بعض المشاهدات المتطرفة (للأعلى و الأسفل) ثم يحسب المدى من القيم المتبقية.

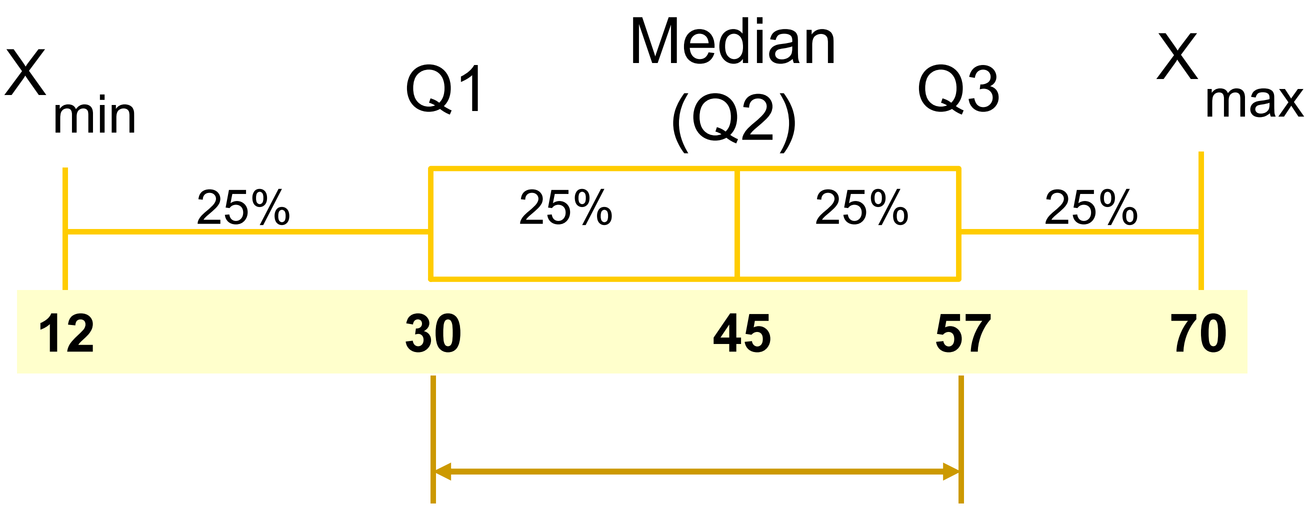
Interquartile range = 3rd quartile – 1st quartile

أو (بالعربية)

المدى الربيعي = الرُبيْع الثالث – الرُبيْع الأول

المدى الرُبَيْعي

مثال:



المدى الربيعي

= 57 – 30 = 27

التباين

- معدل مربعات الانحرافات عن الوسيط
  - تباين العينة:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

  - تباين المجتمع:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

الانحراف المعياري

- أكثر مقاييس التباين استخداماً.
- يظهر التباينات حول الوسيط.
- له نفس وحدات البيانات الأصلية.

  - الانحراف المعياري للعينة:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

  - الانحراف المعياري للمجتمع:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

مثال حسابي: الانحراف المعياري للعينة

بيانات العينة (X_i) :

10	12	14	15	17	18	18	24

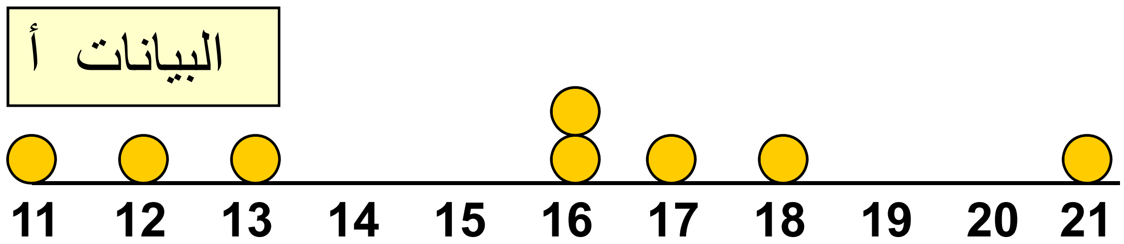
$n = 8$ الوسط $= \bar{x} = 16$

$$s = \sqrt{\frac{(10-\bar{x})^2 + (12-\bar{x})^2 + (14-\bar{x})^2 + \cdots + (24-\bar{x})^2}{n-1}}$$

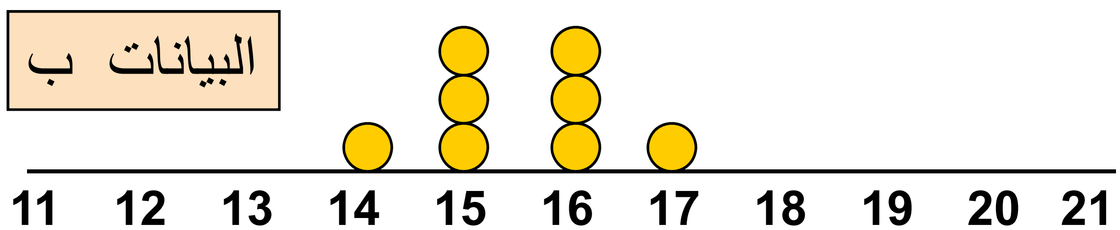
$$= \sqrt{\frac{(10-16)^2 + (12-16)^2 + (14-16)^2 + \cdots + (24-16)^2}{8-1}}$$

$$= \sqrt{\frac{126}{7}} = \boxed{4.2426}$$

مقارنة الانحرافات المعيارية



Mean = 15.5
s = 3.338



Mean = 15.5
s = .9258

Mean = 15.5
s = 4.57

معامل التباين:

- يقيس التباين النسبي
- يحسب كنسبة مئوية
- يظهر التباين نسبة إلى الوسط الحسابي
- يستخدم لمقارنة مجموعتين من البيانات مقيستين بوحدات مختلفة.

Population المجتمع

$$CV = \left(\frac{\sigma}{\mu}\right) \cdot 100\%$$

Sample العينة

$$CV = \left(\frac{s}{\bar{x}}\right) \cdot 100\%$$

مقارنة معاملات التباين

- السهم A :
  - السعر الوسطي للسهم السنة الماضية = $50
  - الانحراف المعياري = $5

$$CV_A = \left(\frac{s}{\bar{x}}\right)\cdot 100\% = \frac{\$5}{\$50}\cdot 100\% = 10\%$$

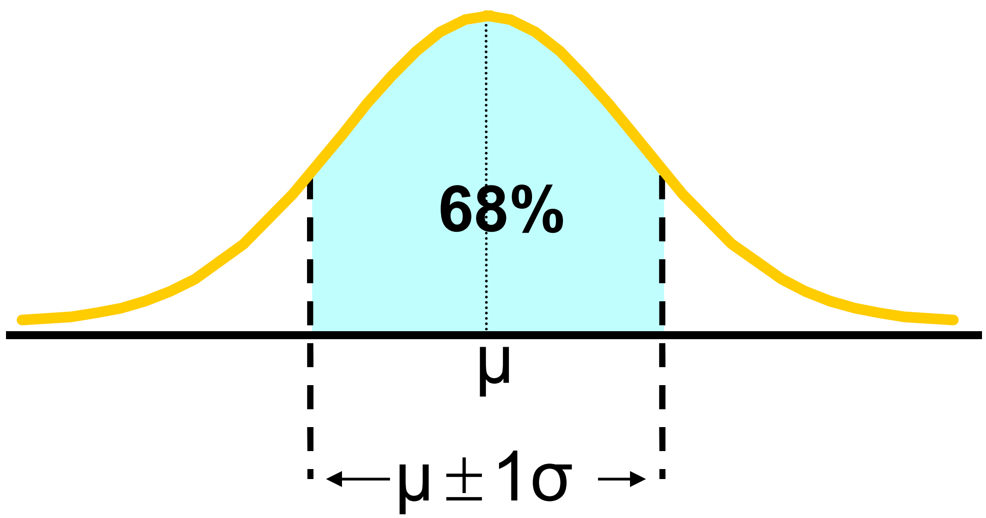
- السهم B:
  - Average price last year = $100
  - Standard deviation = $5

كلا السهمين يظهران نفس الانحراف المعياري و لكن ثانيهما أقل تقلباً نسبياً (مقارنة بالسهم الآخر).

$$CV_B = \left(\frac{s}{\bar{x}}\right)\cdot 100\% = \frac{\$5}{\$100}\cdot 100\% = 5\%$$

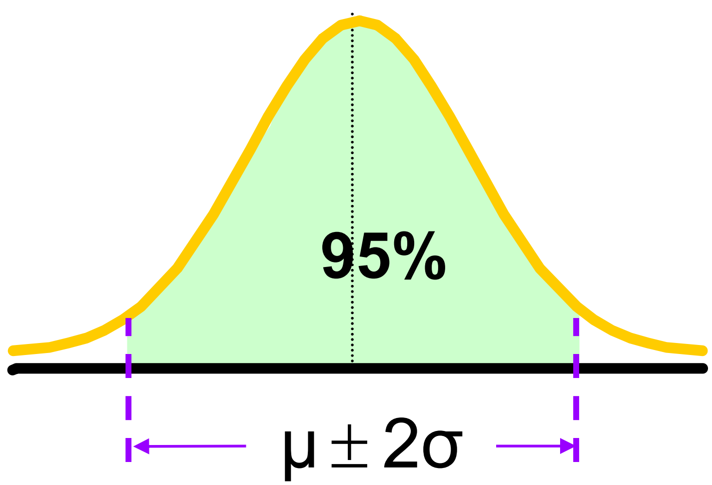
القاعدة النظرية 1

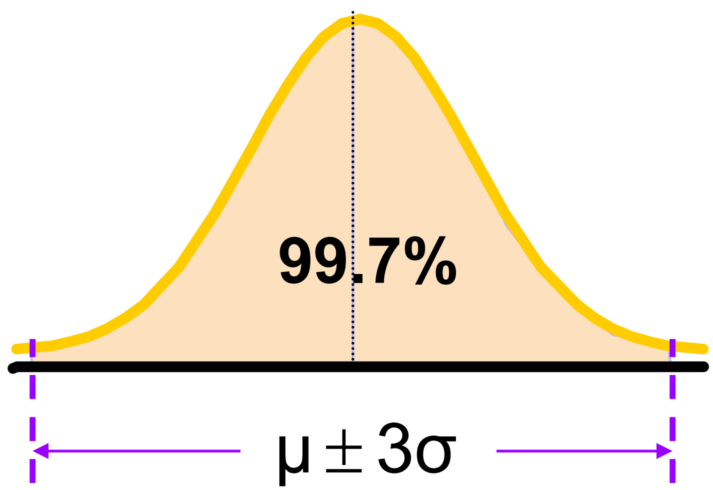
- إذا كانت البيانات على شكل جرس فإن المجال $\mu \pm 1\sigma$ يحتوي على 68% من بينانات المجتمع أو العينة.



القاعدة النظرية 2

- إذا كانت البيانات على شكل جرس فإن المجال $\mu \pm 2\sigma$ يحتوي على أو يحصر 95% من بينانات المجتمع أو العينة.
- إذا كانت البيانات على شكل جرس فإن المجال $\mu \pm 3\sigma$ يحتوي أو يحصر 99% من بينانات المجتمع أو العينة.





مبرهنة تشيبيشيف Tchebysheff's Theorem

- بغض النظر عن توزع البيانات، فإن $(1 - 1/k^2)$ على الأقل من البيانات سوف يقع ضمن k انحرافاً معيارياً من الوسط.

- أمثلة:

$(1 - 1/1^2) =$ 0% ....$k=1$ $(\mu \pm 1\sigma)$

$(1 - 1/2^2) =$ 75% ....$k=2$ $(\mu \pm 2\sigma)$

$(1 - 1/3^2) =$ 89% .... $k=3$ $(\mu \pm 3\sigma)$

القيم المُعَيَّرَة للبيانات Standardized Data Values

- القيمة المعيارية تشير إلى عدد الانحرافات المعيارية عن الوسط الحسابي للبيانات.

- يشار في كثير من الأحيان إلى القيم المعيرة للبيانات بـ: z-scores. أو "زي سكورز"

القيم المُعَيَّرة للمجتمع

$$z = \frac{x - \mu}{\sigma}$$

حيث:

- x = القيمة الأصلية للبيانات
- μ = وسط المجتمع
- σ = الانحراف المعياري للمجتمع
- z = الـ "سكور" المعياري = عدد الانحرافات المعيارية التي تبعدها قيمة المفردة x عن وسط المجتمع

القيم المُعَيَّرة للعينة

$$Z = \frac{x - \bar{x}}{s}$$

حيث:

- x = القيمة الأصلية للبيانات
- $\bar{x}$ = وسط المجتمع
- s = الانحراف المعياري للعينة
- z = الـ "سكور" المعياري = عدد الانحرافات المعيارية التي تبعدها قيمة المفردة x عن $\bar{x}$

استخدام مايكروسوفت أكسل

- يمكن الحصول على الإحصاءات الوصفية من أكسل بسهولة.
  - من القوائم المنسدلة في MS Office 2003 أو من صندوق البيانات في MS Office 2007 ، خيار تحليل البيانات قم باختيار:
  - tools / data analysis / descriptive statistics
  - ثم أدخل البيانات المطلوبة في الصندوق الحواري.
- ملاحظة: على مستخدمي MS Office 2007 أن يقوموا بتحميل حزمة Data Analysis إن لم تكن موجودة باختيار:

Microsoft Office Button/ Excel Options/Add-ins/ Excel Add-ins./Go /Add-Ins available box/ Analysis ToolPak/ OK.

- ملاحظة: على مستخدمي MS Office 2003 أن يقوموا بتحميل حزمة Data Analysis إن لم تكن موجودة باختيار:

tools / Add-Ins/ Analysis Toolpak

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الرابعة /القسم النظري/

الاحتمالات 1: مبادئ أساسية

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

مخرجات المحاضرة الثالثة:

تأكد بنهاية هذه المحاضرة أنك تستطيع:

- تمييز و إيضاح ثلاث طرائق لتقدير الاحتمالات.
- تطبيق القواعد الأساسية للاحتمالات.
- استخدام نظرية بيز للاحتمالات الشرطية.
- التمييز بين التوزيعات الاحتمالية المتقطعة و المستمرة..
- حساب القيمة المتوقعة و الانحراف المعياري للتوزيع الاحتمالي.

اصطلاحات أساسية

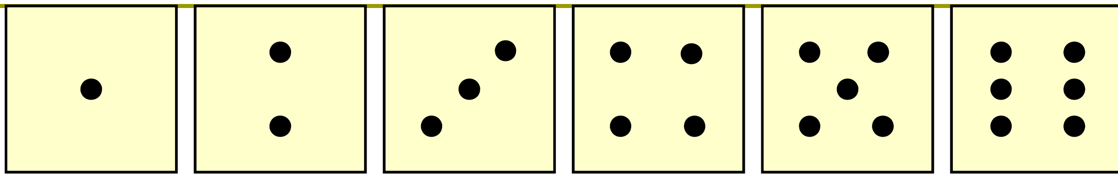
الاحتمال: الفرصة في حصول حدث غير مؤكد و هي دائماً بين الصفر و الواحد.

التجربة: عملية للحصول على نتائج من أحداث غير مؤكدة.

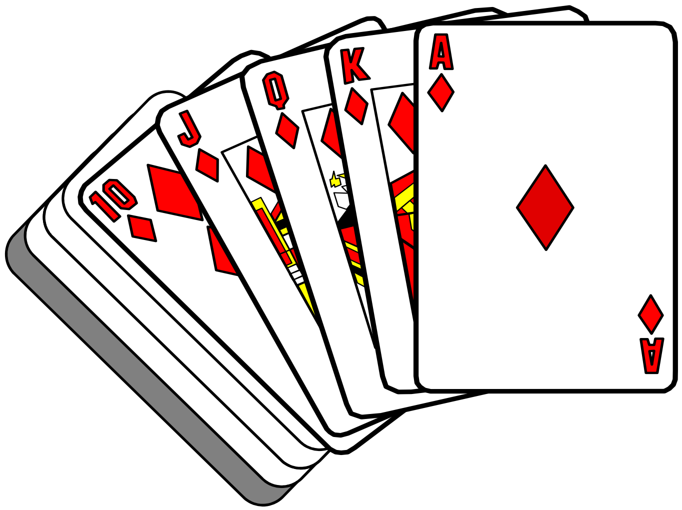
الحدث البسيط (الأولي): أبسط أنواع النتائج الممكن الحصول عليها من تجربة بسيطة.

فضاء العينة: المجموعة التي تضم كافة النتائج الممكنة من الأحداث البسيطة.

أمثلة على فضاء العينة و الأحداث



جميع و جوه حجر النرد



جميع أوراق الشدة الكاملة

- **الحدث البسيط:** نتيجة تم الحصول عليها من فضاء عينة يمتلك صفة (سمة) واحدة.
  - مثال: كرت أحمر من شدة كاملة.

- **الحدث:** قد ينطوي على نتيجتين أو أكثر بآن معاً.
  - مثال: كرت الملك من اللون الأحمر (ديناري أو كبة).

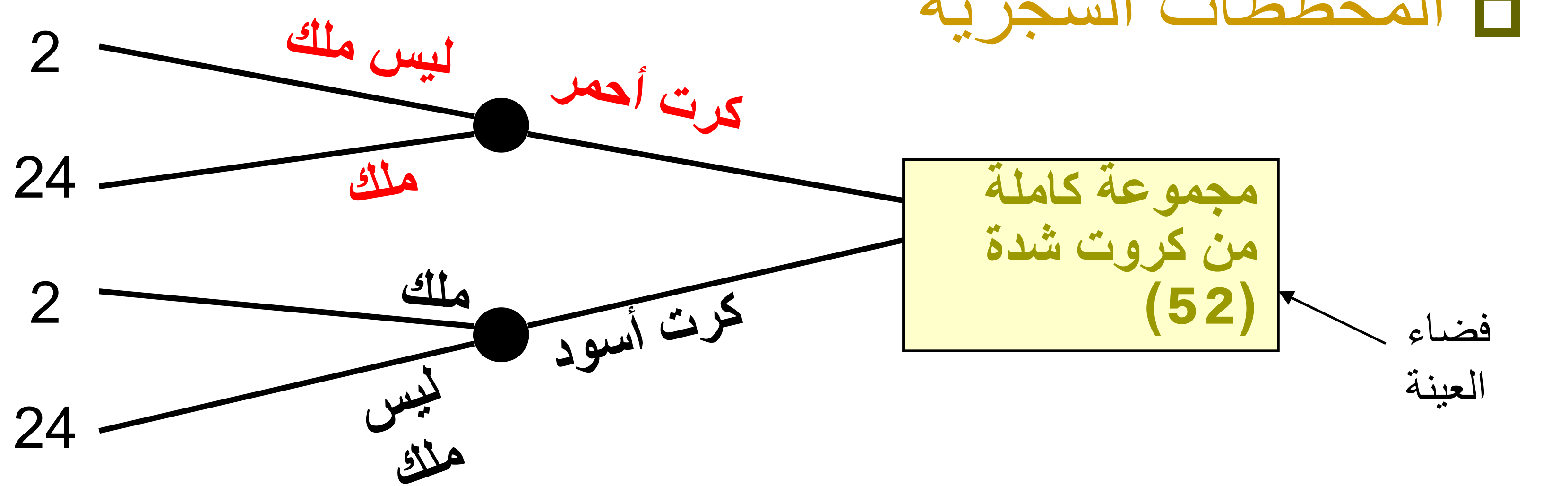
التمثيل البياني للأحداث

- جداول الاستقلال:

	ملك	ليس ملك	Total
أسود	2	24	26
أحمر	2	24	26
المجموع	4	48	52

فضاء العينة

- المخططات الشجرية

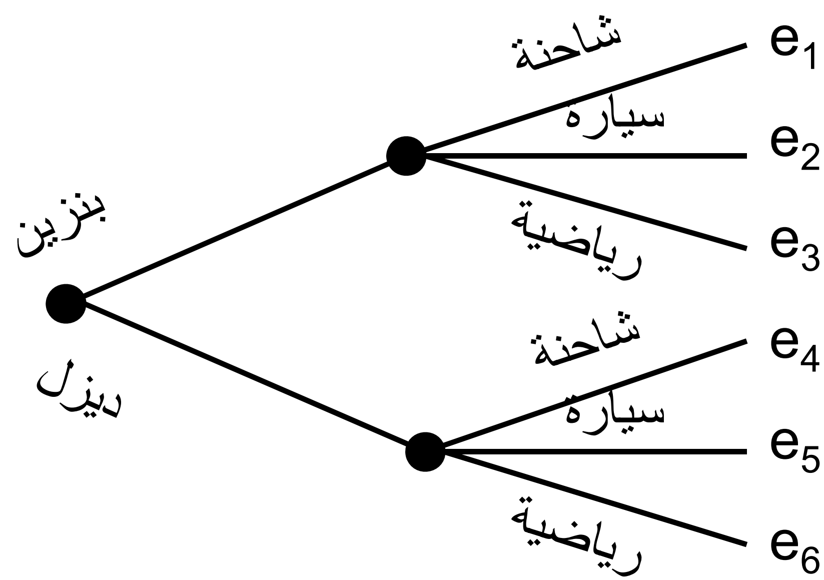


الأحداث البسيطة

- عادة ما تقسم المركبات بحسب نوع السيارة و نوع الوقود.
  - نوعان من الوقود: بنزين و ديزل
  - ثلاثة أنواع من المركبات: شاحنة، سيارة، سيارة رياضية.

- لدينا بالتالي ست أحداث بسيطة:

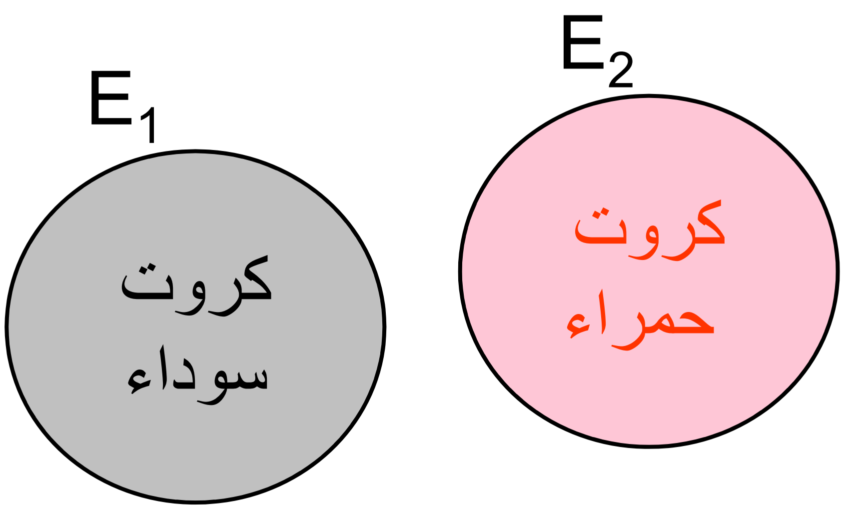
e_1 بنزين, شاحنة
e_2 بنزين, سيارة
e_3 بنزين, رياضية
e_4 ديزل, شاحنة
e_5 ديزل, سيارة
e_6 ديزل, رياضية



مفاهيم احتمالية

الأحداث المتنافية تبادلياً

- وقوع أحدهما E1 ينفي وقوع الآخر E2
- ليس بينهما أي عنصر مشترك



لا يمكن للكرت أن يكون أحمر و أسود اللون في آن معاً.

مفاهيم احتمالية: الاستقلال و التبعية

الاستقلال: حدوث أحدهما لا يؤثر على احتمال حدوث الآخر.

مثال:

E_1 = الحصول على نقش من الرمية الأولى لقطعة نقدية

E_2 = الحصول على نقش من الرمية الثانية لنفس القطعة

التبعية: حدوث أحدهما يؤثر على احتمال حدوث الآخر.

مثال:

E_1 = التنبؤ بطقس سيء في نشرة الأنباء

E_2 = الذهاب في نزهة إلى حديقة عامة

تقدير الاحتمالات:

- المدخل الكلاسيكي

$$\text{احتمال حدوث } E_i = P(E_i) = \frac{\text{عدد الطرائق التي يمكن أن يحدث فيها } E_i}{\text{العدد الإجمالي للأحداث البسيطة}}$$

- التكرار النسبي للحدوث

$$\text{التكرار النسبي لـ } E_i = P(E_i) = \frac{\text{عدد مرات تكرار } E_i}{\text{العدد الإجمالي للتجارب}}$$

- التخمين الذاتي للاحتمال

يعود إلى راي صانع القرار أو حكمه الشخصي.

قواعد الاحتمالات

قواعد الجمع
و القيم المحتملة

مجموع القيم الفردية

$$\sum_{i=1}^{k} P(e_i) = 1$$

القيم الفردية

$$0 \le P(e_i) \le 1$$

من أجل أي e_i

حيث:

k: عدد الأحداث البسيطة في فضاء العينة

e_i: الحدث البسيط رقم i

قاعدة الجمع من أجل الأحداث البسيطة

- احتمال حدوث E_i يساوي مجموع احتمالات الأحداث البسيطة التي تشكل E_i

- أي إذا كان:

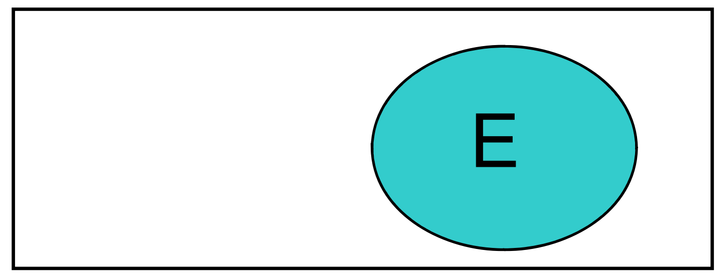
$$E_i = \{e_1, e_2, e_3\}$$

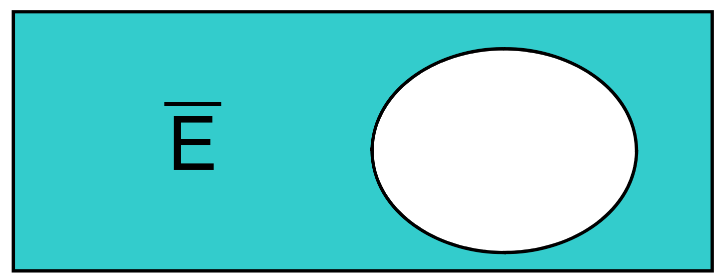
فإن:

$$P(E_i) = P(e_1) + P(e_2) + P(e_3)$$

قاعدة المكمل

- مكمل الحدث E هو المجموعة المشكلة من كافة الاحتمالات البسيطة غير المحتواة في الحدث E.





$$P(\bar{E}) = 1 - P(E)$$

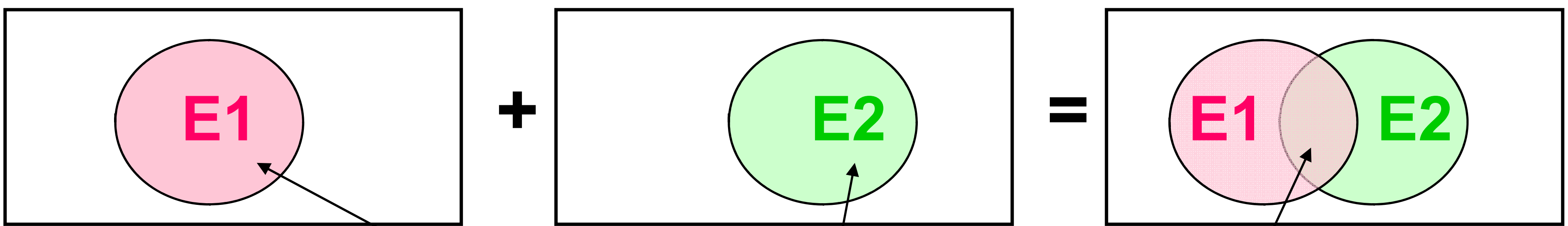
أو

$$P(E) + P(\bar{E}) = 1$$

قاعدة الجمع من أجل حدثين

قاعدة الجمع:

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - P(E_1 \text{ and } E_2)$$



$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - \boxed{P(E_1 \text{ and } E_2)}$$

لا تحتسب العناصر المشتركة

مرتين و إنما مرة واحدة فقط !

مثال على قاعدة الجمع

P(Red or Ace) = P(Red) +P(Ace) - P(Red and Ace)

= 26/52 + 4/52 - 2/52 = 28/52

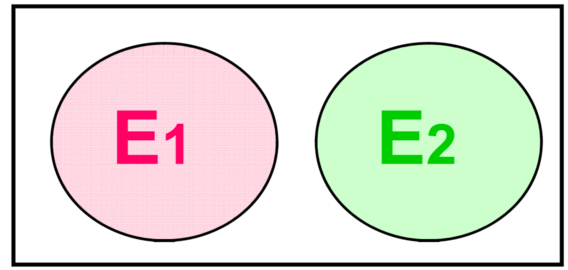
Type	Color		Total
	Red	Black	
Ace	2	2	4
Non-Ace	24	24	48
Total	26	26	52

تعد الآسين من اللون الأحمر مرتين!

قاعدة الجمع من أجل الأحداث المتنافية بشكل متبادل

- إذا كان E_1 و E_2 متنافيين تبادلياً فإن:

$$P(E_1 \text{ and } E_2) = 0$$



- و منه:

$$P(E_1 \text{ or } E_2) = P(E_1) + P(E_2) - \underbrace{P(E_1 \text{ and } E_2)}_{=0}$$

إذا كانا متنافيين تبادلياً

$$= P(E_1) + P(E_2)$$

الاحتمال الشرطي

- الاحتمال الشرطي من أجل أي حدثين: $E_1 \ , \ E_2$

$$P(E_1 | E_2) = \frac{P(E_1 \text{ and } E_2)}{P(E_2)}$$

حيث: $P(E_2) > 0$

مثال على الاحتمال الشرطي

- وجد في معرض للسيارات المستعملة أن 70% من السيارات مزودة بمكيف (AC) و 40% من السيارات مزود بمشغل أقراص (CD) و 20% من السيارات مزود بكليهما.

- ماهو احتمال أن تكون السيارة مزودة بمشغل أقراص علماً أنها مزودة بمكيف؟

بكلمة أخرى، أوجد: $P(CD \mid AC)$

مثال على الاحتمال الشرطي: تابع

- 70% من السيارات مزودة بمكيف (AC) و 40% من السيارات مزود بمشغل أقراص (CD) و 20% من السيارات مزود بكليهما.

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

$$P(CD \mid AC) = \frac{P(CD \text{ and } AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

مثال على الاحتمال الشرطي: تابع

- بما أن الحدث AC قد وقع فإننا ننظر للسطر الأول و الذي يتعلق بـ 70% من السيارات فقط، و من هذه نريد 20% تحتوي على مشغل أقراص (أي تحقق الحدث CD) و هكذا، فـ 20% من 70% يساوي تقريباً 28.75%.

	CD	No CD	Total
AC	.2	.5	.7
No AC	.2	.1	.3
Total	.4	.6	1.0

$$P(CD \mid AC) = \frac{P(CD \text{ and } AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

الاحتمال الشرطي و الأحداث المستقلة

- الاحتمال الشرطي للحدثين المستقلين : E_1 , E_2

$$\boxed{P(E_1 | E_2) = P(E_1)} \qquad P(E_2) > 0$$

$$\boxed{P(E_2 | E_1) = P(E_2)} \qquad P(E_1) > 0$$

قواعد الضرب

- قاعدة الضرب من أجل حدثين: E_1 و E_2

$$\boxed{P(E_1 \text{ and } E_2) = P(E_1)P(E_2 \mid E_1)}$$

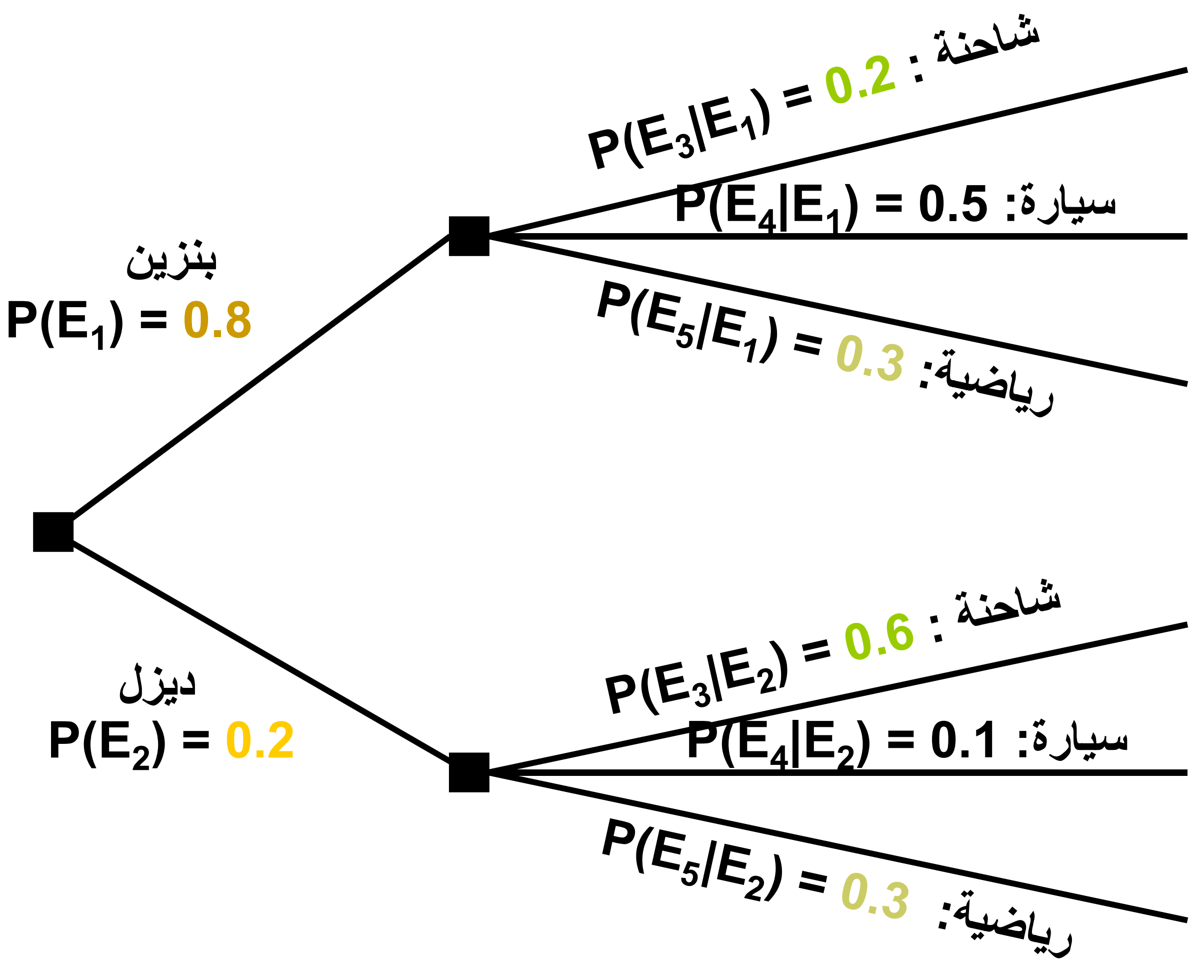
- ملاحظة: إذا كان E_1 و E_2 مستقلان فإن: $P(E_2 \mid E_1) = P(E_2)$

و في هذه الحالة تصبح قاعدة الضرب أبسط، حيث:

$$P(E_1 \text{ and } E_2) = P(E_1)P(E_2)$$

مثال على المخططات الشجرية

- بالعودة لمثال المخطط الشجري للسيارات، لاحظ الفرق بين الاحتمال الشرطي و الاحتمال المشترك!



$P(E_1 \text{ and } E_3) = 0.8 \times 0.2 = 0.16$

$P(E_1 \text{ and } E_4) = 0.8 \times 0.5 = 0.40$

$P(E_1 \text{ and } E_5) = 0.8 \times 0.3 = 0.24$

$P(E_2 \text{ and } E_3) = 0.2 \times 0.6 = 0.12$

$P(E_2 \text{ and } E_4) = 0.2 \times 0.1 = 0.02$

$P(E_3 \text{ and } E_4) = 0.2 \times 0.3 = 0.06$

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الخامسة /القسم النظري/

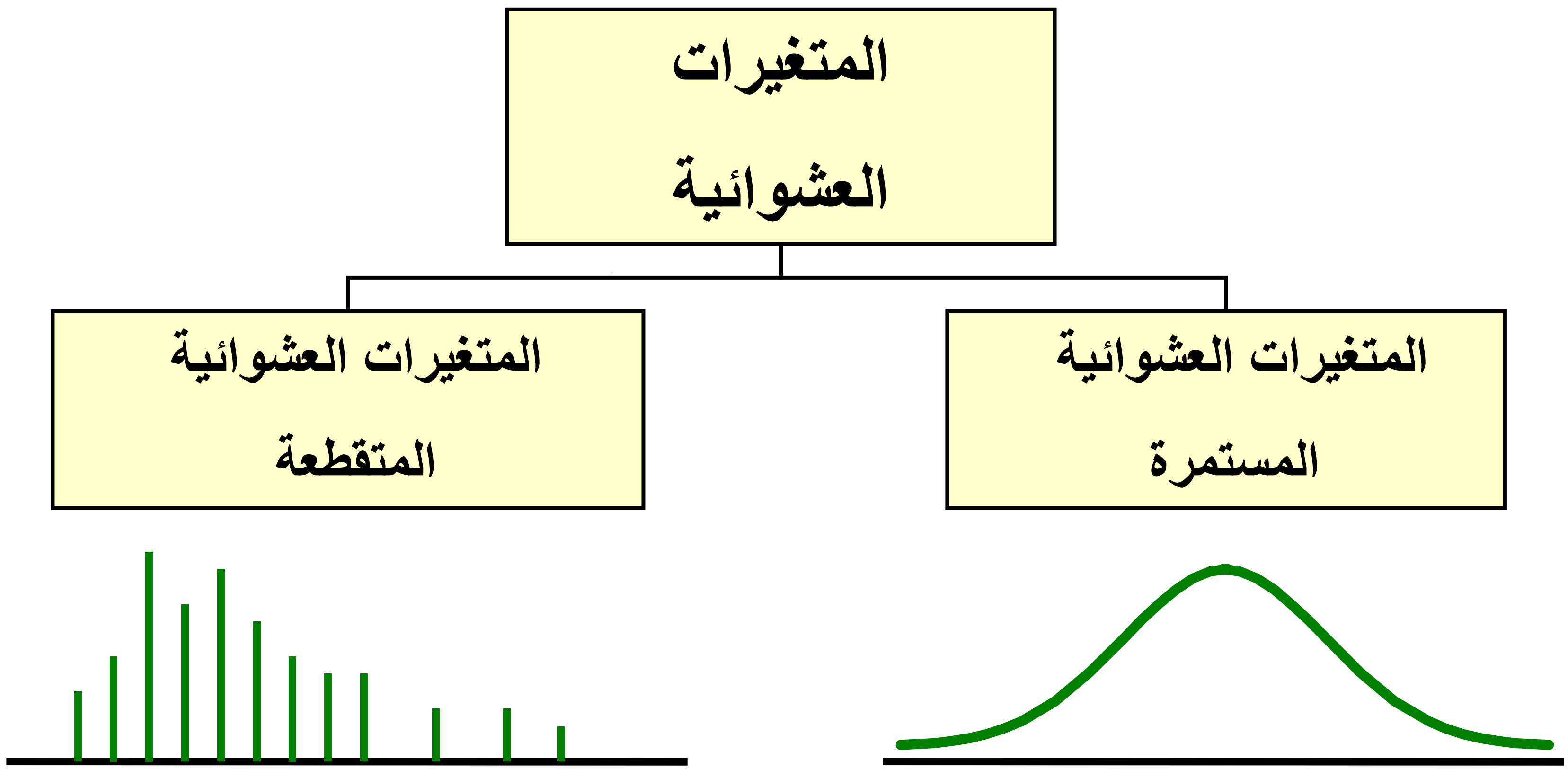
الاحتمالات 2: مقدمة للتوزيعات الاحتمالية

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

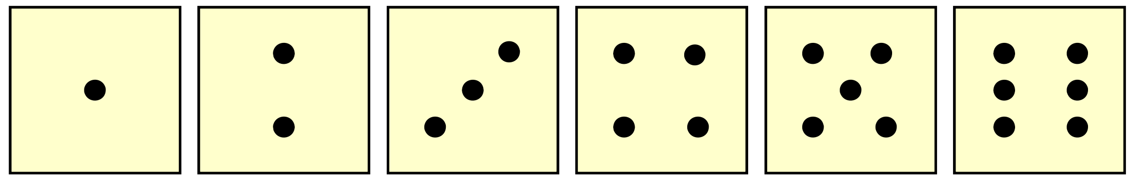
مقدمة إلى التوزيعات الاحتمالية

- **المتغير العشوائي**
  - يمثل قيمة رقمية محتملة من حدث عشوائي



Discrete Probability Distribution التوزيعات الاحتمالية المتقطعة

تتميز بأنها لا تأخذ إلا قيماً تمثل أرقاماً قابلة للعد.



- ارم حجر نرد مرتين، و ارمز بـ x لعدد المرات التي يظهر فيها الرقم 4. و هكذا فالقيم التي يمكن أن يأخذها متغيرنا العشوائي هذا محصورة بالمجموعة {0, 1, 2}

- ارم قطعة نقدية خمس مرات.

  ارمز بـ x لعدد المرات التي تظهر فيها الطرّة، و بالتالي فإن القيمة التي يمكن أن يأخذها متغيرنا هذا هي 1 أو 2 أو 3 أو 4 أو 5.

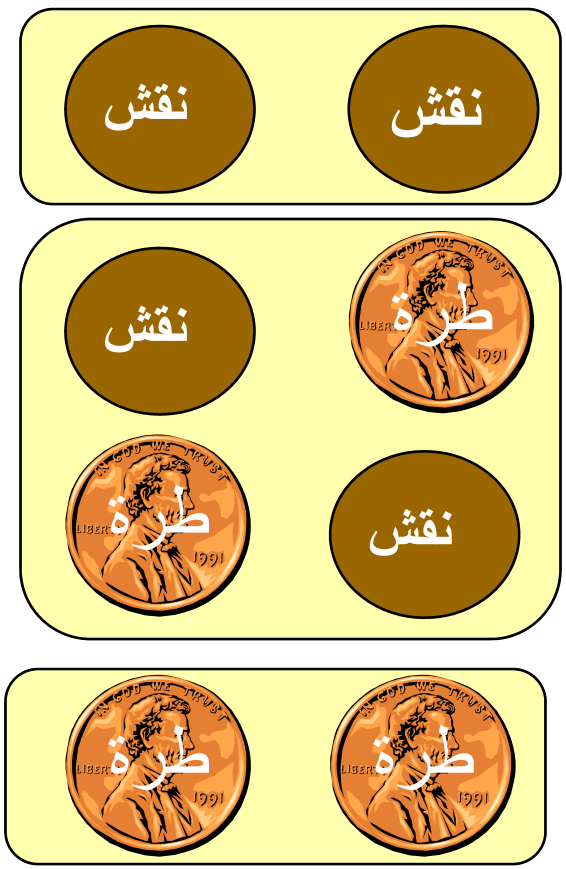
  ويمكن اختصار " أو 2 أو 3 أو 4 أو 5" بأن نكتب:

  (x = 0, 1, 2, 3, 4, or 5)

التوزيعات الاحتمالية المتقطعة Discrete Probability Distribution

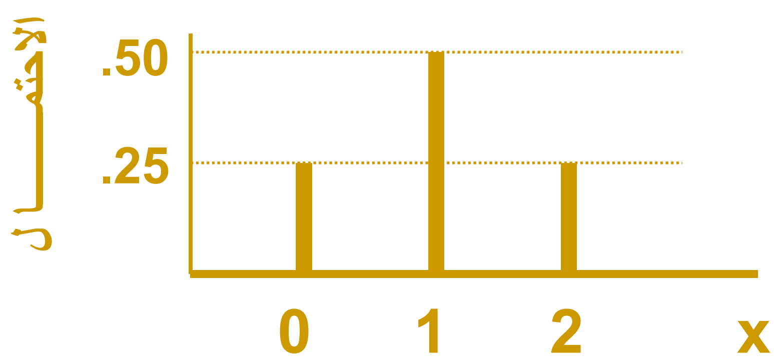
تجربة: ارم قطعتين نقديتين. و ليكن عدد مرات ظهور الطرة x

أربع نتائج محتملة



التوزيع الاحتمالي الموافق

x	قيمة الاحتمال
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25



التوزيع الاحتمالي المتقطع

- قائمة بكافة الثنائيات $[\ x_i\ ,\ P(x_i)\ ]$ الممكنة.

قيمة المتغير العشوائي (النتيجة) $= x_i$

الاحتمال المرتبط بالقيمة $= P(x_i)$

- الـ x_i's متنافية تبادلياً، أي لا يوجد تراكب.
- الـ x_i's تغطي كافة القيم الممكنة (لا تهمل أية قيمة محتملة لـ x_i)
- $0 \le P(x_i) \le 1$ من أجل أي x_i.
- $\Sigma\ P(x_i) = 1$

التوزيع الاحتمالي المتقطع
قياسات تلخيصية

- **القيمة المتوقعة** لتوزيع متقطع:

(معدل مثقل)

$$E(x) = \Sigma x_i P(x_i)$$

مثال: بالعودة لمثال القطعة النقدية، أحسب القيمة المتوقعة لـ x

x	P(x)
0	.25
1	.50
2	.25

E(x) = (0 x .25) + (1 x .50) + (2 x .25) = 1.0

التوزيع العشوائي المتقطع
قياسات تلخيصية (تابع)

- الانحراف المعياري لتوزيع متقطع:

$$\sigma_x = \sqrt{\sum [x - E(x)]^2 P(x)}$$

حيث:

$E(x)$ = القيمة المتوقعة للمتغير العشوائي

x = قيم المتغير العشوائي

$P(x)$ = احتمال أن يأخذ المتغير العشوائي القيمة x

التوزيع العشوائي المتقطع
قياسات تلخيصية *(تابع)*

- ارم قطعتين نقديتين و ليكن لديك متغير عشوائي يمثل القيم المحتملة لـ "كم طرة" يمكن أن تظهر من تجربة الرميتين. تذكر مما سبق أن القيمة المتوقعة لهذا المتغير هي 1.

$$\sigma_x = \sqrt{\sum [x - E(x)]^2 P(x)}$$

$$\sigma_x = \sqrt{(0-1)^2(.25) + (1-1)^2(.50) + (2-1)^2(.25)} = \sqrt{.50} = .707$$

العدد الممكن الحصول عيه من "الطرات"

= 0, 1, or 2

متغيرين عشوائيين متقطعين

- القيمة المتوقعة لمجموع متغيرين عشوائيين متقطعين:

$$E(x + y) = E(x) + E(y)$$
$$= \Sigma\, x\, P(x) + \Sigma\, y\, P(y)$$

لاحظ أنه يمكنك تذكر هذه القاعدة بالقول:

توقع المجموع = مجموع التوقعات

التباين المشترك

- **التباين المشترك بين متغيرين عشوائيين متقطعين :**

$$\sigma_{xy} = \Sigma\ [x_i - E(x)][y_j - E(y)]P(x_iy_j)$$

حيث:

القيم الممكنة للمتغير العشوائي المتقطع$(x_i)= x_i$

القيم الممكنة للمتغير العشوائي المتقطع $(y_j)= y_j$

الاحتمال المشترك للحصول على القيم x_i و y_j (بآن معاً)= $P(x_i\ ,y_j)$

تفسير التباين المشترك

- **التباين المشترك بين متغيرين عشوائيين متقطعين:**

$\sigma_{xy} > 0 \longrightarrow$ المتغيرين يميلان للتحرك بـ **نفس الاتجاه**

$\sigma_{xy} < 0 \longrightarrow$ المتغيرين يميلان للتحرك بـ **اتجاهين متعاكسين**

$\sigma_{xy} = 0 \longrightarrow$ المتغيرين **لا تربطهما علاقة** مشتركة

معامل الارتباط

- **يعبر عن قوة الارتباط الخطي بين متغيرين:**

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

حيث:

ρ = correlation coefficient	معامل الارتباط
σ_{xy} = covariance between x and y	التباين المشترك لـ x و y
σ_x = standard deviation of variable x	الانحراف المعياري لـ x
σ_y = standard deviation of variable y	الانحراف المعياري لـ y

تفسير معامل الارتباط

- **معامل الارتباط يقع بين -1 و +1**

لا ارتباط خطي بين المتغيرين ⟵ $\rho = 0$

الخطية بين المتغيرين أقوى.كلما ابتعدت قيمة هذا المعامل عن الصفر كلما كانت العلاقة

علاقة خطية إيجابية تامة بين المتغيرين ⟵ $\rho = +1$

علاقة خطية سلبية تامة بين المتغيرين ⟵ $\rho = -1$

في نهاية المحاضرة:

- تأكد أنك تفهم تماماً المقصود بكل مما يلي:
  - المتغير العشوائي.
  - التوزيع الاحتمالي.
  - توزيع المتغير العشوائي المتقطع.
  - القياسات التلخيصية للتوزيع المتقطع.
- تأكد أنك تستطيع حساب و تفسير:
  - القيمة المتوقعة لتوزيع متقطع.
  - الانحراف المعياري لتوزيع متقطع.
  - التباين المشترك لمتغيرين عشوائيين.
  - معامل الارتباط الخطي بين متغيرين.

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة السادسة /القسم النظري/

الاحتمالات 3: أهم التوزيعات الاحتمالية المتقطعة

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

التوزيعات الاحتمالية

- التوزيعات الاحتمالية
  - التوزيعات الاحتمالية المستمرة
    - الطبيعي
    - المنتظم
    - الأسّي
  - التوزيعات الاحتمالية المتقطعة
    - ثنائي الحد (بينوميال)
    - بواسون
    - Hypergeometric

توزيع بينوميال

- التوزيعات الاحتمالية
  - التوزيعات الاحتمالية المستمرة
    - الطبيعي
    - المنتظم
    - الأسّي
  - التوزيعات الاحتمالية المتقطعة
    - ثنائي الحد (بينوميال)
    - بواسون
    - Hypergeometric

The Binomial Distribution توزيع بينوميال

خصائص توزيع بينوميال:

- المحاولة الواحدة إما أن "تفشل" أو أن "تنجح"، أي هناك نتيجتين اثنتين محتملتين.
- هناك عدد ثابت n من المحاولات المتطابقة .
- محاولات التجربة مستقلة عن بعضها البعض.
- احتمال النجاح P يبقى ثابتاً من محاولة لأخرى.
- بالتالي احتمال الفشل هو $(1 – p)$ و عادة ما يرمز له بـ q .

أمثلة على حالات استخدام توزيع بينوميال

- طلبات العمل إما أن تقبل أو أن ترفض.
- عروض الأسعار إما أن تفوز بالمناقصة أو أن تفشل.
- استبيانات التسويق تتضمن أسئلة من نوع "نعم أم لا"
- طلبات الحصول على إقامة مؤقتة في بلد أجنبي إما أن تقبل أو أن ترفض.

قاعدة العد: حالة التوافيق Combinations

- التوفيق هو نتيجة تجربة تنطوي على اختيار x عنصراً من أصل مجموعة مكونة من n عنصراً.

$$\boxed{C_x^n = \frac{n!}{x!(n-x)!}}$$

حيث:

$n! = n(n - 1)(n - 2) \ldots (2)(1)$

$x! = x(x - 1)(x - 2) \ldots (2)(1)$

$0! = 1$

الصيغة الرياضية لتوزيع بينوميال

$$P(x) = \frac{n!}{x!(n-x)!} p^{x} q^{n-x}$$

P(x) = احتمال x نجاحاً من n محاولة مع احتمال نجاح p لكل محاولة.

x = عدد "النجاحات" في العينة.

(x = 0, 1, 2, ..., n)

n = عدد المحاولات (حجم العينة)

q = (1 – p) = احتمال "الفشل"

مثال: ارم قطعة نقدية أربع مرات و ليكن:

x = عدد مرات ظهور الطرّة = # طرة

n = 4

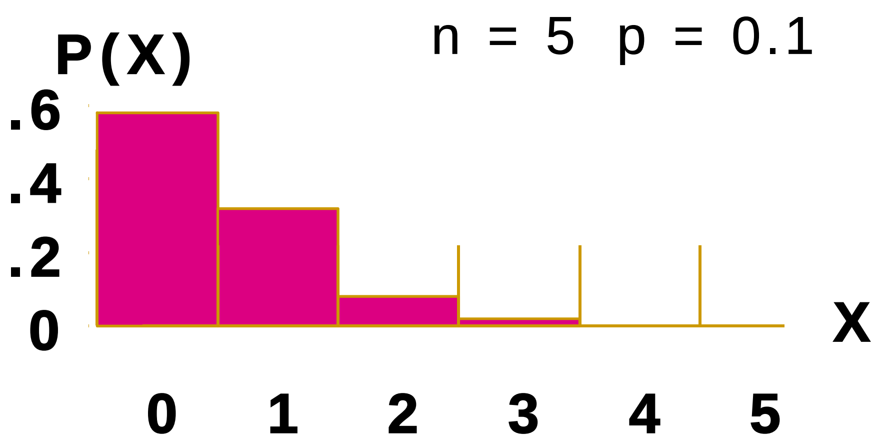
p = 0.5

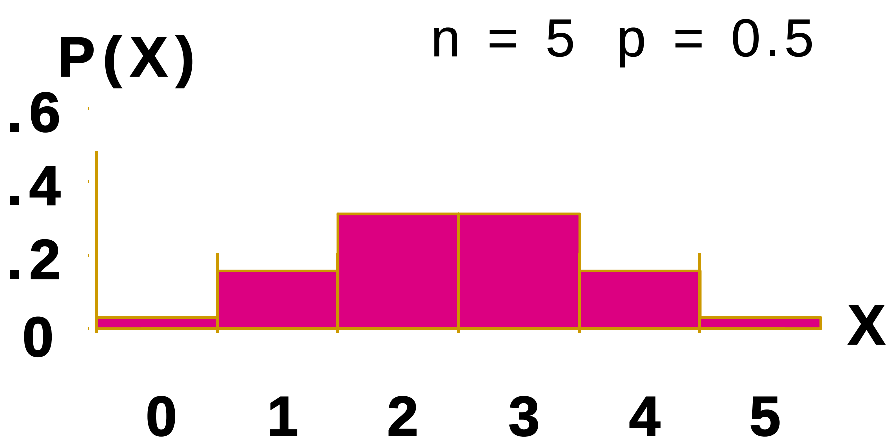
q = (1 - .5) = .5

x = 0, 1, 2, 3, 4

توزيع بينوميال

- يعتمد شكل توزيع بينوميال على قيمتي p و n





خصائص توزيع بينوميال

- الوسط $\mu = E(x) = np$
- التباين $\sigma^2 = npq$
- الانحراف المعياري $\sigma = \sqrt{npq}$

$P(x)$ = احتمال x نجاحاً من n محاولة مع احتمال نجاح p لكل محاولة.

x = عدد "النجاحات" في العينة $(x = 0, 1, 2, \ldots, n)$

n = عدد المحاولات (حجم العينة)

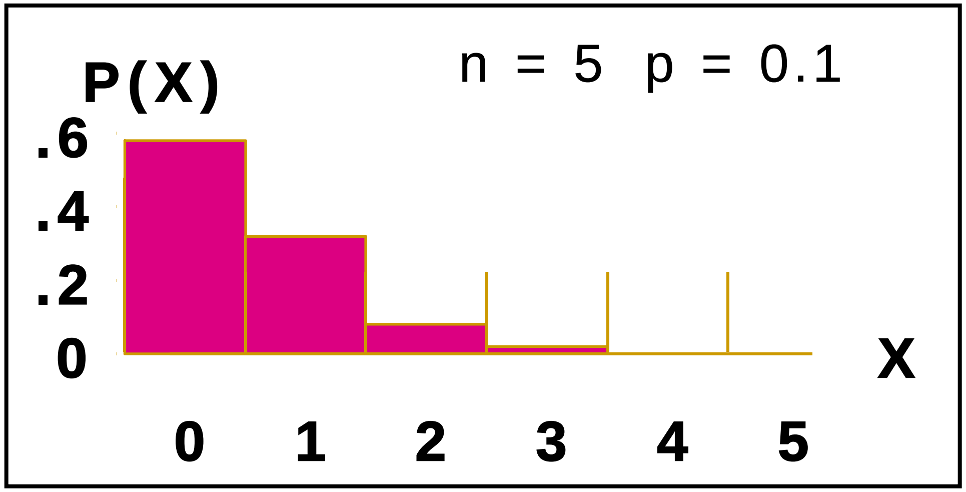
q = $(1 - p)$ = احتمال "الفشل"

خصائص توزيع بينوميال

Examples

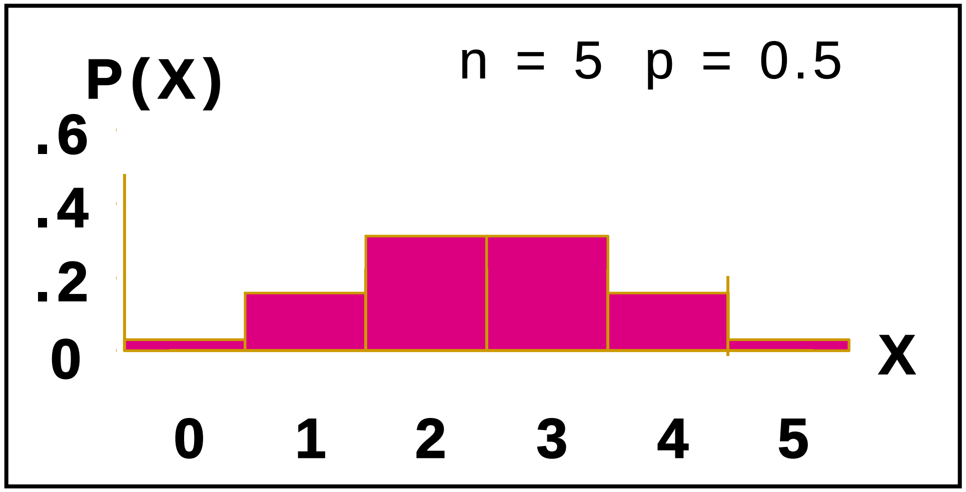
$$\mu = np = (5)(.1) = 0.5$$

$$\sigma = \sqrt{npq} = \sqrt{(5)(.1)(1-.1)}$$
$$= 0.6708$$



$$\mu = np = (5)(.5) = 2.5$$

$$\sigma = \sqrt{npq} = \sqrt{(5)(.5)(1-.5)}$$
$$= 1.118$$



استخدام جداول بينوميال

n = 10									
x	p=.15	p=.20	p=.25	p=.30	**p=.35**	p=.40	p=.45	p=.50	
0	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010	10
1	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098	9
2	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439	8
3	0.1298	0.2013	0.2503	0.2668	**0.2522**	0.2150	0.1665	0.1172	7
4	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051	6
5	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461	5
6	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051	4
7	0.0001	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172	3
8	0.0000	0.0001	**0.0004**	0.0014	0.0043	0.0106	0.0229	0.0439	**2**
9	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098	1
10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0
	p=.85	p=.80	**p=.75**	p=.70	p=.65	p=.60	p=.55	p=.50	**x**

مثلة:

n = 10, p = .35, x = 3: $P(x = 3|n = 10, p = .35) = .2522$

n = 10, p = .75, x = 2: $P(x = 2|n = 10, p = .75) = .0004$

The Poisson Distribution توزيع بواسون

- التوزيعات الاحتمالية
  - التوزيعات الاحتمالية المستمرة
    - الطبيعي
    - المنتظم
    - الأسّي
  - التوزيعات الاحتمالية المتقطعة
    - ثنائي الحد (بينوميال)
    - بواسون
    - Hypergeometric

The Poisson Distribution توزيع بواسون

خصائص توزيع بواسون:

- النتائج التي نهتم بها هي نتائج نادرة بالمقارنة مع النتائج المحتملة.
- متوسط عدد النتائج (محل الاهتمام) في وحدة الزمن أو المكان هو λ
- عدد النتائج (محل الاهتمام) عشوائي و الحصول على نتيجة معينة لا يؤثر فرص بقية النتائج في الحدوث.
- احتمال حدوث نتيجة معينة ضمن قطعة معينة (فترة زمنية محددة / مسافة مكانية محددة) هو نفسه من أجل كل القطع.

الصيغة الرياضية توزيع بواسون

$$P(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

حيث:

t = size of the segment of interest حجم القطعة محل الاهتمام

x = number of successes in segment of interest عدد النجاحات في القطعة محل الاهتمام

λ = expected number of successes in a segment of unit size

توقع عدد النجاحات أو العدد المتوقع للنجاحات في القطعة محل الاهتمام

e = base of the natural logarithm system 2.71828...)

أساس النظام العددي المبني على اللوغارتم الطبيعي

خصائص توزيع بواسون

- الوسط $\mu = \lambda t$
- التباين $\sigma^2 = \lambda t$
- الانحراف المعياري $\sigma = \sqrt{\lambda t}$

λ = number of successes in a segment of unit size

توقع عدد النجاحات أو العدد المتوقع للنجاحات في القطعة محل الاهتمام

t = the size of the segment of interest

حجم القطعة محل الاهتمام

استخدام جداول بواسون

x	λt								
	0.10	0.20	0.30	0.40	**0.50**	0.60	0.70	0.80	0.90
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	**0.0758**	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

$P(x = 2)$; $\lambda = .05$; $t = 100$

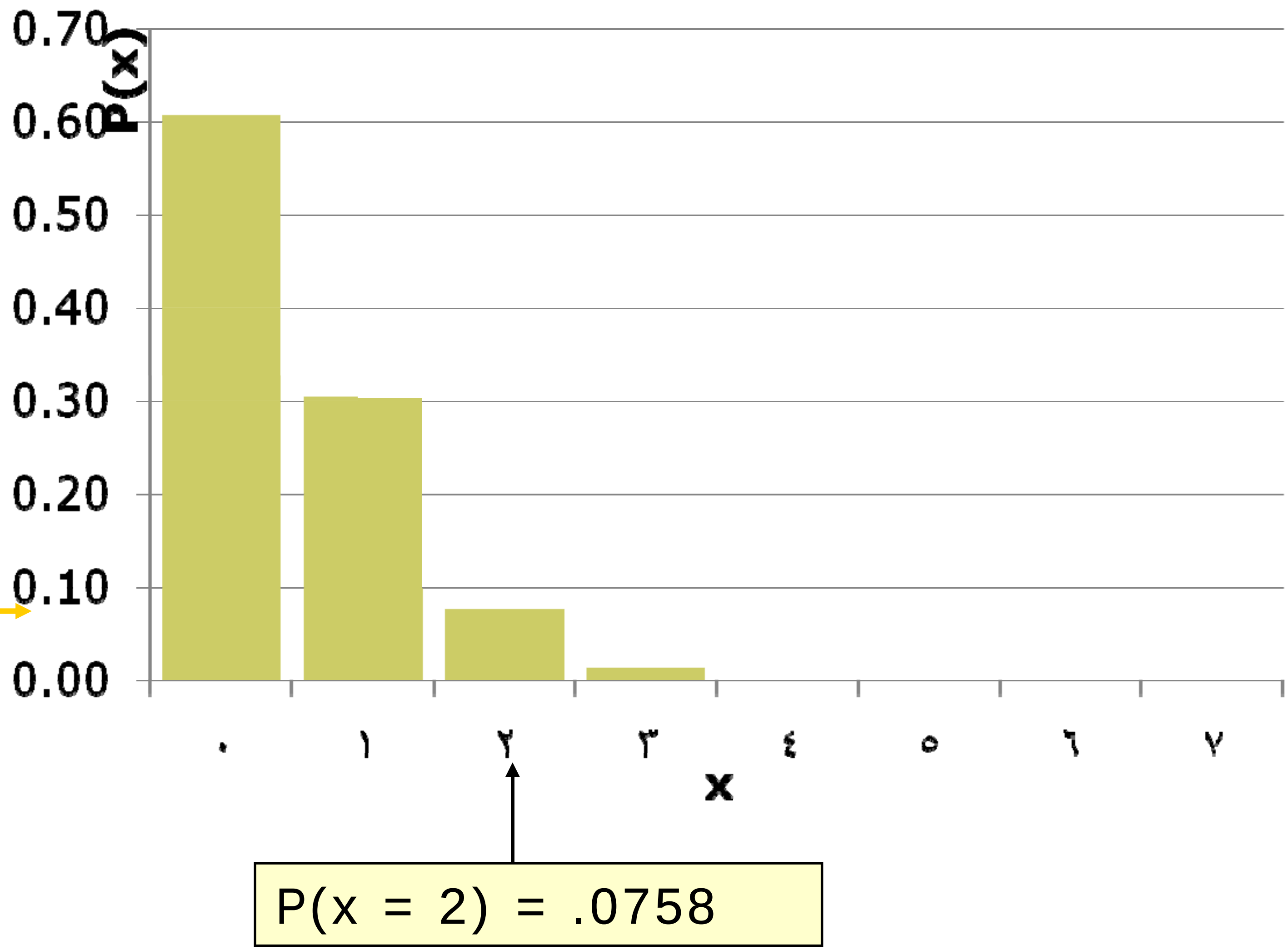
$$P(x=2) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} = \frac{(0.50)^2 e^{-0.50}}{2!} = .0758$$

التمثيل البياني لاحتمالات بواسون

Graphically:

λ = .05 and t = 100

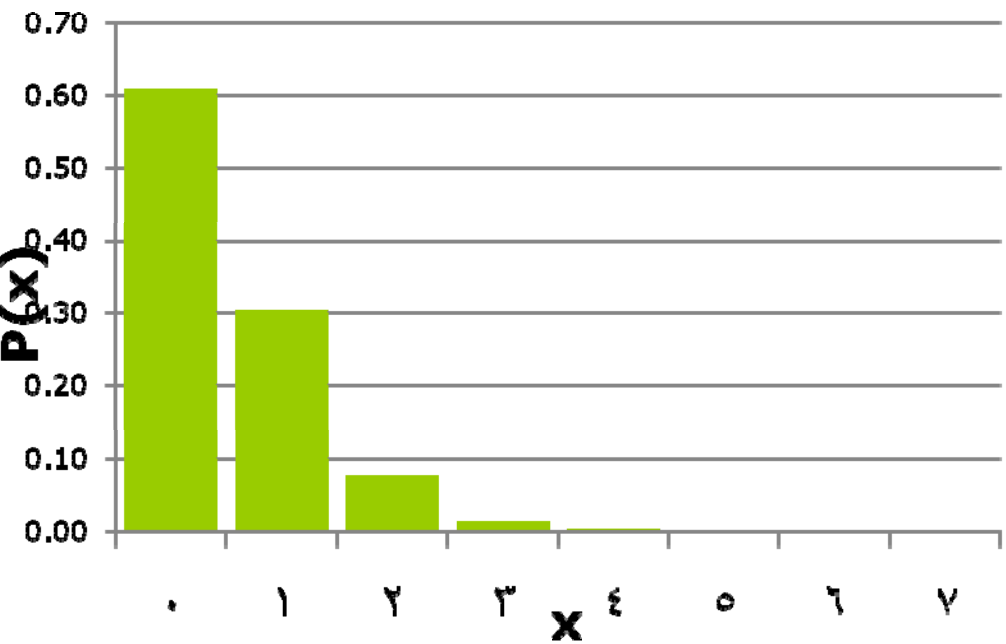
X	λt = 0.50
0	0.6065
1	0.3033
2	0.0758
3	0.0126
4	0.0016
5	0.0002
6	0.0000
7	0.0000



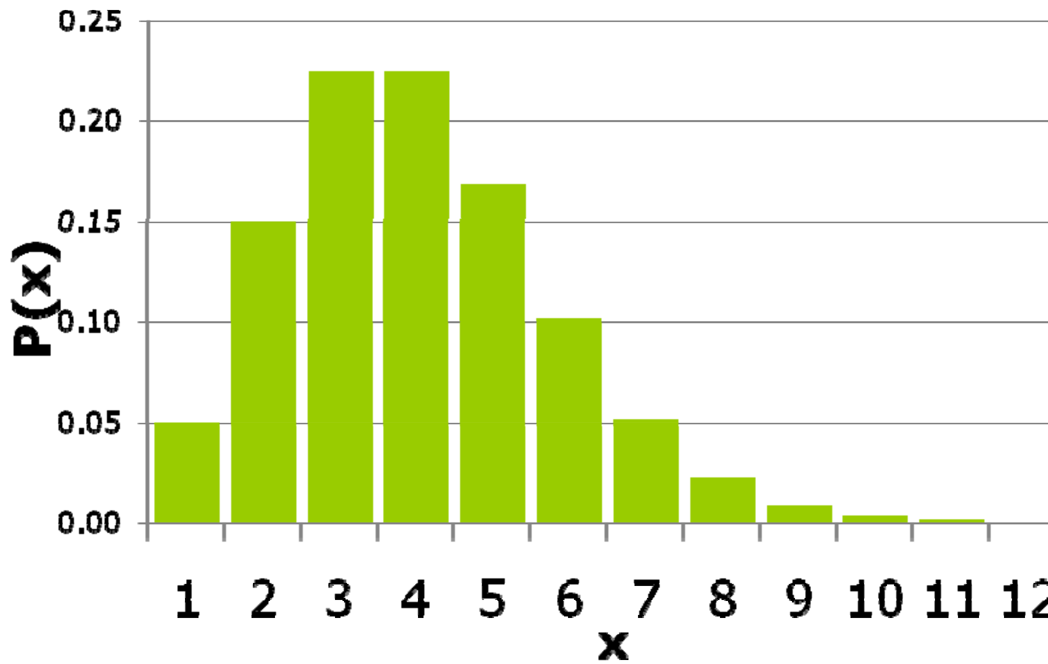
شكل توزيع بواسون

- يعتمد شكل التوزيع على البارامترات λ و t

$\lambda t = 0.50$



$\lambda t = 3.0$



Hypergeometric التوزيع الهندسي

التوزيعات الاحتمالية

- التوزيعات الاحتمالية المتقطعة
  - ثنائي الحد (بينوميال)
  - بواسون
  - Hypergeometric
- التوزيعات الاحتمالية المستمرة
  - الطبيعي
  - المنتظم
  - الأسّي

Hypergeometric التوزيع الهندسي

- لدينا n محاولة لعينة مسحوبة من مجتمع منتهٍ من الحجم N.
- السحب بدون إعادة.
- المحاولات غير مستقلة.
- الاهتمام منصب على إيجاد احتمال الحصول على x نجاحاً من العينة حيث المجتمع ينطوي على X نجاحاً.

الصيغة الرياضية للتوزيع الهندسي

ثمة نتيجتين اثنتين محتملتين من كل محاولة: نجاح أو فشل.

$$P(x) = \frac{C_{n-x}^{N-X} \cdot C_{x}^{X}}{C_{n}^{N}}$$

حيث:

N = population size حجم المجتمع

X = number of successes in the population عدد النجاحات في المجتمع

n = sample size حجم العينة

x = number of successes in the sample عدد النجاحات في العينة

n – x = number of failures in the sample عدد حالات الفشل في العينة

الصيغة الرياضية للتوزيع الهندسي

- مثال:

تم انتقاء 3 من 10مصابيح بينها 4 معطلة، فما هو احتمال أن يكون اثنان من المصابيح الثلاثة معطلين؟

$$N = 10 \qquad n = 3$$
$$X = 4 \qquad x = 2$$

$$P(x=2) = \frac{C_{n-x}^{N-X} \; C_x^X}{C_n^N} = \frac{C_1^6 \; C_2^4}{C_3^{10}} = \frac{(6)(6)}{120} = 0.3$$

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة السابعة /القسم النظري/

الاحتمالات 4: أهم التوزيعات الاحتمالية المستمرة

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

مخرجات المحاضرة

تهدف هذه المحاضرة إلى تطوير المهارات التالية:

- تمييز و استخدام أشهر التوزيعات الاحتمالية المستمرة:
  - الأسّي، المنتظم، الطبيعي
  - normal, uniform, exponential
- إيجاد الاحتمالات باستخدام الصيغ الرياضية و الجداول المعيارية.
- تمييز الحالات المختلفة لتطبيق التوزيعات الاحتمالية المختلفة.
- تطبيق التوزيعات الاحتمالية في اتخاذ القرار.

التوزيعات الاحتمالية

- التوزيعات الاحتمالية
  - التوزيعات الاحتمالية المستمرة
    - الطبيعي
    - المنتظم
    - الأسّي
  - التوزيعات الاحتمالية المتقطعة
    - ثنائي الحد (بينوميال)
    - بواسون
    - Hypergeometric

التوزيع الطبيعي

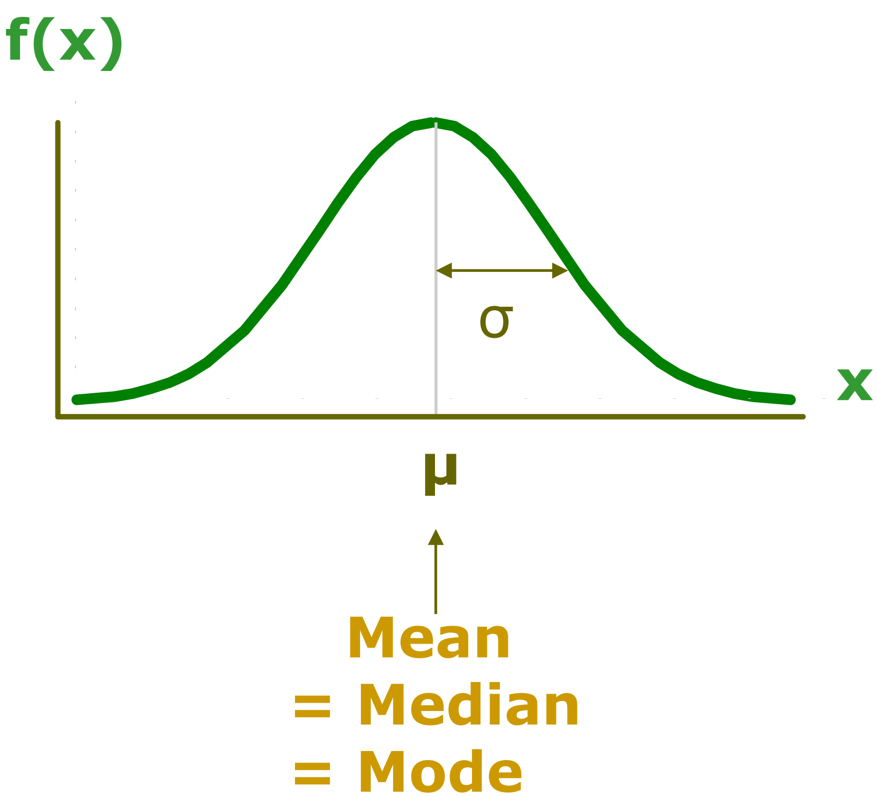
- ذو شكل جرسي
- متناظر
- الوسط = الوسيط = المنوال

μ يحدد الموضع

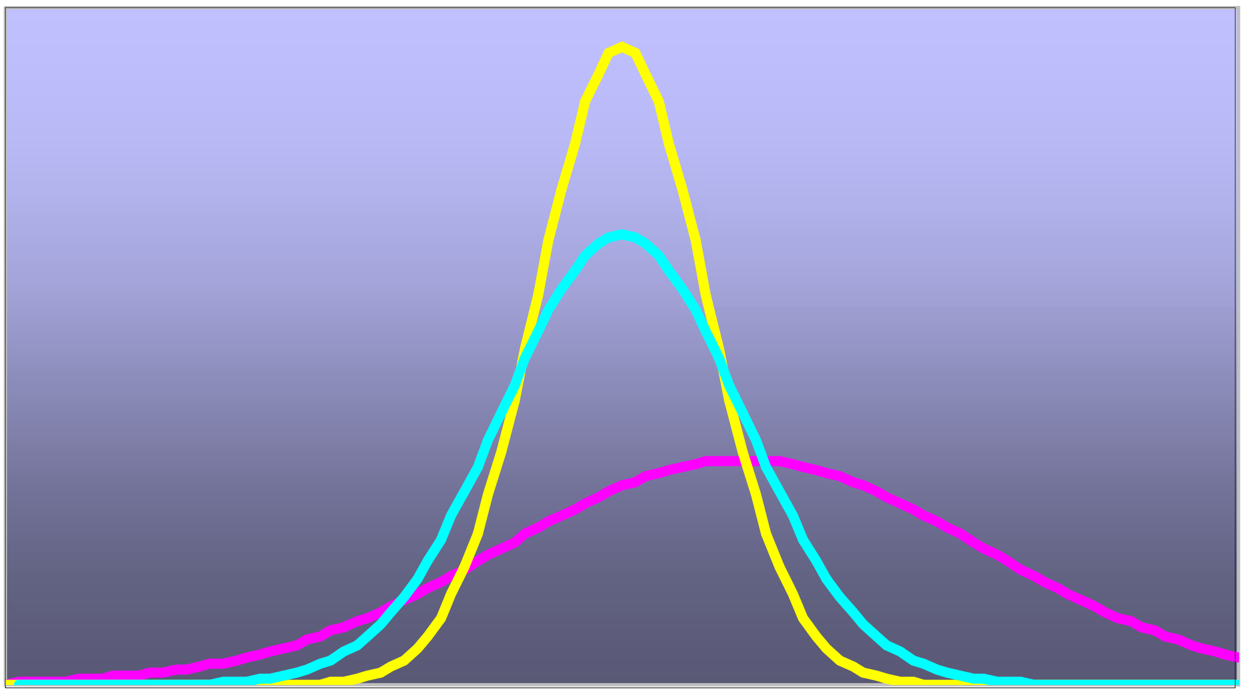
σ يحدد الانتشار

للمتغير العشوائي مدى نظري غير منتهٍ :

$-\infty$ to $+\infty$

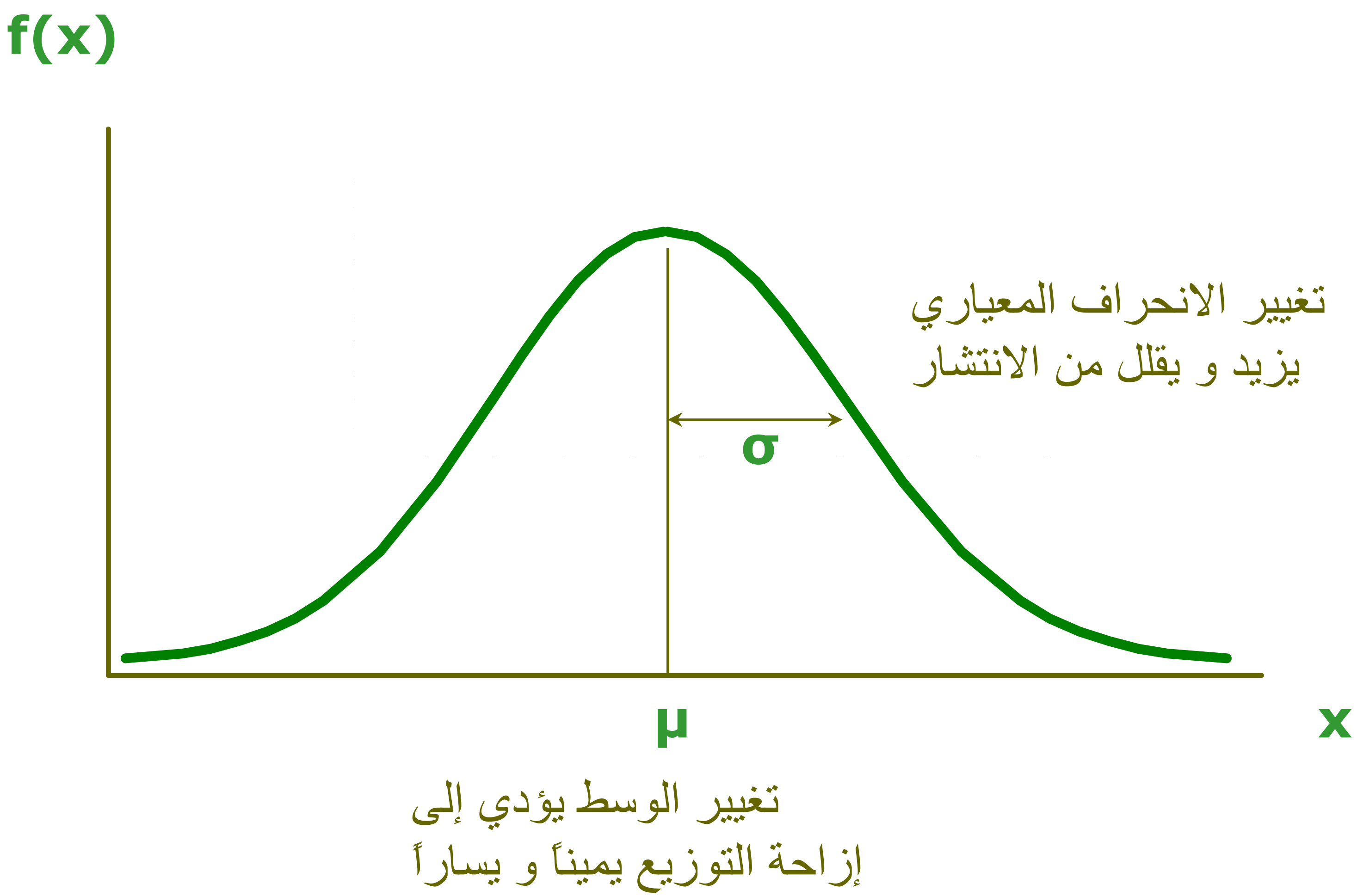


هناك توزيعات طبيعية كثيرة !



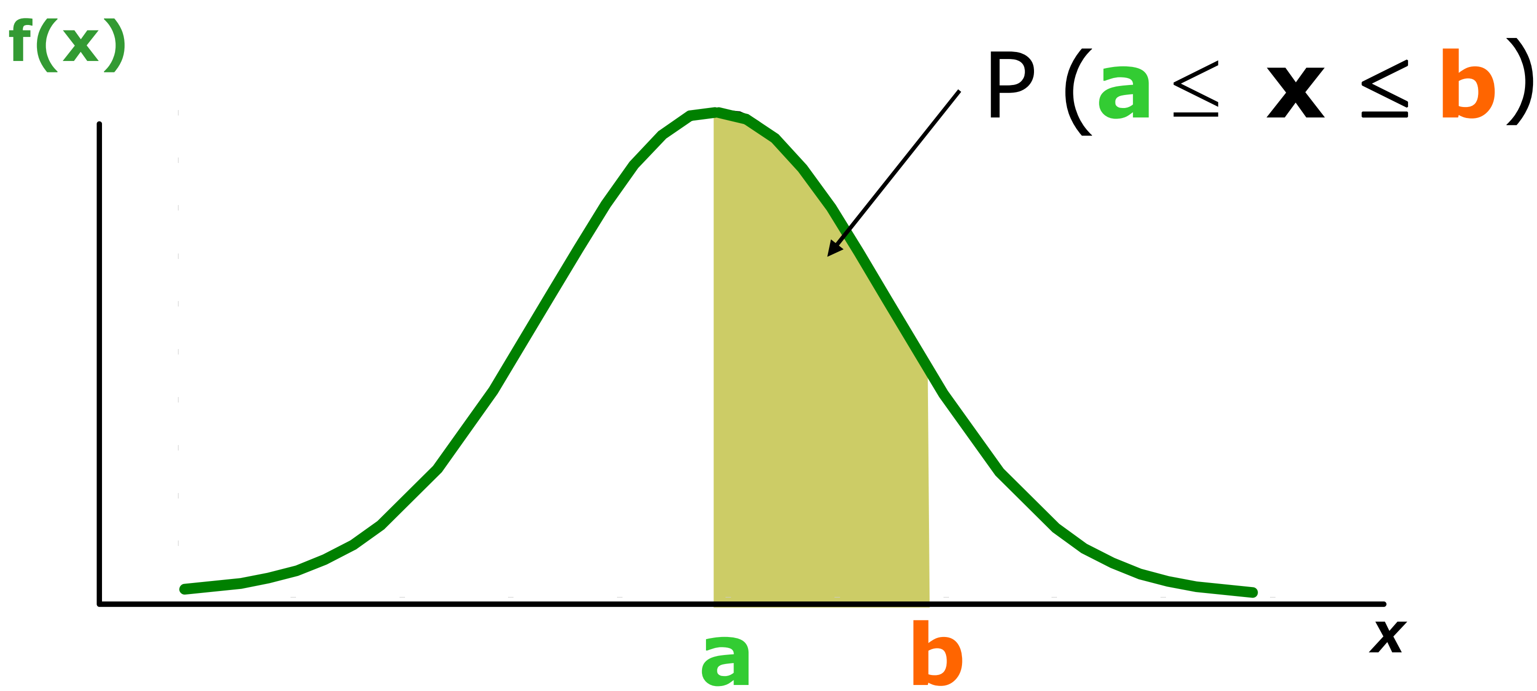
نحصل على أشكال عديدة للتوزيع الطبيعي عن طريق تغيير قيم الوسط و الانحراف المعياري

شكل التوزيع الطبيعي



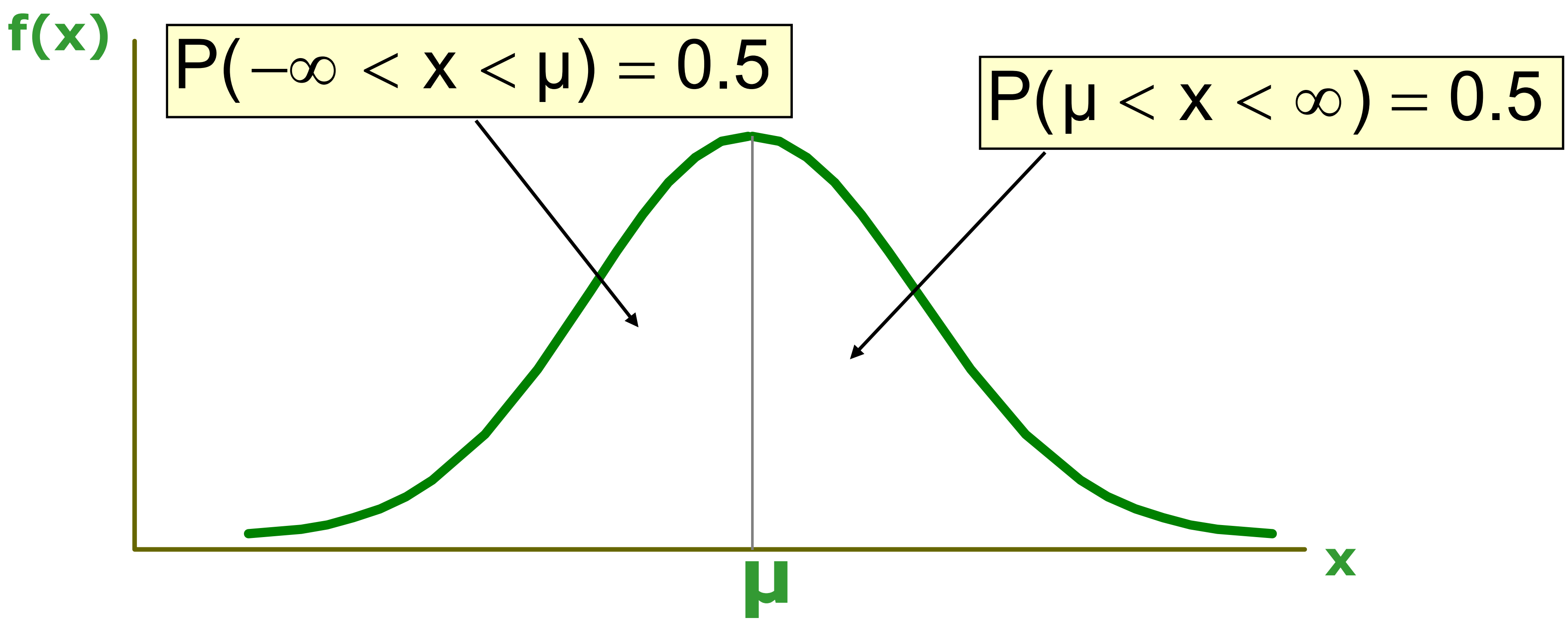
إيجاد الاحتمالات الطبيعية

يقاس الاحتمال بمقدار المساحة تحت المنحنى



الاحتمال كمساحة تحت منحنى

المساحة الكلية تحت المنحنى تساوي الواحد. و بما أن المنحنى متناظر فإن نصف المساحة يقع أسفل الوسط و نصفها الأخر فوقه

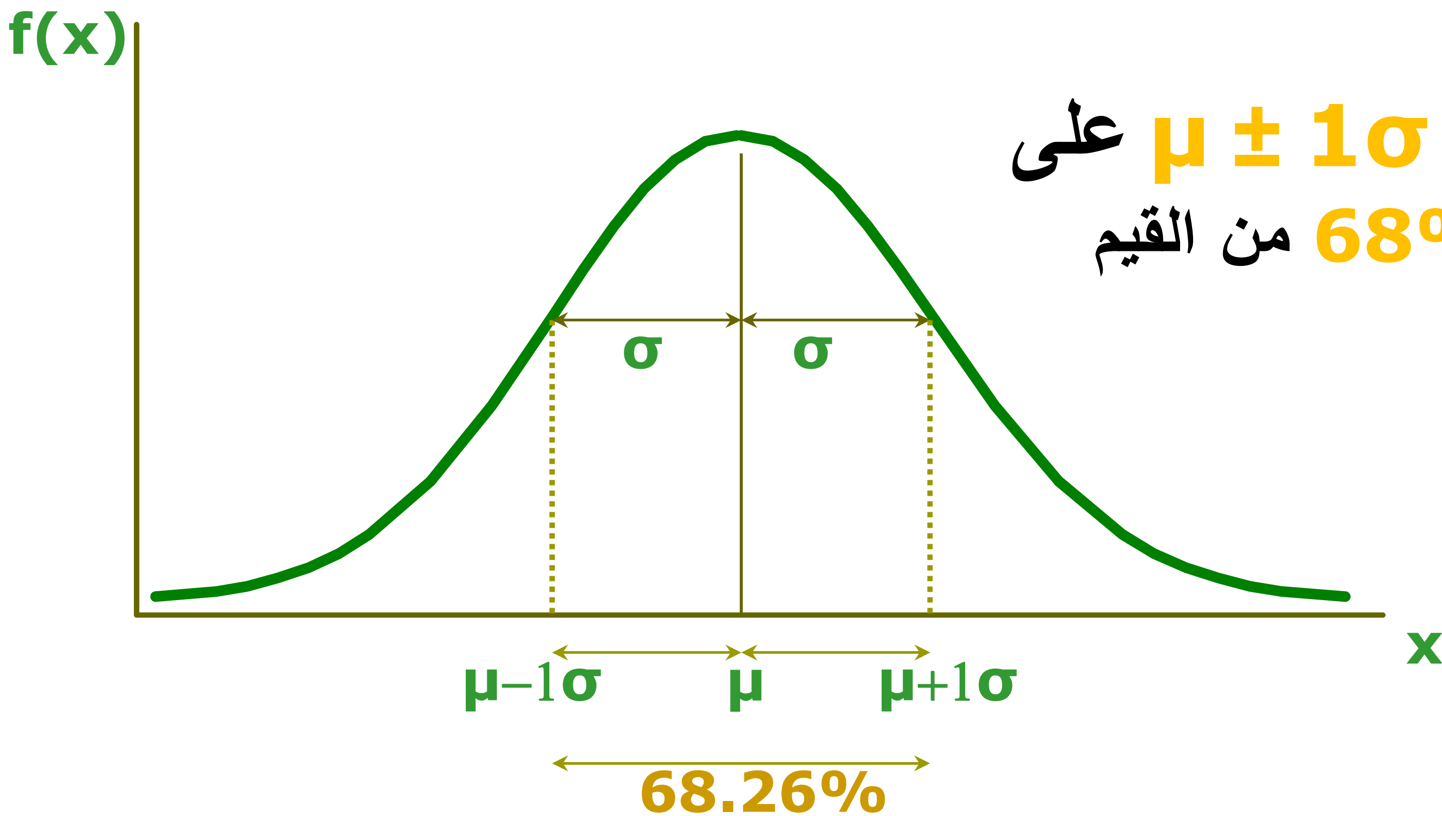


$$P(-\infty < x < \infty) = 1.0$$

القاعدة النظرية

ماالذي يمكننا قولة حول توزّع القيم حول الوسط؟

ثمة قواعد عامة يمكن الاعتماد عليها للإجابة على هذا السؤال.

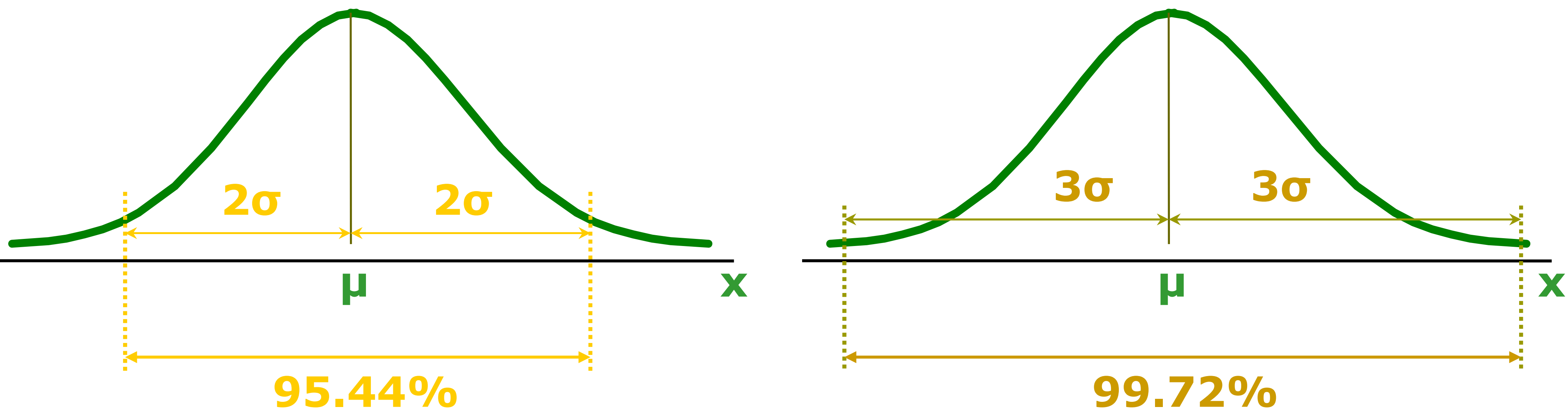


يحتوي المجال $\mu \pm 1\sigma$ على حوالي 68% من القيم

القاعدة النظرية

تتمة

- يحتوي المجال $\mu \pm 2\sigma$ على حوالي **95.44%** من القيم
- يحتوي المجال $\mu \pm 3\sigma$ على حوالي **99.72%** من القيم



أهمية القاعدة النظرية

- إذا كانت القيمة أبعد من انحرافين معياريين عن وسط التوزيع الطبيعي فإن هذه القيمة تعد:

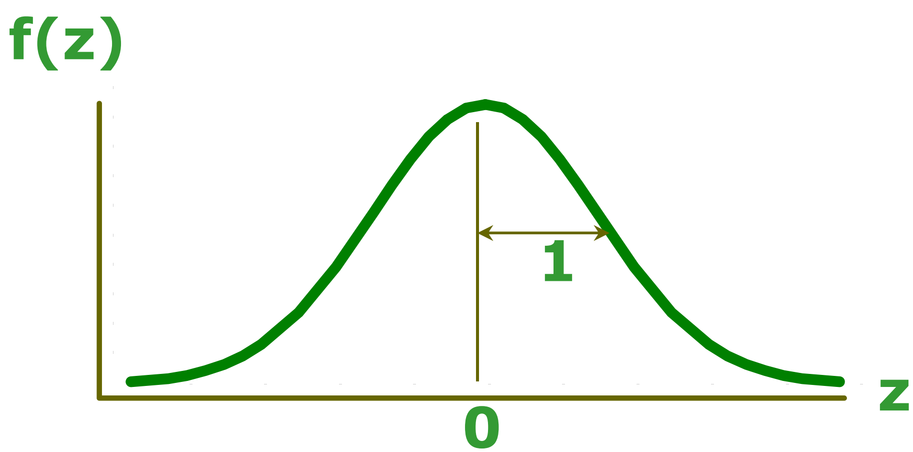
  بعيدة عن الوسط **far from the mean**

- إن فرصة (احتمال) أن تقع قيمة ما أبعد من ثلاثة انحرافات معيارية عن وسط التوزيع الطبيعي تعد:

  ضئيلة أو **مستبعدة** **highly unlikely**

التوزيع الطبيعي المعياري

- و يُعرف أيضاً بـ توزيع "زي" z distribution
- الوسط يساوي الصفر بالتعريف
- الانحراف المعياري يساوي الواحد



القيم الواقعة فوق الوسط هي قيم "زي" موجبة

و أما الواقعة تحت الوسط فهي قيم "زي" سالبة

الترجمة إلى الطبيعي المعياري

- يمكن تحويل أي توزيع طبيعي إلى توزيع طبيعي معياري.
- للقيام بذلك نقوم بتحويل قيم x إلى قيم z
- لترجمة قيم x إلى قيم z نطبق صيغة المعايرة التالية:

$$z = \frac{x - \mu}{\sigma}$$

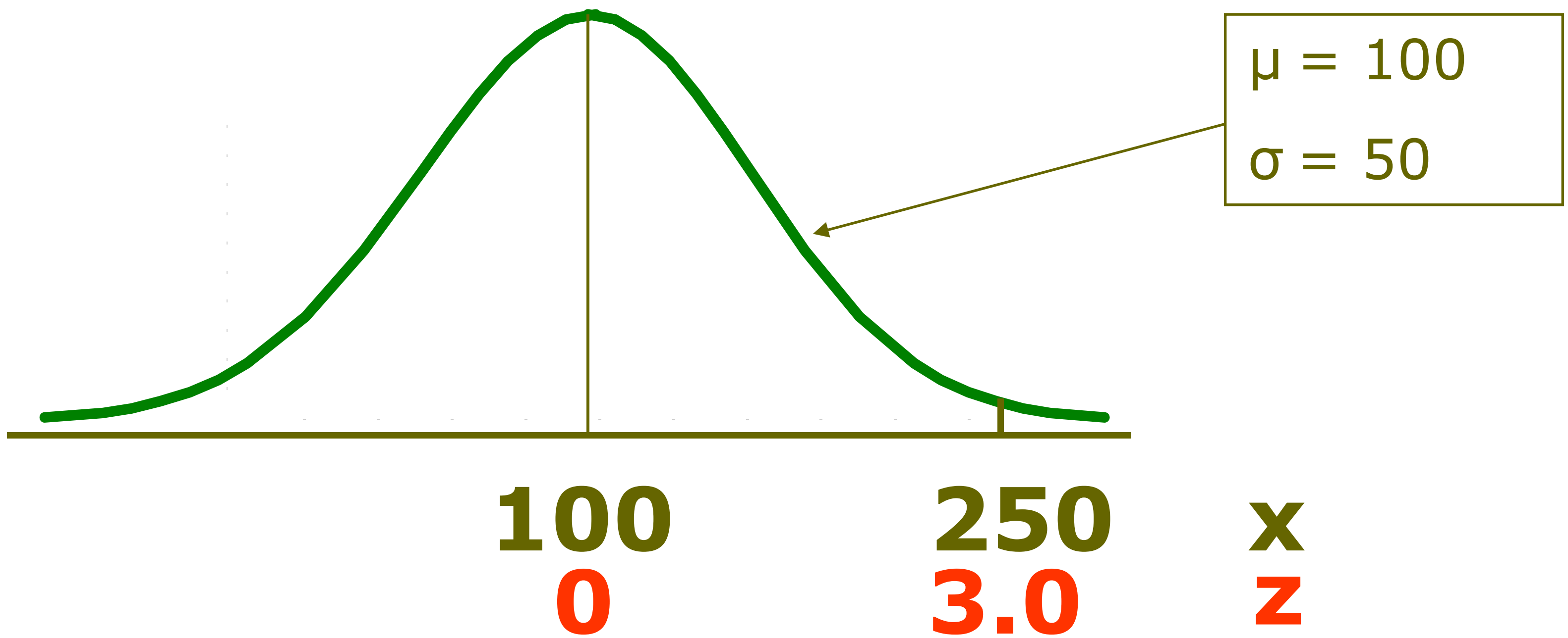
مثال

- إذا كان x موزعاً وفق التوزيع الطبيعي بوسط 100 و انحراف معياري 50 فما هي قيمة z من أجل x = 250

$$z = \frac{x - \mu}{\sigma} = \frac{250 - 100}{50} = 3.0$$

- أي أن z تقع على بعد ثلاثة انحرافات معيارية فوق (يمين) الوسط

مقارنة وحدات x مع وحدات z



لاحظ أن التوزيع الطبيعي لا يتغير بعد تحويله إلى الطبيعي المعياري.

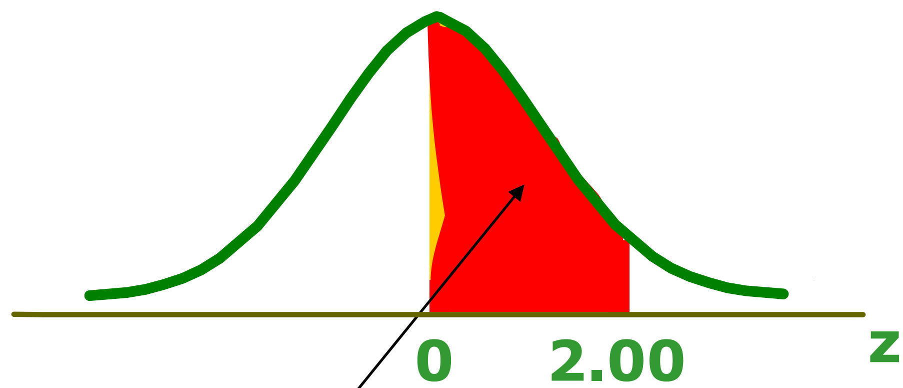
ما يتغير هو وحدات القياس و حسب.

أي أننا نستطيع التعبير عن المشكلة المطروحة باستخدام الوحدات الأصلية أو المعيارية.

الجدول الطبيعي المعياري

- يعطي الجدول الطبيعي المعياري قيم الاحتمالات المحددة بين الوسط (صفر) و بين قيمة مرغوبة من z. لاحظ أن قيمة z محددة بدقة مقدارها خانتين بعد الفاصلة، حيث تستخدم الأسطر للبحث عن الخانة العشرية و الأعمدة للبحث عن المئوية.

z	0.00	0.01	0.02	…
0.1				
0.2				
⋮				
2.0	.4772			



Example:

$P(0 < z < 2.00) = .4772$

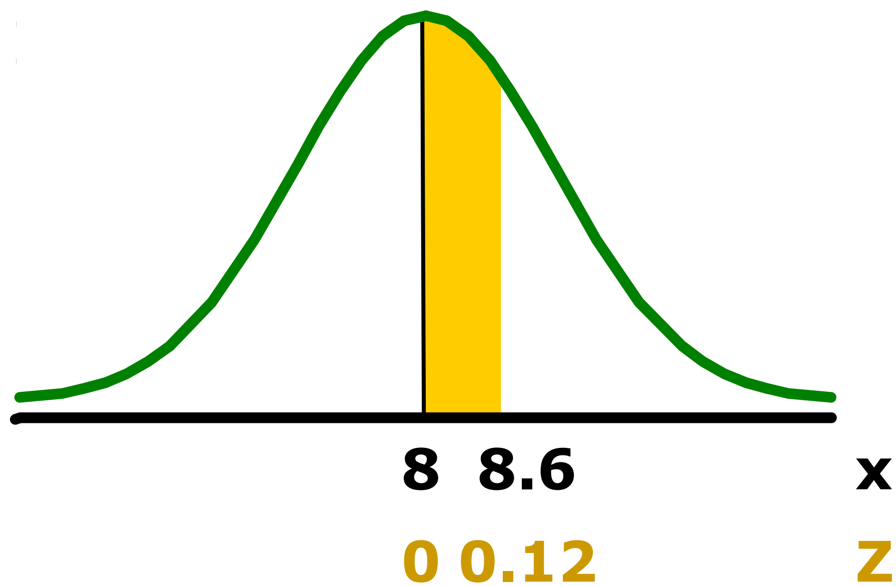
مثال على جدول Z

- افترض أن x يتبع التوزيع الطبيعي بوسط يساوي خمسة و انحراف معياري يساوي ثمانية.

أحسب قيم Z

$$z = \frac{x - \mu}{\sigma} = \frac{8 - 8}{5} = 0$$

$$z = \frac{x - \mu}{\sigma} = \frac{8.6 - 8}{5} = 0.12$$



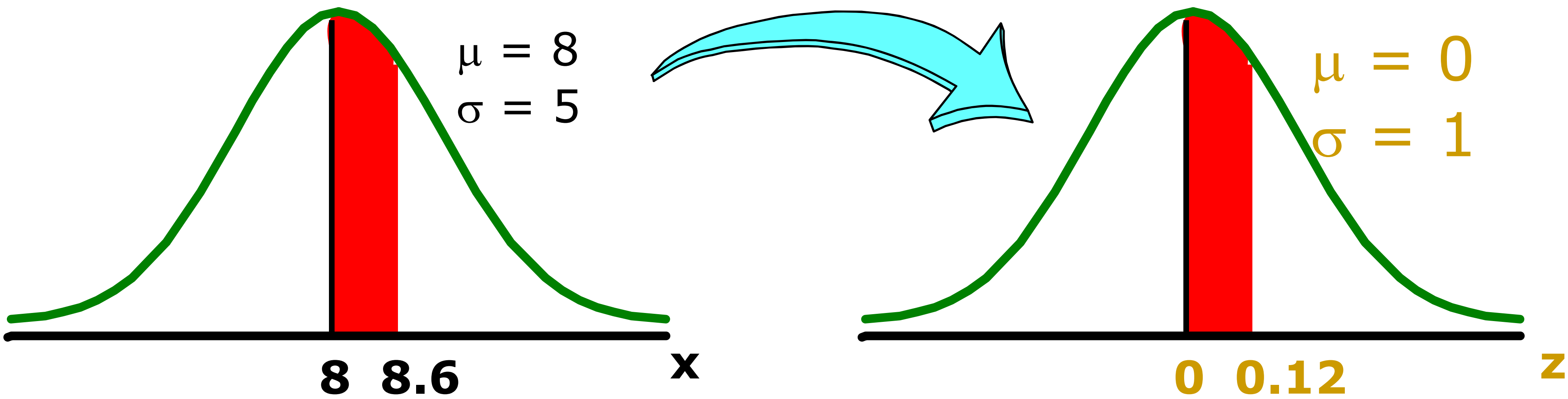
$P(8 < x < 8.6)$

$= P(0 < z < 0.12)$

مثال على جدول Z

تتمة

- افترض أن x يتبع التوزيع الطبيعي بوسط يساوي خمسة و انحراف معياري يساوي ثمانية.
- أوجد $P(8 < x < 8.6)$



$P(8 < x < 8.6)$ $P(0 < z < 0.12)$

Solution: Finding P(0 < z < 0.12)

Standard Normal Probability Table

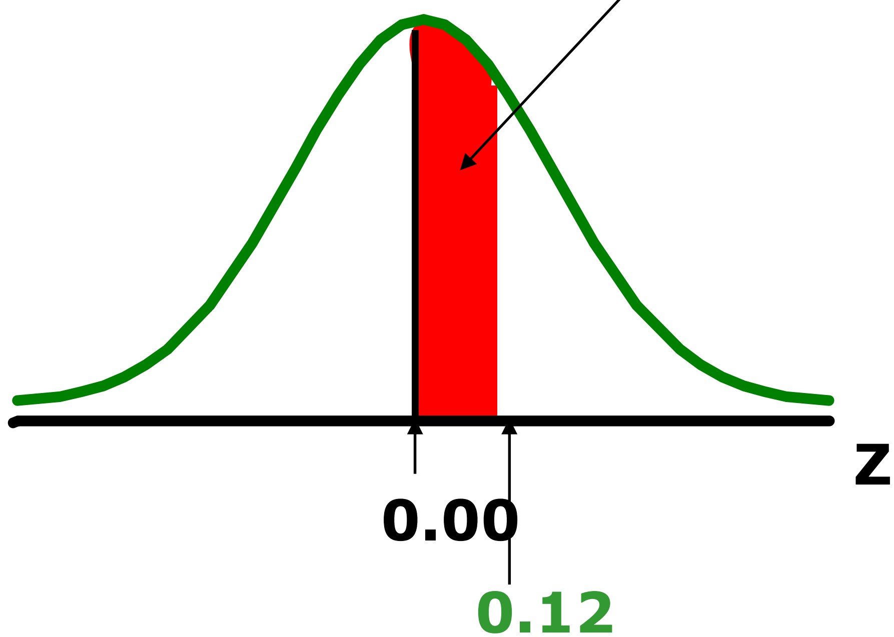
جدول الاحتمالات الطبيعية المعيارية

z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255

$P(8 < x < 8.6)$

$= P(0 < z < 0.12)$

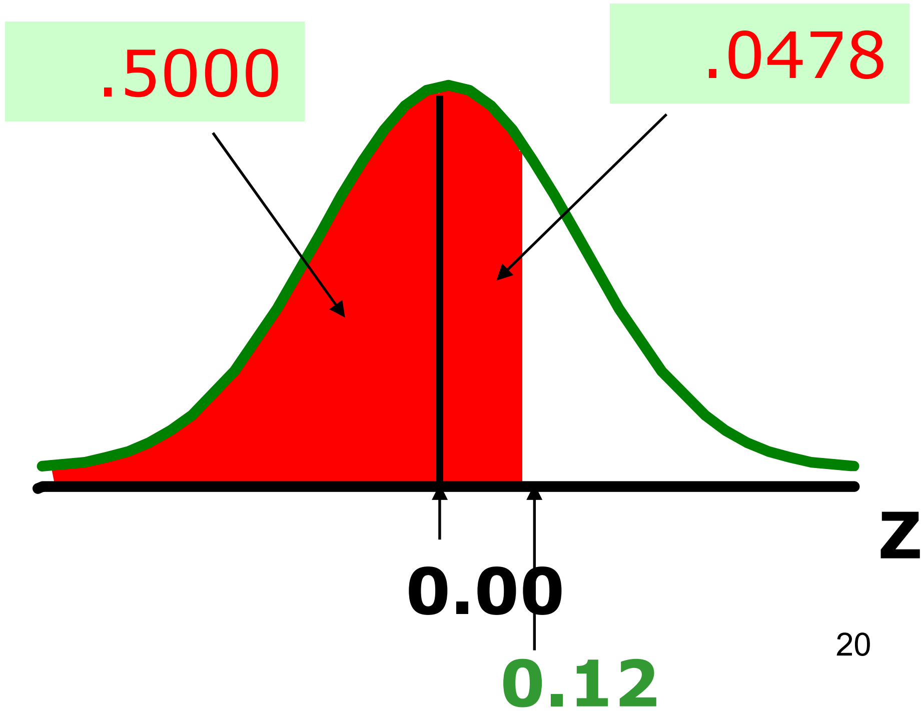
.0478



إيجاد الاحتمالات الطبيعية

- افترض أن x يتبع التوزيع الطبيعي بوسط يساوي خمسة و انحراف معياري يساوي ثمانية.
- أوجد $P(x < 8.6)$

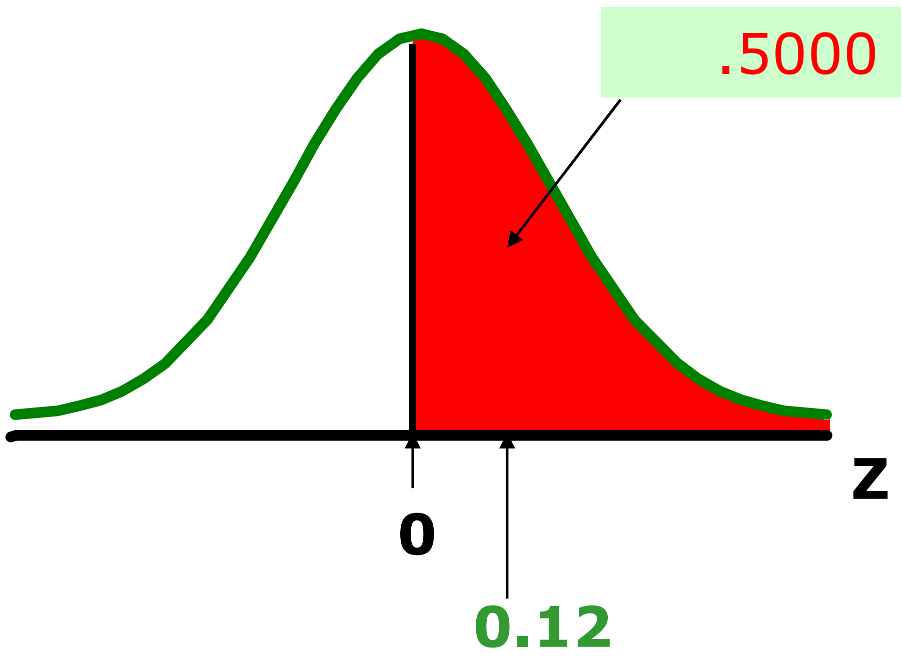
$$\begin{aligned} & P(x < 8.6) \\ &= P(z < 0.12) \\ &= P(z < 0) + P(0 < z < 0.12) \\ &= .5 + .0478 = .5478 \end{aligned}$$

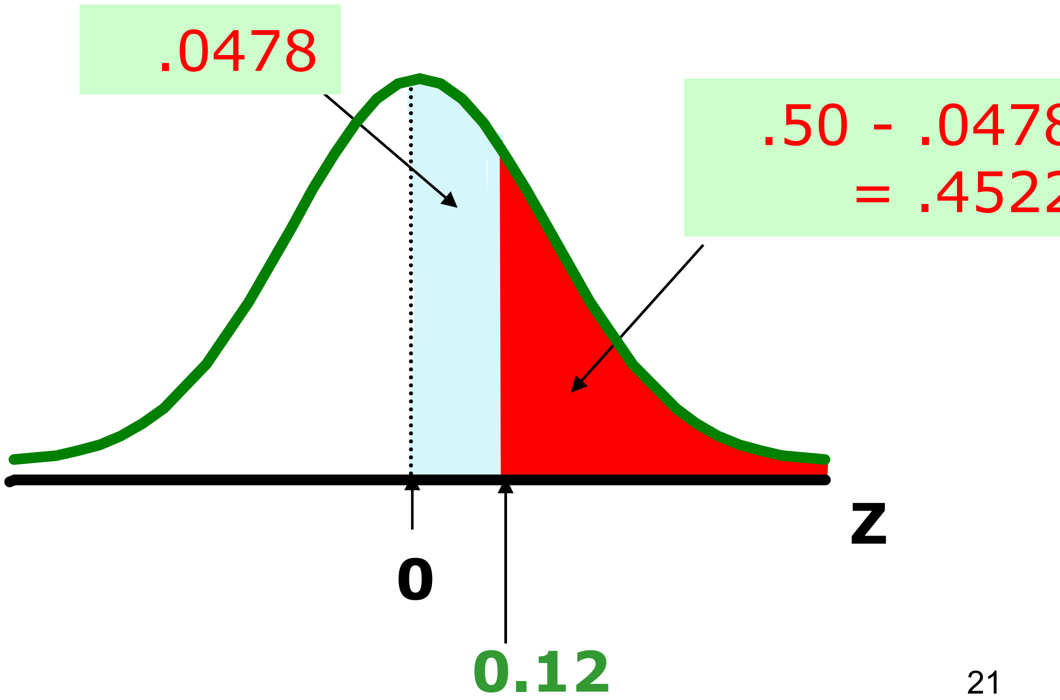


احتمالات الذيل الأعلى

أوجد P(x > 8.6)

$$P(x > 8.6) = P(z > 0.12) = P(z > 0) - P(0 < z < 0.12)$$
$$= .5 - .0478 = .4522$$

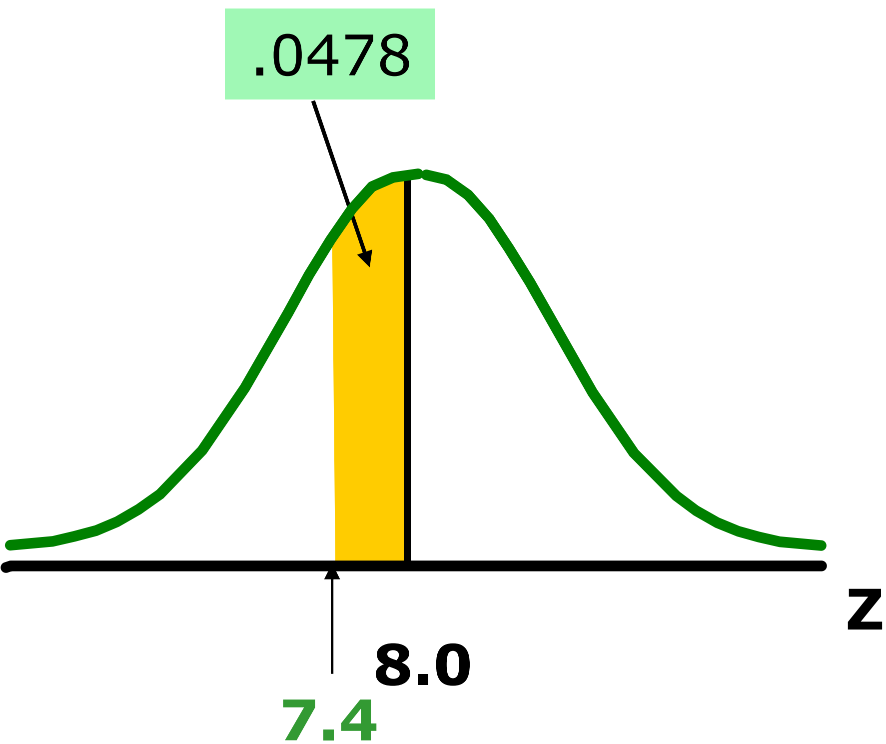




احتمالات الذيل الأدنى

أوجد $P(7.4 < x < 8)$

بما أن التوزيع الطبيعي متناظر، فإننا نستطيع أن نستخدم نفس الجدول حتى و إن كانت قيم "زي" سالبة



$P(7.4 < x < 8)$

$= P(-0.12 < z < 0)$

$= .0478$

التوزيع المنتظم

هو توزيع احتمالي يعطي احتمالات متساوية لكافة القيم المحتملة للمتغير العشوائي

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

where

f(x) = قيمة تابع الكثافة عند أية قيمة لـ x

a = النهاية الدنيا للمجال

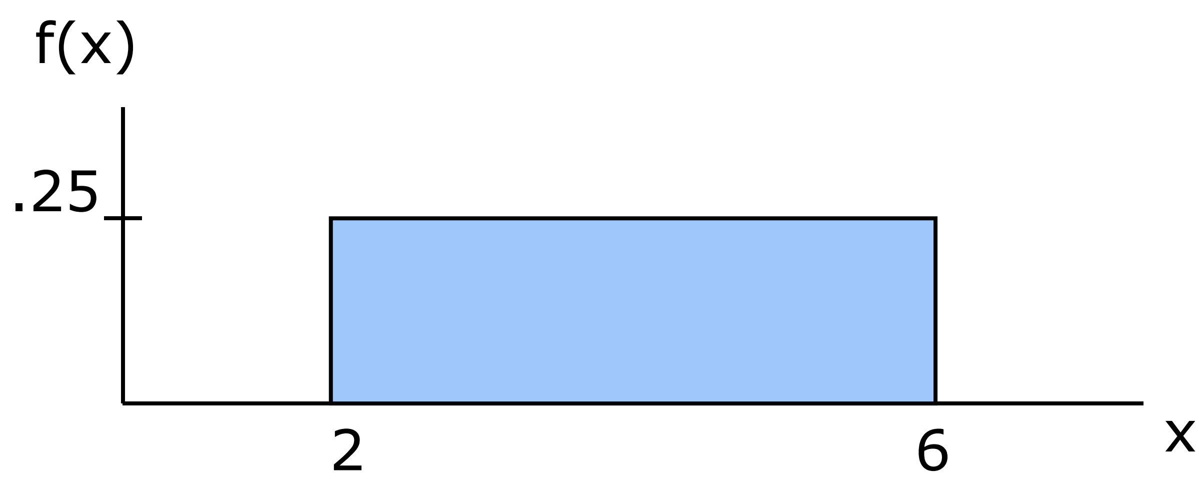
b = النهاية العليا للمجال

التوزيع المنتظم: مثال

التوزيع المنتظم على المجال:

$$2 \le x \le 6$$

$$f(x) = \frac{1}{6-2} = .25 \quad ; \quad 2 \le x \le 6$$



التوزيع الأسّي

- يستخدم لقياس الزمن الفاصل بين "أوقات الوصول" أي الزمن الذي يمرّ بدءً من لحظة حصول الحدث و حتى تكرر حصوله مرة أخرى.

- أمثلة:
  - الزمن الفاصل بين وصول شاحنتين إلى موقع التحميل.
  - الزمن الفاصل بين استخدامات الصرّاف الآلي.
  - الزمن الفاصل بين المكالمات الواصلة للمبدّلة الرئيسية.

التوزيع الأسّي

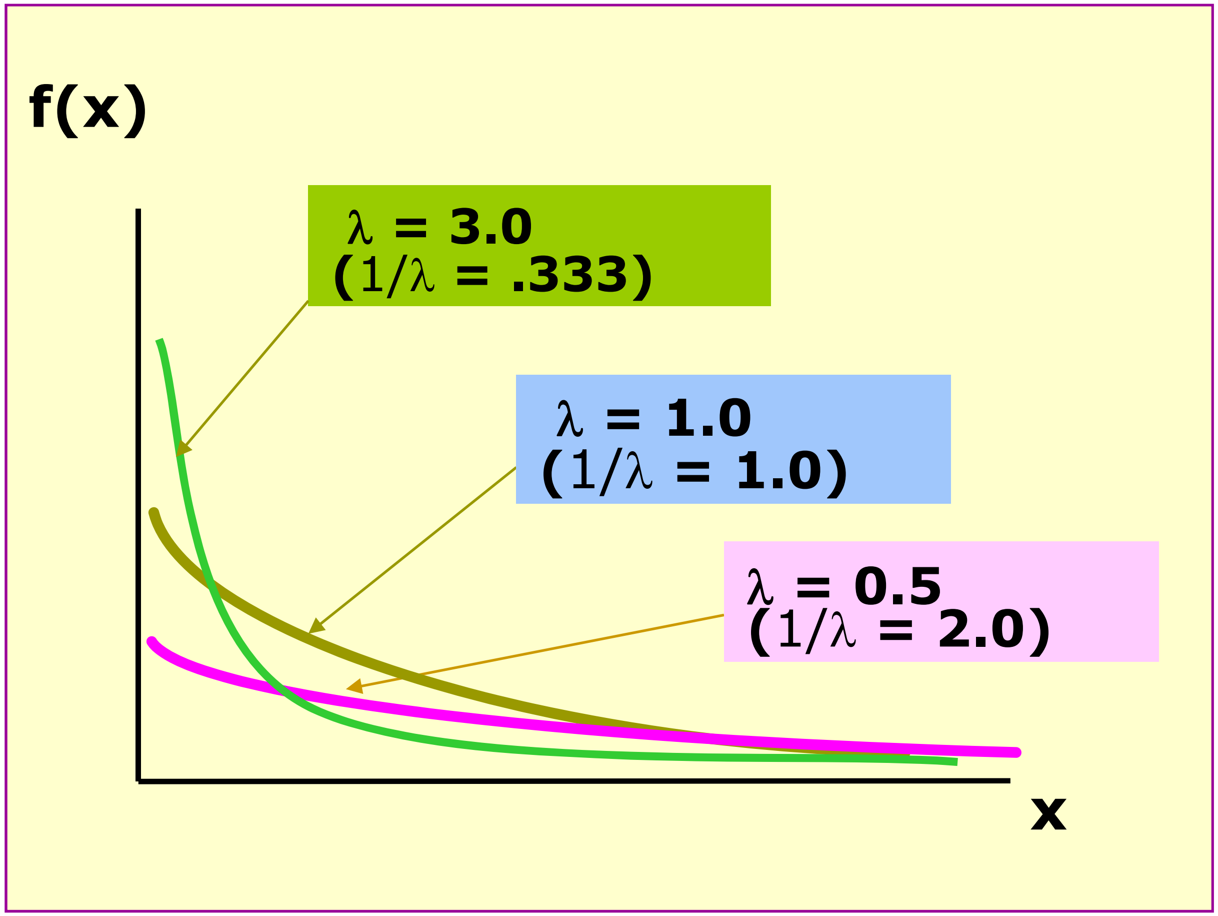
إن احتمال أن يكون زمن الوصول مساوياً أو أصغر من زمن محدَّد a هو:

$$P(0 \le x \le a) = 1 - e^{-\lambda a}$$

حيث $1/\lambda$ هو الزمن الوسطي بين الأحداث

لاحظ أنه إذا كان عدد مرات حصول الحدث في الفترة الزمنية المعينة موزعاً وفق بواسون بوسط λ، فإن الزمن الفاصل بين حصول الأحداث يتبع التوزيع الأسي بوسط $1/\lambda$.

شكل التوزيع الأسّي



مثال

يصل الزبائن عند نافذة الاستلام بمعدل 15 زبوناً في الساعة (وفق توزيع بواسون). ما هو احتمال أن يكون زمن الوصول بين زبونين متتاليين أقل من خمس دقائق؟

- الزمن بين الوصولات موزع أسّيّاً بزمن وسطي بين الوصولات مقداره أربع دقائق (15 كل 60 دقيقة، وسطيّاً)

$1/\lambda = 4.0$, so $\lambda = .25$

$P(x < 5) = 1 - e^{-\lambda a} = 1 - e^{-(.25)(5)} = .7135$

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الثامنة /القسم النظري/

توزيعات المعاينة: مقدمة

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

مخرجات المحاضرة

تأكد بنهاية المحاضرة أنك تستطيع:

- تعريف مفهوم خطأ المعاينة
- تحديد الوسط و الانحراف المعياري لتوزيع معاينة وسط عينة
- تحديد الوسط و الانحراف المعياري لتوزيع معاينة نسبة محسوبة من عينة
- نظرية النزعة المركزية و أهميتها.
- تطبيق توزيع المعاينة من أجل الوسط و النسب.

خطأ المعاينة

- تستخدم إحصاءات العينة لتقدير بارامترات المجتمع

$\bar{X}$ = an estimate of μ

- و لكن المشكلة هي أن:
  - العينات المختلفة تعطي تقديرات مختلفة لبارامترات المجتمع
  - قابلية النتائج المأخوذة من العينات للتغير تعني أن هنالك "خطأ معاينة"
  - يعرّف خطأ المعاينة بأنه الفرق بين قيمة الإحصاءة و قيمة البارامتر الذي تستخدم لتقديره، أي:

$$\boxed{\text{Sampling Error} = \bar{x} - \mu}$$

  - يمكن أن يكون خطأ المعاينة موجباً أو سالباً، و تتناقص قيمته بشكل عام مع تزايد حجم العينة.

تذكّر

- وسط العينة

$$\bar{x} = \frac{\sum x_i}{n}$$

وسط المجتمع

$$\mu = \frac{\sum x_i}{N}$$

حيث:

μ = Population mean وسط المجتمع

$\bar{x}$ = sample mean وسط العينة

x_i = قيم مفردات من المجتمع أو العينة

N = Population size حجم المجتمع

n = sample size حجم العينة

توزيع المعاينة

- توزيع المعاينة: هو عبارة عن توزيع لكافة القيم المحتملة لإحصاءة من عينة سحبت من مجتمع معين، و ذات حجم محدد.

- A sampling distribution is a distribution of the possible values of a statistic for a given size sample selected from a population

بناء توزيع معاينة مبسط

- افترض مجتمعاً صغيراً للتبسيط:
- المجتمع مكون من أربعة أفراد:
- N=4
- عمر الفرد =X
- قيم X : 18, 20, 22, 24

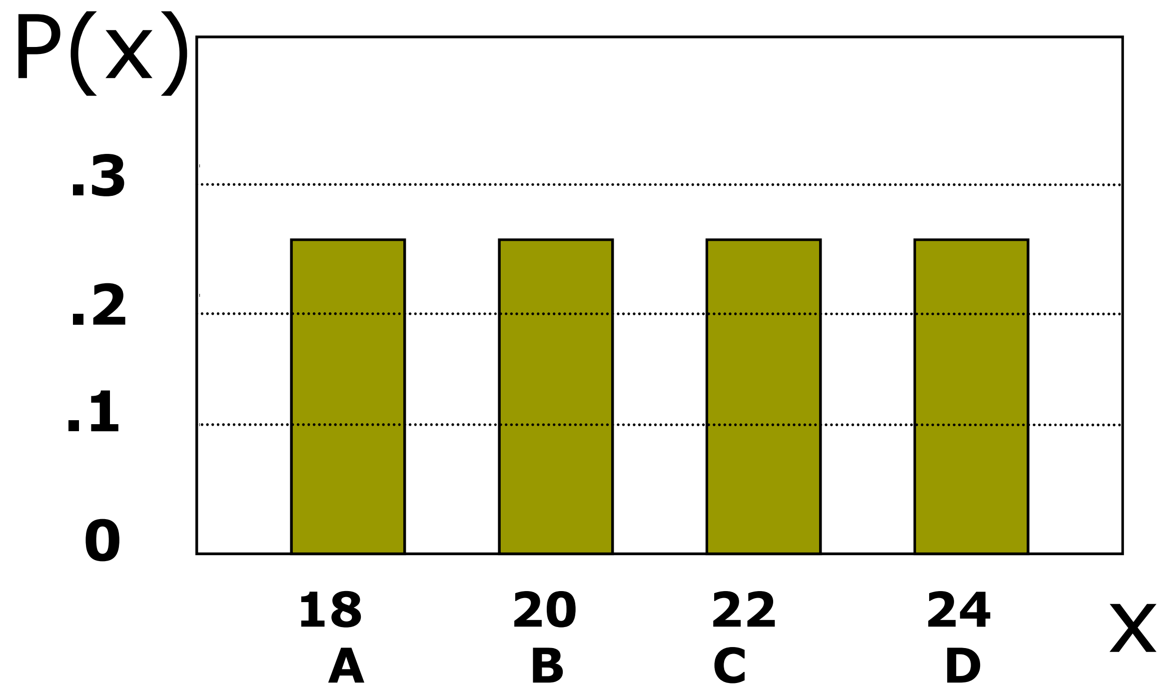
بناء توزيع معاينة مبسط

تتمة

قياسات تلخيصية لتوزيع المجتمع

$$\mu = \frac{\sum x_i}{N} = \frac{18+20+22+24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = 2.236$$



توزيع منتظم

بناء توزيع معاينة مبسط

تتمة

كافة العينات من الحجم n=2 الممكن سحبها من المجتمع

	2nd Obs			
1st Obs	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

هناك 16 عينة محتملة و
السحب بدون إعادة

ينتج لدينا 16 وسطاً
محسوباً من 16 عينة

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

بناء توزيع معاينة مبسط

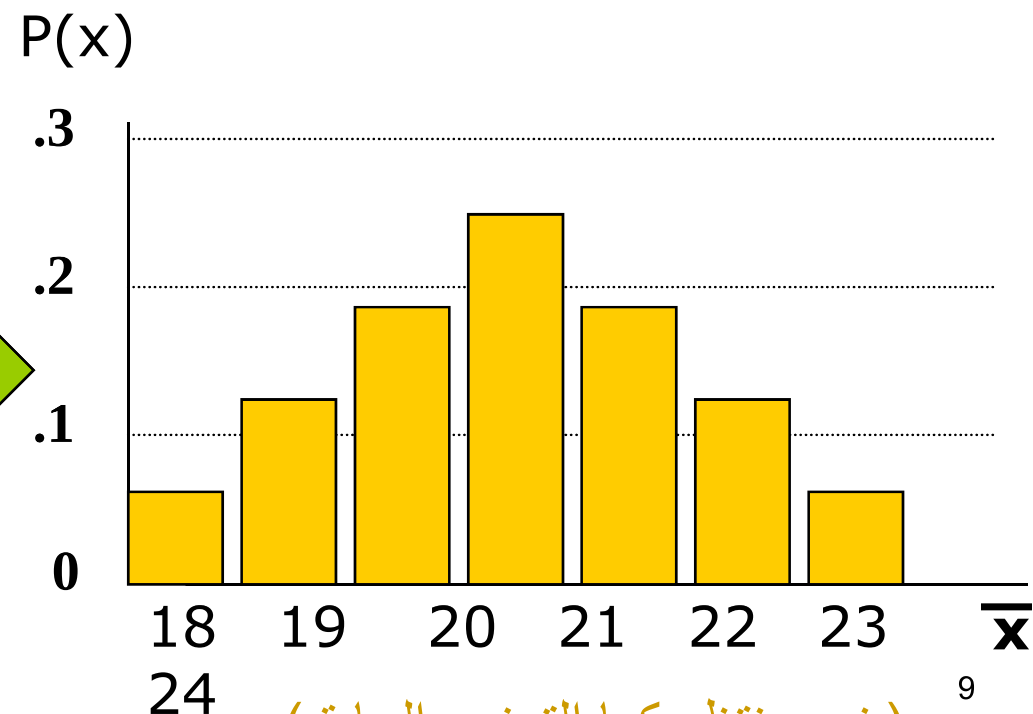
تتمة

توزيع المعاينة لكافة الأوساط الحسابية

16 وسطاً حسابياً

1st Obs	2nd Observation			
	18	**20**	**22**	**24**
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

توزيع أوساط المعاينة



(غير منتظم كما التوزيع السابق)

بناء توزيع معاينة مبسط

تتمة

قياسات مختصرة لتوزيع المعاينة

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu_{\bar{x}})^2}{N}}$$

$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

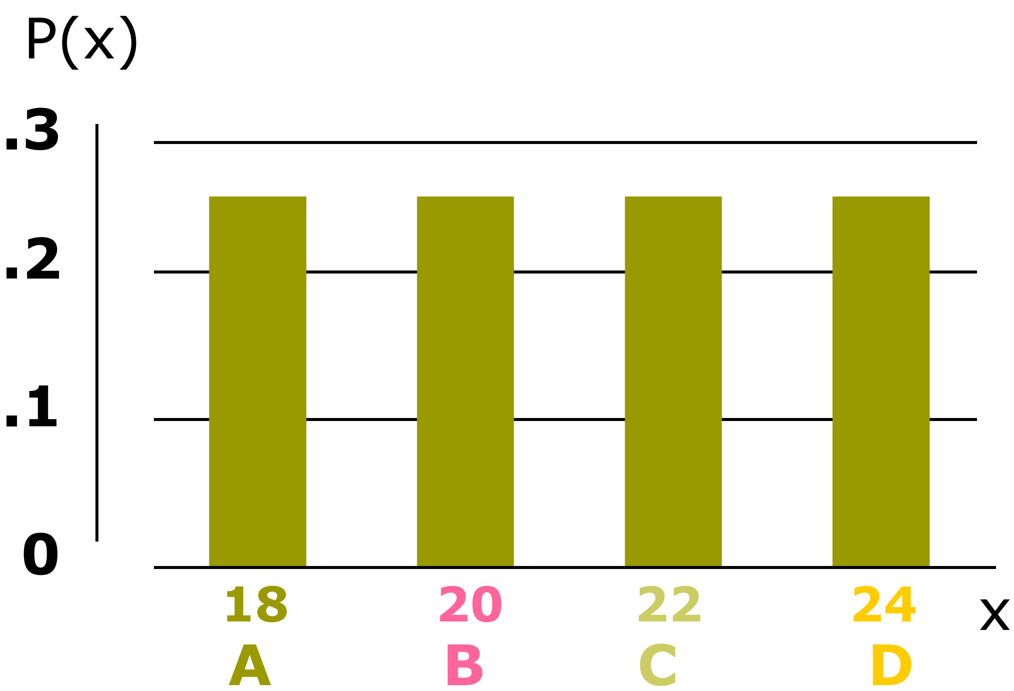
مقارنة المجتمع مع توزيع معاينته

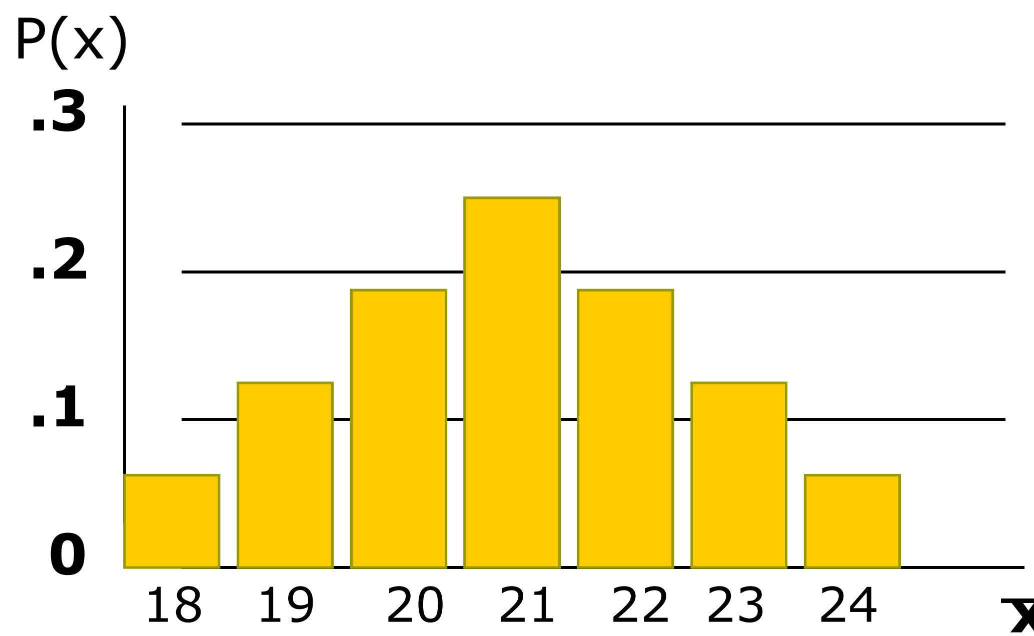
المجتمع
N = 4

$\mu = 21 \qquad \sigma = 2.236$

توزيع المعينة للأوساط
n = 2

$\mu_{\bar{x}} = 21 \qquad \sigma_{\bar{x}} = 1.58$





إذا كان المجتمع طبيعياً

مبرهنة: إذا كان المجتمع طبيعياً ذا وسط μ و انحراف معياري σ فإن توزيع معاينة $\bar{X}$ هوتوزيع طبيعي ذي وسط $\boxed{\mu_{\bar{x}} = \mu}$ و انحراف معياري $\boxed{\sigma_{\bar{x}} = \sigma / \sqrt{n}}$

If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with $\boxed{\mu_{\bar{x}} = \mu}$ and $\boxed{\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}}$

القيم المعيّرة لتوزيع معاينة $\bar{x}$

و تحسب كالتالي:

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x}$ = sample mean وسط العينة

μ = population mean وسط المجتمع

σ = population standard deviation الانحراف المعياري للمجتمع

n = sample size حجم العينة

تصحيح المجتمع المنتهي

- ويطبق هذا التصحيح عندما يكون:
  - حجم العينة كبيراً نسبياً بالنسبة للمجتمع:

    $(n > 5\% \text{ of } N)$

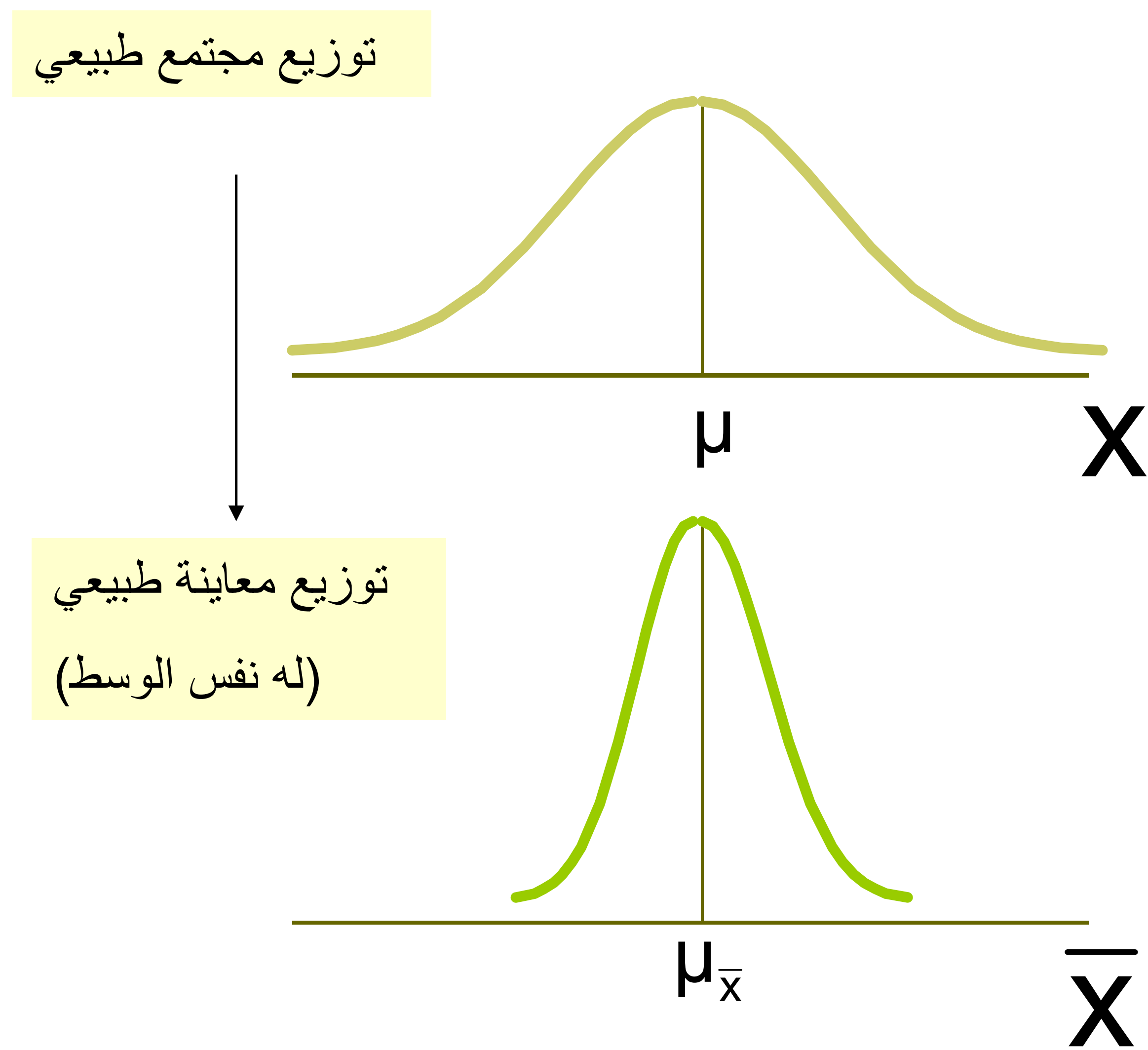
    و...
  - المعاينة (السحب) بدون إعادة

حيث نستخدم:

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}}$$

خصائص توزيع المعاينة

$$\boxed{\mu_{\bar{x}} = \mu}$$

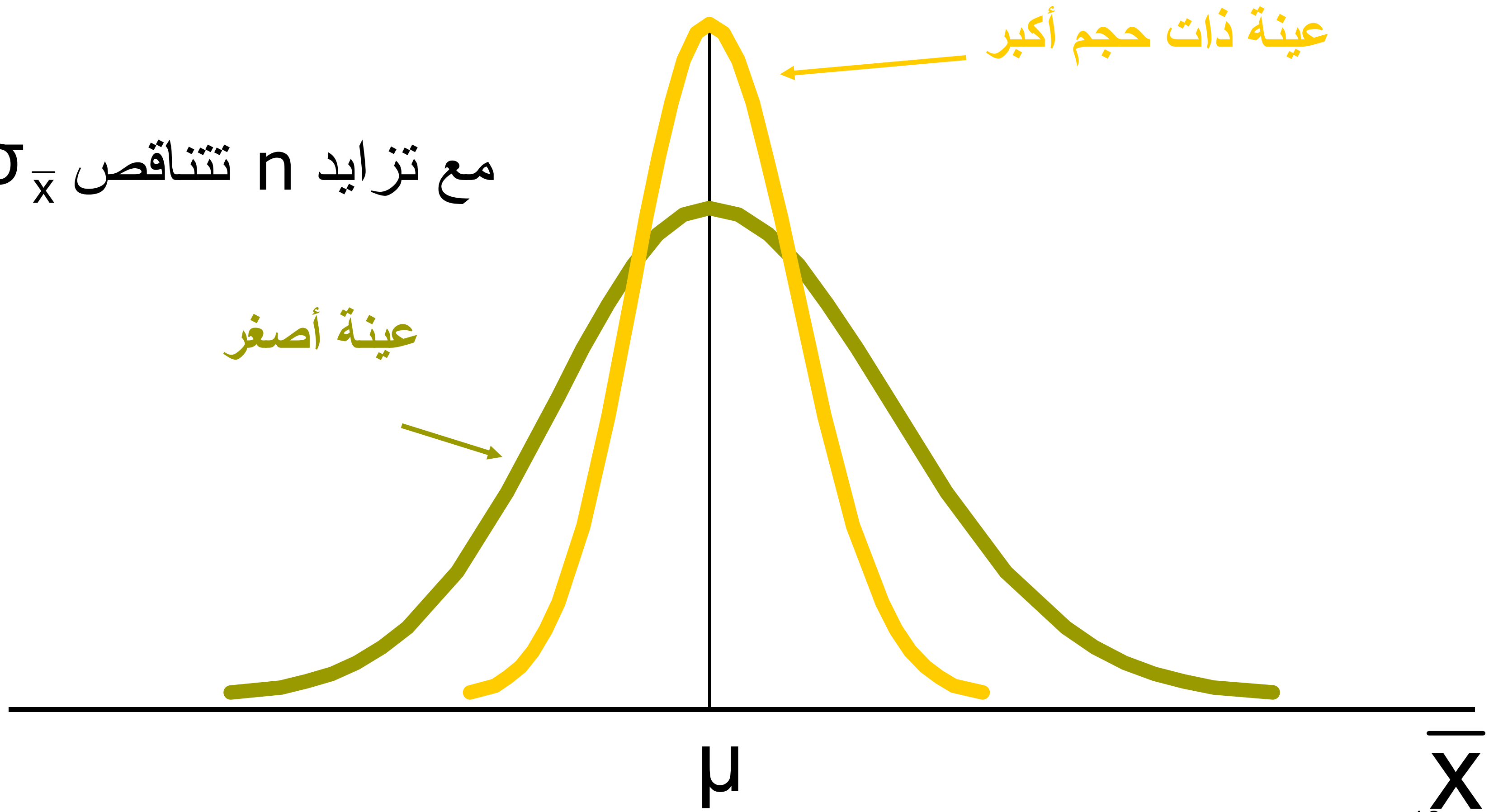


خصائص توزيع المعاينة

تتمة

- من أجل المعاينة / السحب مع إعادة:

مع تزايد n تتناقص $\sigma_{\bar{x}}$



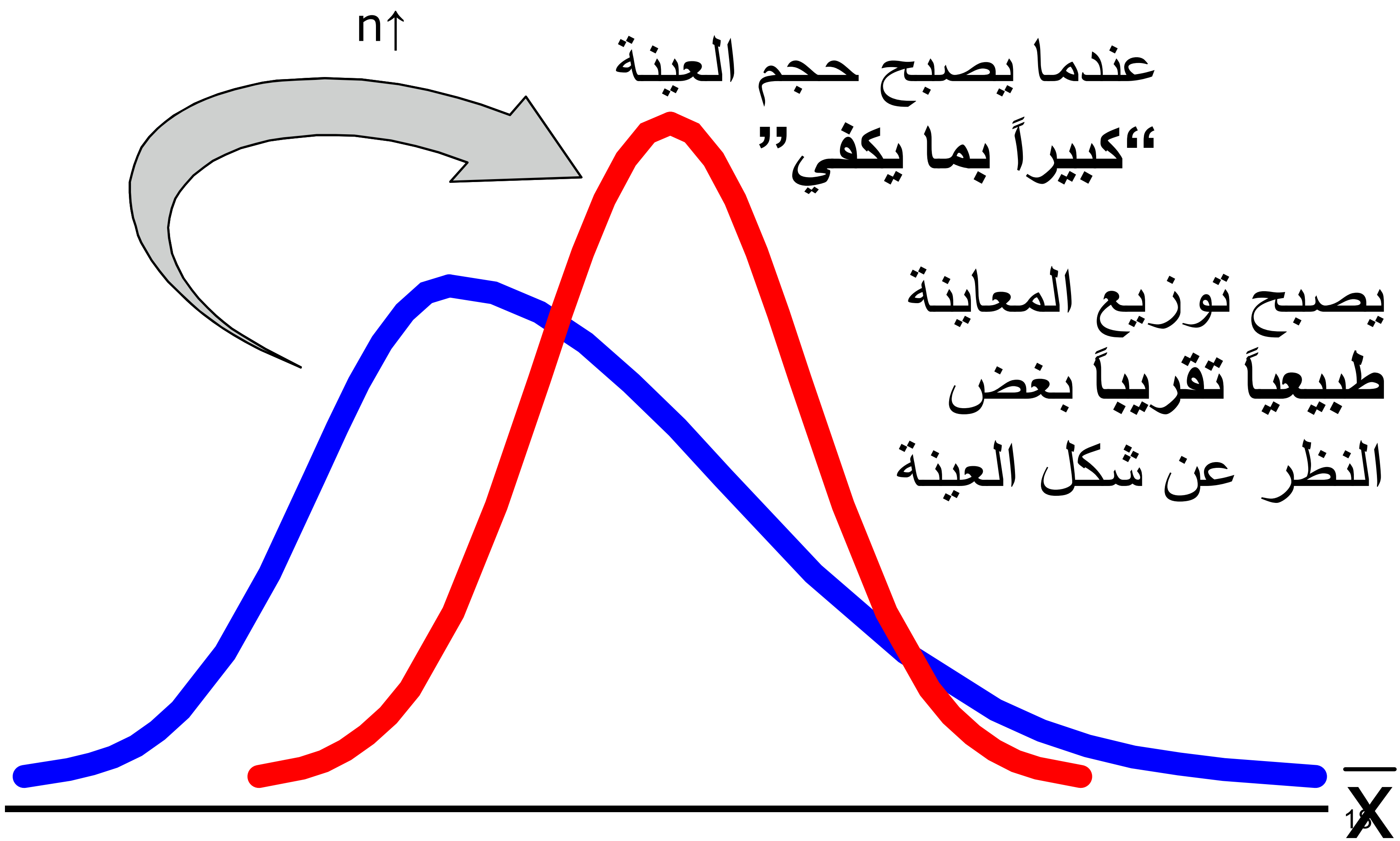
إذا كان لم يكن المجتمع طبيعياً

- نستطيع تطبيق نظرية النهاية المركزية:
  - حتى إذا كان المجتمع غير طبيعي:
  - سوف تكون أوساط المعاينة طبيعية تقريباً ما دام حجم العينة كبيراً
  - حيث يكون لتوزيع المعاينة الوسط و الانحراف المعياري التاليين:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem نظرية النهاية المركزية



عندما يصبح حجم العينة

"كبيراً بما يكفي"

يصبح توزيع المعاينة

طبيعياً تقريباً بغض

النظر عن شكل العينة

إذا كان لم يكن المجتمع طبيعياً

تتمة

خصائص

توزيع المعاينة:

النزعة المركزية

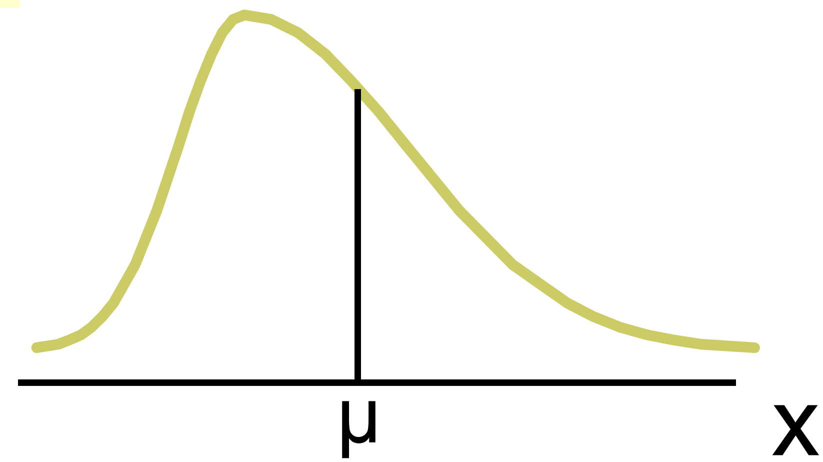
$$\mu_{\bar{x}} = \mu$$

التباين

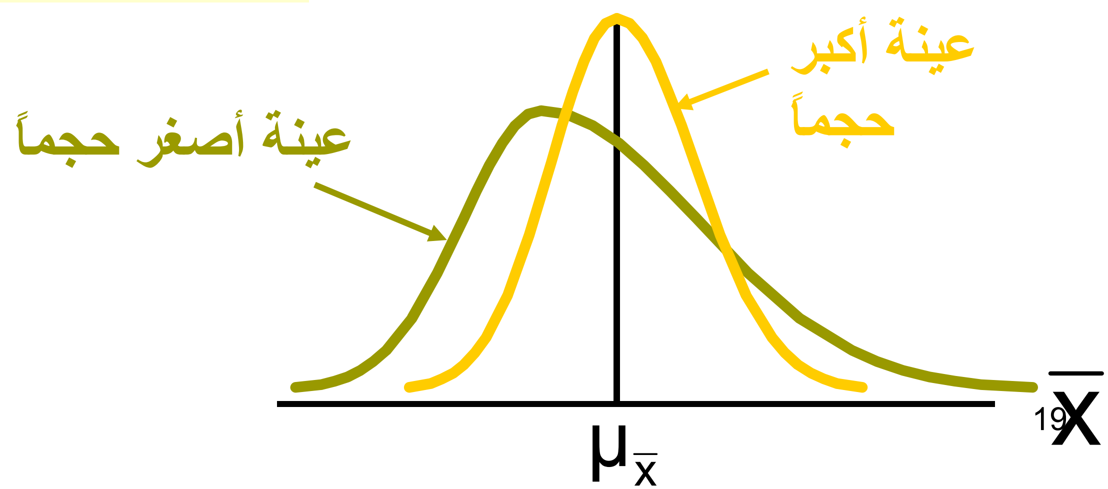
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

السحب مع إعادة

توزيع المجتمع



توزيع المعاينة
يصبح طبيعياً مع تزايد n



ما هو الحجم "الكبير بما يكفي" ؟

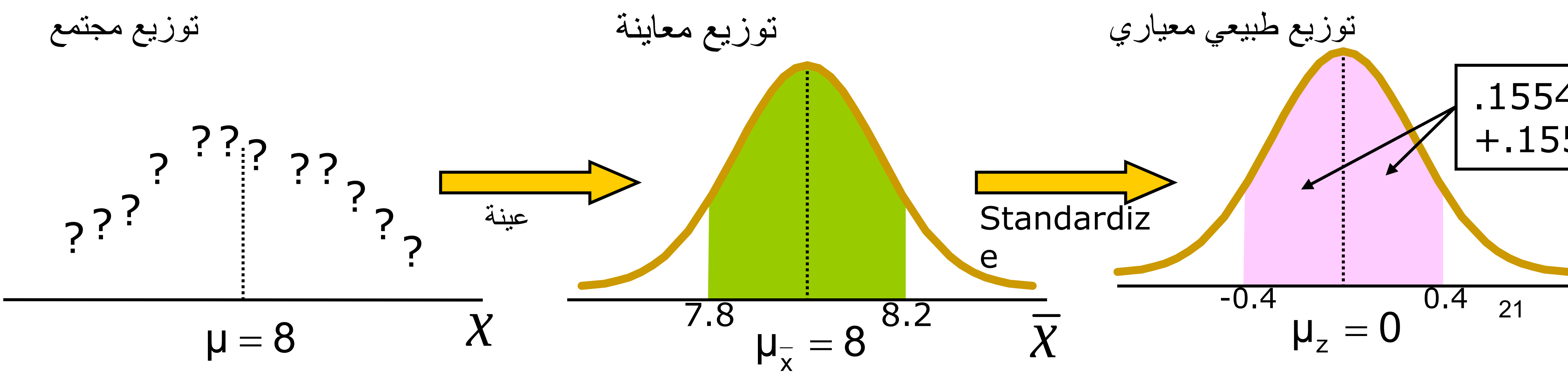
- أكبر من 30 مفردة من أجل معظم التوزيعات.
- أكبر من 15 من أجل التوزيعات المتناظرة إلى حد معقول.
- و أما من أجل التوزيعات الطبيعية فإن توزيع المعاينة سيكون طبيعياً في كل الأحوال

- مثال: ليكن لديك عينة عشوائية $n = 36$ سحبت من مجتمع ذي $\mu = 8$ و $\sigma = 3$ فما هو احتمال أن يقع وسط العينة بين القيمتين 7.8 و 8.2؟

حل المثال:

- لأن حجم العينة كبير بما يكفي، ووفقاً لمبرهنة النهاية المركزية: حتى و إن لم يكن المجتمع طبيعياً فإن توزيع المعاينة للوسط سيكون طبيعياً تقريباً، مع وسط يساوي 8، و انحراف معياري يساوي: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$
- وعليه يمكننا أن نحسب:

$$P(7.8 < \mu_{\bar{x}} < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\mu_{\bar{x}}-\mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}}\right) = P(-0.4 < z < 0.4) = 0.310$$



Population Proportions, p نسب المجتمع

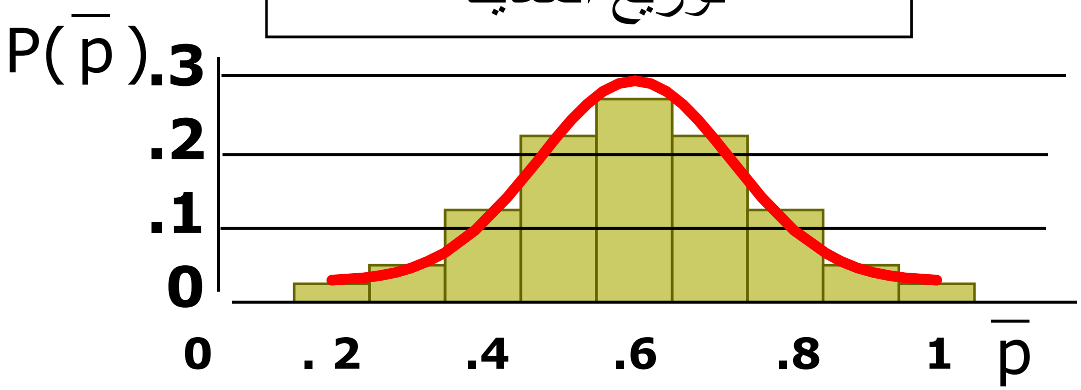
نسبة من المجتمع تمتلك نفس السمة = p

Sample proportion = $(\bar{p}) = p$ تقدير = نسبة العينة

إذا كان لدينا ناتجين (نتيجتين) يتبعان توزيع بينوميال فإنّ:

$$\bar{p} = \frac{x}{n} = \frac{\text{number of successes in the sample}}{\text{sample size}}$$

توزيع المعاينة لنسب المجتمع



يصبح طبيعياً تقريباً عندما
يكون لدينا:

$$np \geq 5$$

$$n(1-p) \geq 5$$

و عند ذلك فإن:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

و

$$\mu_{\bar{p}} = p$$

حيث p هي نسبة المجتمع population proportion

قيم "زي" المعيارية للنسب:

يمكن تعيير قيم نسب العينات $\bar{p}$ (تحويلها إلى الطبيعي المعياري) كما يلي:

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

- إذا كانت المعاينة **بدون إعادة** و كانت n أكبر من 5% من حجم المجتمع فلا بد لـ $\sigma_{\bar{p}}$ من استخدام عامل التصحيح للمجتمع المنهي:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

مثال

- إذا كانت النسبة الحقيقة للناخبين الداعمين للاقتراح A هي 40%، فما هو احتمال أن تعطي عينة من 200 مفردة نسبة عينة تتراوح بين 40% و 45% ؟

- بكلمة أخرى، إذا كانت $n = 200$ و $p = .4$ ، فما هو:

$$P(.40 \le \bar{p} \le .45) \ ?$$

مثال

تتمة

- if p = .4 and n = 200, what is $P(.40 \le \bar{p} \le .45)$?

أوجد $\sigma_{\bar{p}}$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.4(1-.4)}{200}} = .03464$$

حوّل إلى الطبيعي المعياري:

$$P(.40 \le \bar{p} \le .45) = P\left(\frac{.40-.40}{.03464} \le z \le \frac{.45-.40}{.03464}\right)$$

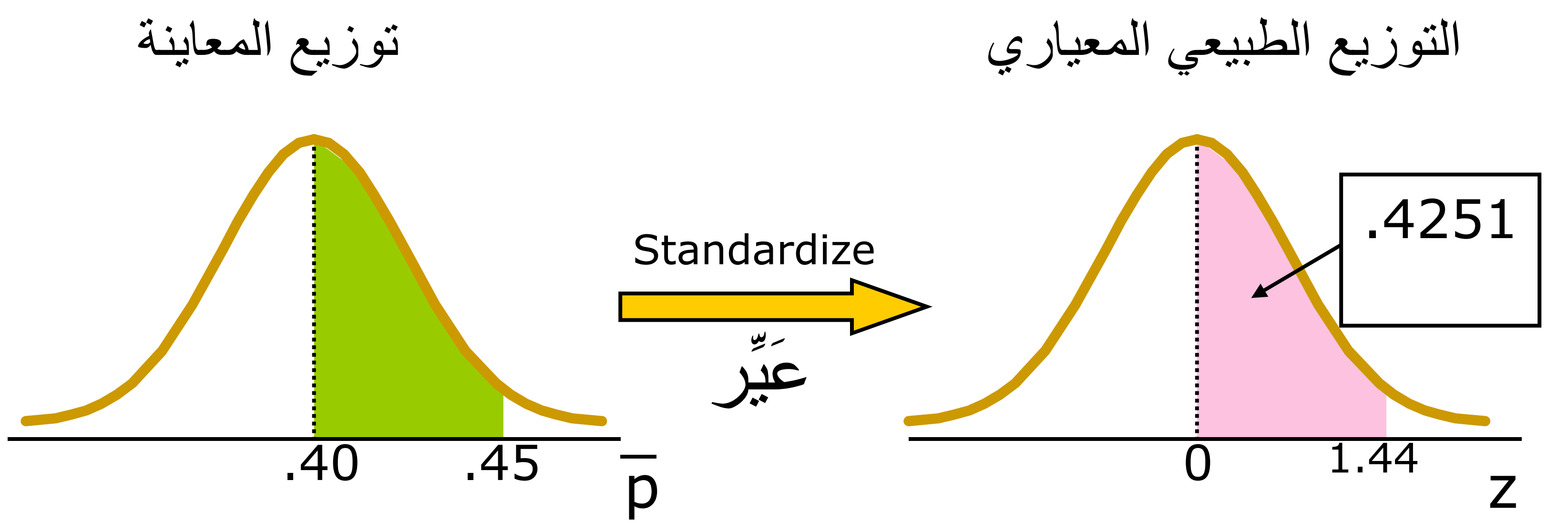
$$= P(0 \le z \le 1.44)$$

مثال

تتمة

- if p = .4 and n = 200, what is
$P(.40 \le \bar{p} \le .45)$?

استخدم الجدول الطبيعي المعياري للحصول على: $P(0 \le z \le 1.44) = \boxed{.4251}$



الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة التاسعة /القسم النظري/

تقدير قيم المجتمع الإحصائي

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

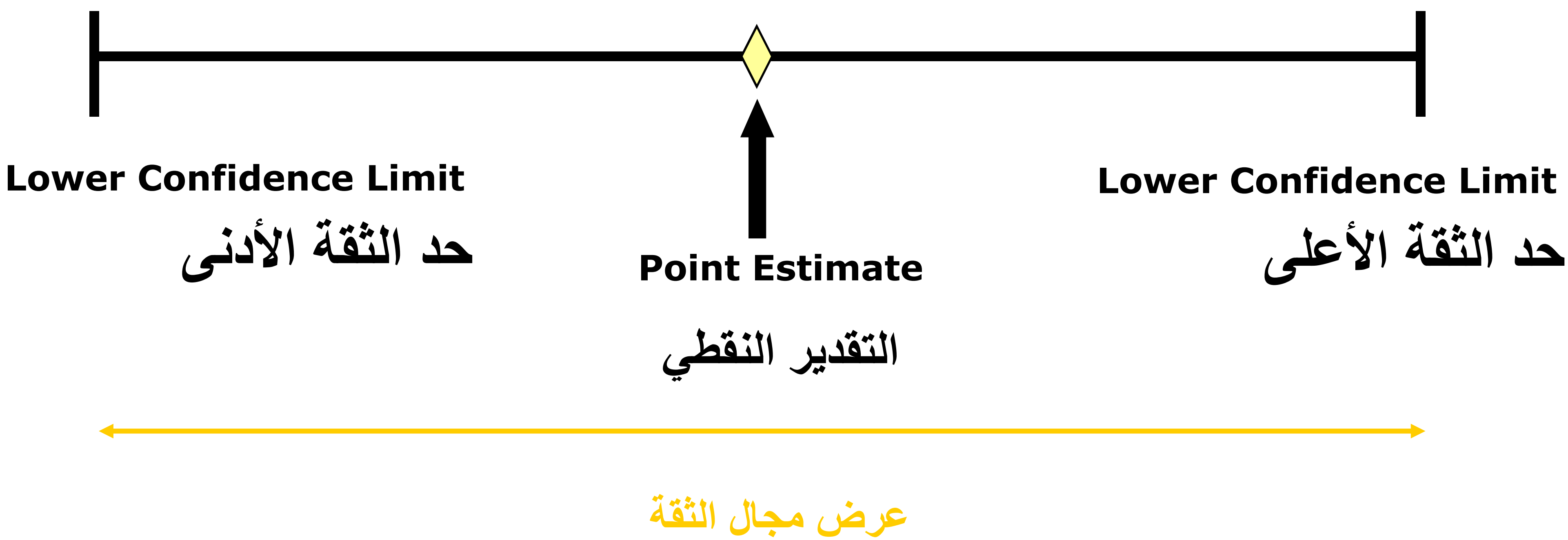
Chapter Goals

يفترض بنهاية هذه المحاضرة أن تصبح قادراً على القيام بما يلي:

- التمييز بين "تقدير النقطة" (أو التقدير النقطي) و "تقدير المجال" (أو تقدير الجوار)
- إنشاء تقدير لـ "مجال ثقة" و تفسيره من أجل وسط مجتمع وحيد باستخدام توزيع "زي" و توزيع "تي"
- تحديد حجم العينة اللازم لتقدير وسط وحيد لمجتمع ضمن هامش معين للخطأ.
- صياغة و تفسير تقدير لمجال ثقة من أجل نسبة مجتمع وحيدة.

Point and Interval Estimates التقديرات النقطية و تقديرات المجال

- التقدير النقطي هو رقم وحيد
- مجال الثقة يعطي معلومات إضافية عن مستوى التشتت



Point Estimates التقدير النقطي

نسطيع باستخدام إحصاءة العينة (تقدير نقطي)	أن نقوم بتقدير بارامتر المجتمع	
$\bar{x}$	μ	وسط
$\bar{p}$	p	نسبة

Confidence Intervals مجالات الثقة

- مدى عدم التأكد المرتبط بتقدير نقطي لبارامتر مجتمع

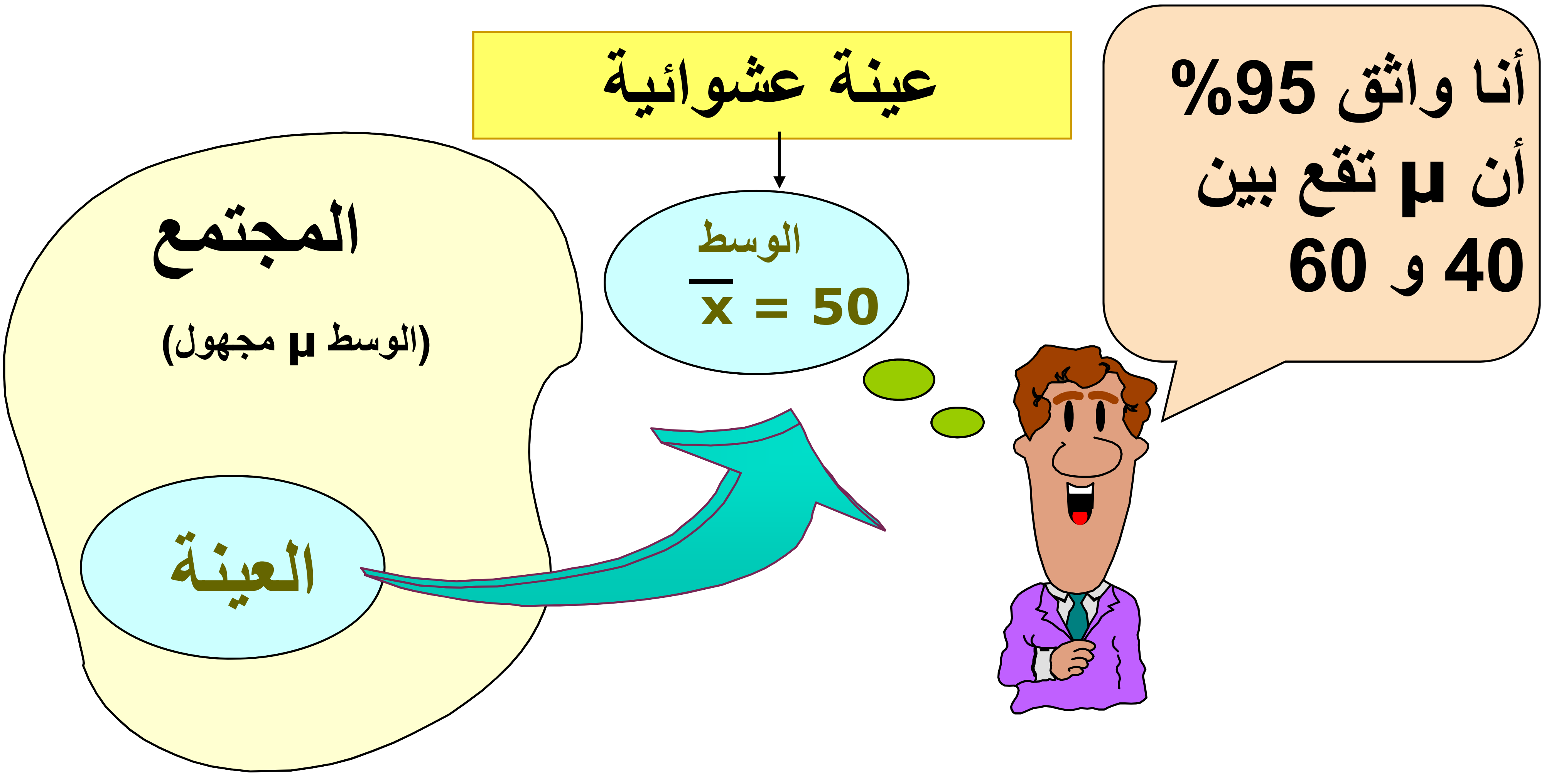
- تقدير المجال يعطي معلومات أكثر عن خواص المجتمع بالمقارنة مع التقدير النقطي

- تقديرات المجال هذه تسمى **مجالات الثقة**

Confidence Interval Estimate تقدير مجال الثقة

- يعطي المجال **مدىً** من القيم:
  - يأخذ بعين الاعتبار قابلية الإحصاءة للتغير من عينة لأخرى
  - يبنى على أساس المشاهدة من عينة واحدة فقط
  - يعطي معلومات عن مدى القرب من بارامترات المجتمع
  - يحسب على أساس مستوى محدد من الثقة و لا يمكن أن يصل هذا المستوى إلى 100% من التأكّد.

Estimation Process عملية التقدير



General Formula الصيغة العامة للتقدير

- الصيغة الهامة لتقدير كافة مجالات الثقة هي:

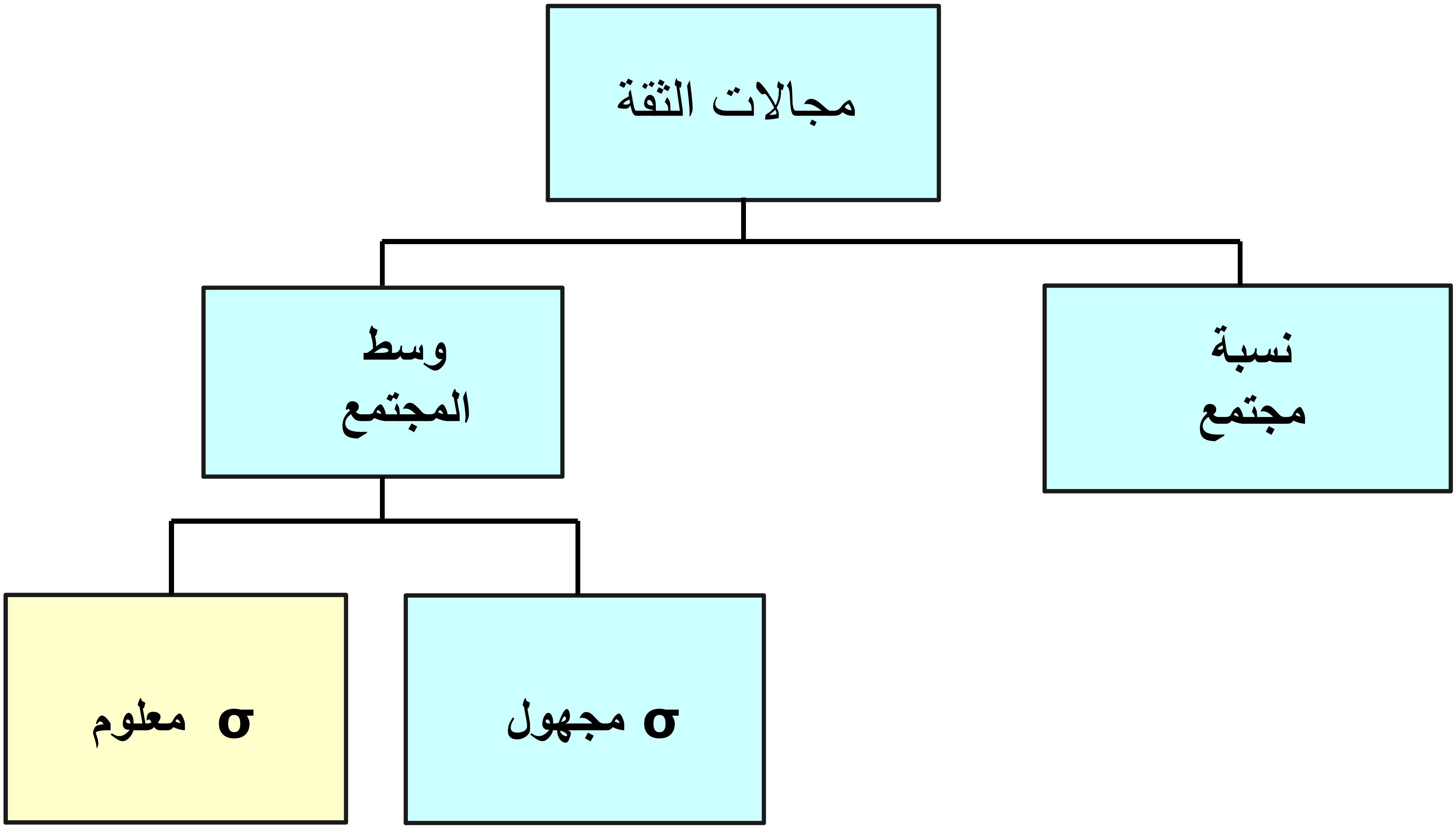
Point Estimate ± (Critical Value)(Standard Error)

التقدير النقطي	±	**(القيمة الحرجة)**	**(الخطأ المعياري)**

Confidence Level, $(1-\alpha)$ مستوى الثقة

- مستوى الثقة يمثّل درجة الثقة بأن المجال سوف يحتوي البارامتر المجهول للمجتمع معبَّراً عنها بنسبة مئوية.
- كأن تقول أن مستوى الثقة = 95%
- و يكتب أيضاً بالصيغة $(1 - \alpha) = .95$
- تفسير مبني على مفهوم التكرار النسبي:
  - على الأمد الطويل، 95% من كافة مجالات الثقة التي يمكن إنشاؤها سوف تحتوي البارمتر الحقيقي المجهول.
- إن مجالاً معيناً يمكن أن / ألا يحتوي البارامتر الحقيقي.
  - فمن أجل مجال بعينه ليس هنالك ما يقال بلغة احتمالية.

Confidence Intervals مجالات الثقة



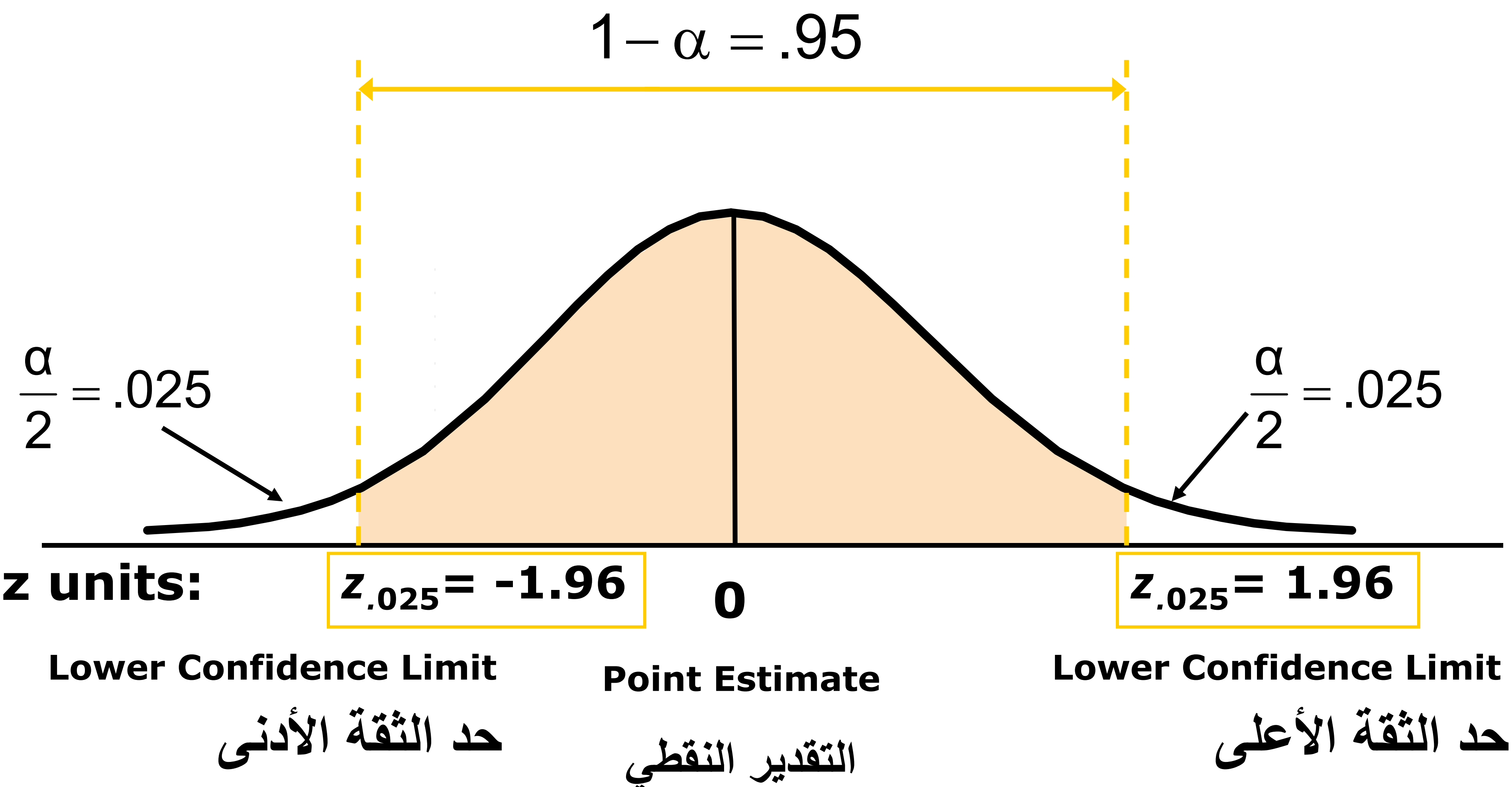
مجالات الثقة لـ μ (σ معلوم)

- افتراضات
  - الانحراف المعياري للمجتمع σ معلوم
  - المجتمع يتبع التوزيع الطبيعي.
  - استخدم عينة كبيرة إذا لم يكن المجتمع طبيعياً.
- تقدير مجال الثقة (أو مُقَدَّر مجال الثقة)

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

إيجاد القيمة الحرجة

- ليكن لدينا مجال ثقة بمستوى 95% $\boxed{z_{\alpha/2} = \pm 1.96}$



درجات الثقة الشائعة

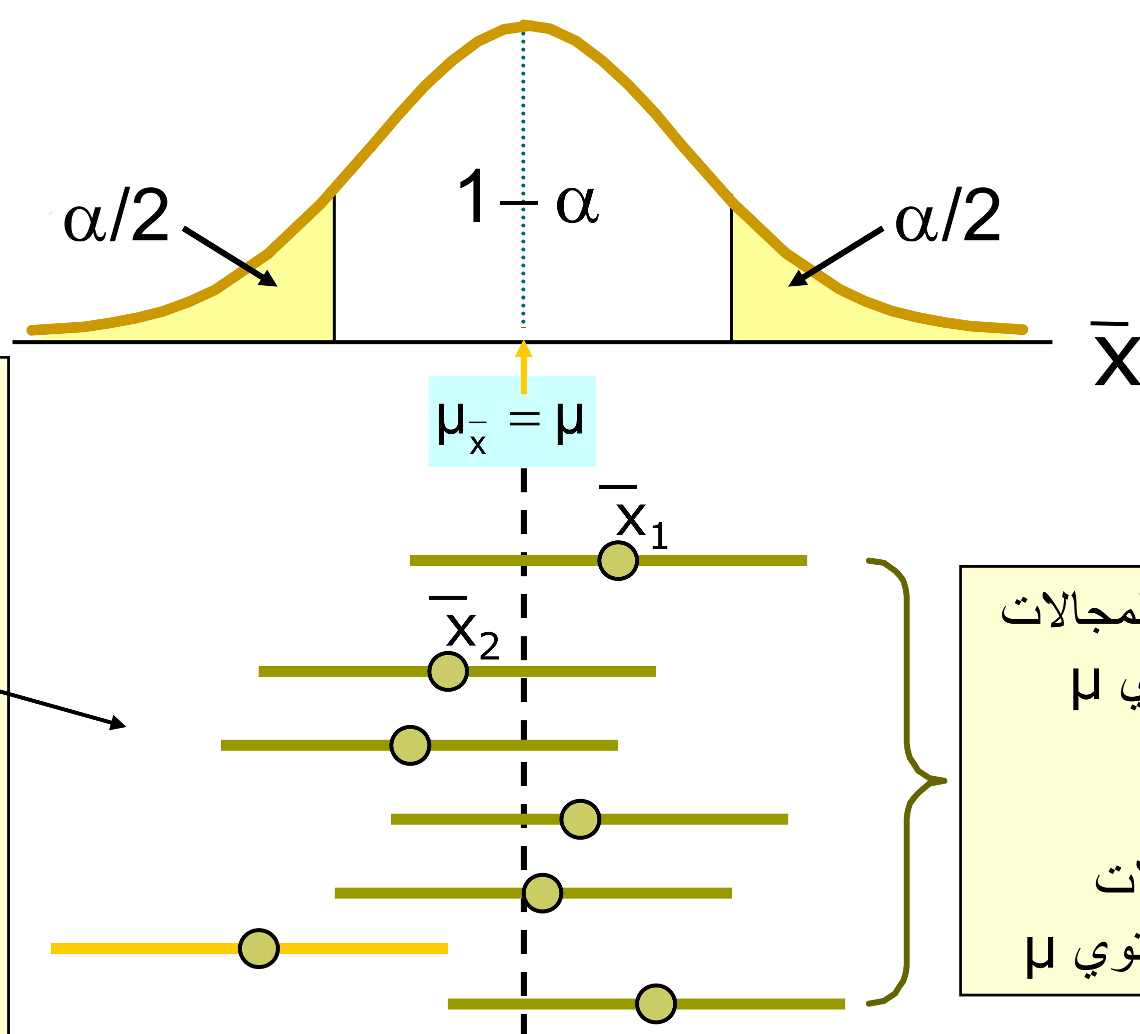
- مستويات الثقة الشائعة هي:

90%, 95%, 99%

مستوى الثقة Confidence level	$1-\alpha$ معامل الثقة Confidence Coefficient	قيمة z $Z_{\alpha/2}$
80%	**.80**	**1.28**
90%	**.90**	**1.645**
95%	**.95**	**1.96**
98%	**.98**	**2.33**
99%	**.99**	**2.58**
99.8%	**.998**	**3.08**
99.9%	**.999**	**3.27**

مجال و درجة الثقة

توزيع معاينة الوسط



يمتد المجال من

$$\bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

إلى

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$100(1\text{-}\alpha)\%$ من المجالات الممكن إنشاؤها تحتوي μ

$100\alpha\%$ من المجالات الممكن إنشاؤها لا تحتوي μ

مجالات الثقة

Margin of Error هامش الخطأ

- هامش الخطأ Margin of Error (e):

الكمية التي تضاف إلى و تطرح من التقدير النقطي بهدف صياغة مجال الثقة

مثال: هامش الخطأ في تقدير μ مع σ معلوم

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$e = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

العوامل المؤثرة في هامش الخطأ

$$e = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$e \downarrow$ مع $\sigma \downarrow$

$e \downarrow$ مع $n \uparrow$

$e \downarrow$ إذا $1 - \alpha \downarrow$

مثال

- عينة من 11 دارة ألكترونية مسحوبة من مجتمع طبيعي كبير تتصف بمقاومة وسطية قدرها 2.20 أوم. نعلم من اختبارات سابقة أن الانحراف المعياري للمجتنع هو 0.35 أوم.

- أنشئ مجالاً بمستوى ثقة 95% من أجل المقاومة الوسطية الحقيقية للمجتمع.

مثال: تتمة

- الحل:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$= 2.20 \pm 1.96\ (.35/\sqrt{11})$$

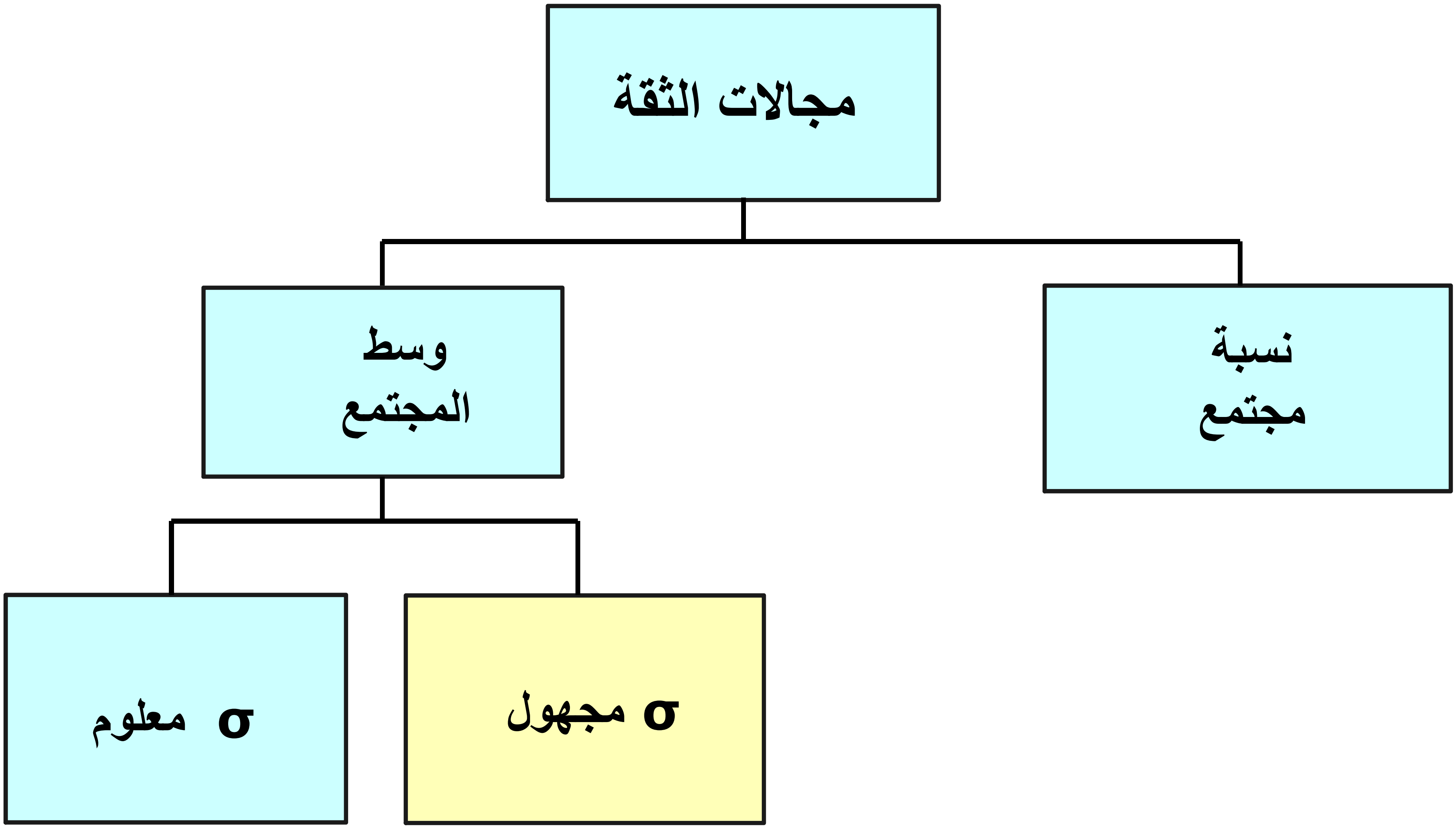
$$= 2.20 \pm .2068$$

$$\boxed{1.9932 \quad \ldots\ldots\ldots\ldots\ldots \quad 2.4068}$$

التفسير

- نحن واثقون بنسبة 95% أن الوسط الحقيق للمقاومة يقع بين 1.9932 و 2.4068 أوم
- على الرغم من أن هذا المجال يمكن أن أو ألا يحتوي الوسط الحقيقي فإن 95% من المجالات المنشأة بهذه الطريقة سوف تحتوي الوسط الحقيقي.
- من الخطأ تفسير المجال المذكور بالقول أن: هناك احتمال مقداره 95% لأن يقع الوسط الحقيق للمجتمع ضمن المجال.
- إذاً: هذا المجال يمكن أن **<u>يحتوي</u>** و يمكن **<u>ألا يحتوي</u>** الوسط الحقيقي و لا يمكن ربط مجال وحيد بأي احتمال.

Confidence Intervals مجالات الثقة



مجالات الثقة لـ μ (σ مجهول)

- نضطر لاستخدام الانحراف المعيلري للعينة
- الأمر الذي يزيد من درجة عدم التأكد كون الانحراف المعياري للعينة يتغير من عينة لأخرى
- و لذلك نلجأ لاستخدام توزيع t بدلاً من اعتماد التوزيع الطبيعي

مجالات الثقة لـ μ (σ مجهول)

تتمة

- افتراضات:
  - الانحراف المعياري للمجتمع مجهول
  - المجتمع موزع طبيعياً
  - إذا لم يكن المجتمع طبيعياً، استخدك عينة كبيرة
- استخدم توزيع *t*
- مجال الثقة المقدَّر:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

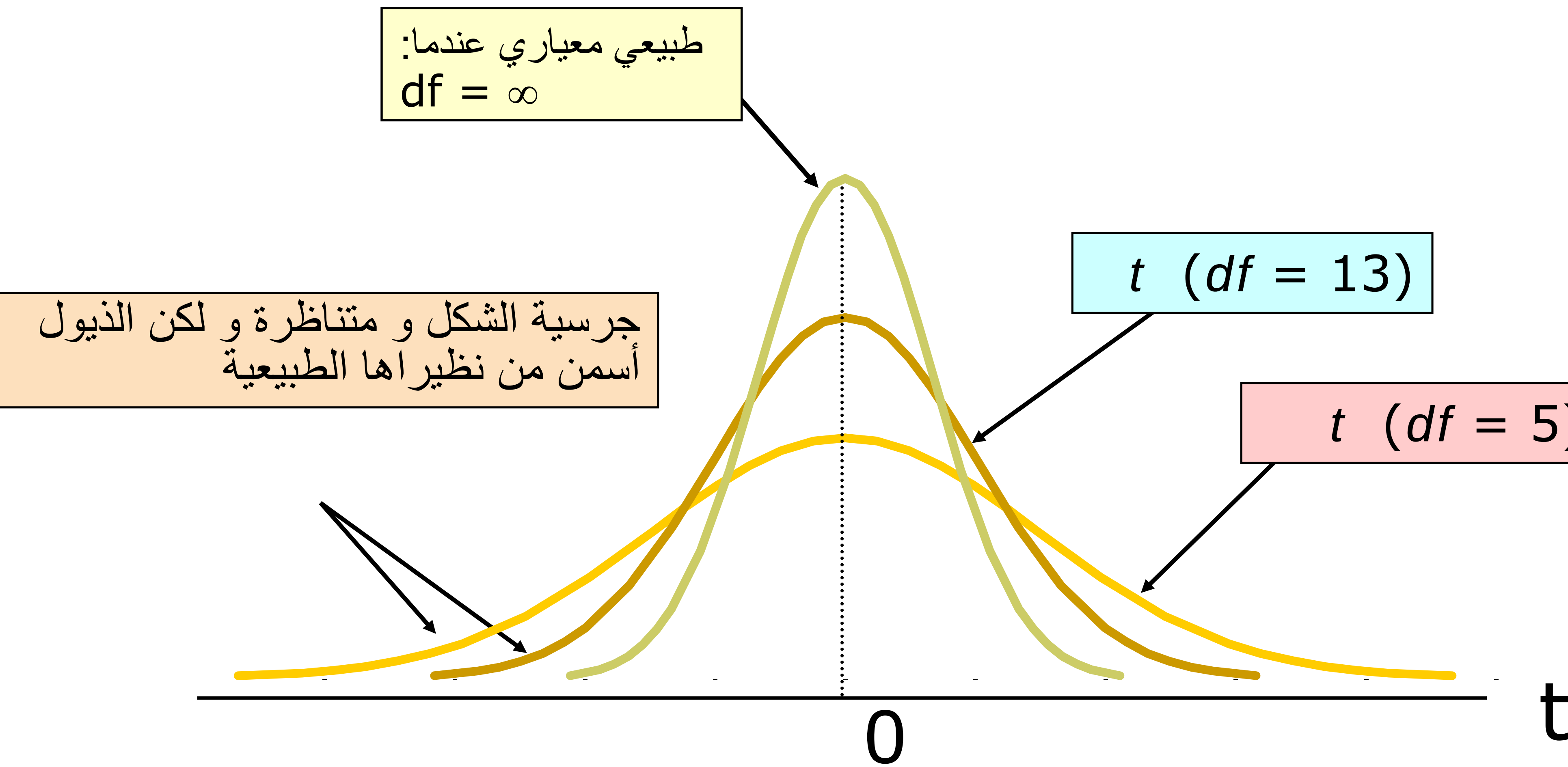
Student's t Distribution توزيع تي ستودنت

- في الحقيقة *t* يمثل عائلة من التوزيعات
- تعتمد قيمة *t* على عدد درجات الحرية d.f.
  - عدد المشاهدات المستقلة، أي الحرة في التغير بعد حساب وسط العينة

$$\boxed{d.f. = n - 1}$$

توزيع تي ستودنت

لاحظ أن *t* يقترب من z مع ازدياد n

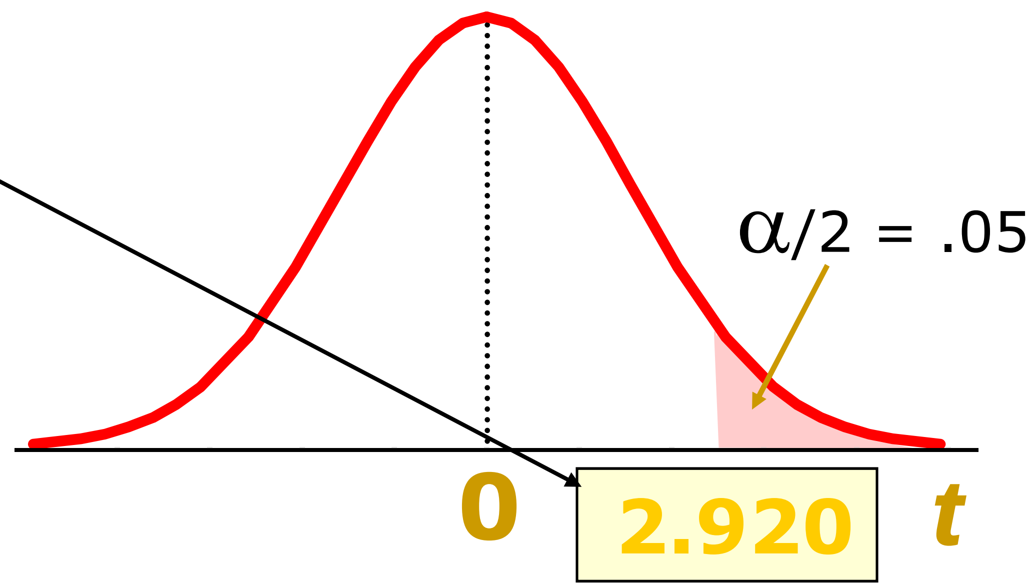


جدول تي ستودنت

	مساحة الذيل الأعلى **Upper Tail Area**		
df	.25	.10	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	**2.920**
3	0.765	1.638	2.353

لا يحتوي جسم الجدول على احتمالات و إنما على قيم t

Let: n = 3
df = n - 1 = 2
α = .10
$\alpha/2$ =.05



t distribution values قيم توزيع تي *t*

بالمقارنة مع قيم Z

Confidence Level	t (10 d.f.)	t (20 d.f.)	t (30 d.f.)	z
.80	1.372	1.325	1.310	1.28
.90	1.812	1.725	1.697	1.64
.95	2.228	2.086	2.042	1.96
.99	3.169	2.845	2.750	2.58

لاحظ أن *t* يقترب من z مع ازدياد n

مثال

ليكن لديك عينة عشوائية بالخصائص التالية

$s = 8 \qquad n = 25 \qquad \bar{x} = 50$

قم بإنشاى مجال ثقة 95% من أجل μ

- بما أن: $d.f. = n - 1 = 24$ فإن:

$$t_{\alpha/2\,,\,n-1} = t_{.025,24} = 2.0639$$

و منه:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 50 \pm (2.0639) \frac{8}{\sqrt{25}}$$

46.698 .... 53.302

التقريب من أجل العينات الكبيرة

- إن حقيقة أن t يقترب من z مع ازدياد n يمكننا من استخدام التقريب الطبيعي من أجل $n \geq 30$

التقريب من أجل n كبيرة	الصيغة الصحيحة
$\bar{x} \pm z_{\alpha/2} \dfrac{s}{\sqrt{n}}$	$\bar{x} \pm t_{\alpha/2} \dfrac{s}{\sqrt{n}}$

تحديد حجم العينة

- يمكن إيجاد حجم العينة المطلوب من أجل مستوى مرغوب من هامش الخطأ و مستوى الثقة $(1 - \alpha)$.

  - حجم العينة المطلوب: σ معلوم

$$n = \frac{z_{\alpha/2}^{2}\,\sigma^{2}}{e^{2}} = \left(\frac{z_{\alpha/2}\,\sigma}{e}\right)^{2}$$

مثال

إذا كان 45 = σ فما هو حجم العينة الذي نحتاجه لكي نكون واثقين 90% بأننا على صواب ضمن 5 ± ؟

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{e}\right)^2 = \left(\frac{1.645(45)}{5}\right)^2 = 219.19$$

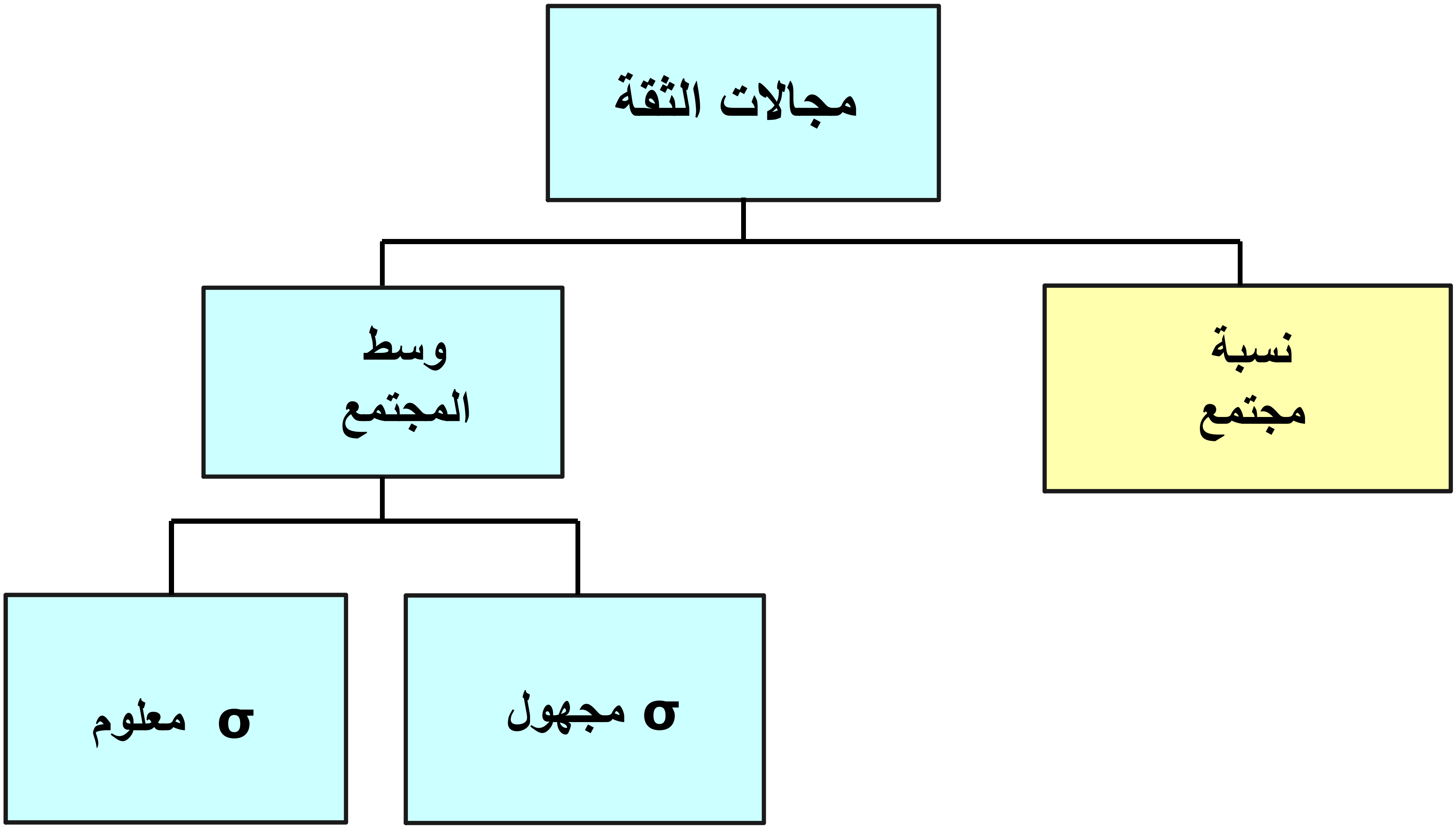
حجم العينة اللازم هو إذاً **n = 220**

(قرّب للأعلى دائماً)

إذا كانت σ مجهولة

- عند حساب حجم العينة المطلوب يمكن استخدام تقريب لـ σ عندما يكون σ مجهولاً.
  - استخدم قيمة من المتوقع ألا تقل عن القيمة الحقيقية لـ σ
  - قم بجمع عينة مبدئية و استخدام الانحراف المعياري المحسوب من هذه العينة كتقريب.

Confidence Intervals مجالات الثقة



مجالات ثقة من أجل نسبة المجتمع p

- يمكن تقدير مجال لنسبة مجتمع (p)عن طريق إضافة درجة من السماح بعدم التأكّد لنسبة العينة ($\bar{p}$).

مجالات ثقة من أجل نسبة المجتمع p

تتمة

- تذكر أن توزيع نسبة العينة عندما يكون حجم العينة كبيراً يصبح طبيعياً تقريباً بانحراف معياري:

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

- وسنقوم بتقدير هذا الأخير باستخدام

$$s_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

حدود مجال الثقة

- يمكن حساب الحدود الدنيا و العليا للثقة من أجل نسبة مجتمع باستخدام الصيغة:

$$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

- حيث:
  - z هي القيمة الطبيعية المعيارية من أجل المستوى المرغوب من الثقة.
  - $\bar{P}$ هي نسبة المجتمع.
  - n هو حجم العينة.

مثال

- أظهرت عينة عشوائية مكونة من 100 فرد أن 25 منهم كانوا يساريين (يعتمدون على اليد اليسرى!)
- قم بإنشاء مجال ثقة 95% من أجل النسبة الحقيقية لليساريين.

مثال

تتمة

- أظهرت عينة عشوائية مكونة من 100 فرد أن 25 منهم كانوا يساريين (يعتمدون على اليد اليسرى!) قم بإنشاء مجال ثقة 95% من أجل النسبة الحقيقية لليساريين.

1. $\bar{p} = 25/100 = .25$

2. $S_{\bar{p}} = \sqrt{\bar{p}(1-\bar{p})/n} = \sqrt{.25(.75)/n} = .0433$

3. $$.25 \pm 1.96\ (.0433)$$

$$\boxed{0.1651\ \ \ldots\ldots 0.3349}$$

التفسير

- نحن واثقون 95% بأن النسبة الحقيقة لليساريين في المجتمع هي بين: 16.51% و 33.49%

- على الرغم من أن هذا المجال يمكن أن و يمكن ألا يحتوي النسبة الحقيقية، فإن 95% من المجالات المنشأة من عينات من الحجم 100 و بهذهالطريقة سوف تحتوي على النسبة الحقيقية.

تغيير حجم العينة

- الزيادات في حجم العينة تنقِص من عرض مجال الثقة.
  - Increases in the sample size reduce the width of the confidence interval.

مثال:

إذا ما ضاعفنا حجم العينة في المثال السابق إلى 200، و إذا كان هناك 50 يساري في العينة، فإن المجال يبقى متمركزاً حول 25. و لكن عرضه يتقلص إلى:

.19 ...... .31

إيجاد حجم العينة اللازم من أجل مسائل النسب

$$e = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

عرف مجال الخطأ

$$n = \frac{z_{\alpha/2}^{2}\, p(1-p)}{e^{2}}$$

أحسب n

يمكن تقدير p باستخدام عينة مبدئية إذا ما دعت الضرورة (أو استخدم النسبة المحافظة $p = .50$)

كم هو حجم العينة ...؟

- ما هو مقدار ضخامة حجم العينة الضرورية لتقدير النسبة الحقيقية من القطع المعطوبة في مجتمع كبير ضمن 3% و بثقة 95%

افترض أن العينة المبدئية أعطت $\bar{p} = .12$

كم هو حجم العينة ...؟

تتمة

الحل:

استخدم $Z = 1.96$ من أجل مستوى ثقة 95%

لدينا $E = .03$ و $\bar{p} = .12$ و بالتالي نستخدمها لتقدير p

$$\boxed{n = \frac{z_{\alpha/2}^{2}\, p(1-p)}{e^{2}} = \frac{(1.96)^{2}(.12)(1-.12)}{(.03)^{2}} = 450.74}$$

استخدم إذاً $n = 451$

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة العاشرة /القسم النظري/

اختبار الفرضيات 1

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 - 2009

الدكتور معاذ الشرفاوي الجزائرلي

أهداف المحاضرة

يفترض بنهاية المحاضرة أن تصبح قادراً على:

- صياغة فرضية العدم و الفرضية البديلة في سياق تطبيقات تنطوي على وسط مجتمع أو نسبة مجتمع.
- صياغة قاعدة اتخاذ قرار لاختبار فرضية.
- استخدام إحصاءة الاختبار، القيمة الحرجة، و قيمة P لاختبار فرضية العدم
- تمييز الخطأ من النوع الأول Type I error

  و الخطأ من النوع الثاني Type II errors.
- حساب احتمال الوقوع في الخطأ من النوع الثاني Type II errors.

What is a Hypothesis?

ما هو المقصود بالفرضية؟

الفرضية هي عبارة عن إدعاء (افتراض)
حول بارامتر المجتمع

- وسط مجتمع

مثال: إن وسط فاتورة الهاتف الجوال في هذه المدينة هو

$\mu = \$42$

- نسبة مجتمع

مثال: إن نسبة البالغين الذين يملكون هاتفاً جوالاً في هذه المدينة هي:

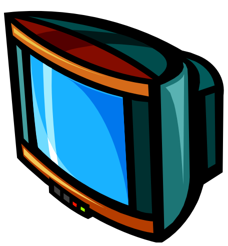
$p = .68$

The Null Hypothesis, H_0 فرضية العدم

- تحدد الفرض (الرقمي) المطلوب اختباره

مثال: العدد الوسطي لأجهزة التلفزيون في بيوت هذه المدينة لا يقل عن ثلاثة أجهزة. أو: $H_0 : \mu \geq 3$

- فرضية العدم يتم إنشاؤه دائماً من أجل بارامتر المجتمع و ليس من أجل إحصاءة من عينة.



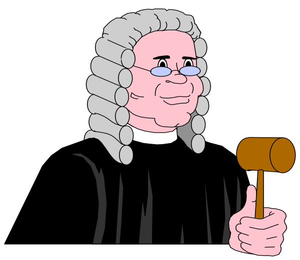
$$H_0 : \mu \geq 3$$

$$\cancel{H_0 : \bar{x} \geq 3}$$

The Null Hypothesis, H_0 فرضية العدم

تتمة

- تنطلق من اعتبار أن فرضية العدم صحيحة مبدئياً.
  - تماماً كالقول "بريئة حتى تثبت إدانتها"

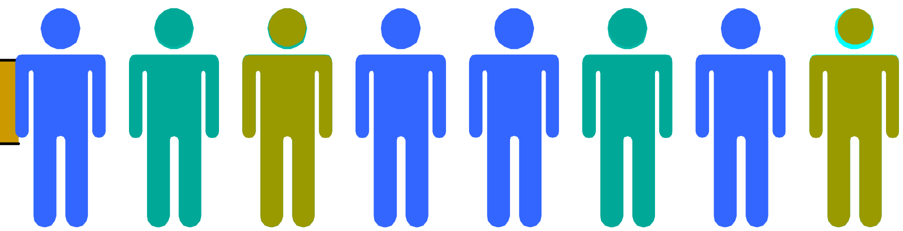


- تشير عادةً إلى الحالة الراهنة.
- تنطوي صياغتها دائماً على أحد الإشارات:
  - "$=$" أو "$\leq$" أو "$\geq$"
- يمكن أن تُرفَض أو ألا ترفض.

The Alternative Hypothesis, H_A الفرضية البديلة

- عكس فرضية العدم
  - مثال (تابع المثال السابق): العدد الوسطي لأجهزة التلفزيون في بيوت هذه المدينة يزيد عن ثلاثة أجهزة. أو: ($H_A: \mu < 3$)
- تتحدى الحالة الراهنة
- لا يمكن أبداً أن تحتوي على الإشارات التالية:
  - "=" أو "≤" أو "≥"
- يمكن أن تُقبل أو ألا تُقبل.
- هي بشكل عام الفرضية التي يُؤمن الباحث بصحتها أو يحتاج إلى إثباتها.

عملية اختبار الفرضية



مجتمع

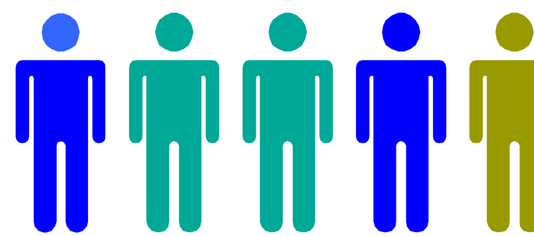
إدعاء:

وسطي عمر
المجتمع هو 50

فرضية العدم:

H_0:

اسحب عينة عشوائية



عينة

افترض أن وسطي
عمر العينة هو 20

$\bar{x} = 20$

هل من المرجح ان يكون $x = 20$
إذا كان $\mu = 50$ ؟

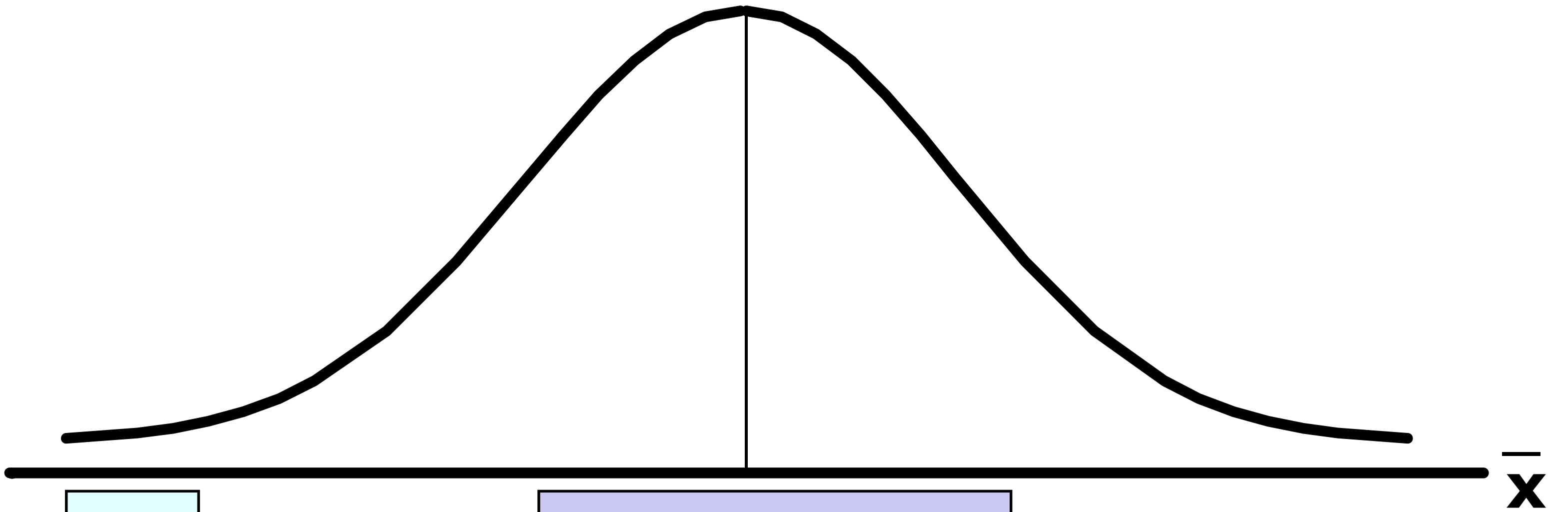
إن لم يكن ذلك مرجحاً

ارفض

فرضية العدم

سبب رفض H_0

توزيع معاينة الأوساط



20

$\mu = 50$

إذا كانت

H_0صحيحة

إذا كان من غير المرجّح أن نحصل على وسط عينة بهذه القيمة

و إذا كان وسط المجتمع

فإننا نرفض فرضية العدم التي تقول بأن:

$\mu = 50.$

Level of Significance, α مستوى الدلالة

- **يحدد القيم المستبعدة لإحصاءة العينة إذا كانت فرضية العدم صحيحة.**
  - **يحدد منطقة الرفض في توزيع المعاينة.**
  - **يرمز له بـ α**
  - **القيم النموذجية (الشائعة) هي :** .01, .05, or .10
- يتم اختياره من قبل الباحث منذ البدء.
- يعطي القيمة (أو ربما القيم) الحرجة للاختبار

مستوى الدلالة و منطقة الرفض

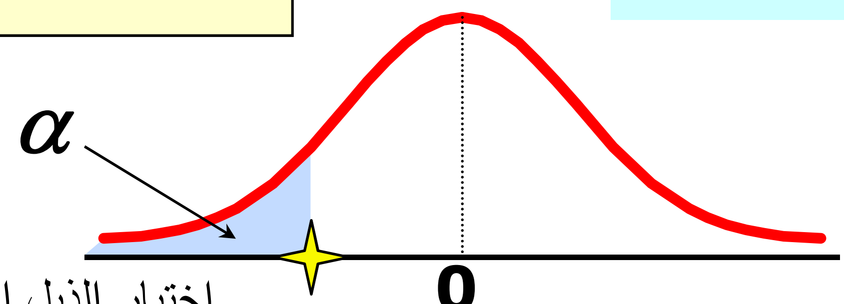
مستوى الدلالة = α

تمثل القيمة الحرجة

المنطقة المظللة هي منطقة الرفض

$H_0: \mu \geq 3$

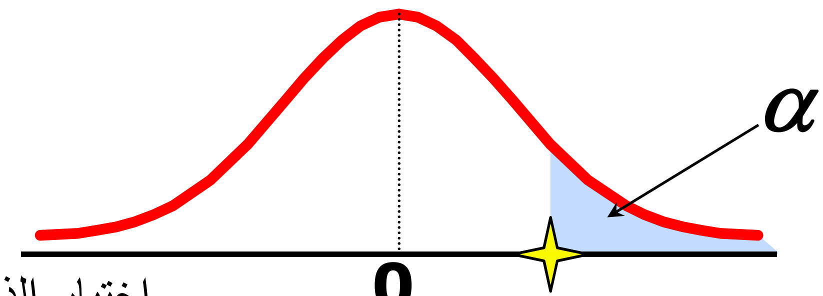
$H_A: \mu < 3$



اختبار الذيل الأدنى (اختبار من اليسار)

$H_0: \mu \leq 3$

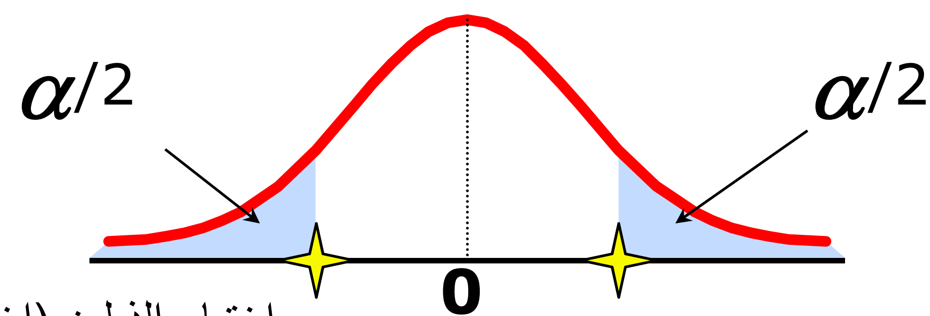
$H_A: \mu > 3$



اختبار الذيل الأعلى (اختبار من اليمين)

$H_0: \mu = 3$

$H_A: \mu \neq 3$



اختبار الذيلين (اختبار من الطرفين)

أخطاء في اتخاذ القرار

- **الخطأ من النوع الأول Type I Error**
  - رفض فرضية عدم صحيحة.
  - و يعتبر نوعاً مهمّاً من الأخطاء

احتمال الوقوع في الخطأ من النوع الأول = α

- **و يطلق عليه اسم مستوى دلالة الاختبار**
- يتم تحديده من الباحث بشكل مسبق

أخطاء في اتخاذ القرار

تتمة

- **الخطأ من النوع الثاني Type II Error**
  - الفشل في رفض فرضية عدم خاطئة

احتمال الوقوع في الخطأ من النوع الأول = β

Outcomes and Probabilities

النواتج و الاحتمالات

النتائج المحتملة لاختبار فرضية

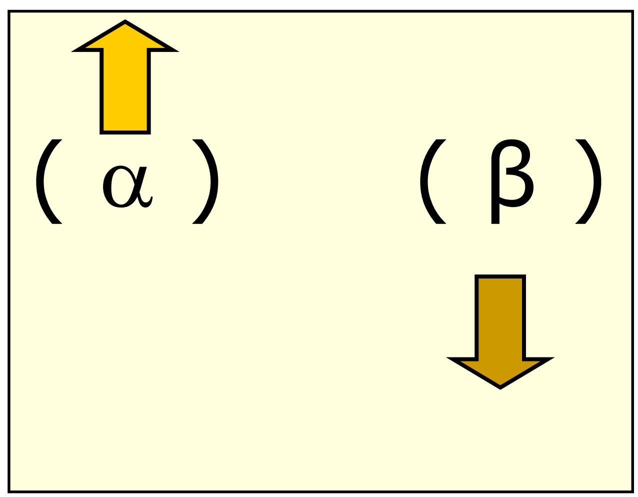
	الحالة	
القرار	H_0 صحيحة	H_0 خطأ
لا ترفض H_0	لا يوجد خطأ $(1 - \alpha)$	خطأ من النوع الثاني (β)
ارفض H_0	خطأ من النوع الأول (α)	لا يوجد خطأ $(1 - \beta)$

الناتج

(الاحتمال)

العلاقة بين الخطأ الأول و الخطأ الثاني Type I & II Error Relationship

- لا يمكن أن يحدثا بآن معاً.
- لا يمكن أن يحدث الخطأ الأول إلا إذا كانت فرضية العدم صحيحة.
- لا يمكن أن يحدث الخطأ الثاني إلا إذا كانت فرضية العدم خاطئة.



العوامل المؤثرة في الخطأ الثاني

- بفرض بقاء العوامل الأخرى ثابتة

- β الفرق بين البارامتر تحت الفرضية و قيمته الحقيقية

$\beta\uparrow$ $\alpha\downarrow$

$\beta\uparrow$ $\sigma\uparrow$

$\beta\uparrow$ $n\downarrow$

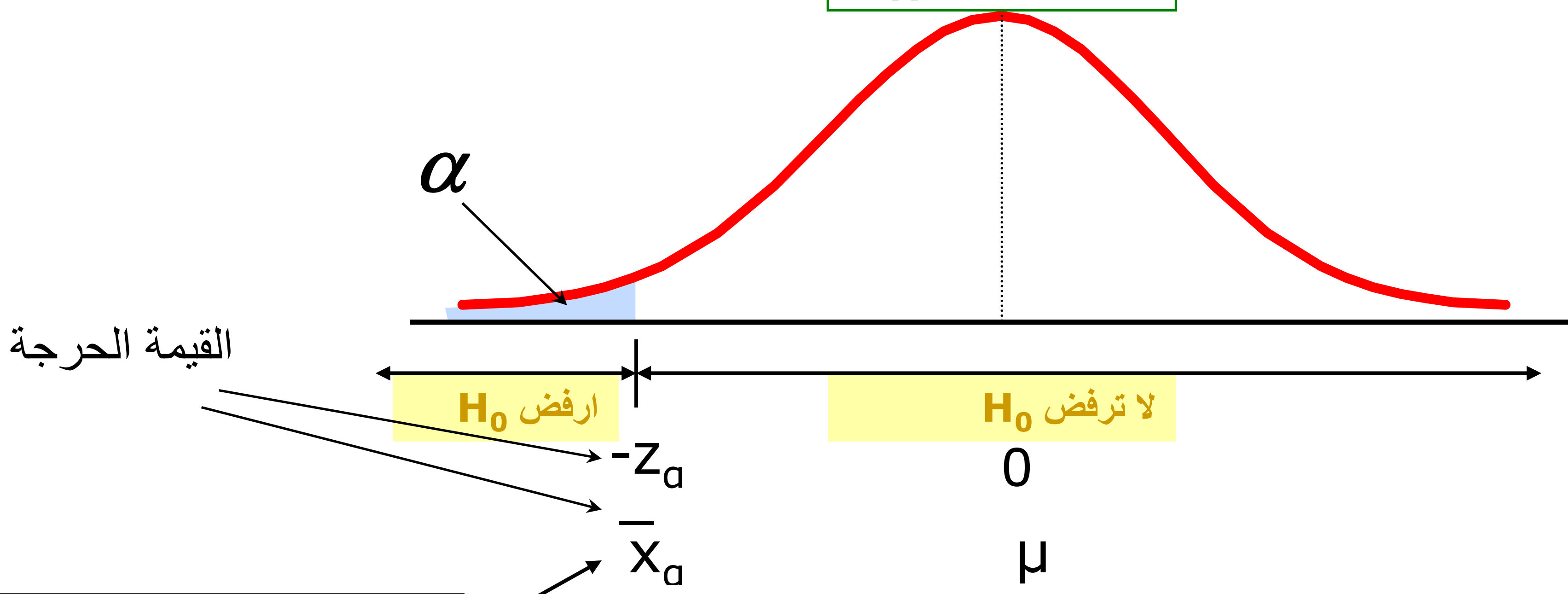
مدخل الاختبار بالقيمة الحرجة

- حول إحصاءة العينة sample statistic ($\bar{x}$ مثلاً) إلى إحصاءة اختبار test statistic (z أو t)
- حدد القيمة أو القيم الحرجة الخاصة بمستوى الدلالة المعتمد و ذلك من جدول أو من الحاسب.
- إذا وقعت إحصاءة الاختبار في منطقة الرفض ارفض فرضية العدم، و إن لم يكن كذلك فلا ترفضها.

Lower Tail Tests اختبارات الذيل الأدنى

$H_0: \mu \geq 3$

$H_A: \mu < 3$

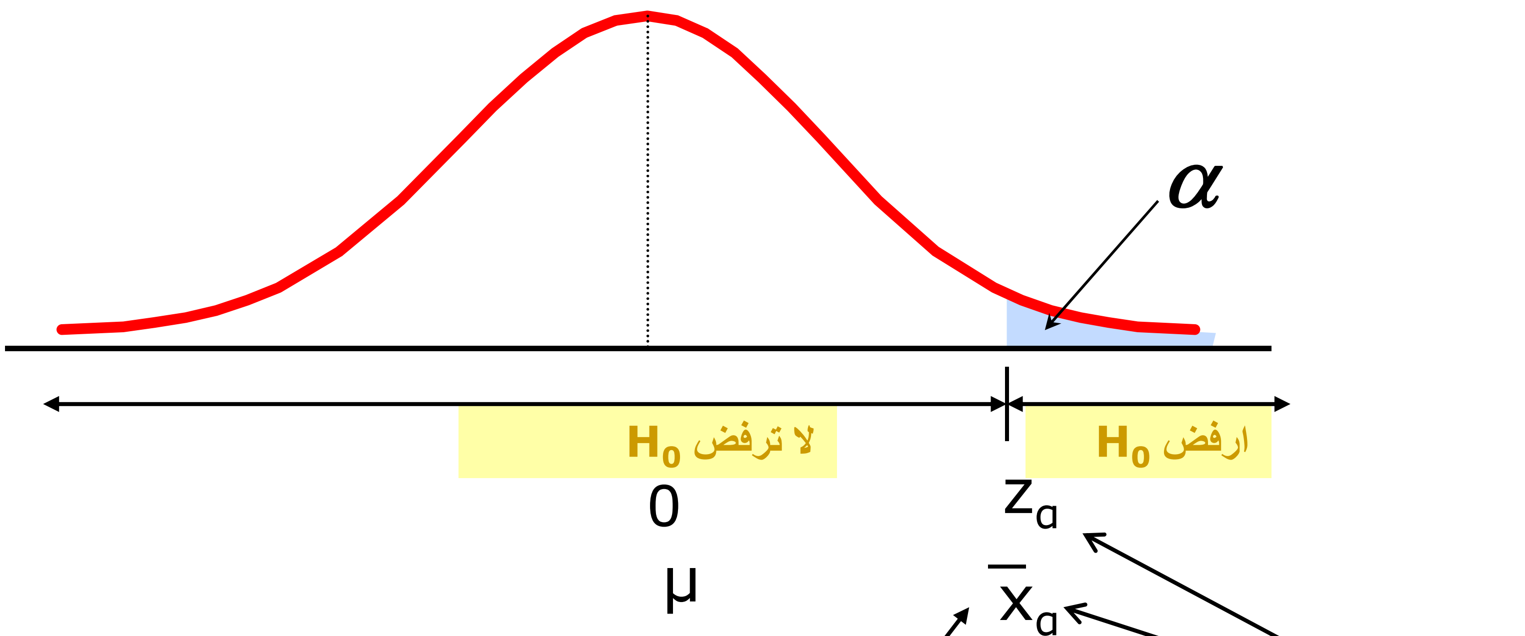


$$\bar{x}_\alpha = \mu - z_\alpha \frac{\sigma}{\sqrt{n}}$$

Upper Tail Tests اختبارات الذيل الأعلى

$H_0: \mu \leq 3$

$H_A: \mu > 3$



$$\bar{x}_\alpha = \mu + z_\alpha \frac{\sigma}{\sqrt{n}}$$

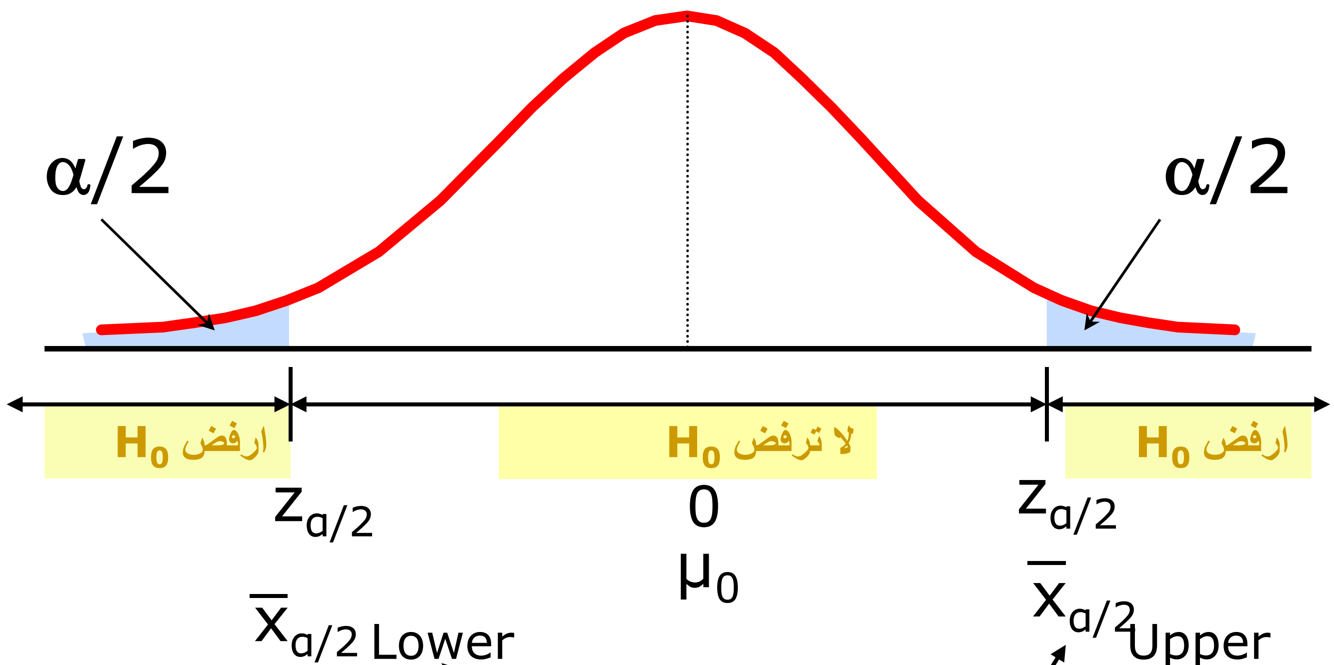
Two Tailed Tests اختبارات الذيلين (من الطرفين)

$\pm z_{\alpha/2}$: هناك قيمتان حرجتان

$H_0: \mu = 3$
$H_A: \mu \neq 3$

$\bar{x}_{\alpha/2}$ دنيا

$\bar{x}_{\alpha/2}$ عليا



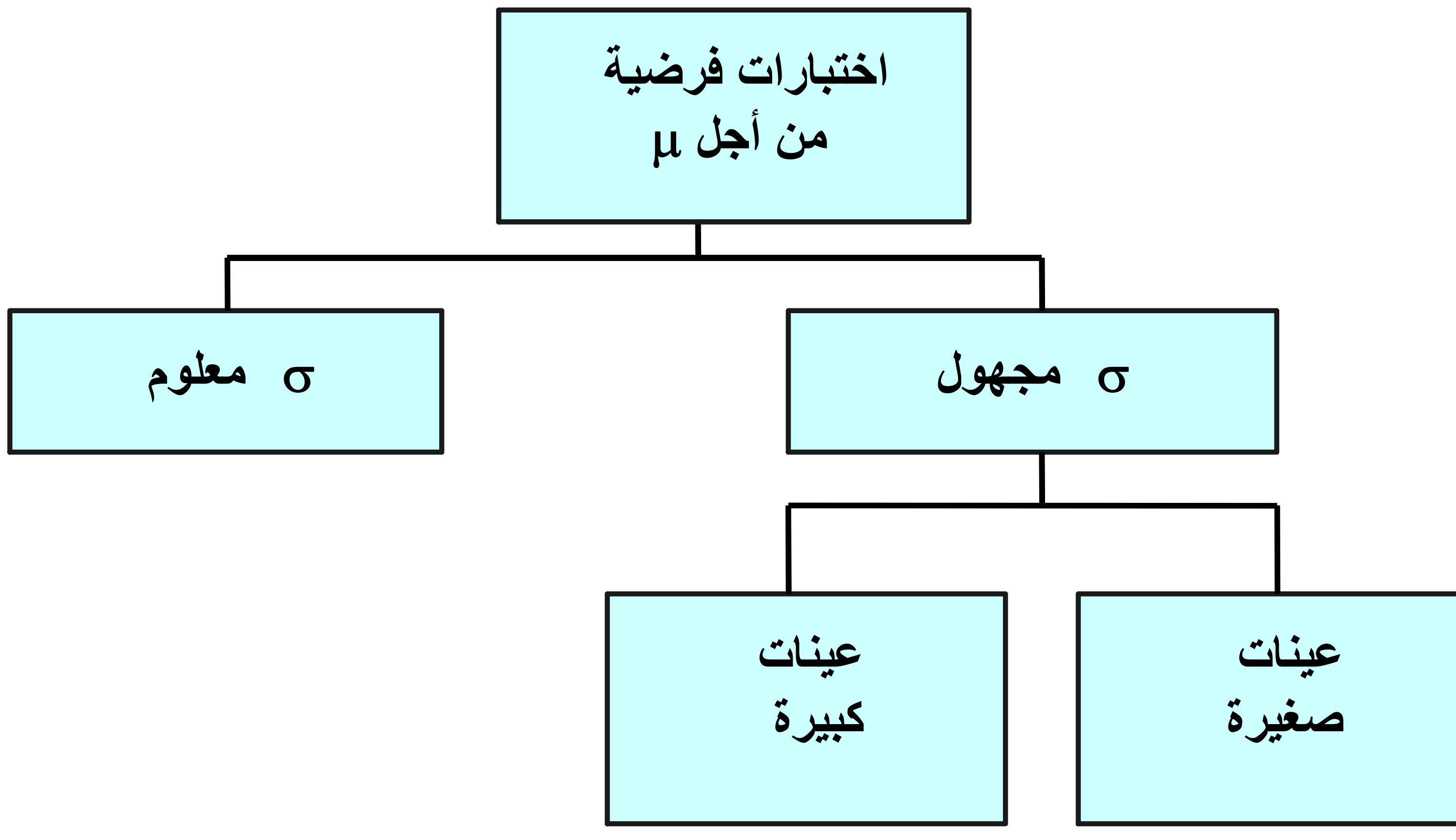
$\bar{x}_{\alpha/2}$ دنيا

$\bar{x}_{\alpha/2}$ عليا

$$\bar{x}_{\alpha/2} = \mu \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

مدخل الاختبار بالقيمة الحرجة

حول إحصاءة العينة sample statistic إلى إحصاءة اختبار test statistic (t أو z)



حساب إحصاءة الاختبار Test Statistic

اختبارات فرضية
من أجل μ

σ معلوم

σ مجهول

عينات
كبيرة

عينات
صغيرة

إحصاءة الاختبارهي:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

حساب إحصاءة الاختبار Test Statistic

تتمة

اختبارات فرضية
من أجل μ

σ مجهول

σ معلوم

عينات
صغيرة

عينات
كبيرة

على أنها تقرَّب
أحياناً إلى z

إحصاءة الاختبارهي:

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

حساب إحصاءة الاختبار Test Statistic

تتمة

اختبارات فرضية
من أجل μ

σ معلوم

σ مجهول

عينات
كبيرة

عينات
صغيرة

إحصاءة الاختبار هي:

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

(يجب أن يكون المجتمع طبيعي تقريباً)

مراجعة: خطوات اختبار الفرضية

1. حدد القيمة المجتمعية population value محل الاهتمام
2. قم بصياغة فرضية العدم المناسبة و البديلة الموافقة
3. حدد مستوى الدلالة المرغوب significance level
4. عيّن منطقة الرفض rejection region
5. اسحب العينة و احسب إحصاءة الاختبار test statistic
6. اتخذ القرار reach a decision
7. فسّر النتيجة interpret the result

مثال على اختبار الفرضية

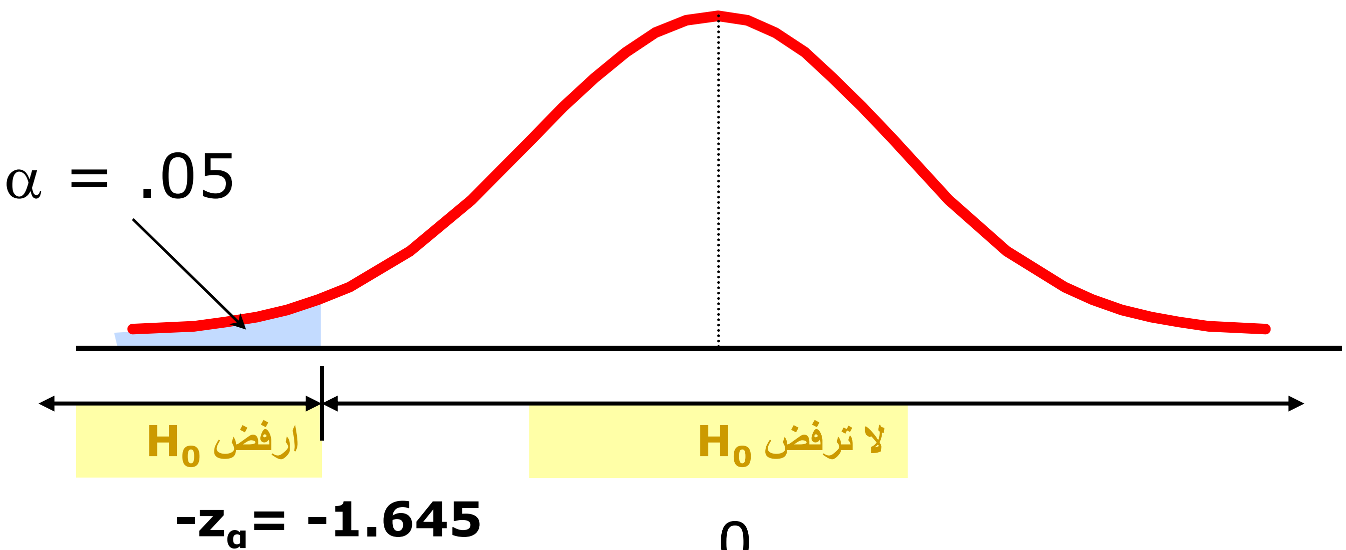
قم باختبار الادعاء القائل بأن العدد الحقيق لأجهزة التلفزيون في منازل الولايات المتحدة يبلغ 3 على الأقل
(افترض أن $\sigma = 0.8$)

- 1.حدد القيمة المجتمعية population value محل الاهتمام
  - العدد الوسطي لأجهزة التلفزيون في المنازل
- 2. قم بصياغة فرضية العدم المناسبة و البديلة الموافقة
  - (اختبار ذيل أدنى / من الطرف الأيسر)
  - $H_0: \mu \geq 3$ $H_A: \mu < 3$
- 3. حدد مستوى الدلالة المرغوب
  - افترض أن $\alpha = .05$ هي القيمة المرغوبة لمستوى الدلالة

مثال على اختبار الفرضية

تتمة

4. عيّن منطقة الرفض rejection region



هذا الاختبار هو اختبار وحيد الذيل (من طرف واحد) مع $\alpha = .05$

بما أن **σ معلوم** فإن القيمة الحرجة لـ z هي:

ارفض H_0 إذا كان لديك $z < z_\alpha = -1.645$ ، فإن لم يكن لا ترفض H_0

مثال على اختبار الفرضية

تتمة

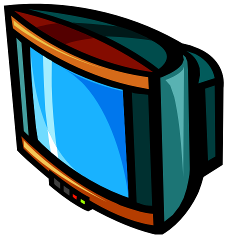
5. اسحب العينة و احسب إحصاءة الاختبار test statistic

افترض أن العينة المسحوبة أعطت النتائج التالية:

$n = 100, \ \bar{x} = 2.84$ ($\sigma = 0.8$ معلوم)

- و بالتالي فإن قيمة إحصاءة الاختبار هي:

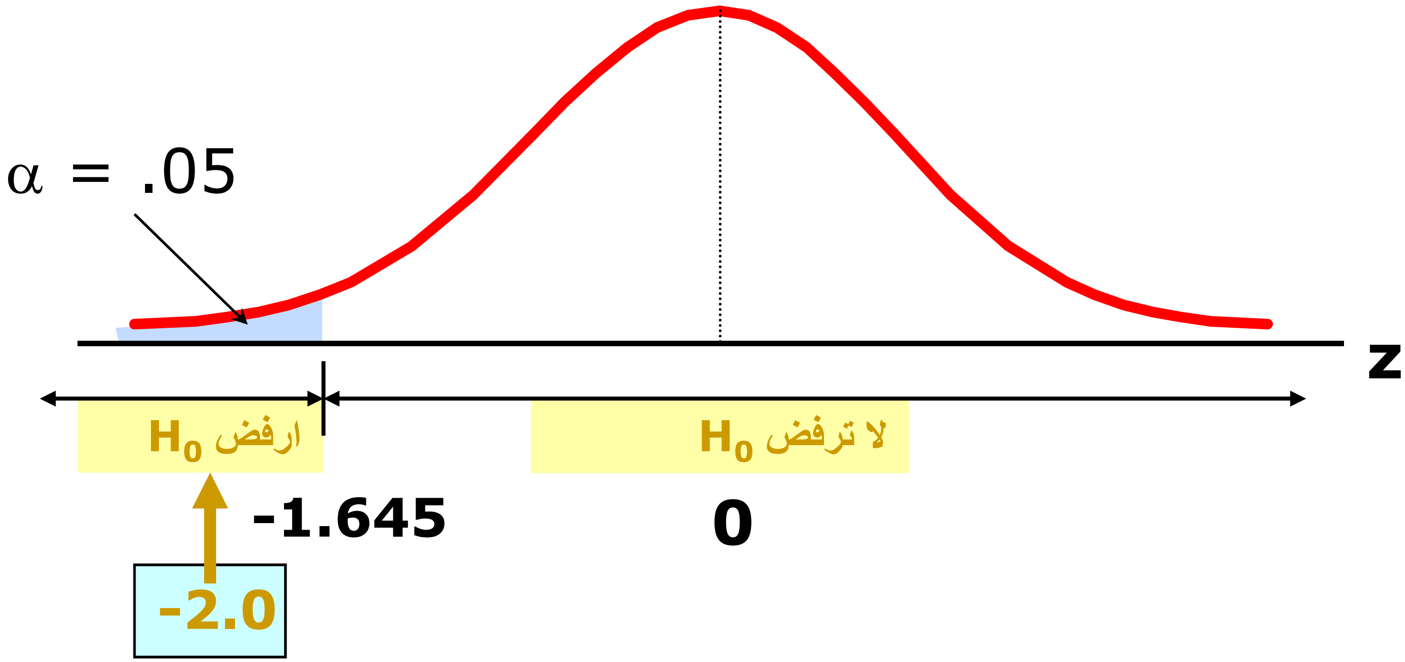
$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.84 - 3}{\frac{0.8}{\sqrt{100}}} = \frac{-.16}{.08} = -2.0$$



مثال على اختبار الفرضية

تتمة

6. اتخذ القرار reach a decision



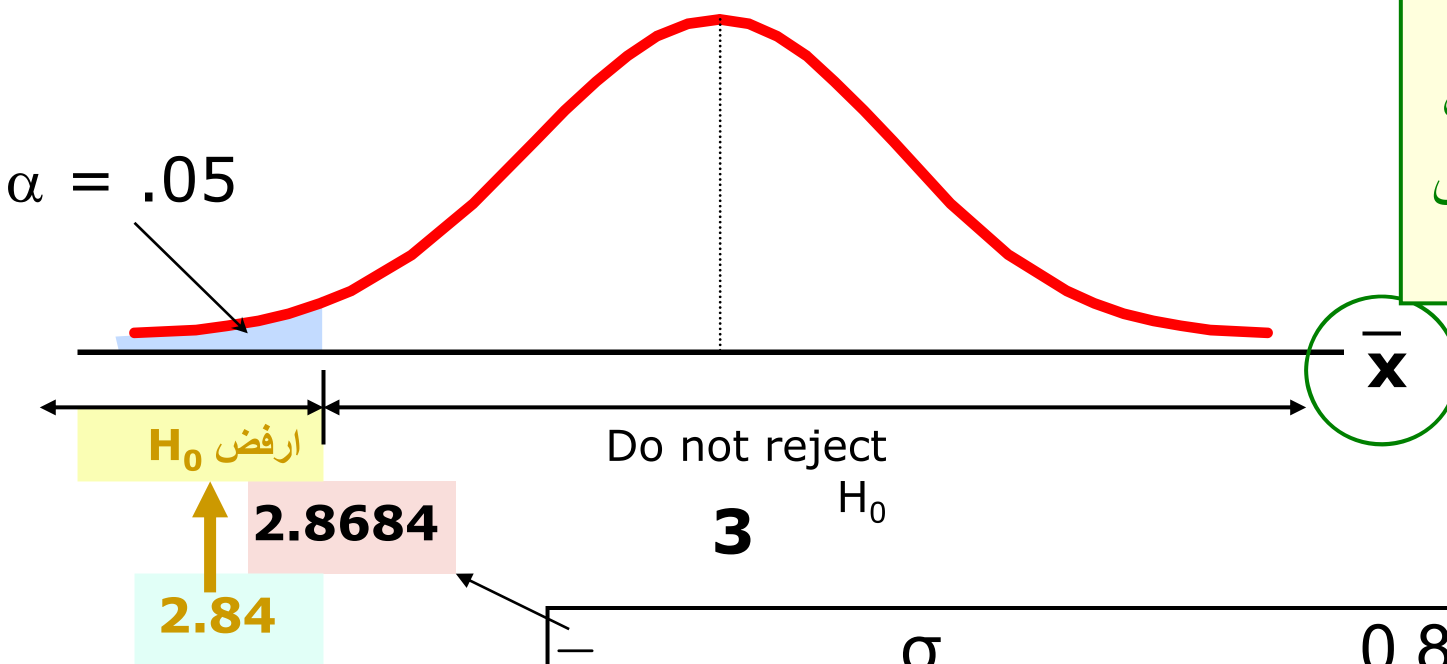
7. **التفسير:** بما أن $z = -2.0 < -1.645$ فإننا **<u>نرفض فرضية العدم</u>** القائلة بأن متوسط عدد التلفزيون في منازل المدينة المعنية لا يقل عن ثلاثة أجهزة.

مثال على اختبار الفرضية

تتمة

- هناك طريقة بديلة لإنشاء منطقة الرفض:

و الآن يعبر عنه بوحدات من $\bar{x}$ و ليس من z



$$\bar{x}_\alpha = \mu - z_\alpha \frac{\sigma}{\sqrt{n}} = 3 - 1.645\frac{0.8}{\sqrt{100}} = 2.868$$

بما أن $\bar{x} = 2.84 < 2.8684$

فإننا نرفض فرضية العدم

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الحادية عشرة /القسم النظري/

اختبار الفرضيات 2

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 - 2009

الدكتور معاذ الشرفاوي الجزائرلي

مدخل p-Value لاختبار الفرضية

- قم بتحويل إحصاءة العينة ($\bar{x}$ مثلاً) إلى إحصاءة اختبار
- إحصل على الـ p-value من الكمبيوتر أو من جدول
- قارن قيمة الاحتمال p-value مع α:
  - ارفض H_0 إذا كانت **p** أصغر من α
  - لا ترفض H_0 إذا كانت p أكبر أو تساوي α

If p-value $< \alpha$, reject H_0

If p-value $\geq \alpha$, do not reject H_0

مدخل p-Value لاختبار الفرضية

- p-value: احتمال الحصول على إحصاءة اختبار أكثر تطرفاً ($\leq$ أو $\geq$) من قيمة العينة المشاهدة بفرض أن H_0 صحيحة.

- و تسمى أيضاً القيمة المشاهدة للدلالة.
  - أصغر قيمة لـ α و التي من أجلها يمكن رفض H_0.

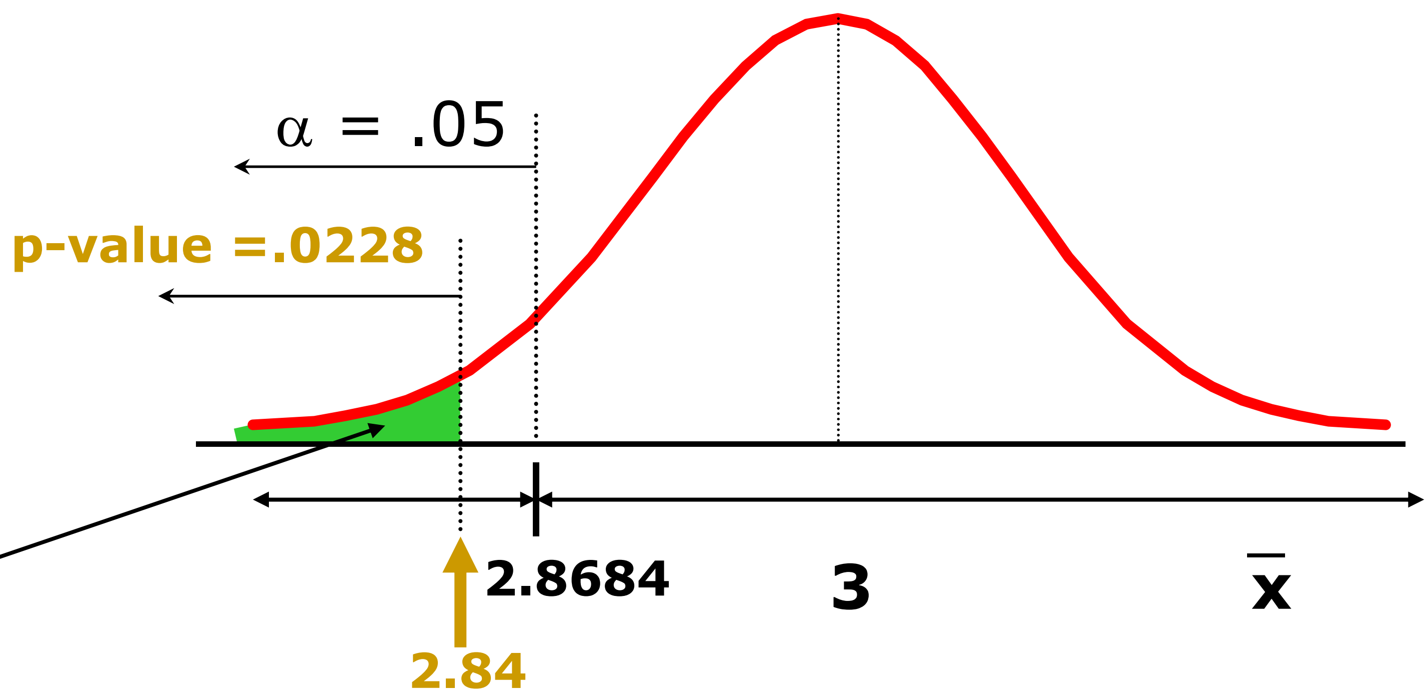
مثال على الاختبار باستخدام p-value

- ما هي فرصة مشاهدة وسط عينة قدره 2.84 أو أدنى، إذا ما كان الوسط الحقيقي $\mu = 3.0$ ؟

$$P(\bar{x} < 2.84 \mid \mu = 3.0)$$

$$= P\left(z < \frac{2.84 - 3.0}{0.8/\sqrt{100}}\right)$$

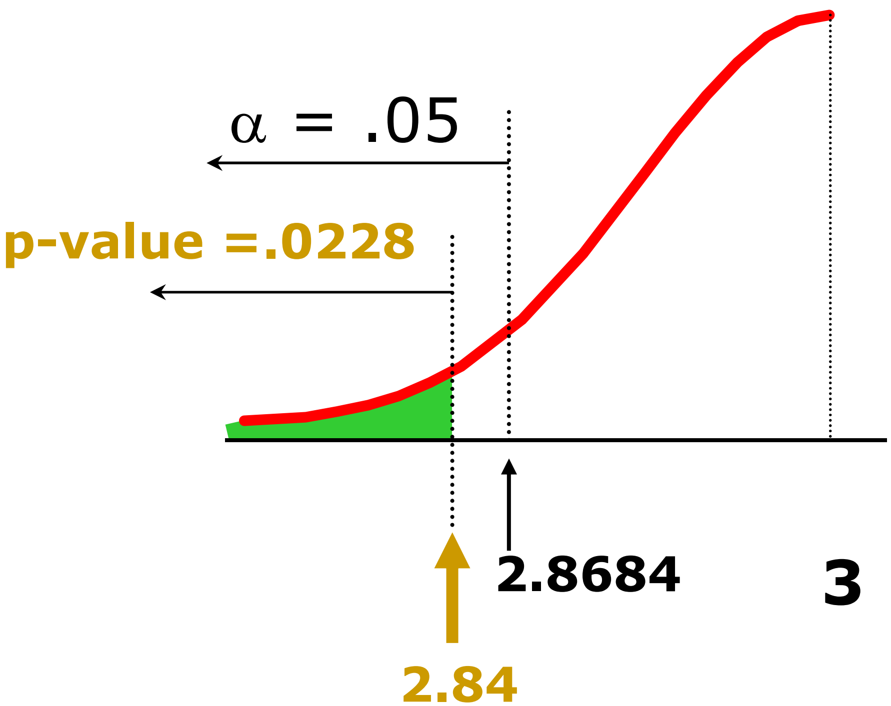
$$= P(z < -2.0) = .0228$$



مثال على الاختبار باستخدام p-value

تتمة

- قارن قيمة الاحتمال p-value مع α:
  - ارفض H_0 إذا كانت **p** أصغر من α
  - لا ترفض H_0 إذا كانت p أكبر أو تساوي α

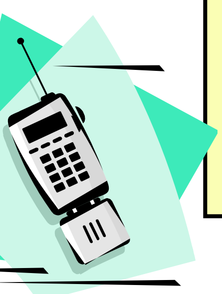


هنا: p-value = .0228

$\alpha = .05$

بما أن **.05 > .0228** فإننا نرفض H_0

مثال: اختبار z ذيل أعلى للوسط (σ معلوم)



يعتقد مدير لشركة هواتف نقالة أن فاتورة الهاتف الشهرية للزبائن قد ارتفعت ووصلت إلى متوسط يفوق $52 في الشهر. ترغب الشركة في اختبار هذا الادعاء. افترض أن الانحراف المعياري للمجتمع معلوم و يساوي $\sigma = 10$

صغ اختبار الفرضية

$H_0: \mu \le 52$ الوسط لا يفوق $52 شهرياً

$H_A: \mu > 52$ الوسط يفوق $52 شهريا

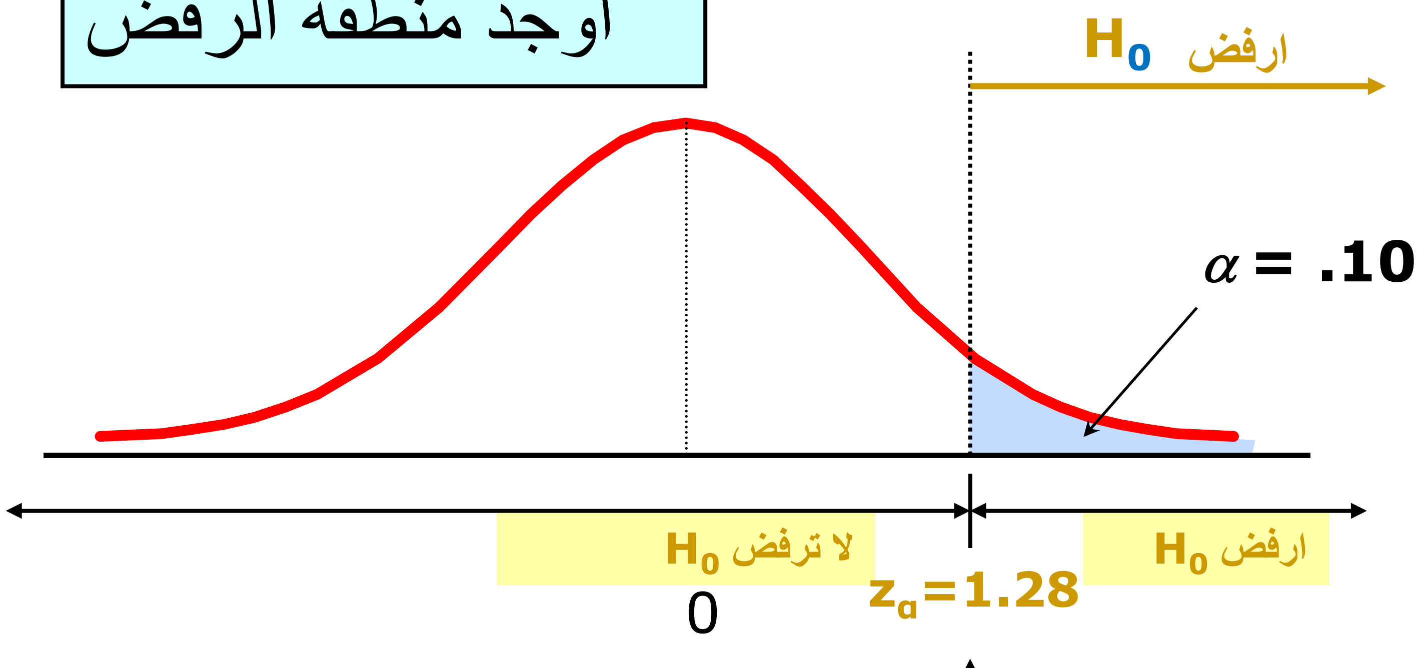
(أي أن لدى المدير من قرائن ما يدفعه لهذا الاعتقاد)

مثال: أوجد منطقة الرفض

تتمة

- افترض أن $\alpha = .10$ هي القيمة المختارة لهذا الاختبار

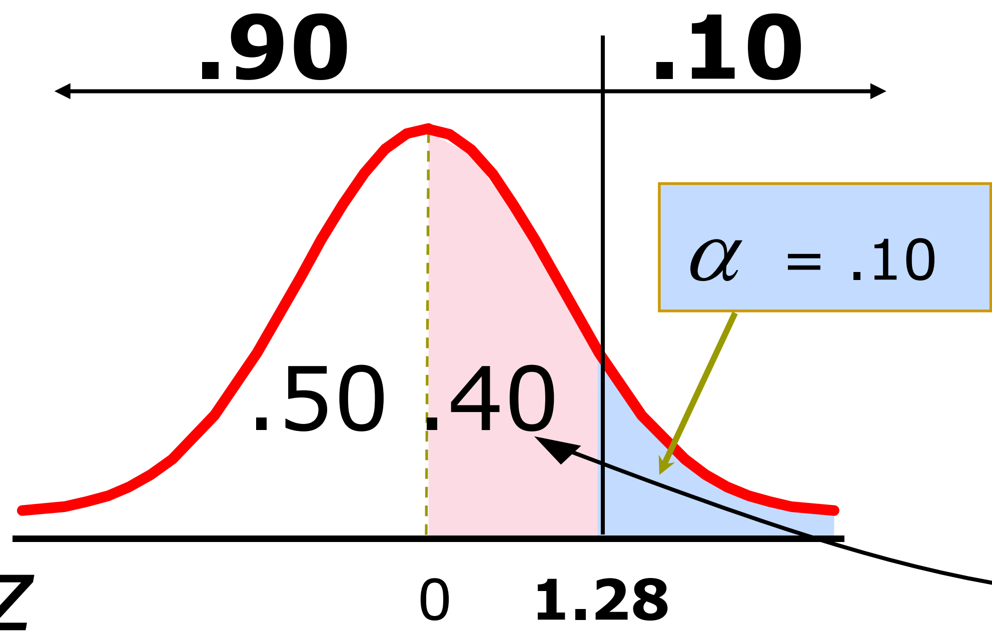
أوجد منطقة الرفض



Reject H_0 if $z > 1.28$

مراجعة: إيجاد القيمة الحرجة ـ ذيل واحد

كم هي z إذا كانت $\alpha = 0.10$



القيمة الحرجة= **1.28**

جزء مقتطع من جدول التوزيع الطبيعي المعياري

Z	.07	.08	.09
1.1	.3790	.3810	.3830
1.2	.3980	.3997	.4015
1.3	.4147	.4162	.4177

مثال: إحصاءة الاختبار

تتمة

قم بسحب العينة و احصل على إحصاءة الاختبار.

افترض أنك حصلت على عينة بالنتائج التالية:

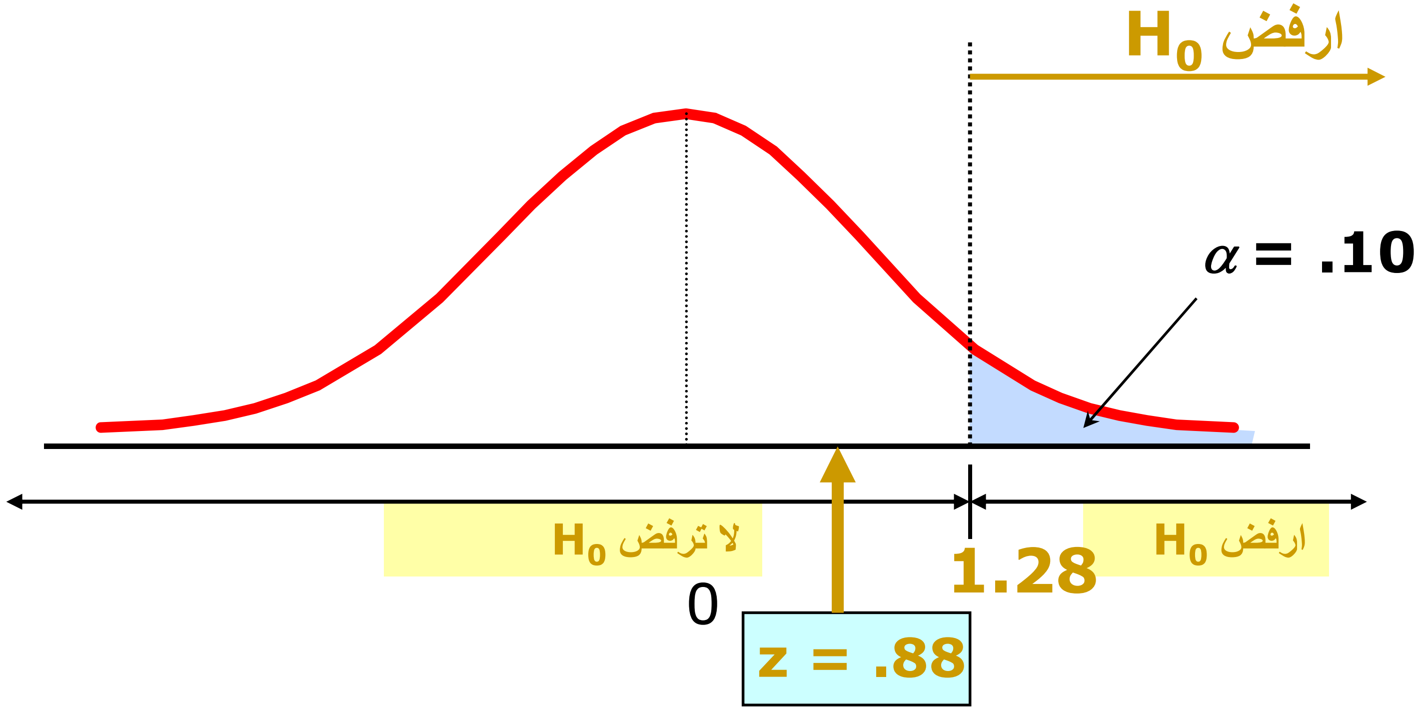
(يفترض أن $\sigma = 10$ مجهول) $n = 64, \quad \bar{x} = 53.1$

و بالتالي فإن إحصاءة الاختبار هي:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{53.1 - 52}{\frac{10}{\sqrt{64}}} = 0.88$$

مثال: القرار

قم باتخاذ القرار و فسّر النتيجة



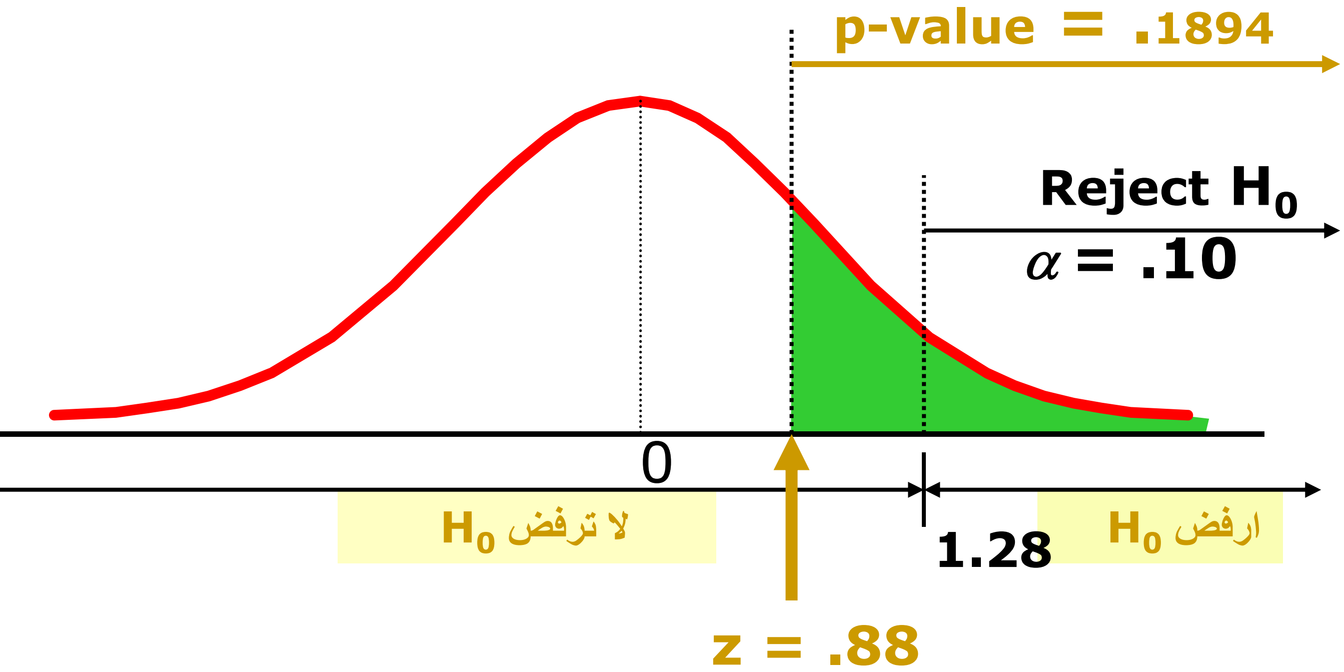
لا ترفض H0 لأن since z = 0.88 ≤ 1.28

لا يوجد قرينة كافية للقول بأن الفاتورة الوسطية تفوق 52$

تتمة

الحل باستخدام p -Value

أحسب p-value و قارنها مع α



$$P(\bar{x} \geq 53.1 \mid \mu = 52.0)$$

$$= P\left(z < \frac{53.1 - 52.0}{10 / \sqrt{64}} \right)$$

$$= P(z \geq 0.88) = .5 - .310$$

$$= .1894$$

لا ترفض **H0** على أساس أن:

p-value = .1894 > α = .10

مثال: اختبار الذيلين أو اختبار من الطرفين (σ مجهول)

يقال أن التكلفة الوسطية لغرفة فندقية في نيويورك هي 168$ في الليلة. و كان الوسط الناتج عن سحب عينة عشوائية من 25 فندقاً:

$\bar{x} = \$172.50$ و

$s = \$15.40$

اختبر عند مستوى $\alpha = 0.05$

(افترض أن توزيع المجتمع طبيعي)

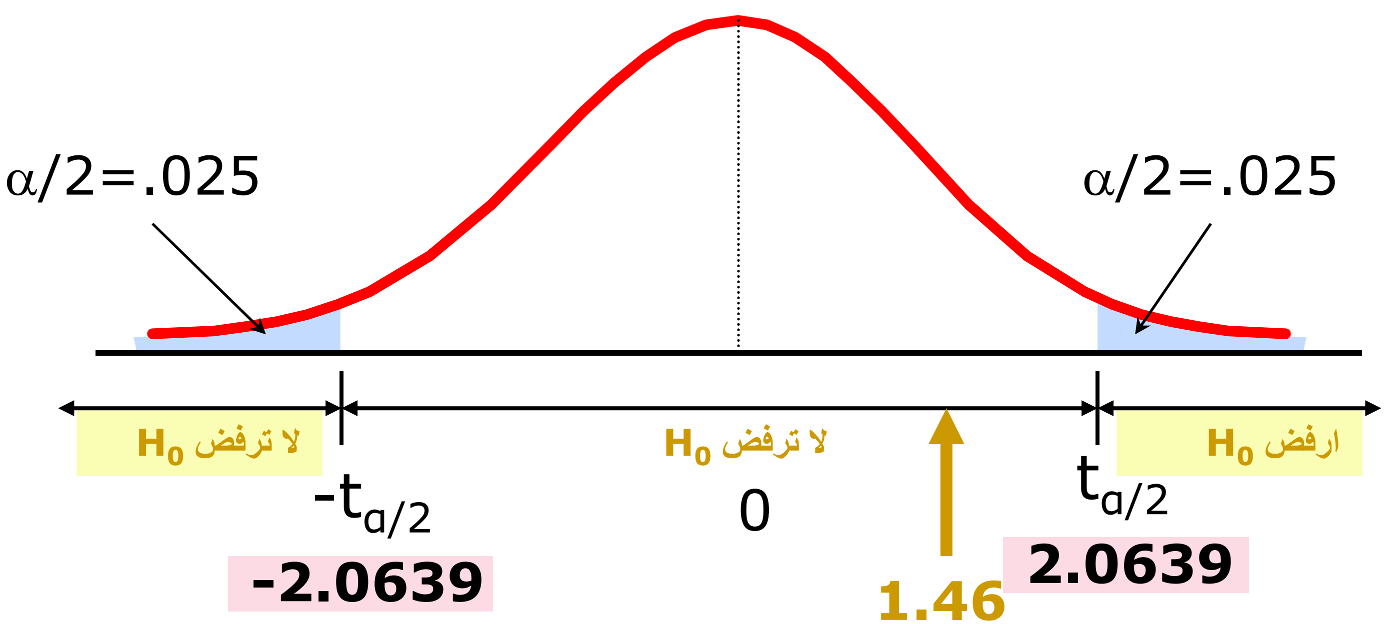
H_0: $\mu = 168$

H_A: $\mu \neq 168$

حل المثال: اختبار من الطرفين

$H_0: \mu = 168$

$H_A: \mu \neq 168$



$\alpha = 0.05$

$n = 25$

بما أن σ مجهول، استخدم t

القيمة الحرجة:

$t_{24} = \pm 2.0639$

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{172.50 - 168}{\frac{15.40}{\sqrt{25}}} = \boxed{1.46}$$

لا ترفض H_0 لانعدام الدليل الكافي على أن التكلفة الوسطية تختلف عن 168$

اختبار الفرضيات من أجل النسب

- ينطوي عل قيم فئوية categorical values
- هناك ناتجان محتملان:
  - "نجاح" و يعني أن المفردة تمتلك صفة معينة.
  - "فشل" و يعني أن المفردة لا تمتلكها.
- يرمز لنسبة (كسر) المجتمع المصنفة تحت فئة "نجاح" بـ p

النسب

تتمة

- يرمز لنسبة العينة المصنفة تحت فئة "نجاح" بـ $\overline{p}$

$$\overline{p} = \frac{x}{n}$$

- عندما يكون كل من np و n(1-p) أقل من 5، فإنه يمكن تقريب p بتوزيع طبيعي ذي وسط و انحراف معياري:

$$\mu_{\overline{P}} = p$$

$$\sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}}$$

اختبار الفرضيات من أجل النسب

- توزيع المعاينة لنسبة العينة طبيعي و بالتالي فإن إحصاءة الاختبار هي Z

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

اختبار الفرضية
من أجل **p**

np ≥ 5
و
n(1-p) ≥ 5

np < 5
أو
n(1-p) < 5

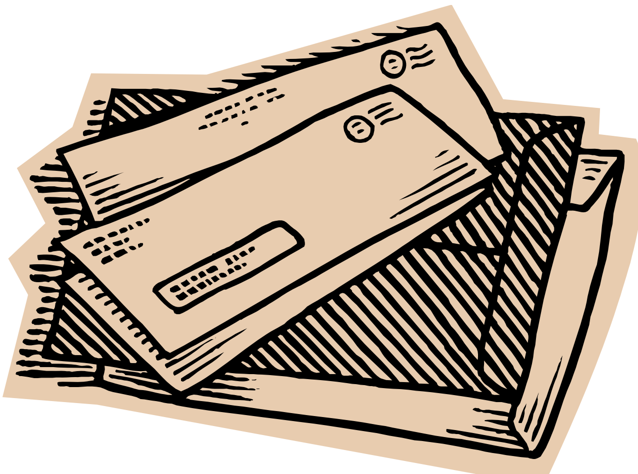
لن تناقش هذه الحالة
في هذه المحاضرة

مثال: اختبار z من أجل نسبة

تدعي شركة للتسويق بأنها تحصل على ردود بريدية بنسبة 8% من البريد المرسل.

و لاختبار هذا الادعاء، تم سحب 500 مفردة عشوائياً فكان عدد الردود 25.

قم بالاختبار مستخدماً مستوى دلالة قدره $\alpha = .05$



لاحظ:

✓ $n\,p = (500)(.08) = 40$

$n(1-p) = (500)(.92) = 460$

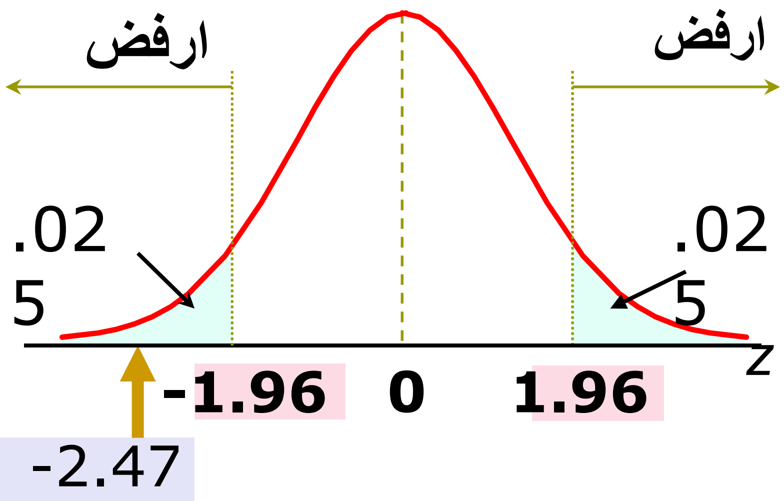
حل مثال: اختبار z من أجل نسبة

$H_0: p = .08$
$H_A: p \neq .08$

$\alpha = .05$

$n = 500, \quad \bar{p} = .05$

القيمة الحرجة: ± 1.96



إحصاءة الاختبار:

$$z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{.05 - .08}{\sqrt{\frac{.08(1-.08)}{500}}} = -2.47$$

القرار:

ارفض H_0 عند $\alpha = .05$

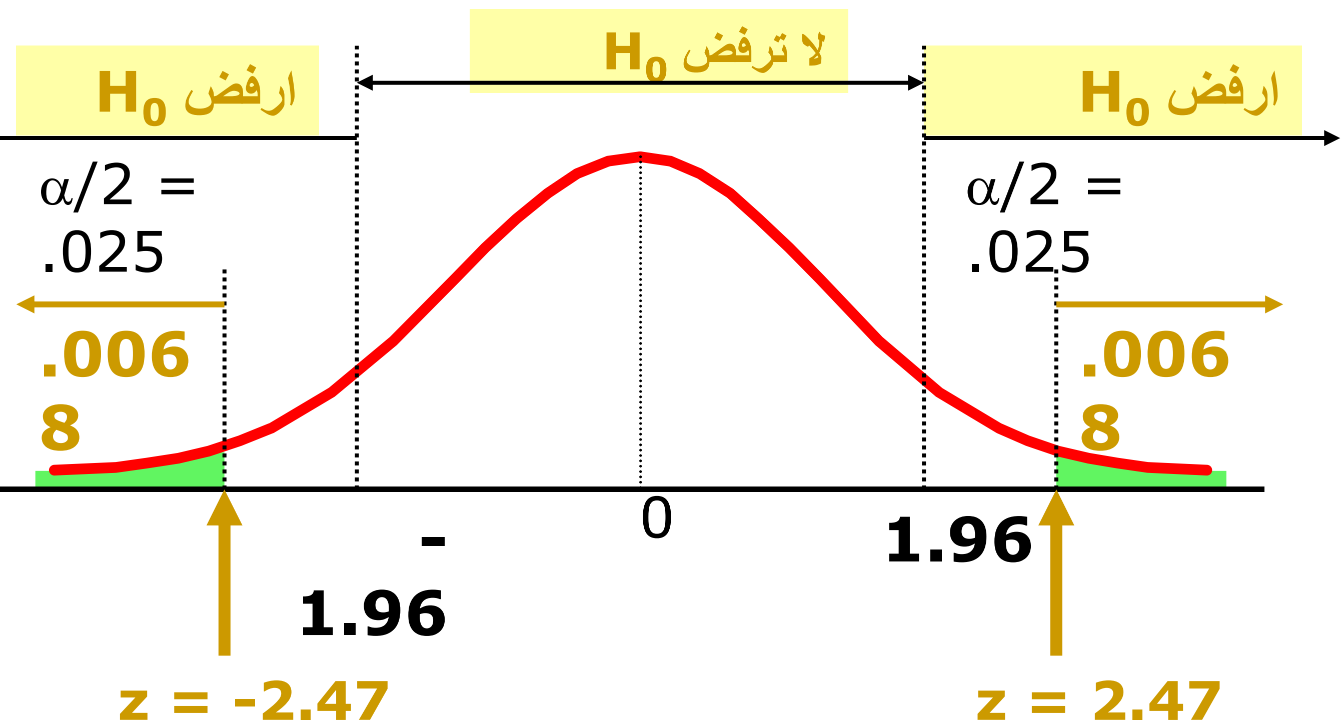
الاستنتاج:

هناك دليل كافٍ لرفض ادعاء الشركة بأن نسبة الاستجابة البريدية هي 8%

الحل باستخدام p -Value

أحسب p-value و قارنها مع α

(إذا كان الاختبار من الطرفين فإن p-value تكون من الطرفين أيضاً)



p-value = .0136:

$$P(z \le -2.47) + P(x \ge 2.47$$

$$= 2(.5 - .4932)$$

$$= 2(.0068) = 0.0136$$

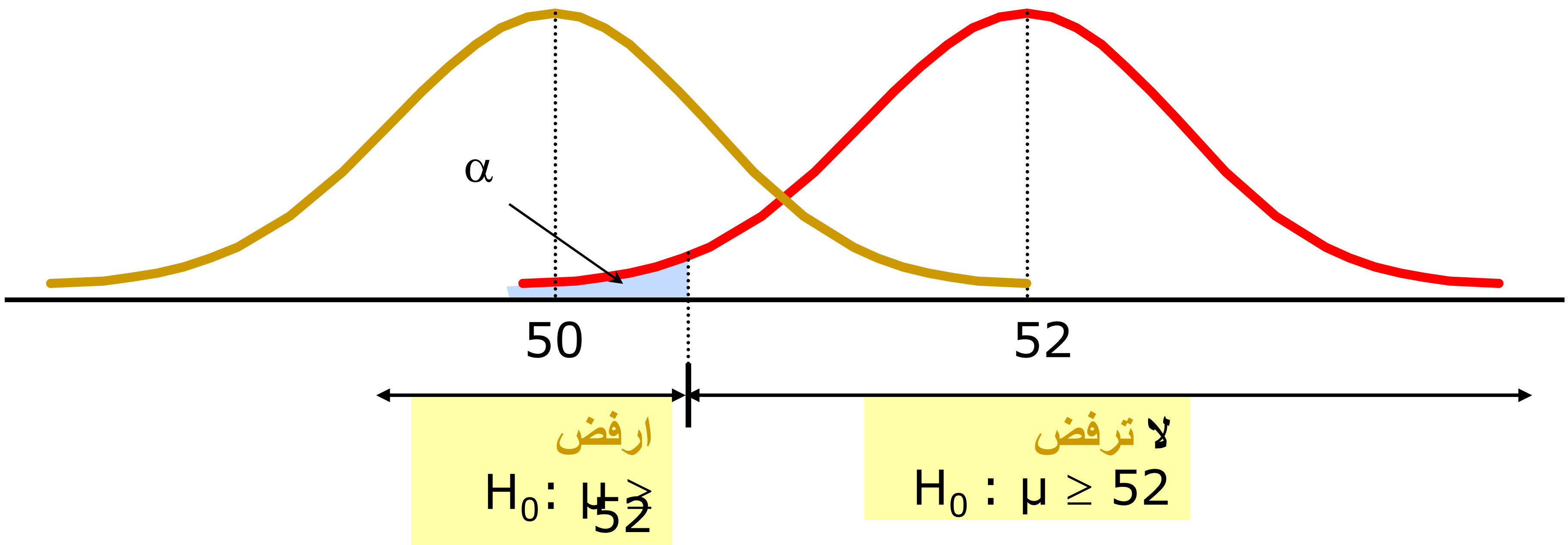
ارفض فرضية العدم لأن p-value = .0136 < α = .05

Type II Error الخطأ من النوع الثاني

- الخطأ من النوع الثاني هو احتمال الفشل في رفض H_0 خاطئة

افترض أننا فشلنا في رفض $H_0: \mu \geq 52$

عندما كان الوسط الحقيقي $\mu = 50$

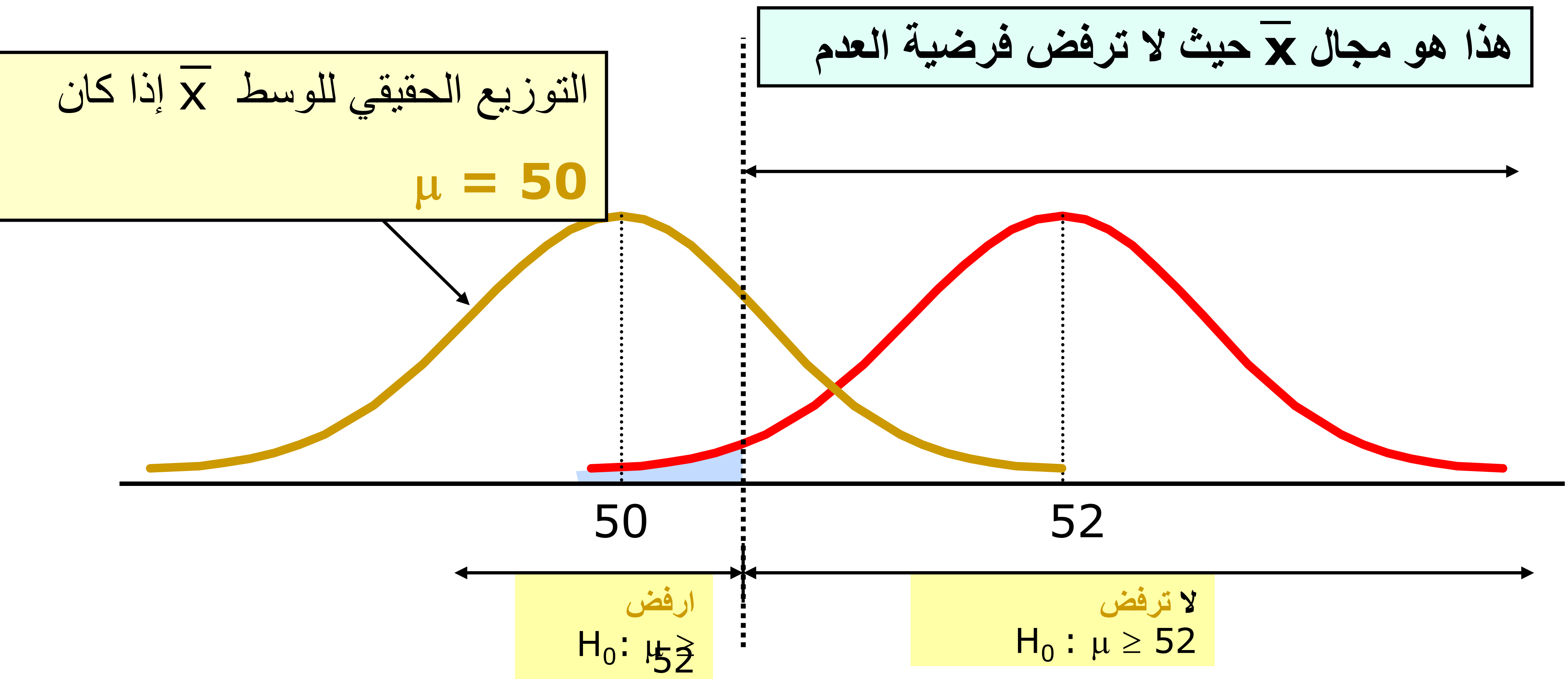


الخطأ من النوع الثاني Type II Error

تتمة

- افترض أننا لا نرفض $H_0: \mu \geq 52$

في حين أن الوسط الحقيقي $\mu = 50$



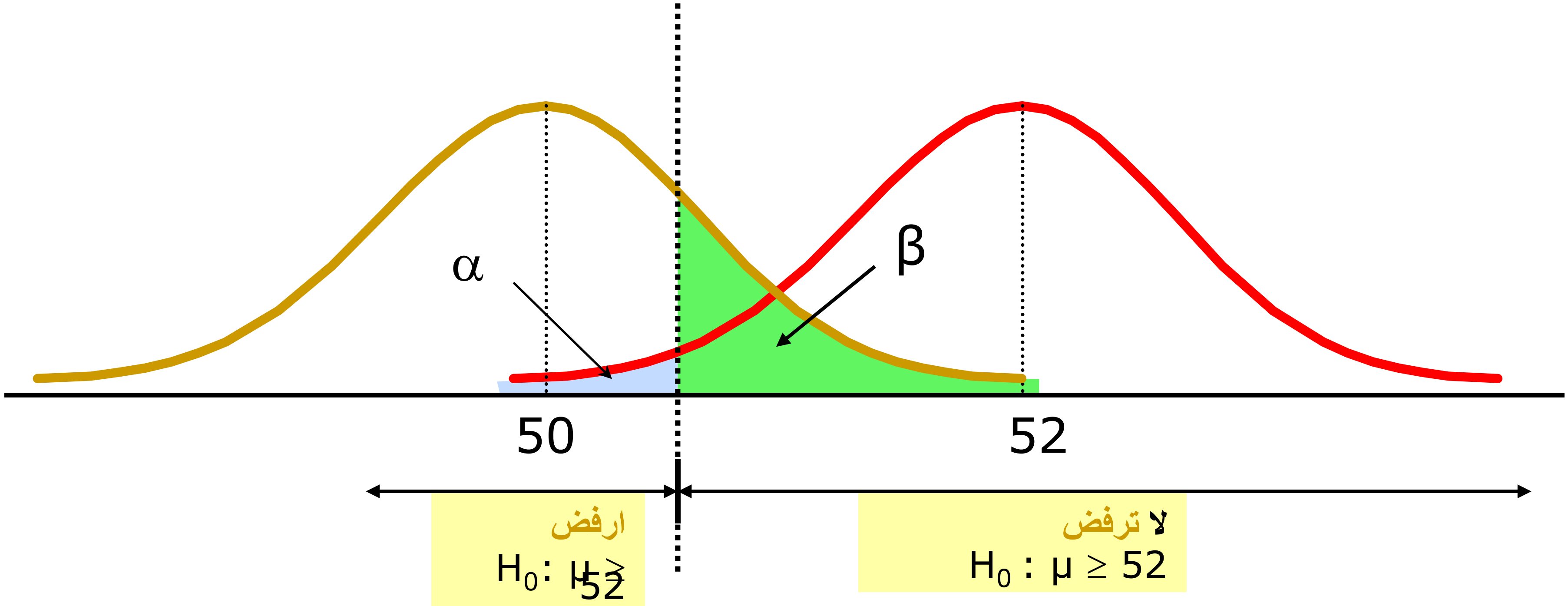
الخطأ من النوع الثاني Type II Error

تتمة

- افترض أننا لا نرفض $H_0: \mu \ge 52$

في حين أن الوسط الحقيقي $\mu = 50$

هنا: $\beta = P(x \ge \text{cutoff})$ إذا كان $\mu = 50$



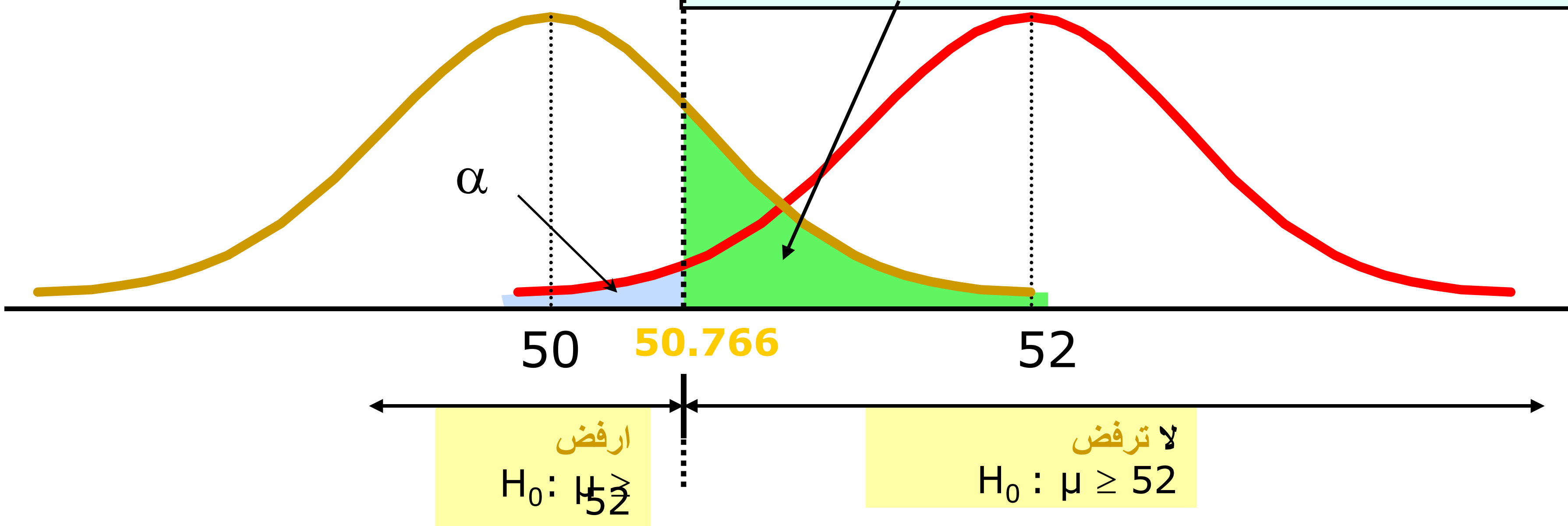
Calculating β حساب بيتا

- افترض أن: $n = 64$, $\sigma = 6$, and $\alpha = .05$

$$\text{cutoff} = \bar{x}_{\alpha} = \mu - z_{\alpha}\frac{\sigma}{\sqrt{n}} = 52 - 1.645\frac{6}{\sqrt{64}} = \boxed{50.766}$$

($H_0 : \mu \geq 52$) من أجل

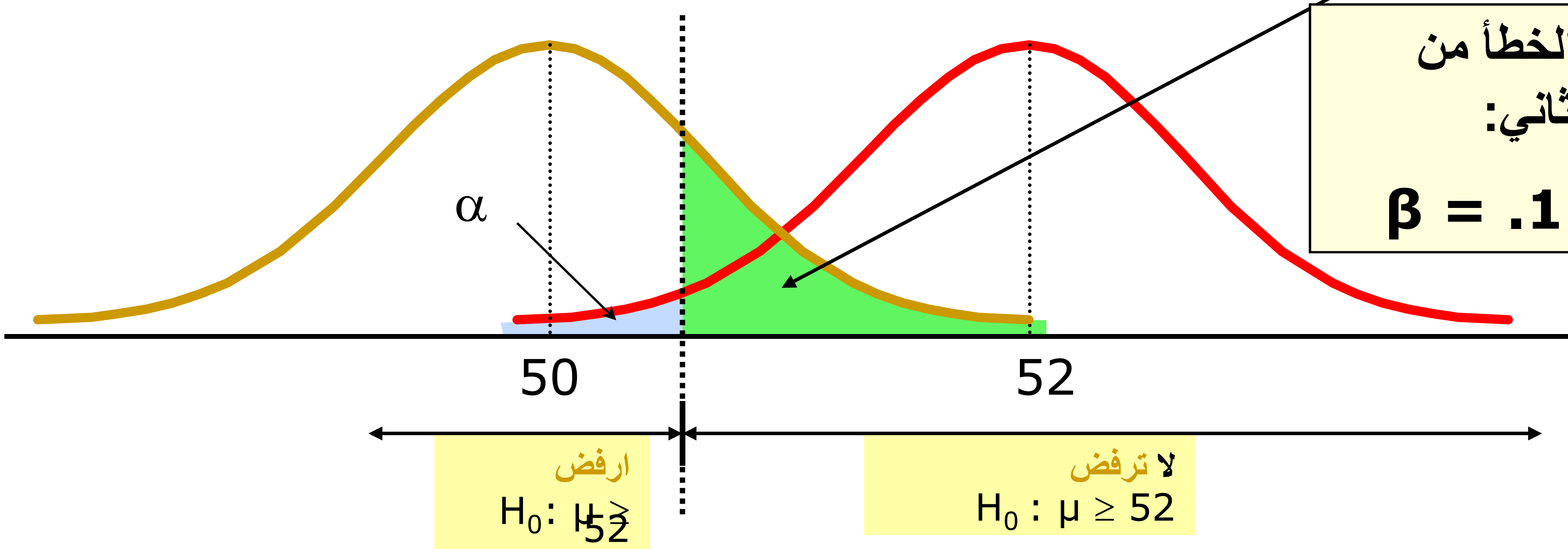
و هكذا فإن $\beta = \overline{P}(x \geq 50.766)$ إذا كان $\mu = 50$



Calculating β حساب بيتا

- افترض أن: n = 64 , σ = 6 , and α = .05

$$P(\bar{x} \geq 50.766 \mid \mu = 50) = P\left(z \geq \frac{50.766 - 50}{6/\sqrt{64}}\right) = P(z \geq 1.02) = .5 - .3461 = \boxed{.1539}$$



احتمال الخطأ من النوع الثاني:

β = .1539

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الثانية عشرة /القسم النظري/

التقدير و اختبار الفرضيات من أجل بارامتري مجتمعين

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

التقدير من أجل مجتمعين

تقدير

قيمتين مجتمعيتين

نسب مجتمع	أزواج العينات	متوسطات مجتمعات، عينات مستقلة

أمثلة:

مجتمع 1 إزاء مجتمع 2	نفس المجموعة before vs. after treatment	مجموعة إزاء مجموعة أخرى مستقلة

الفرق بين متوسطين

الهدف: إنشاء مجال ثقة للفرق بين متوسطي مجتمعين: $\mu_1 - \mu_2$

* متوسطات مجتمعات، عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان
  n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان
  n_1 or $n_2 < 30$

التقدير النقطي للفرق هو

$$\bar{x}_1 - \bar{x}_2$$

عينات مستقلة

- مصادر بيانات مختلفة
  - لا صلة بينها
  - مستقلة
- ليس لعينة مسحوبة من مجتمع ما تأثير على عينة مسحوبة من مجتمع آخر
- يتم استخدام الفرق بين متوسطي عينتين
- يستخدم z test
- أو pooled variance *t* test

متوسطات مجتمعات، عينات مستقلة *

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان
n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان
n_1 or $n_2 < 30$

σ_1 و σ_1 معلومان

متوسطات مجتمعات،
عينات مستقلة

σ_1 و σ_1 معلومان *

σ_1 و σ_1 مجهولان
n_1 and $n_2 \geq 30$

σ_1 و σ_1 معلومان
n_1 or $n_2 < 30$

افتراضات:

- استقلالية و عشوائية السحب.
- توزيعات المجتمعات طبيعية أو كلا الحجمين أكبر من 30
- الانحرافات المعيارية للمجتمع معلومة

σ_1 و σ_1 معلومان

تتمة

متوسطات مجتمعات

عينات مستقلة

- σ_1 و σ_1 معلومان *
- σ_1 و σ_1 مجهولان n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان n_1 or $n_2 < 30$

عندما يكون σ_1 و σ_1 معلومان و توزيعات المجتمعات طبيعية أو كل الأحجام أكبر من 30 فإن إحصاءة الاختبار هي z - value....

و الخطأ المعياري للمقدار: $\bar{x}_1 - \bar{x}_2$ هو

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}$$

σ_1 و σ_1 معلومان

تتمة

متوسطات مجتمعات، عينات مستقلة

- σ_1 و σ_1 معلومان *
- σ_1 و σ_1 مجهولان n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان n_1 or $n_2 < 30$

مجال الثقة من أجل $\mu_1 - \mu_2$ هو:

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}$$

σ_1 و σ_1 مجهولان، و العينات كبيرة

افتراضات:

- استقلالية و عشوائية السحب.
- توزيعات المجتمعات طبيعية أو كلا الحجمين أكبر من 30
- الانحرافات المعيارية للمجتمع مجهولة

متوسطات مجتمعات، عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان n_1 and $n_2 \geq 30$ *
- σ_1 و σ_1 معلومان n_1 or $n_2 < 30$

σ_1 و σ_1 مجهولان، و العينات كبيرة

تتمة

متوسطات مجتمعات، عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان
  n_1 and $n_2 \geq 30$ *
- σ_1 و σ_1 معلومان
  n_1 or $n_2 < 30$

إنشاء تقديرات للمجال:

استخدم s لتقدير σ

إحصاءة الاختبار هي z value

σ_1 و σ_1 مجهولان، و العينات كبيرة

تتمة

متوسطات مجتمعات، عينات مستقلة

مجال الثقة من أجل $\mu_1 - \mu_2$ هو:

σ_1 و σ_1 معلومان

σ_1 و σ_1 مجهولان
n_1 and $n_2 \geq 30$ *

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

σ_1 و σ_1 معلومان
n_1 or $n_2 < 30$

σ_1 و σ_1 مجهولان، و العينات صغيرة

- الافتراضات:
- المجتمعات تتبع التوزيع الطبيعي
- تمتلك المجتمعات نفس التباين
- العينات مستقلّة

متوسطات مجتمعات،

عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان
  n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان
  n_1 or $n_2 < 30$ *

σ_1 و σ_1 مجهولان، و العينات صغيرة

تتمة

متوسطات مجتمعات، عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان n_1 or $n_2 < 30$ *

- إنشاء تقديرات للمجال:
- يفترض أن تباينات المجتمعات متساوية. وعليه، قم باستخدام انحرافات العينات المعيارية مجتمعةً لتقدير σ مستخدماً:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}}$$

- إحصاءة الاختبار هي t value مع $(n_1 + n_2 - 2)$ درجة حرية.

σ_1 و σ_1 مجهولان، و العينات صغيرة

تتمة

متوسطات مجتمعات، عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان
  n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان
  n_1 or $n_2 < 30$ *

مجال الثقة من أجل $\mu_1 - \mu_2$ هو:

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

حيث لـ $t_{\alpha/2}$ ما يساوي $(n_1 + n_2 - 2)$ درجة حرية

و:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2}}$$

أزواج من العينات (العينات الثنائية)

اختبارات على المتوسطات لمجتمعين (بينهما صلة ما)

عينات مأخوذة على شكل أزواج

قياسات متكررة (قبل / بعد)

استخدم الفرق بين أزواج القيم:

$$\boxed{d = x_1 - x_2}$$

أزواج من العينات

- السؤال هنا هو حول جوهرية الفروق.
- الافتراضات:
  - كلا المجتمعين يتبع التوزيع الطبيعي
  - فإن لم يكن الأمر كذلك، يمكن استخدام العينات الكبيرة

أزواج الفروق

أزواج من العينات

ليكن الفرق رقم i معطى بـ:

$$d_i = x_{1i} - x_{2i}$$

فيكون التقدير النقطي للفرق بين زوجي متوسطي المجتمعين $\bar{d}$:

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

و الانحراف المعياري للعينة هو:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

n هو عدد الأزواج في العينات الثنائية (أزواج من العينات)

أزواج الفروق

تتمة

أزواج من العينات

مجال الثقة من أجل $\bar{d}$

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

Where $t_{\alpha/2}$ has $n - 1$ d.f. and s_d is:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

n هو عدد الأزواج في العينات الثنائية (أزواج من العينات)

اختبارات الفرضية من أجل الفرق بين متوسطين

- اختبار الفرضيات حول $\mu_1 - \mu_2$

- يستخدم نفس الحالات التي نوقشت قبل قليل:
  - فهل الانحرافات المعيارية معلومة أم مجهولة
  - و هل أحجام العينات ≥ 30 أم بخلاف ذلك

اختبارات الفرضية من أجل نسبتي مجتمعين

متوسطا مجتمعين، عينات مستقلة

اختبار من الطرفين :

$H_0: \mu_1 = \mu_2$
$H_A: \mu_1 \neq \mu_2$

i.e.,

$H_0: \mu_1 - \mu_2 = 0$
$H_A: \mu_1 - \mu_2 \neq 0$

اختبار الذيل الأعلى :

$H_0: \mu_1 \leq \mu_2$
$H_A: \mu_1 > \mu_2$

i.e.,

$H_0: \mu_1 - \mu_2 \leq 0$
$H_A: \mu_1 - \mu_2 > 0$

اختبار الذيل الأدنى :

$H_0: \mu_1 \geq \mu_2$
$H_A: \mu_1 < \mu_2$

i.e.,

$H_0: \mu_1 - \mu_2 \geq 0$
$H_A: \mu_1 - \mu_2 < 0$

اختبارات الفرضية من أجل $\mu_1 - \mu_2$

متوسطات مجتمع، عينات مستقلة

- σ_1 و σ_1 معلومان → إحصاءة الاختبار هي z value
- σ_1 و σ_1 مجهولان، n_1 and $n_2 \geq 30$ → استخدم s لتقدير σ المجهول، و يستخدم اختبار **z** على وجه التقريب
- σ_1 و σ_1 معلومان، n_1 or $n_2 < 30$ → استخدم s لتقدير σ المجهول، استخدم اختبار t و الانحراف المعياري العام pooled

σ_1 و σ_1 معلومان

متوسطات مجتمع،

عينات مستقلة

σ_1 و σ_1 معلومان *

σ_1 و σ_1 مجهولان
n_1 and $n_2 \geq 30$

σ_1 و σ_1 معلومان
n_1 or $n_2 < 30$

إحصاءة الاختبار من أجل $\mu_1 - \mu_2$:

$$z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

σ_1 و σ_1 مجهولان، و العينات كبيرة

متوسطات مجتمع،

عينات مستقلة

σ_1 و σ_1 معلومان

σ_1 و σ_1 مجهولان

n_1 and $n_2 \geq 30$

*

σ_1 و σ_1 معلومان

n_1 or $n_2 < 30$

إحصاءة الاختبار من أجل $\mu_1 - \mu_2$:

$$z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

σ_1 و σ_1 مجهولان، و العينات صغيرة

متوسطات مجتمع،

عينات مستقلة

- σ_1 و σ_1 معلومان
- σ_1 و σ_1 مجهولان
  n_1 and $n_2 \geq 30$
- σ_1 و σ_1 معلومان *
  n_1 or $n_2 < 30$

إحصاءة الاختبار من أجل $\mu_1 - \mu_2$:

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

حيث تمتلك $t_{\alpha/2}$ $(n_1 + n_2 - 2)$ درجة حرية

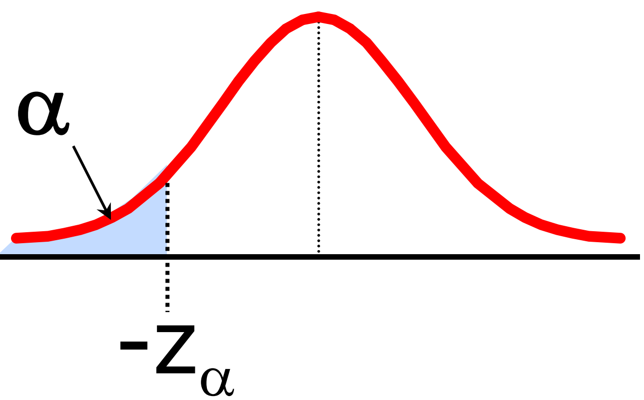
و

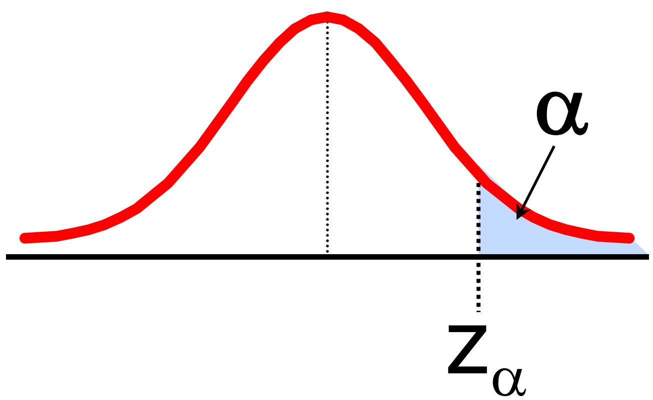
$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

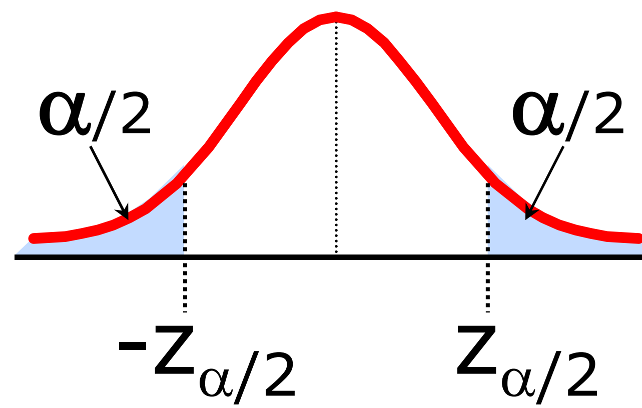
اختبارات الفرضية من أجل $\mu_1 - \mu_2$

متوسطات مجتمع، عينات مستقلة

اختبار الذيل الأدنى :	اختبار الذيل الأعلى:	اختبار من الطرفين:
$H_0: \mu_1 - \mu_2 \geq 0$ $H_A: \mu_1 - \mu_2 < 0$	$H_0: \mu_1 - \mu_2 \leq 0$ $H_A: \mu_1 - \mu_2 > 0$	$H_0: \mu_1 - \mu_2 = 0$ $H_A: \mu_1 - \mu_2 \neq 0$







ارفض H_0 إذا $z < -z_\alpha$

ارفض H_0 إذا $z > z_\alpha$

ارفض H_0 إذا $z < -z_{\alpha/2}$
or $z > z_{\alpha/2}$

مثال على اختبار t من أجل s_p مُجمّع (pooled)

أنت محلل إحصائي لشركة سمسرة. هل هناك إي فرق في عائد الأرباح بين الأسهم المدرجة على NYSE & NASDAQ؟

افترض أنك قمت بجمع البيانات التالية:

	NYSE	**NASDAQ**
الرقم	**21**	**25**
متوسط العينة	**3.27**	**2.53**
الانحراف المعياري للعينة	**1.30**	**1.16**

بفرض أن التباينات متساوية، هل هناك فرق في العائد المتوسطي؟

(α = 0.05)

حساب إحصاءة الاختبار

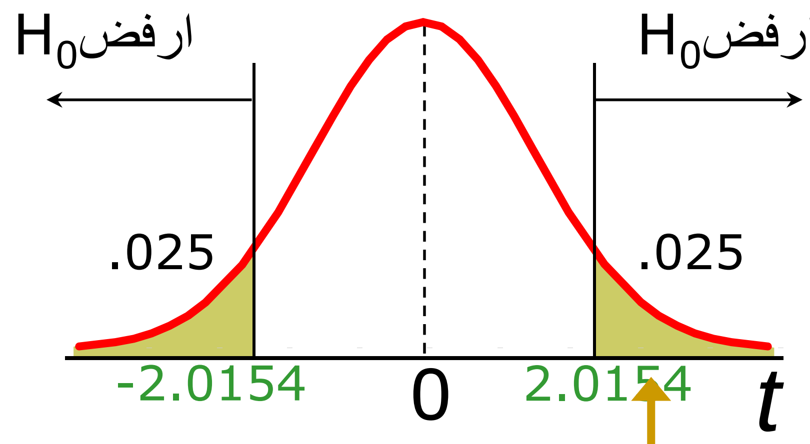
إحصاءة الاختبار هي:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(3.27 - 2.53) - 0}{1.2256\sqrt{\frac{1}{21} + \frac{1}{25}}} = \boxed{2.040}$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{n_1 + n_2 - 2}} = \sqrt{\frac{(21 - 1)1.30^2 + (25 - 1)1.16^2}{21 + 25 - 2}} = 1.2256$$

الحل

$H_0: \mu_1 - \mu_2 = 0$ i.e. $(\mu_1 = \mu_2)$
$H_A: \mu_1 - \mu_2 \neq 0$ i.e. $(\mu_1 \neq \mu_2)$



مستوى الدلالة $\alpha = 0.05$

درجات الحرية $df = 21 + 25 - 2 = 44$

القيمة الحرجة: $t = \pm 2.0154$

2.040

احصاءة الاختبار:

$$t = \frac{3.27 - 2.53}{1.2256\sqrt{\frac{1}{21} + \frac{1}{25}}} = \boxed{2.040}$$

القرار:

ارفض H_0 عند $\alpha = 0.05$

الاستنتاج:

هناك دليل على وجود فرق بين المتوسطات

اختبار الفرضية من أجل أزواج من العينات

أزواج من العينات

إحصاءة الاختبار من أجل $\bar{d}$ is

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

حيث $t_{\alpha/2}$ تمتلك $n - 1$ د.ح. و s_d معطى بـ:

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$$

n هو عدد الأزواج في العينات الثنائية (أزواج من العينات)

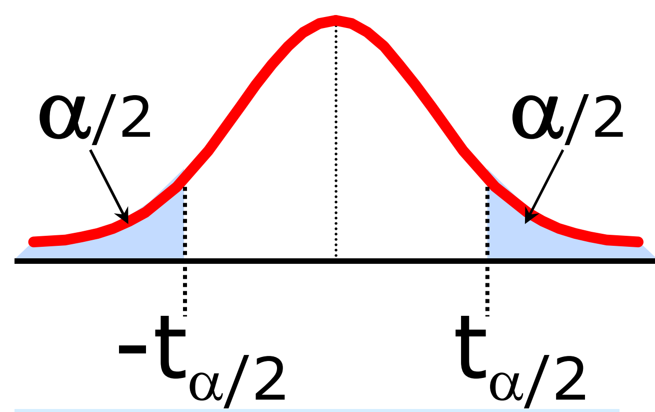
اختبار الفرضية من أجل أزواج من العينات

تتمة

أزواج من العينات

اختبار من الطرفين :

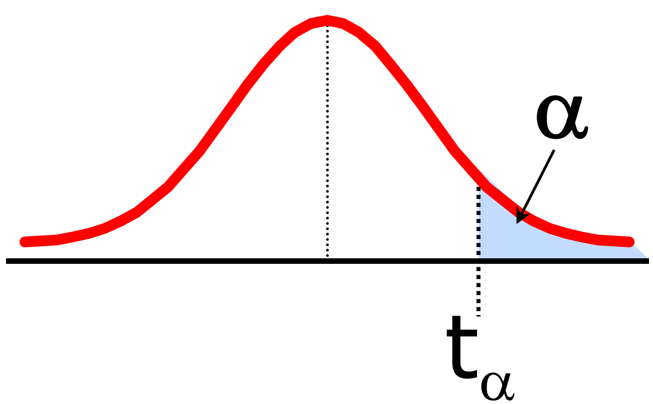
$H_0: \mu_d = 0$
$H_A: \mu_d \neq 0$



ارفض H_0 إذا $t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$

اختبار الذيل الأعلى :

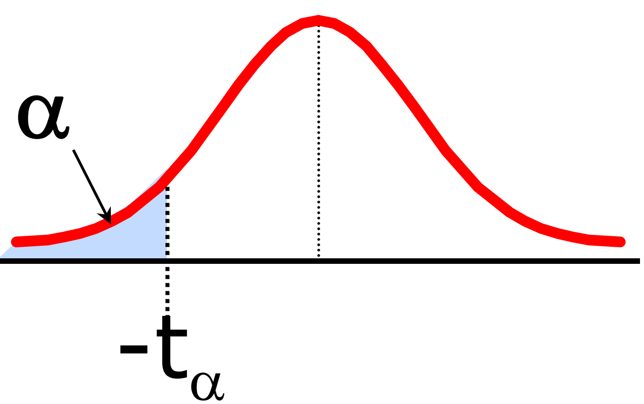
$H_0: \mu_d \leq 0$
$H_A: \mu_d > 0$



ارفض H_0 إذا $t > t_\alpha$

اختبار الذيل الأدنى :

$H_0: \mu_d \geq 0$
$H_A: \mu_d < 0$



ارفض H_0 إذا $t < -t_\alpha$

حيث $t_{\alpha/2}$ تمتلك $n - 1$ د.ح.

أزواج من العينات : مثال

- افترض أنك قمت بإرسال موظي مبيعات إلى دورة تدريبية حول "خدمة الزبون" هل التدريب فعال؟ افترض أنك جمعت البيانات التالية:

موظف البيع	عد الشكاوى قبل (1)	 بعد (2)	الفرق d_i
1#	6	4	- 2
2#	20	6	-14
3#	3	2	- 1
4#	0	0	0
5#	4	0	- 4
			-21

$$\bar{d} = \frac{\sum d_i}{n}$$

$$= -4.2$$

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

$$= 5.67$$

أزواج من العينات: الحل

هل كان للتدريب أثر جوهري على عدد الشكاوى (عند المستوى 0.01 من الدلالة)؟

H_0: $\mu_d = 0$
H_A: $\mu_d \neq 0$

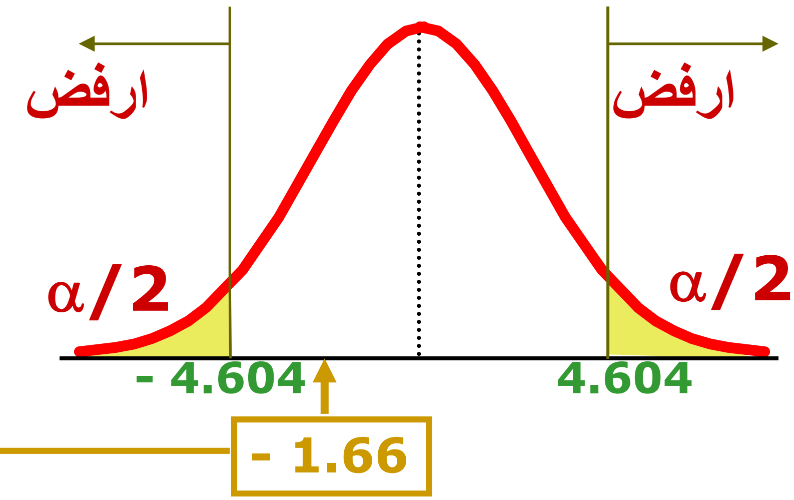
$\alpha = .01$ $\bar{d} = -4.2$

القيمة الحرجة: ± 4.604

درجات الحرية: $n - 1 = 4$

إحصاء الاختبار:

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{-4.2 - 0}{5.67/\sqrt{5}} = \boxed{-1.66}$$



القرار: لا ترفض H_0

الاستنتاج: ليس هناك تغيير جوهري في عدد الشكاوى

نسبتا مجتمعين

الهدف: إنشاء مجال ثقة أو اختبار فرضية حول الفرق بين نسبتي مجتمعين $\boxed{p_1 - p_2}$

نسب مجتمع

الافتراضات:

$$n_1 p_1 \geq 5 \quad , \quad n_1(1-p_1) \geq 5$$

$$n_2 p_2 \geq 5 \quad , \quad n_2(1-p_2) \geq 5$$

التقدير النقطي للفرق هو:

$$\bar{p}_1 - \bar{p}_2$$

مجالات الثقة (حدود الثقة) من أجل نسبتي مجتمعين

نسب مجتمع

مجال الثقة من أجل

$p_1 - p_2$ is:

$$\left(\bar{p}_1 - \bar{p}_2\right) \pm z_{\alpha/2} \sqrt{\frac{\bar{p}_1(1-\bar{p}_1)}{n_1} + \frac{\bar{p}_2(1-\bar{p}_2)}{n_2}}$$

اختبارات الفرضية من أجل نسبتي مجتمعين

نسب مجتمع

اختبار من الطرفين :

$H_0: p_1 = p_2$
$H_A: p_1 \neq p_2$

i.e.,

$H_0: p_1 - p_2 = 0$
$H_A: p_1 - p_2 \neq 0$

اختبار الذيل الأعلى :

$H_0: p_1 \leq p_2$
$H_A: p_1 > p_2$

i.e.,

$H_0: p_1 - p_2 \leq 0$
$H_A: p_1 - p_2 > 0$

اختبار الذيل الأدنى :

$H_0: p_1 \geq p_2$
$H_A: p_1 < p_2$

i.e.,

$H_0: p_1 - p_2 \geq 0$
$H_A: p_1 - p_2 < 0$

نسبتا مجتمعين

بما أننا ننطلق من افتراض أن نظرية العدم صحيحة، فإننا نفترض أن $p_1 = p_2$
و نقوم بتجميع pool تقديرات $\bar{p}$

نسب مجتمع

التقدير المجمع للنسبة الإجمالية:

$$\bar{p} = \frac{n_1\bar{p}_1 + n_2\bar{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

حيث x_1 و x_2 هما الأرقام من العينتين 1 و 2 اللواتي تحملن الصفة محل الاهتمام

نسبتا مجتمعين

تتمة

نسب مجتمع

إحصاءة الاختبار من أجل $p_1 - p_2$:

$$z = \frac{\left(\overline{p}_1 - \overline{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\overline{p}\left(1 - \overline{p}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

اختبارات الفرضية من أجل نسبتي مجتمعين

نسب مجتمع

اختبار من الطرفين :

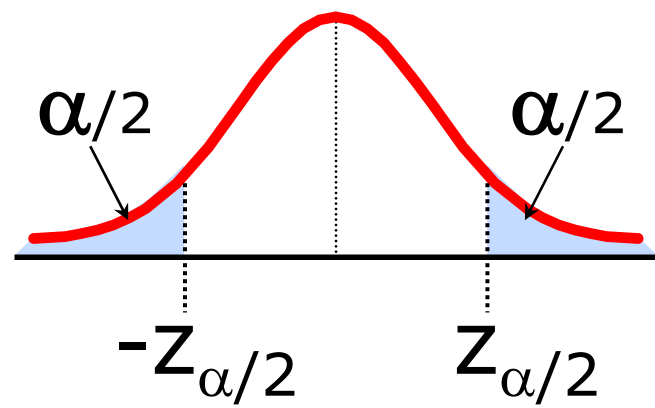
$H_0: p_1 - p_2 = 0$
$H_A: p_1 - p_2 \neq 0$

اختبار الذيل الأعلى :

$H_0: p_1 - p_2 \leq 0$
$H_A: p_1 - p_2 > 0$

اختبار الذيل الأدنى :

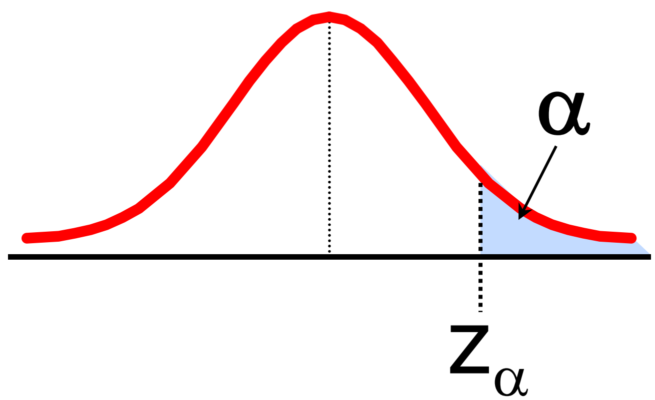
$H_0: p_1 - p_2 \geq 0$
$H_A: p_1 - p_2 < 0$



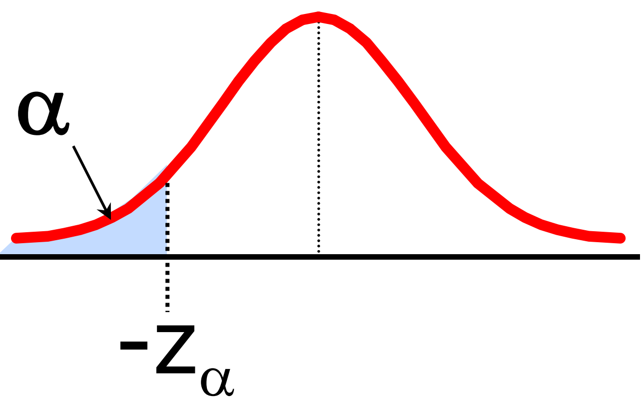
ارفض H_0 إذا $z < -z_{\alpha/2}$

أو

$z > z_{\alpha/2}$



ارفض H_0 إذا $z > z_\alpha$



ارفض H_0 إذا $z < -z_\alpha$

مثال : نسبتا مجتمعين

هل هناك فرق ذو دلالة بين نسبة الرجال و نسبة النساء الذين سيصوتون بـنعم على الاقتراح A

- في عينة عشوائية، صرّح 36 من 72 رجلاً و 31 من 50 امرأة، بأنهم ينوون التصويت بنعم.
- اختبر عند مستوى دلالة قدره .05

مثال: نسبتا مجتمعين

تتمة

▫ اختبار الفرضية هو:

$H_0: p_1 - p_2 = 0$ النسبتان متساويتان

$H_A: p_1 - p_2 \neq 0$ هناك فرق جوهري بين النسب

- نسب العينات هي:
  - رجال: $\bar{p}_1 = 36/72 = .50$
  - نساء: $\bar{p}_2 = 31/50 = .62$
- التقدير المجمّع pooled estimate للنسبة الإجمالية:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{36 + 31}{72 + 50} = \frac{67}{122} = .549$$

مثال : نسبتا مجتمعين

تتمة

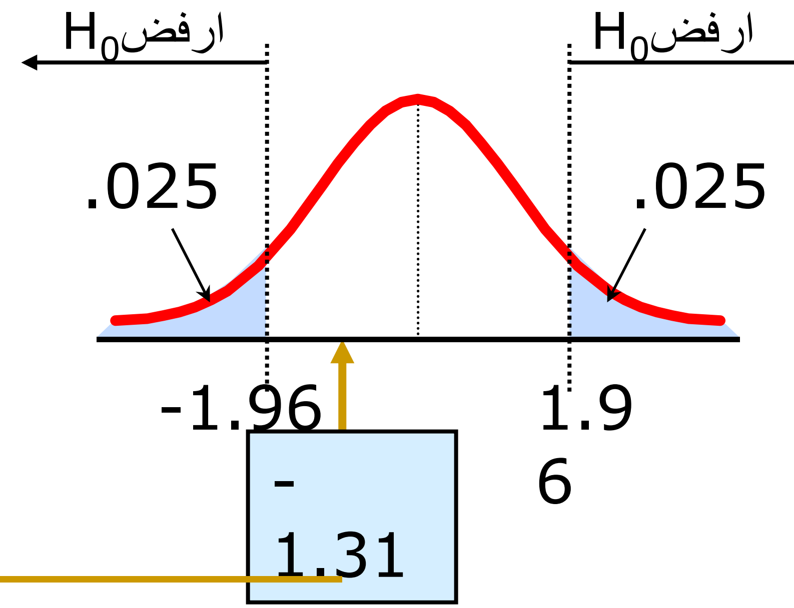
$p_1 - p_2$ is: إحصاءة الاختبار من أجل

$$= \frac{\left(\bar{p}_1 - \bar{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\bar{p}\,(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{\left(.50 - .62\right) - \left(0\right)}{\sqrt{.549\,(1 - .549)\left(\frac{1}{72} + \frac{1}{50}\right)}} = \boxed{-1.31}$$

القيمة الحرجة تساوي: **±1.96**

من أجل: **α = .05**



القرار: لا ترفض H_0

الاستنتاج:

ليس هناك دليل كاف على وجود فرق جوهري بين النسب (التي ستصوت بنعم و تللك التي ستصوت بلا)

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الثالثة عشرة /القسم النظري/

اختبارات الفرضية الخاصة بتباين مجتمع واحد
و تبايني مجتمعين

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

اختبرات الفرضية من أجل التباين

اختبرات الفرضية
من أجل التباينات

اختبارات من أجل
تبايني مجتمعين

إحصاءة اختبار F

اختبارات من أجل
تباين مجتمع واحد

إحصاءة اختبار كاي

مجتمع واحد

اختبارات الفرضية من أجل التباينات

اختبارات من أجل
تباين مجتمع واحد *

إحصاءة الاختبار كاي مربع

اختبار من الطرفين

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_A: \sigma^2 \neq \sigma_0^2$$

اختبار الذيل الأدنى

$$H_0: \sigma^2 \geq \sigma_0^2$$
$$H_A: \sigma^2 < \sigma_0^2$$

اختبار الذيل االأعلى

$$H_0: \sigma^2 \leq \sigma_0^2$$
$$H_A: \sigma^2 > \sigma_0^2$$

إحصاءة اختبار كاي مربع

اختبارات الفرضية من أجل التباينات

اختبارات من أجل
تباين مجتمع واحد

إحصاءة الاختبار كاي مربع

إحصاءة اختبار كاي مربع من أجل تباين مجتمع وحيد هي:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

حيث:

χ^2 = متغير كاي مربع المعياري

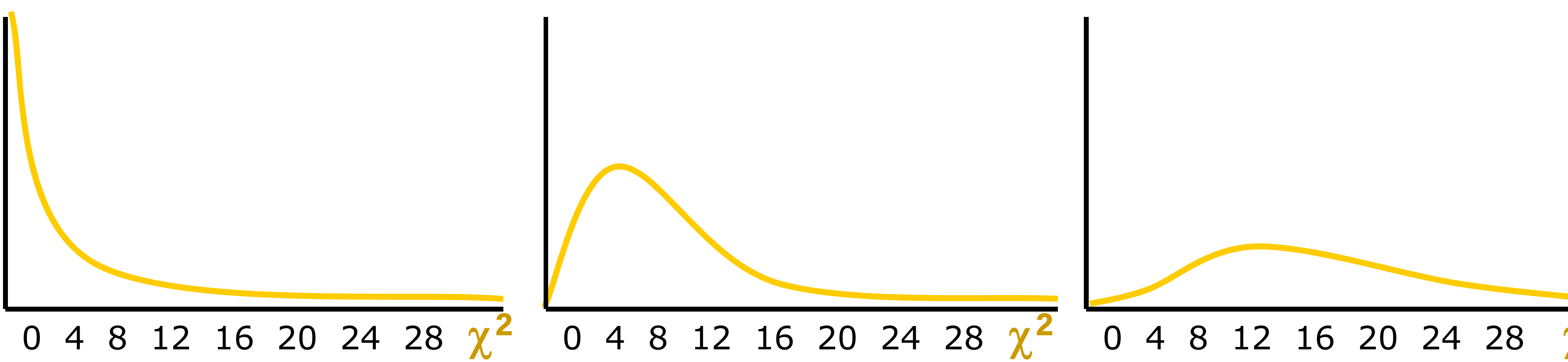
n = حجم العينة

s^2 = تباين العينة

σ^2 = التباين بحسب الفرضية

توزيع كاي مربع

- توزيع كاي مربع هو عائلة من التوزيعات حيث يتوقف شكل التوزيع على عدد درجات الحرية (د.ح.):
- د.ح. = d.f. = n − 1



d.f. = 1

d.f. = 5

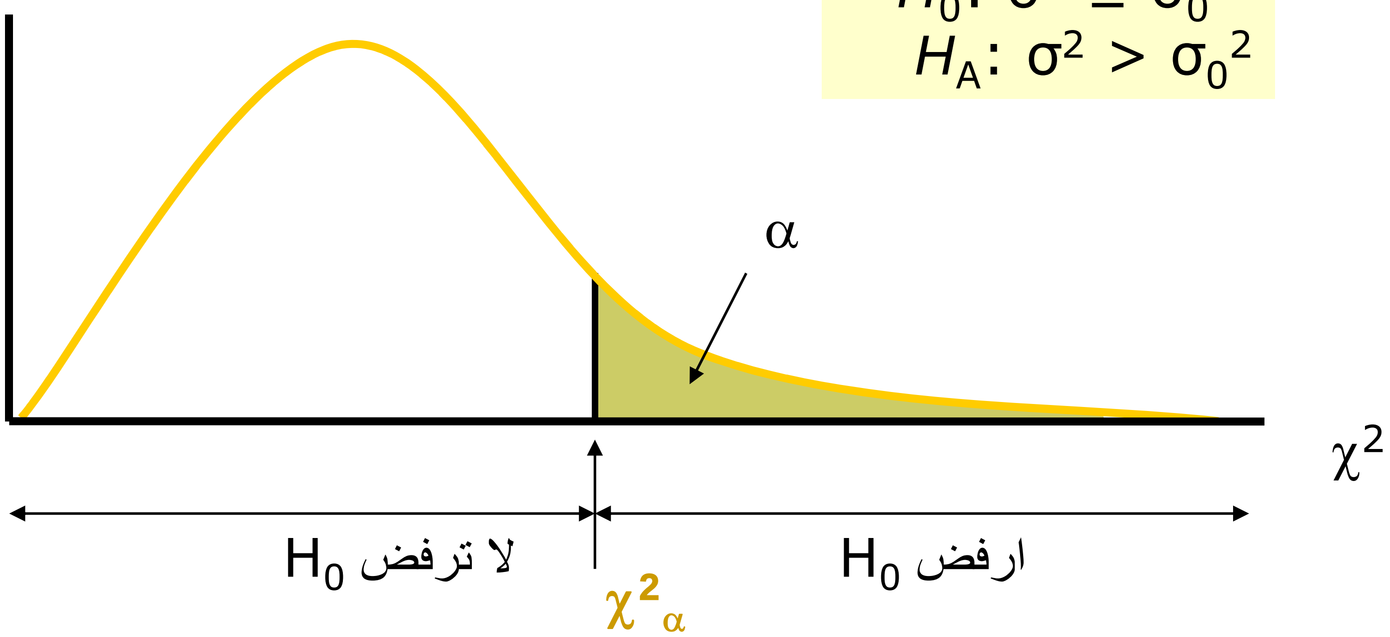
d.f. = 15

إيجاد القيمة الحرجة

- يمكن إيجاد القيمة الحرجة χ^2_α من الجدول:

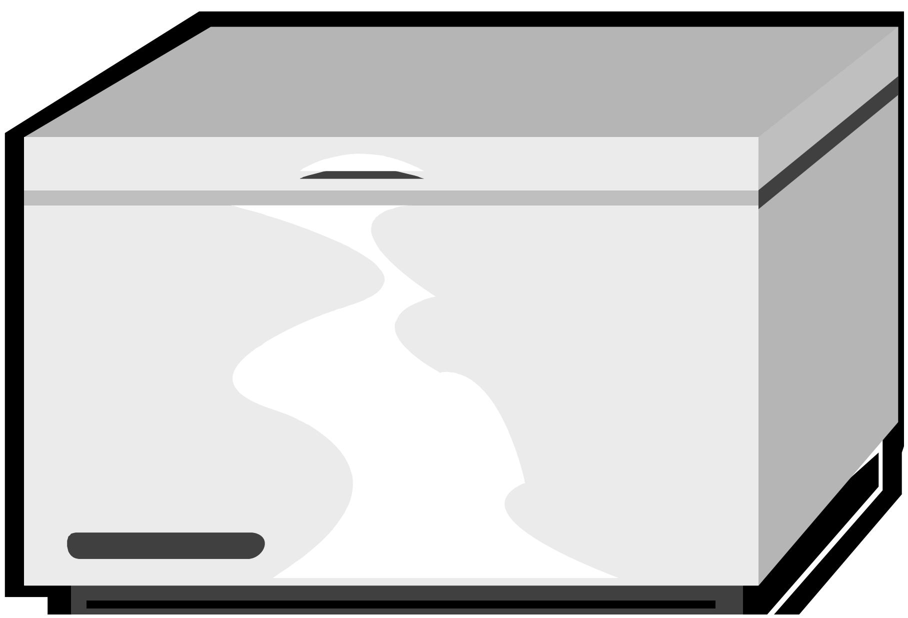
اختبار الذيل االأعلى:

H_0: $\sigma^2 \le \sigma_0^2$
H_A: $\sigma^2 > \sigma_0^2$



مثال

- يجب على الثلاجات التجارية أن تحافظ على مستوى محدد من الحرارة مع القليل من التباين. تتطلب المواصفات انحرافاً معيارياً لا يزيد عن أربع درجات (أو تباين مقداره 16 درجة).



- تم اختبار عينة مكونة من 16 ثلاجة حيث تم الحصول على $s^2 = 24$.
- اختبر ما إذا تم تجاوز مستوى الانحراف المعياري المحدد بالمواصفات مستخدماً $\alpha = .05$.

إيجاد القيمة الحرجة

- من جدول كاي مربع:

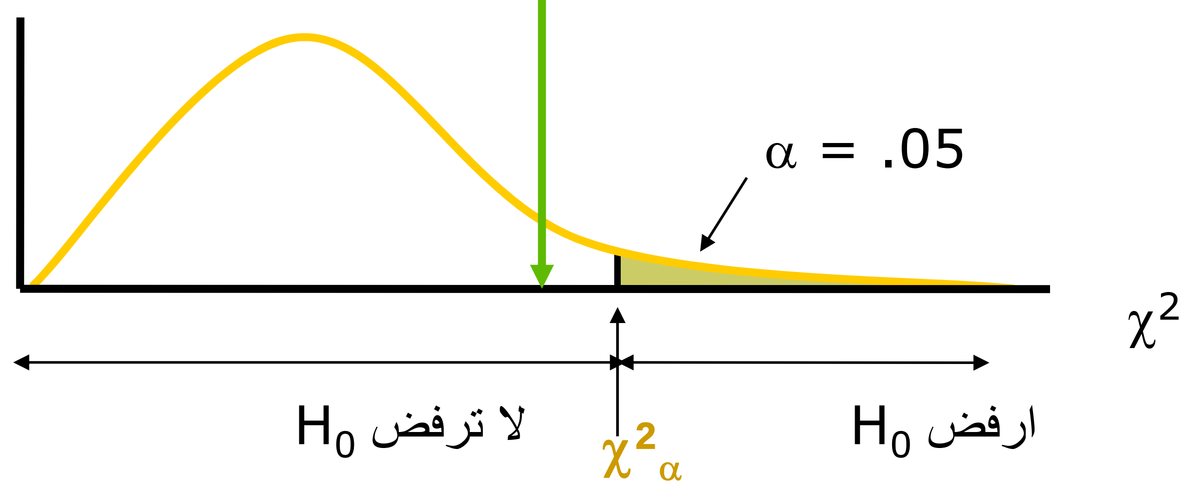
$$\chi^2_\alpha = 24.9958 \quad (\alpha = .05 \; ; \; 16 - 1 = 15 \; d.f.)$$

و إحصاءة الاختبار هي:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(16-1)24}{16} = \boxed{22.5}$$

بما أن $22.5 < 24.9958$

لا ترفض H_0



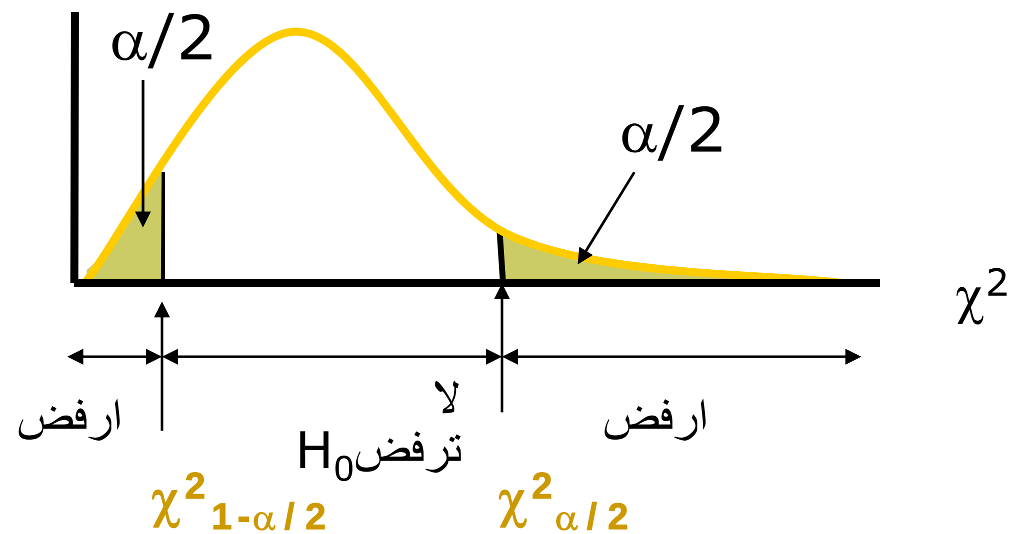
= 24.9958

لا يوجد دليل جوهري عند مستوى الدلالة $\alpha = .05$ على أن الانحراف المعياري يفوق ما حددته المواصفات

اختبارات كاي ذات الطرف الأيسر و ذات الطرفين

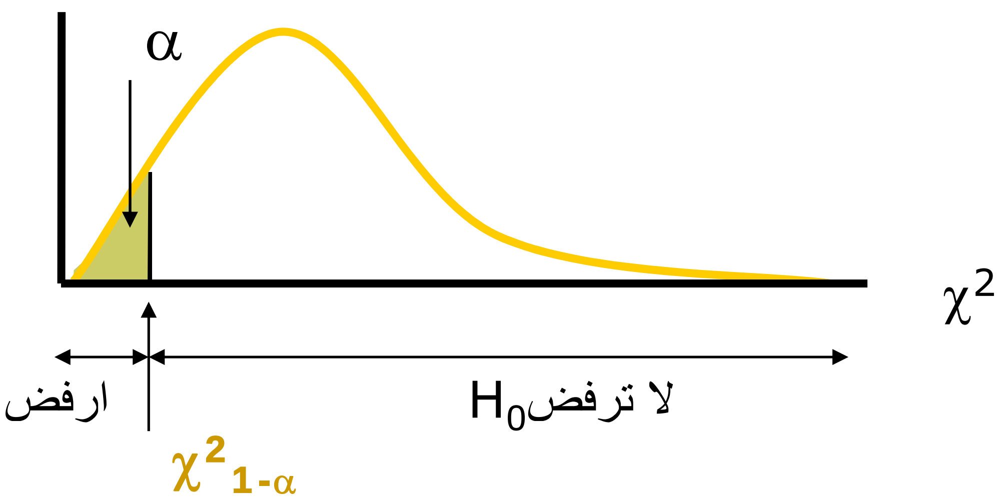
اختبار من الطرفين:

$$H_0: \sigma^2 = \sigma_0^2$$
$$H_A: \sigma^2 \neq \sigma_0^2$$



اختبار الذيل الأدنى:

$$H_0: \sigma^2 \geq \sigma_0^2$$
$$H_A: \sigma^2 < \sigma_0^2$$



اختبار F من أجل الفرق بين تبايني مجتمعين

اختبارات الفرضية من أجل التباينات

اختبارات من
تبايني مجتمعين

إحصاءة الاختبار F

اختبار من الطرفين

$$H_0: \sigma_1^2 - \sigma_2^2 = 0$$
$$H_A: \sigma_1^2 - \sigma_2^2 \neq 0$$

اختبار الذيل الأدنى

$$H_0: \sigma_1^2 - \sigma_2^2 \geq 0$$
$$H_A: \sigma_1^2 - \sigma_2^2 < 0$$

اختبار الذيل الأعلى

$$H_0: \sigma_1^2 - \sigma_2^2 \leq 0$$
$$H_A: \sigma_1^2 - \sigma_2^2 > 0$$

اختبار F من أجل الفرق بين تبايني مجتمعين

اختبارات الفرضية من أجل التباينات

اختبارات من
تبايني مجتمعين

إحصاءة الاختبار F

إحصاءة الاختبارF هي :

$$F = \frac{s_1^2}{s_2^2}$$

التباين
الأكبر
على
البسط

s_1^2 = تباين العينة الأولى

$n_1 - 1$ = درجات حرية البسط

s_2^2 = تباين العينة الثانية

$n_2 - 1$ = درجات حرية المقام

توزيع F

- يمكن الحصول على قيم F الحرجة من جدول F
- هناك عددان من درجات الحرية: البسط و المقام

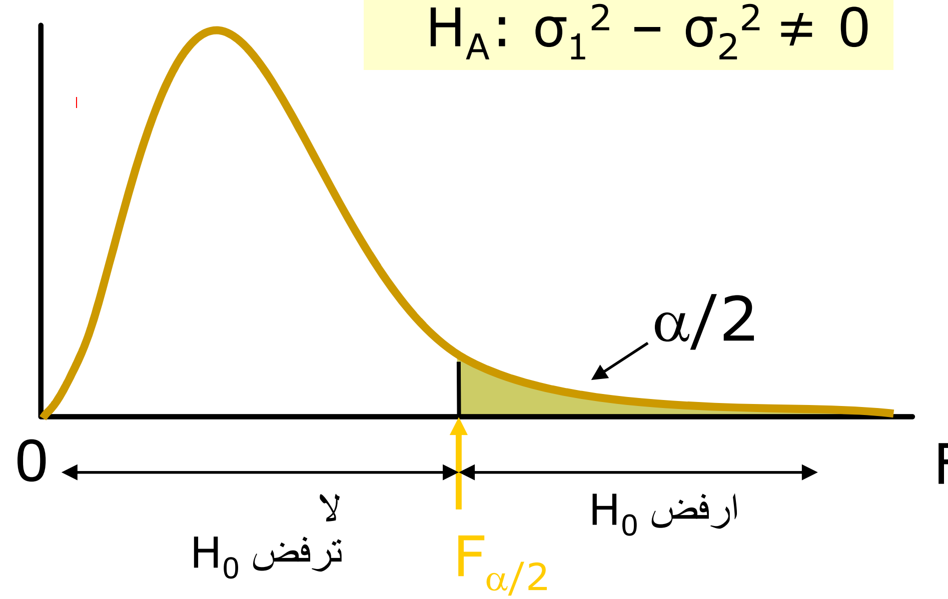
$$F = \frac{s_1^2}{s_2^2}$$

حيث $df_1 = n_1 - 1 \; ; \quad df_2 = n_2 - 1$

- في جدول F
  - درجات حرية البسط تحدد السطر
  - درجات حرية المقام تحدد المقام

إيجاد القيمة الحرجة

$$H_0: \sigma_1^2 - \sigma_2^2 = 0$$
$$H_A: \sigma_1^2 - \sigma_2^2 \neq 0$$

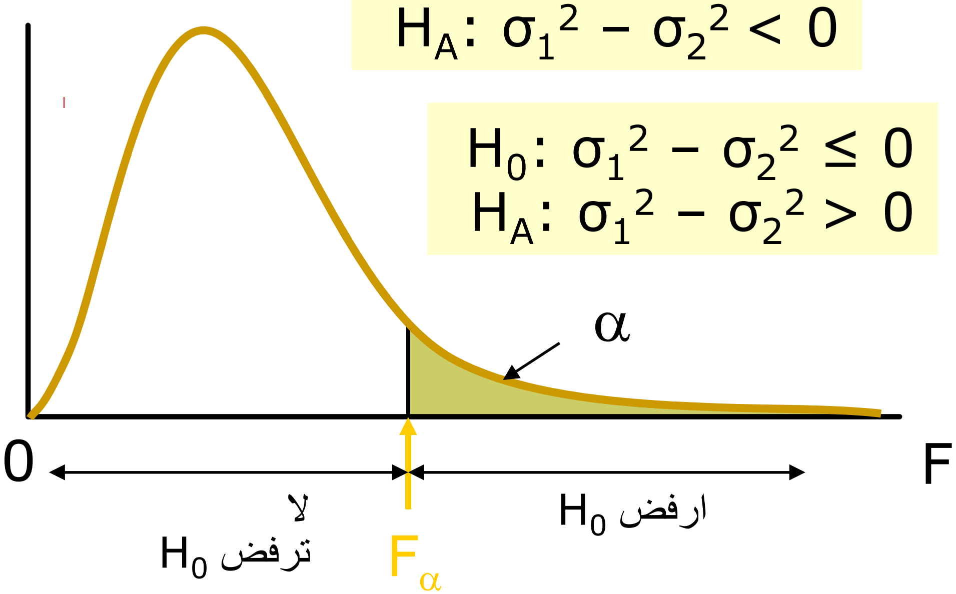


- منطقة الرفض لاختبار من الطرفين

$$F = \frac{s_1^2}{s_2^2} > F_{\alpha/2}$$

$$H_0: \sigma_1^2 - \sigma_2^2 \geq 0$$
$$H_A: \sigma_1^2 - \sigma_2^2 < 0$$

$$H_0: \sigma_1^2 - \sigma_2^2 \leq 0$$
$$H_A: \sigma_1^2 - \sigma_2^2 > 0$$



- منطقة الرفض لاختبار من طرف واحد

$$F = \frac{s_1^2}{s_2^2} > F_\alpha$$

(حيث التباين الأكبر على البسط)

توزيع F: مثال

- أنت محلل مالي في شركة سمسرة. و ترغب بمقارنة عائد الأرباح الموزعة بين سهمين مدرجين على NYSE و NASDAQ حيث قمت بجمع البيانات التالية:

	NYSE	NASDAQ
العدد	21	25
المتوسط	3.27	2.53
الانحراف المعياري	1.30	1.16

هل هناك فرق في التباين بين NYSE و NASDAQ

عند مستوى $\alpha = 0.05$ من الدلالة؟

توزيع F: حل المثال

- صغ اختبار الفرضية:

$$H_0: \sigma^2_1 - \sigma^2_2 = 0$$ لا فرق بين التباينات

$$H_A: \sigma^2_1 - \sigma^2_2 \neq 0$$ هناك فرق بين التباينات

- أوجد قيمة F الحرجة من أجل $\alpha = .05$
  - البسط:
    - $df_1 = n_1 - 1 = 21 - 1 = 20$
  - المقام:
    - $df_2 = n_2 - 1 = 25 - 1 = 24$

$$\boxed{F_{.05/2,\ 20,\ 24} = 2.327}$$

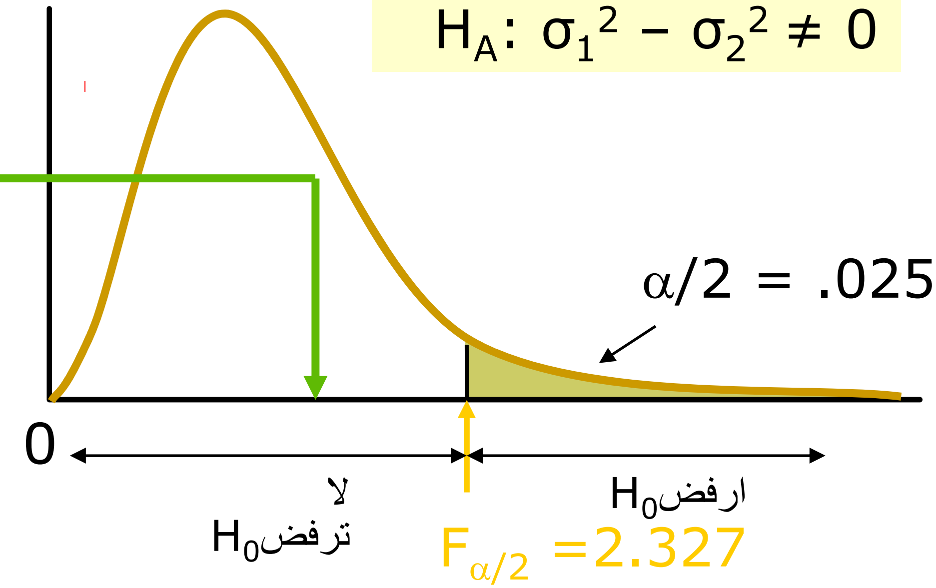
توزيع F: حل المثال

تتمة

- إحصاءة الاختبار هي:

$H_0: \sigma_1^2 - \sigma_2^2 = 0$
$H_A: \sigma_1^2 - \sigma_2^2 \neq 0$

$$F = \frac{s_1^2}{s_2^2} = \frac{1.30^2}{1.16^2} = \boxed{1.256}$$



- $F = 1.256 > 2.327$ إذاً لا ترفض H_0

- الاستنتاج: لا يوجد دليل على وجود فرق بين التباينات عند $\alpha = .05$

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الرابعة عشرة /القسم النظري/

تحليل التباين 2 / 1

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

محتويات المحاضرة

- تحليل التباين (ANOVA)
  - تحليل التباين: اتجاه واحد (وحيد الاتجاه)
    - اختبار F
    - اختبار TK
  - ANOVA بلوك عشوائي تام
    - اختبار F
    - اختبار فيشر للفرق الأقل دلالة
  - ANOVA عاملين مع التكرار

التوصيف العام لتحليل التباين

- يتحكم الباحث بمتغير أو عدة متغيرات مستقلة
  - تسمى "عوامل" أو "متغيرات المعاملة"
  - كل عامل يحتوي اثنين أو أكثر من المستويات (الفئات / التصنيفات)
- تتم ملاحظة الآثار على المتغير التابع
  - الاستجابة لمختلف مستويات المتغير المستقل
- التصميم التجريبي: الخطة المستخدمة لاختبار الفرضية

تحليل التباين وحيد الاتجاه

- يقوم بتقييم الفرق بين متوسطات ثلاث فأكثر من المجتمعات

مثال :معدلات الحوادث في الورديات الأولى و الثانية و الثالثة ...
العمر المتوقع لخمسة أنواع من إطارات السيارات

- الافتراضات
  - المجتمعات تتبع التوزيع الطبيعي
  - تساوي تباينات المجتمعات
  - العينات مسحوبة عشوائياً و بشكل مستقل

التصميم العشوائي تماماً

- الوحدات التجريبية (العناصر الخاضعة للتجرية) مخصصة للمعاملات عشوائياً
- عامل واحد أو متغير مستقل واحد
  - مع اثنين أو أكثر من المستويات
- يتم التحليل عبر:
  - تحليل تباين ذو عامل واحد
- يسمى تصميماً متوازناً إذا كان لجميع مستويات العامل نفس حجم العينة.

فرضيات تحليل التباين وحيد الاتجاه

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

- كل متوسطات المجتمعات متساوية
- أي ليس هناك "أثر معاملة" أو لا تغاير في المتوسطات عبر المجموعات

H_A: ليس هناك تساوي بين كافة المتوسطات المجتمعية

- متوسط أحد المجتمعات على الأقل مختلف
- أي أن هناك أثر معاملة
- و لا يعني بالضرورة اختلاف كافة المتوسطات عن بعضها البعض

تحليل التباين وحيد الاتجاه

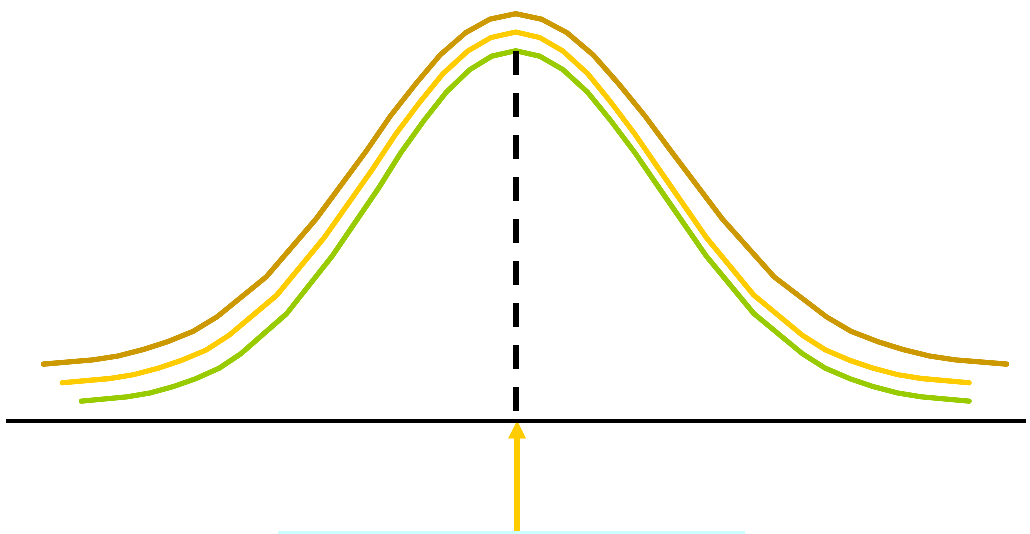
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

$$H_A : \text{Not all } \mu_i \text{ are the same}$$

المتوسطات متساوية:

فرضية العدم صحيحة

(لا يوجد “أثر معاملة”)



$$\mu_1 = \mu_2 = \mu_3$$

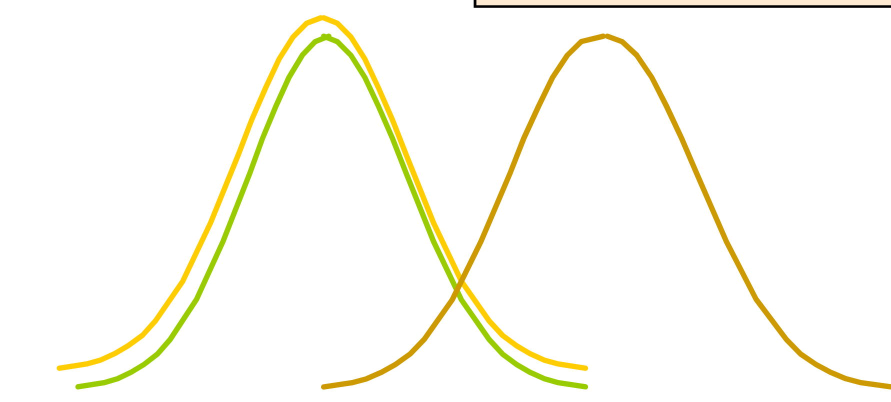
تحليل التباين وحيد الاتجاه

تتمة

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

$$H_A : \text{Not all } \mu_i \text{ are the same}$$

متوسط واحد على الأقل مختلف:
فرضية العدم ليست صحيحة
(هناك أثر معاملة)



$$\mu_1 = \mu_2 \neq \mu_3$$

أو

$$\mu_1 \neq \mu_2 \neq \mu_3$$

تجزيء التباين

- يمكن تجزيء التباين إلى جزأين:

$$\boxed{SST = SSB + SSW}$$

SST = المجموع الكلي للمربعات
SSB = مجموع المربعات البيني
SSW = مجموع المربعات الداخلي

تجزيء التباين

تتمة

$$SST = SSB + SSW$$

التباين الكلي = التشتت الكلي لقيم المفردات عبر كافة المعاملات (SST)

التباين بين العينات = تشتت متوسطات العينات (SSB)

التباين داخل العينات = تشتت قيم المفردات داخل مستوى عامل (SSW)

تجزيء التباين الكلي

Total Variation
(SST)

=

تشتت ناجم عن العوامل

(SSB)

+

تشتت ناجم عن عشوائية سحب العينات

(SSW)

و يسمى:

- مجموع المربعات البيني (بين)
- مجموع المربعات فيما بين
- مجموع المربعات المفسّر
- التباين فيما بين المجموعات

و يسمى:

- مجموع المربعات الداخلي (في)
- خطأ مجموع المربعات
- مجموع المربعات غير المفسّر
- التباين داخل المجموعات

مجموع المربعات الكلي

$$SST = SSB + SSW$$

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{\bar{x}})^2$$

حيث:

SST= مجموع المربعات الكلي

k = (المستويات) عدد المجتمعات

n_i = حجم العينة من المجتمع i

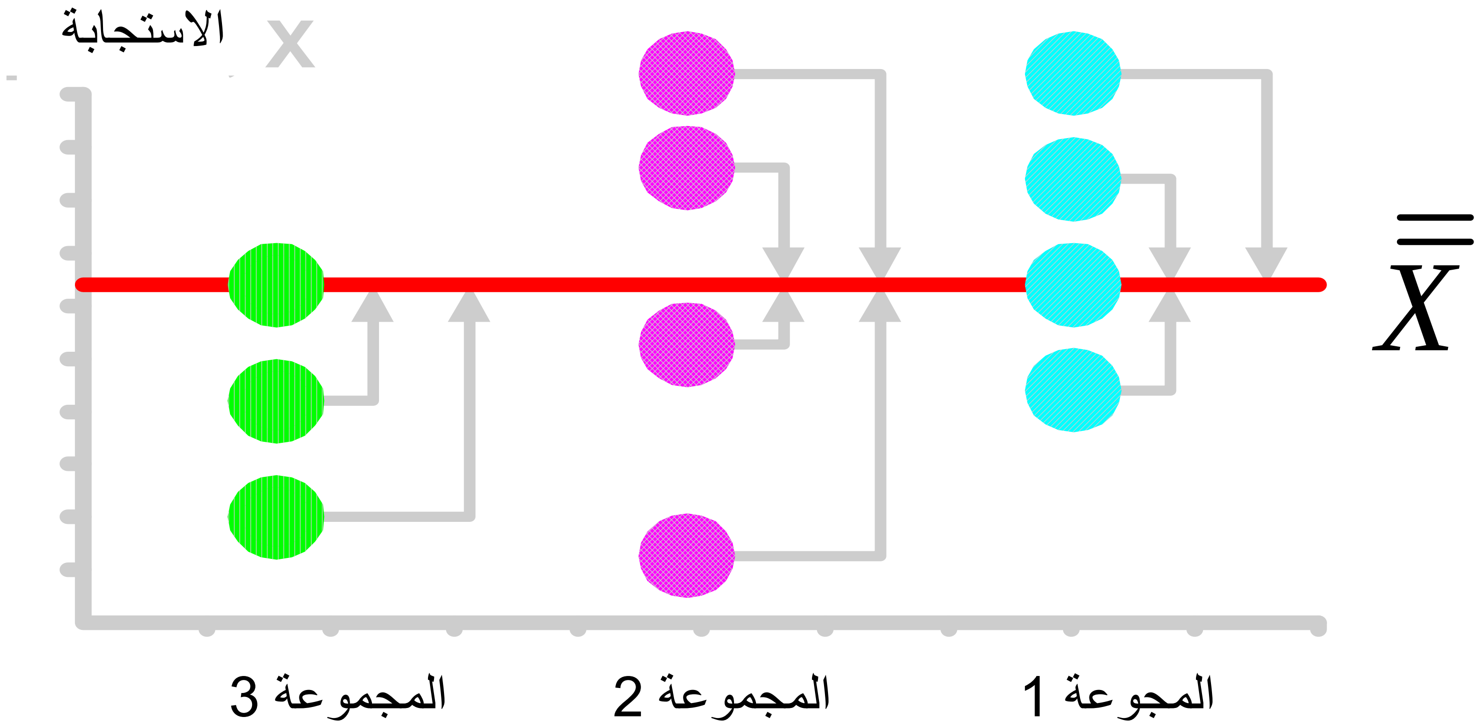
x_{ij} = القياس j من المجتمع i

$\bar{\bar{x}}$ = المتوسط العام

التباين الكلي

تتمة

$$SST = (x_{11} - \bar{\bar{x}})^2 + (x_{12} - \bar{\bar{x}})^2 + \ldots + (x_{kn_k} - \bar{\bar{x}})^2$$



مجموع المربعات البيني

$$SST = \boxed{SSB} + SSW$$

$$\boxed{SSB = \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{\bar{x}})^2}$$

حيث:

SSB = مجموع المربعات البيني

k = عدد المجتمعات (المستويات)

n_i = حجم العينة من المجتمع i

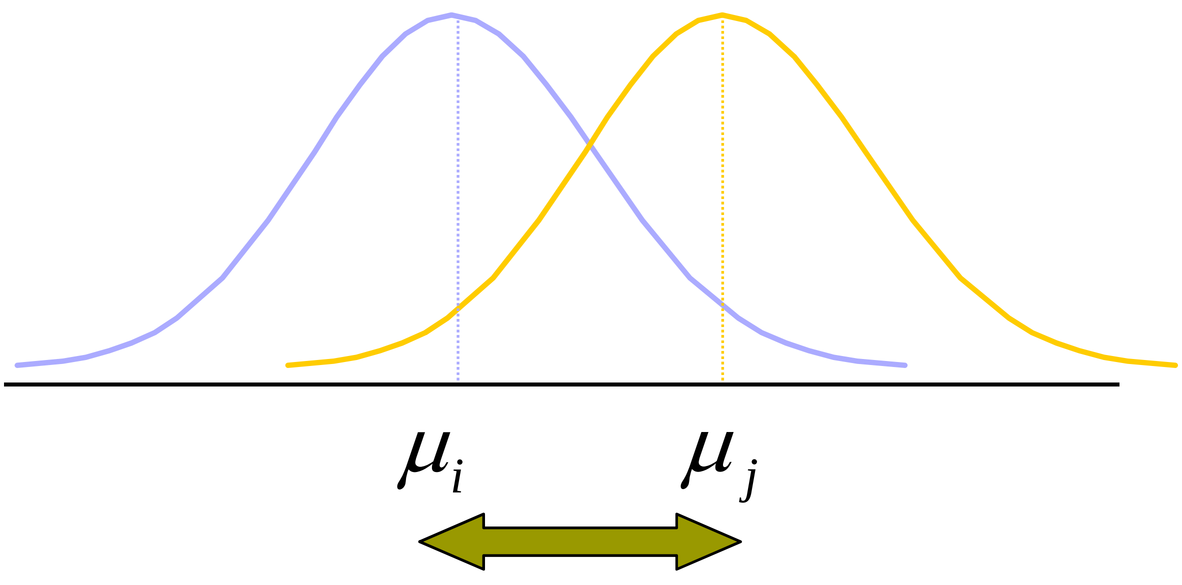
$\bar{x}_i$ = متوسط المجتمع i

$\bar{\bar{x}}$ = المتوسط العام

التباين بين المجموعات

$$SSB = \sum_{i=1}^{k} n_i(\overline{x}_i - \overline{\overline{x}})^2$$

تباين ناجم عن فروقات بين المجموعات



$$MSB = \frac{SSB}{k-1}$$

Mean Square Between = SSB/degrees of freedom

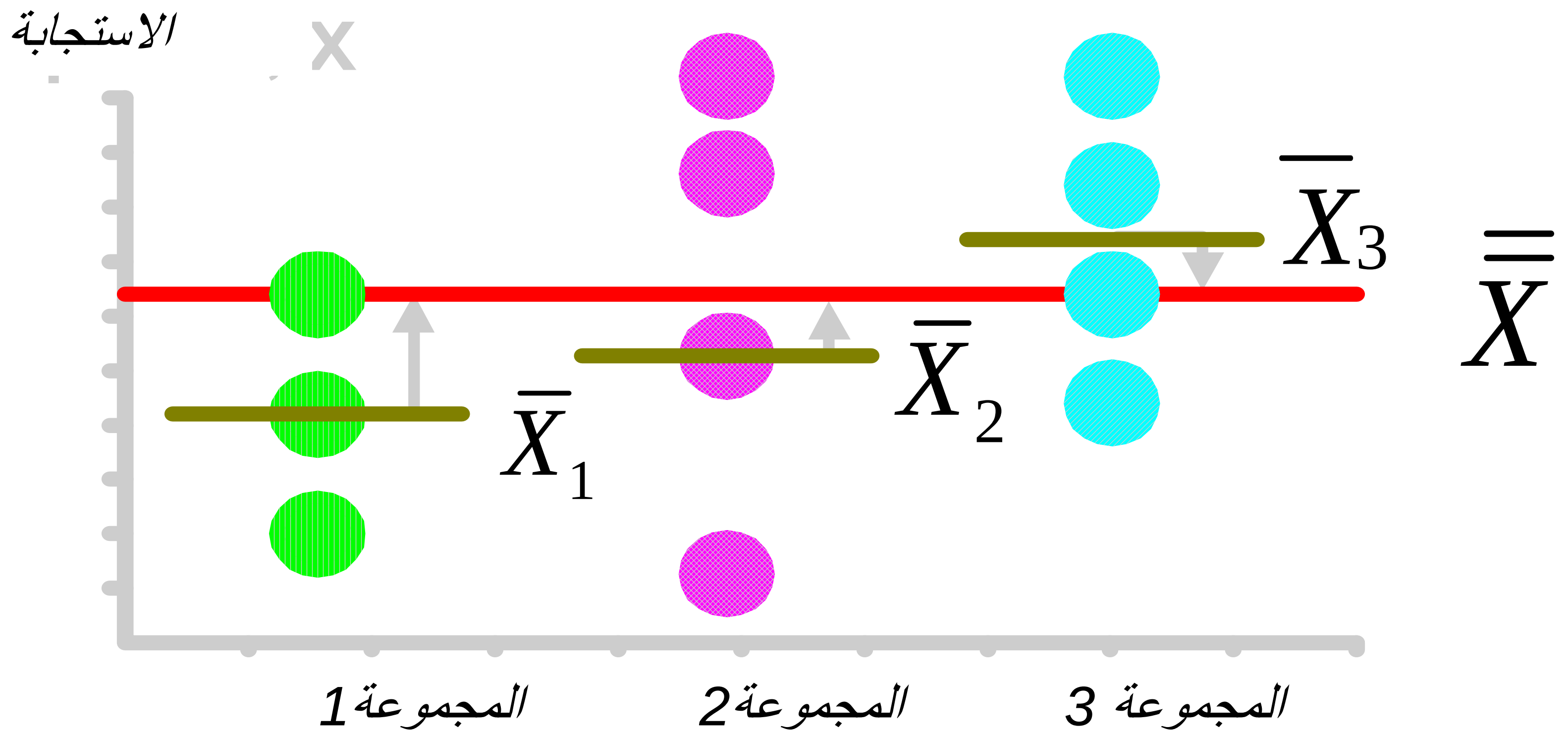
متوسط التباين البيني =

التباين البيني تقسيم عدد درجات الحرية

التباين بين المجموعات

تتمة

$$SSB = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \ldots + n_k(\bar{x}_k - \bar{\bar{x}})^2$$



مجموع المربعات الداخلي

$$SST = SSB + \boxed{SSW}$$

$$\boxed{SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_j} (x_{ij} - \bar{x}_i)^2}$$

حيث:

SSW = مجموع المربعات الداخلي

k = عدد المجتمعات (المستويات)

n_i = حجم العينة من المجتمع i

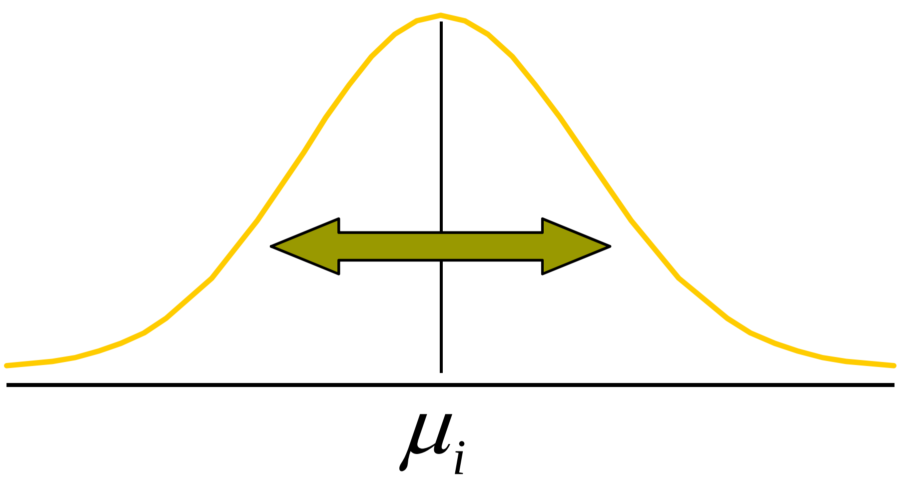
$\bar{x}_i$ = متوسط المجتمع i

$\bar{\bar{x}}$ = المتوسط العام

التباين داخل المجموعات

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_j} (x_{ij} - \bar{x}_i)^2$$

جمع التباين داخل كل مجموعة و من ثم
الجمع على كافة التباينات



$$MSW = \frac{SSW}{N - k}$$

Mean Square Within =
SSW/degrees of freedom

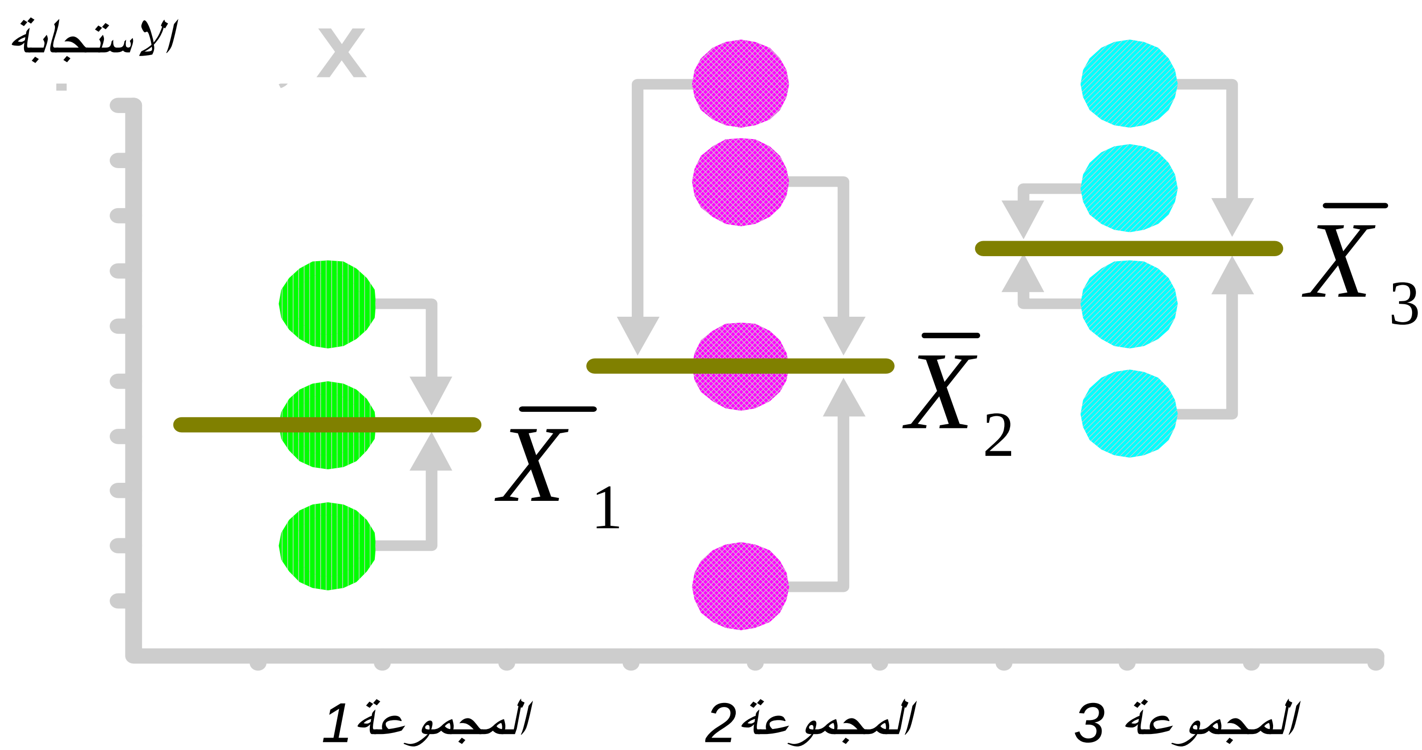
متوسط التباين الداخلي =

التباين الداخلي تقسيم عدد د.ح.

التباين داخل المجموعات

تتمة

$$SSW = (x_{11} - \bar{x}_1)^2 + (x_{12} - \bar{x}_2)^2 + \ldots + (x_{kn_k} - \bar{x}_k)^2$$



جدول تحليل تباين وحيد الاتجاه

مصدر التباين	SS	df	MS	F ratio
بين العينات	SSB	k - 1	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSW}$
في العينات	SSW	N - k	$MSW = \frac{SSW}{N-k}$	
الكلي	SST = SSB+SSW	N - 1		

k = عدد المجتمعات

N = مجموع أحجام العينات من كافة المجتمعات

df = درجات الحرية

جدول تحليل تباين وحيد الاتجاه (مكرر)

Source	SS	df	MS	F ratio
Between	SSB	k - 1	MSB	F
Within	SSW	N - k	MSW	
Total	SST	N - 1		

تحليل التباين وحيد الاتجاه
إحصاءة الاختبار F

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$

H_A: أحد المتوسطات المجتمعية على الأقل مختلف عن باقي المتوسطات

- إحصاءة الاختبار

$$F = \frac{MSB}{MSW}$$

- درجات الحرية
  - $df_1 = k - 1$ (k = عدد المجتمعات)
  - $df_2 = N - k$ (N = مجموع أحجام العينات)

تفسير إحصاءة الاختبار F
تحليل التباين وحيد الاتجاه

- إحصاءة الاختبار F هي نسبة التباين بين العينات إلى التباين داخل العينات.
  - $df_1 = k - 1$ و تكون صغيرة في العادة
  - $df_2 = N - k$ و هي كبيرة عموماً

يفترض أن تكون هذه النسبة قريبة من الواحد عندما تكون $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ صحيحة

يفترض أن تكون هذه النسبة أكبرمن الواحد عندما تكون $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ خاطئة

تحليل التباين وحيد الاتجاه
مثال على الاختبار F

Club 1	**Club 2**	**Club 3**
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

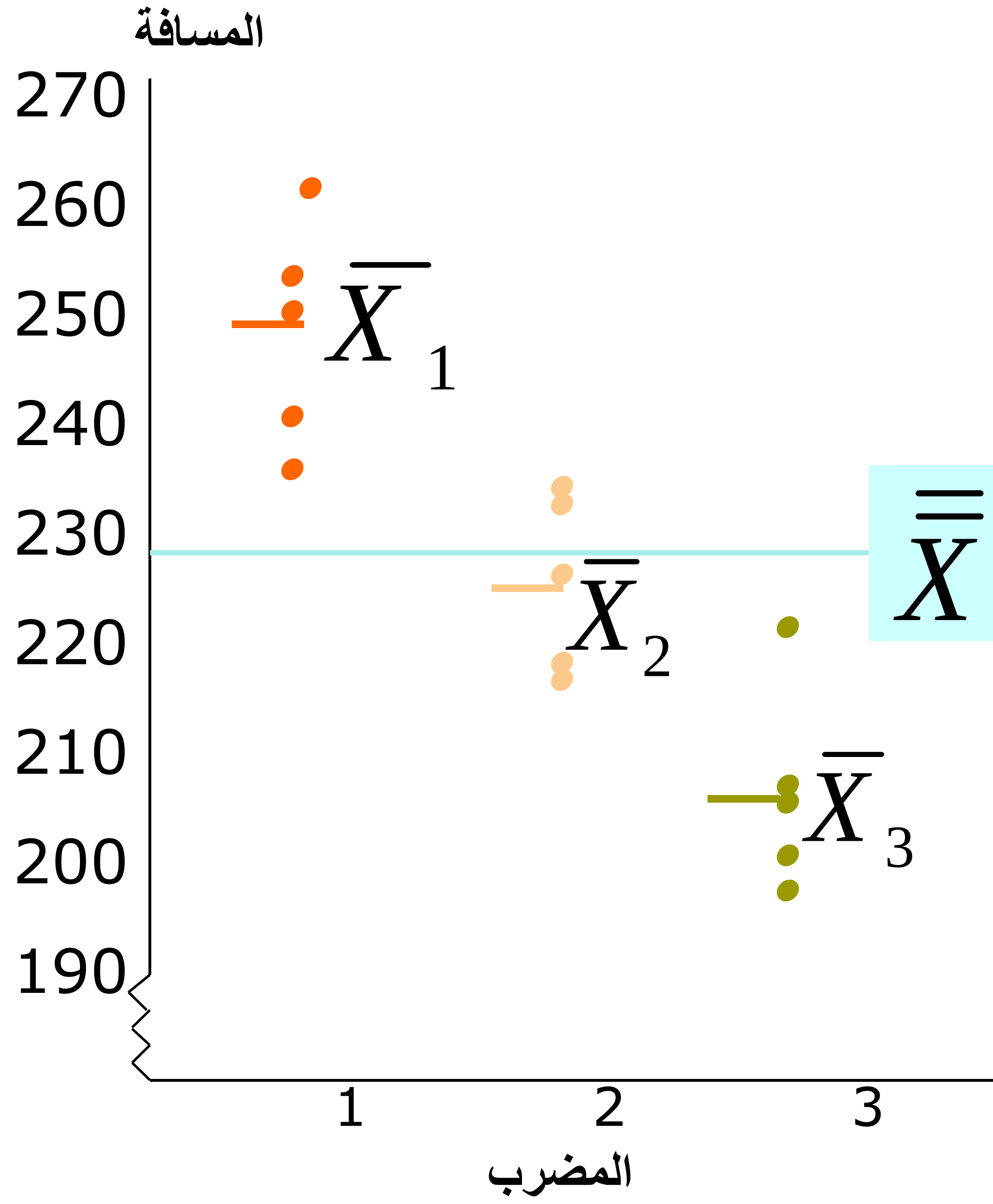
أنت ترغب برؤية ما إذا كان هنالك فرق جوهري في أداء ثلاثة مضارب غولف من حيث مسافة الرمية. قمت بسحب خمسة قياسات عشوائية من محاولات على آلة رمي مؤتمتة و من أجل كل مضرب. هل هناك فرق جوهري في متوسطات المسافة للمضارب الثلاثة عند مستوى 0.05 من الدلالة؟

تحليل التباين وحيد الاتجاه
مثال على الاختبار F: مخطط تبعثر

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$	$\bar{x}_2 = 226.0$	$\bar{x}_3 = 205.8$

$\bar{\bar{x}} = 227.0$



مثال على تحليل التباين وحيد الاتجاه
حسابات

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$ $\quad n_1 = 5$

$\bar{x}_2 = 226.0$ $\quad n_2 = 5$

$\bar{x}_3 = 205.8$ $\quad n_3 = 5$

$\bar{\bar{x}} = 227.0$ $\quad N = 15$

$k = 3$

$$SSB = 5\,[\,(249.2 - 227)^2 + (226 - 227)^2 + (205.8 - 227)^2\,] = 4716.4$$

$$SSW = (254 - 249.2)^2 + (263 - 249.2)^2 + \ldots + (204 - 205.8)^2 = 1119.6$$

$$MSB = 4716.4 / (3\text{-}1) = 2358.2$$

$$MSW = 1119.6 / (15\text{-}3) = 93.3$$

$$F = \frac{2358.2}{93.3} = 25.275$$

مثال على تحليل التباين وحيد الاتجاه
الحل

$H_0: \mu_1 = \mu_2 = \mu_3$

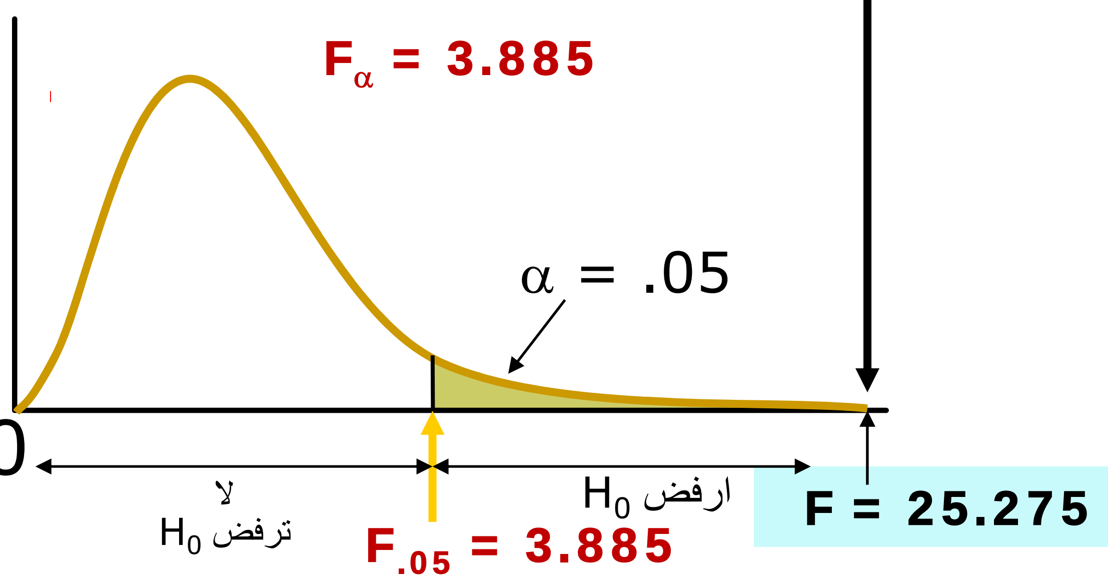
$H_A: \mu_i$ not all equal

$\alpha = .05$

$df_1 = 2 \qquad df_2 = 12$

القيمة الحرجة:

$F_\alpha = 3.885$



إحصاءة الاختبار :

$$F = \frac{MSB}{MSW} = \frac{2358.2}{93.3} = 25.275$$

القرار:

ارفض H_0 عند $\alpha = 0.05$

الاستنتاج:

هناك دليل على وجود μ_i على الأقل مختلف عن البقية

مثال على تحليل التباين وحيد الاتجاه

مخرجات Excel

EXCEL: tools | data analysis | ANOVA: single factor

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Club 1	5	1246	249.2	108.2		
Club 2	5	1130	226	77.5		
Club 3	5	1029	205.8	94.2		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	**4716.4**	**2**	**2358.2**	**25.275**	4.99E-05	**3.885**
Within Groups	**1119.6**	**12**	**93.3**			
Total	5836.0	14				

الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة الخامسة عشرة /القسم النظري/

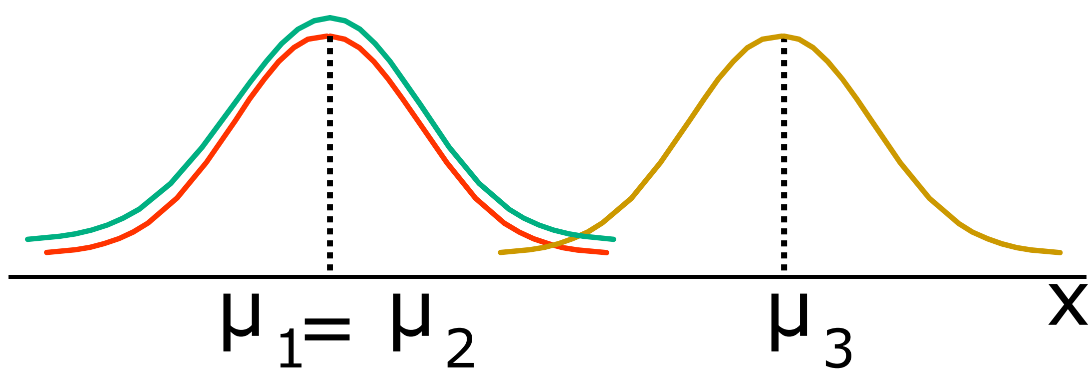
تحليل التباين 2 / 2

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

إجراء تُكي كرامر Tukey-Kramer

- يخبرنا أي المتوسطات المجتمعية مختلف جوهرياً
  - $\mu_1 = \mu_2 \neq \mu_3$
  - يستخدم في تحليل التباين بعد رفض تساوي المتوسطات
- يسمح بمقارنات مثنى مثنى
  - يقارن فروق المتوسطات المطلقة مع المجال الحرج



مجال تُكي كرامرالحرج

$$CR = q_{\alpha}\sqrt{\frac{MSW}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

حيث:

CR	=	المجال الحرج
q_{α}	=	قيمة مأخوذة من جدول المجال المعياري مع
		k و $N - k$ درجة حرية
		و مستوى دلالة مرغوب α
MSW	=	متوسط التباين الداخلي
n_i و n_j	=	أحجام العينات من المجتمعات (المستويات) i و j

إجراء تُكي كرامر Tukey-Kramer
مثال

. أحسب فروق المتوسطات المطلقة:

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$$\left|\bar{x}_1 - \bar{x}_2\right| = \left|249.2 - 226.0\right| = 23.2$$

$$\left|\bar{x}_1 - \bar{x}_3\right| = \left|249.2 - 205.8\right| = 43.4$$

$$\left|\bar{x}_2 - \bar{x}_3\right| = \left|226.0 - 205.8\right| = 20.2$$

2. جد قيمة q value من الجدول من أجل مستوى الدلالة المرغوب α

$$\boxed{q_\alpha = 3.77}$$

إجراء تُكي كرامر Tukey-Kramer
مثال

3. أحسب المجال الحرج:

$$\text{Critical Range} = q_\alpha \sqrt{\frac{MSW}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = 3.77\sqrt{\frac{93.3}{2}\left(\frac{1}{5}+\frac{1}{5}\right)} = 16.285$$

4. قارن:

$$|\bar{x}_1 - \bar{x}_2| = 23.2$$

$$|\bar{x}_1 - \bar{x}_3| = 43.4$$

$$|\bar{x}_2 - \bar{x}_3| = 20.2$$

5. كافة فروق المتوسطات المطلقة أكبر من المجال الحرج. مما يعني أن هناك فرقاً جوهرياً بين كل زوجين من المتوسطات عند مستوى 5% من الدلالة.

تحليل التباين : بلوك كامل عشوائي

- هنا أيضاً نختبر تساوي المتوسطات المجتمعية (من أجل مستويات مختلفة للعامل على سبيل المثال.
- و لكننا نرغب بضبط تأثير عامل ثان (مع مستويين أو أكثر)
- يستخدم عندما يؤثر أكثر من عامل واحد على قيمة المتغير التابع، في الوقت الذي نهتم فيه بعامل مفتاحي واحد.
- تدعى مستويات العامل الثانوي بلوكات blocks

تجزيء التباين

- يمكن تجزيء التباين الكلي إلى ثلاثة أجزاء:

$$SST = SSB + SSBL + SSW$$

SST = المجموع الكلي للمربعات
SSB = مجموع المربعات بين مستويات العامل
SSBL= مجموع المربعات بين البلوكات
SSW = مجموع المربعات داخل المستويات

مجموع مربعات للبلوكات

$$SST = SSB + \boxed{SSBL} + SSW$$

$$\boxed{SSBL = \sum_{j=1}^{b} k(\bar{x}_j - \bar{\bar{x}})^2}$$

حيث:

$k =$ عدد مستويات العامل المعني

$b =$ عدد البلوكات

$\bar{x}_j =$ متوسط العينة من البلوك j

$\bar{\bar{x}} =$ المتوسط العام (من كل البيانات)

تجزيء التباين

- يمكن تجزيء التباين الكلي الآن إلى ثلاثة أجزاء:

$$\boxed{SST} = \boxed{SSB} + SSBL + \boxed{SSW}$$

و هذان يحسبان كما في حالة
تحليل التباين وحيد الاتجاه

$$SSW = SST - (SSB + SSBL)$$

Mean Squares

$$\text{MSBL} = \text{Mean square blocking} = \frac{\text{SSBL}}{\text{b}-1}$$

$$\text{MSB} = \text{Mean square between} = \frac{\text{SSB}}{\text{k}-1}$$

$$\text{MSW} = \text{Mean square within} = \frac{\text{SSW}}{(\text{k}-1)(\text{b}-1)}$$

جدول تحليل التباين : بلوك كامل عشوائي

مصدر التباين	SS	df	MS	نسبة F
بين البلوكات	SSBL	b - 1	MSBL	$\frac{MSBL}{MSW}$
بين العينات	SSB	k - 1	MSB	$\frac{MSB}{MSW}$
في العينات	SSW	(k−1)(b-1)	MSW	
الإجمالي	SST	N - 1		

N = عدد أحجام العينات من كل المجتمعات

df = درجات الحرية

k = عدد المجتمعات

b = عدد البلوكات

جدول تحليل التباين العشوائي: بلوك كامل (مكرر)

Randomized Block ANOVA Table

Source of Variation	SS	df	MS	F ratio
Between Blocks	SSBL	b - 1	MSBL	$\frac{MSBL}{MSW}$
Between Samples	SSB	k - 1	MSB	$\frac{MSB}{MSW}$
Within Samples	SSW	(k–1)(b-1)	MSW	
Total	SST	N - 1		

k = number of populations

b = number of blocks

N = sum of the sample sizes from all populations

df = degrees of freedom

اختبار البلوكات

$$H_0 : \mu_{b1} = \mu_{b2} = \mu_{b3} = \ldots$$

H_A : متوسطات البلوكات ليست جميعها متساوية

اختبار البلوكات: $F = \dfrac{\mathbf{MSBL}}{\mathbf{MSW}}$

$df_1 = b - 1$

$df_2 = (k - 1)(b - 1)$

ارفض H_0 إذا كان $F > F_\alpha$

اختبار العامل الرئيسي

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$$

H_A : المتوسطات المجتمعية ليست جميعها متساوية

اختبار البلوكات:

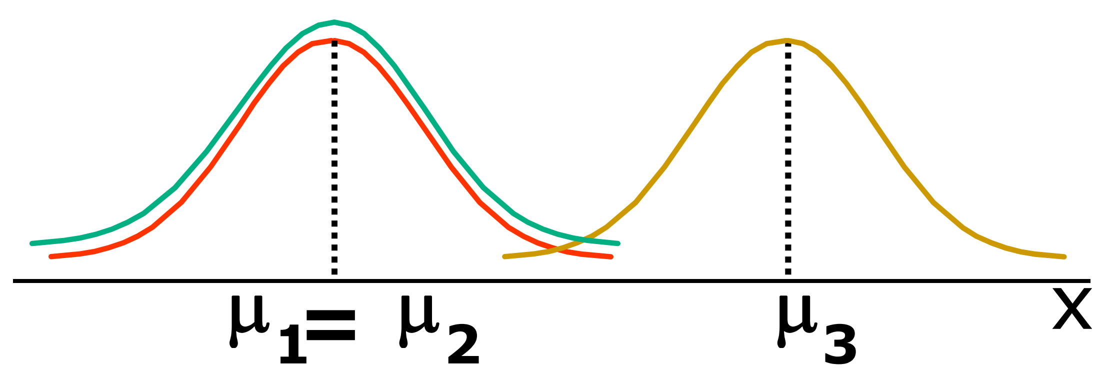
$$\mathbf{F = \frac{MSB}{MSW}}$$

$df_1 = b - 1$
$df_2 = (k - 1)(b - 1)$

ارفض H_0 إذا كان $F > F_\alpha$

اختبار فيشر للفرق الأقل دلالة

- يستخدم لاختبار أي المتوسطات المجتمعية مختلف جوهرياً
  - $\mu_1 = \mu_2 \neq \mu_3$
  - يستخدم بعد رفض تساوي المتوسطات في تحليل التباين من نمط البلوك العشوائي
  - يسمح بمقارنات مثنى مثنى
  - يقارن الفروق المطلقة للمتوسطات مع المجال الحرج



اختبار فيشر للفرق الأقل دلالة LSD

$$LSD = t_{\alpha/2}\sqrt{MSW}\sqrt{\frac{2}{b}}$$

حيث:

$t_{\alpha/2}$ = قيمة من توزيع تي (الذيل الأعلى)
من أجل $\alpha/2$ و $(k-1)(n-1)$ درجة حرية

MSW = متوسط المربعات الداخلي من جدول تحليل التباين

b = عدد البلوكات

k = عدد مستويات العامل الرئيسي

اختبار فيشر للفرق الأقل دلالة

تتمة

$$LSD = t_{\alpha/2}\sqrt{MSW}\sqrt{\frac{2}{b}}$$

قارن:

$$\left|\bar{x}_1 - \bar{x}_2\right|$$

$$\left|\bar{x}_1 - \bar{x}_3\right|$$

$$\left|\bar{x}_2 - \bar{x}_3\right|$$

etc...

$$\left|\bar{x}_i - \bar{x}_j\right| \overset{?}{>} LSD$$

إذا كانت القيمة المطلقة للفرق أكبر من LSD فإن هنالك فرق جوهري بين أزواج المتوسطات عند المستوى المستخدم للدلالة.

تحليل التباين باتجاهين

- يقوم بفحص تأثير:

  - عاملين فأكثر على المتغير التابع
    - النسبة المئوية للكربنة و سرعة خط الانتاج في عملية تعبئة مياه غازية
  - التفاعل بين المستويات المختلفة لتلك العوامل
    - هل يعتمد تأثير نسبة كربنة % محددة على مستوى سرعة الخط؟

تحليل التباين باتجاهين

تتمة

- افتراضات
  - المجتمعات تتبع التوزيع الطبيعي
  - تباينات المجتمعات متساوية
  - العينات مستقلة و مسحوبة عشوائياً

تحليل التباين باتجاهين: مصدر التباين

هناك عاملان نهتم بهما: A و B

$a =$ عدد مستويات العامل **A**

$b =$ عدد مستويات العامل **B**

$N =$ العدد الإجمالي لمشاهدات في كافة الخلايا

تحليل التباين باتجاهين: مصدر التباين

تتمة

$$SST = SS_A + SS_B + SS_{AB} + SSE$$

درجات الحرية:

SST
التباين الكلي
$N - 1$

SS_A
التباين العائد للعامل A
$a - 1$

SS_B
التباين العائد للعامل B
$b - 1$

SS_{AB}
التباين العائد للتفاعل بين العاملين A و B
$(a - 1)(b - 1)$

SSE
التباين الملازم (الخطأ)
$N - ab$

معادلات تحليل التباين باتجاهين

المجموع الكلي للمربعات:

$$SST = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n'} (x_{ijk} - \bar{\bar{x}})^2$$

مجموع المربعات للعامل A

$$SS_A = bn' \sum_{i=1}^{a} (\bar{x}_i - \bar{\bar{x}})^2$$

مجموع المربعات للعامل B

$$SS_B = an' \sum_{j=1}^{b} (\bar{x}_j - \bar{\bar{x}})^2$$

معادلات تحليل التباين باتجاهين

تتمة

مجموع مربعات التفاعل بين A و B:

$$SS_{AB} = n' \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$$

مجموع مربعات الخطأ:

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n'} (x_{ijk} - \bar{x}_{ij})^2$$

معادلات تحليل التباين باتجاهين

تتمة

حيث:

المتوسط العام
$$\bar{\bar{x}} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n'} x_{ijk}}{abn'} =$$

$$\bar{x}_i = \frac{\sum_{j=1}^{b}\sum_{k=1}^{n'} x_{ijk}}{bn'} =$$ **متوسط كل مستوى من العامل A**

$$\bar{x}_j = \frac{\sum_{i=1}^{a}\sum_{k=1}^{n'} x_{ijk}}{an'} =$$ **متوسط كل مستوى من العامل B**

$$\bar{x}_{ij} = \sum_{k=1}^{n'} \frac{x_{ijk}}{n'} =$$ **متوسط كل خلية**

a = A عدد مستويات العامل
b = B عدد مستويات العامل
n′ = number of replications in each cell

حسابات وسطي المربعات

$$MS_A = \text{Mean square factor } A = \frac{SS_A}{a-1}$$

المتوسط التربيعي للعامل A

$$MS_B = \text{Mean square factor } B = \frac{SS_B}{b-1}$$

المتوسط التربيعي للعامل B

$$MS_{AB} = \text{Mean square interaction} = \frac{SS_{AB}}{(a-1)(b-1)}$$

متوسط مربع التفاعل

$$MSE = \text{Mean square error} = \frac{SSE}{N-ab}$$

متوسط مربع الخطأ

تحليل التباين باتجاهين :
إحصاءة الاختبار F

إختبار F للأثر الرئيسي للعامل A

$H_0: \mu_{A1} = \mu_{A2} = \mu_{A3} = \cdots$

H_A: Not all μ_{Ai} are equal

$$F = \frac{MS_A}{MSE}$$

ارفض H_0 إذا كان $F > F_\alpha$

إختبار F للأثر الرئيسي للعامل B

$H_0: \mu_{B1} = \mu_{B2} = \mu_{B3} = \cdots$

H_A: ليست كل الـ μ_{Bi} متساوية

$$F = \frac{MS_B}{MSE}$$

ارفض H_0 إذا كان $F > F_\alpha$

اختبار F للتفاعل بين العاملين

H_0: العاملان لا يتفاعلان في التأثير على الاستجابة الوسطية

H_A: يوجد تفاعل بين العاملين

$$F = \frac{MS_{AB}}{MSE}$$

ارفض H_0 إذا كان $F > F_\alpha$

جدول تحليل التباين باتجاهين

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	إحصاءة F
العامل A	SS_A	$a - 1$	$MS_A = SS_A/(a - 1)$	$\frac{MS_A}{MSE}$
العامل B	SS_B	$b - 1$	$MS_B = SS_B/(b - 1)$	$\frac{MS_B}{MSE}$
AB (تفاعل)	SS_{AB}	$(a - 1)(b - 1)$	$MS_{AB} = SS_{AB}/[(a - 1)(b - 1)]$	$\frac{MS_{AB}}{MSE}$
خطأ	SSE	$N - ab$	$MSE = SSE/(N - ab)$	
مجموع	SST	$N - 1$		

جدول تحليل التباين باتجاهين (مكرر)

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Statistic
Factor A	SS_A	$a - 1$	$MS_A = SS_A/(a - 1)$	$\frac{MS_A}{MSE}$
Factor B	SS_B	$b - 1$	$MS_B = SS_B/(b - 1)$	$\frac{MS_B}{MSE}$
AB (Interaction)	SS_{AB}	$(a - 1)(b - 1)$	$MS_{AB} = SS_{AB} / [(a - 1)(b - 1)]$	$\frac{MS_{AB}}{MSE}$
Error	SSE	$N - ab$	$MSE = SSE/(N - ab)$	
Total	SST	$N - 1$		

تحليل التباين باتجاهين : خصائص إحصاءة الاختبار F

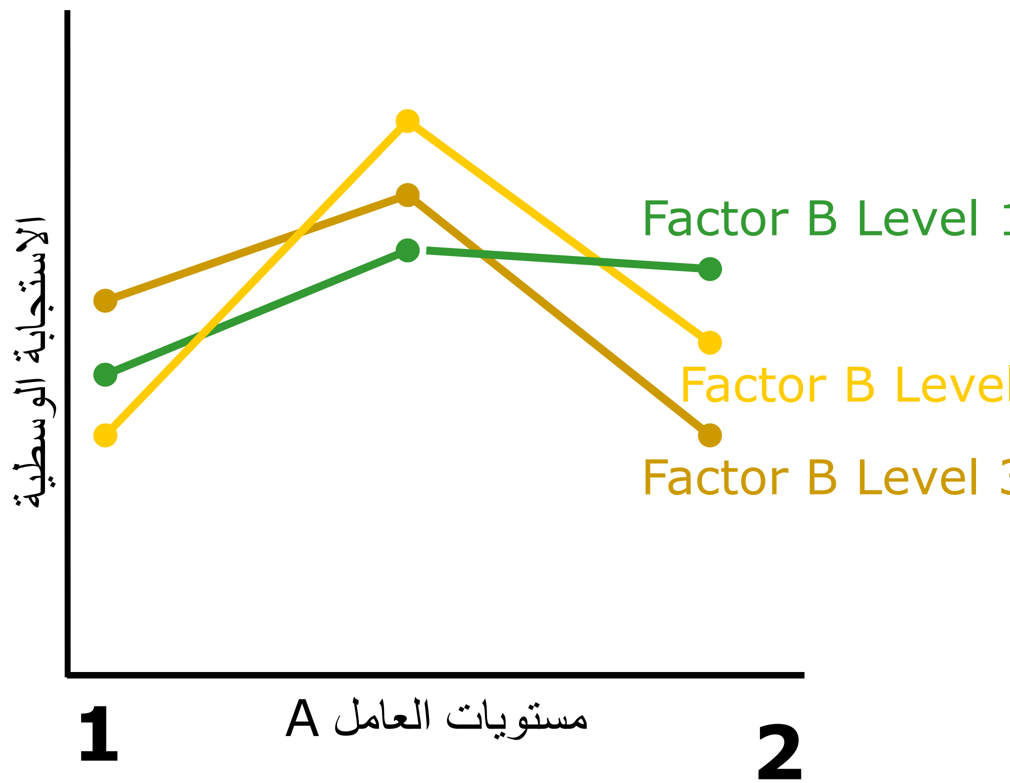
- عدد درجات الحرية هو دائماً المجموع التالي:
  - $N-1 = (N-ab) + (a-1) + (b-1) + (a-1)(b-1)$
  - Total = error + factor A + factor B + interaction

- مقام اختبار F هو نفسه دائماً و لكن البسط يتغير

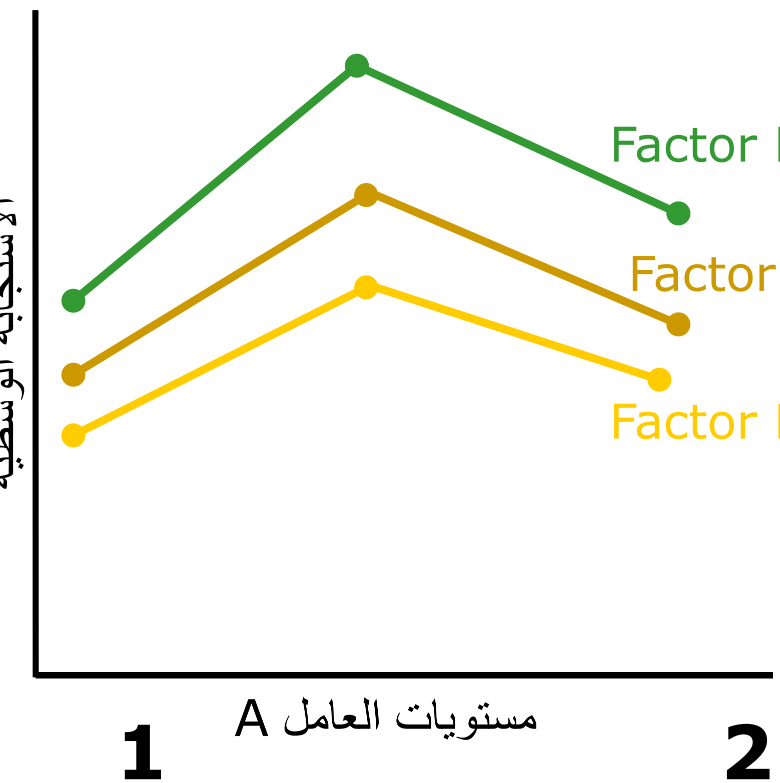
- يبلغ مجموع المربعات المجموع التالي دائماً
  - $SST = SSE + SS_A + SS_B + SS_{AB}$
  - Total = error + factor A + factor B + interaction

أمثلة:
المقارنة بين "تفاعل: و "لا تفاعل"

يوجد تفاعل



لا يوجد تفاعل



الطرق الكمية الإحصائية

مدخل صنع القرار

المحاضرة السادسة عشرة /القسم النظري/

اختبارات جودة التوفيق

و

اختبارات الاستقلال

جامعة دمشق، المعهد العالي للتنمية الإدارية
ماجستير الأعمال الدولية 2008 – 2009

الدكتور معاذ الشرفاوي الجزائرلي

إختبار كاي مربع لجودة التوفيق

- هل تتبع بيانات المعاينة توزيعاً مفترضاً؟
  - أمثلة:
- هل اتصالات طلبات الصيانة متساوية في العدد على مدى أيام الأسبوع؟ (أي هل تتبع التوزيع الطبيعي؟)
- هل تتبع القياسات المأخوذة من عملية إنتاجية ما التوزيع الطبيعي؟

إختبار كاي مربع لجودة التوفيق

(تتمة)

- هل اتصالات (مكالمات) الصيانة متساوية في العدد على مدى أيام الأسبوع؟ (أي هل تتبع التوزيع الطبيعي؟)
  - بيانات عينة لعشرة أيام مأخوذة من أجل كل يوم من أيام الأسبوع

	عدد الاتصالات في اليوم:
السبت	230
الأحد	192
الإثنين	290
الثلاثاء	250
الأربعاء	238
الخميس	257
الجمعة	265
	$\Sigma = 1722$

منطق إختبار جودة التوفيق

- إذا كانت المكالمات موزعة توزيعاً منتظماً فإنه من النتوقع أن تتوزع المكالمات الـ 1722 بشكل متساوٍ على أيام الأسبوع السبعة:

$$\frac{1722}{7} = 246$$

- إختبار كاي مربع لجودة التوفيق:

يقوم باختبار ما إذا كانت نتائج العينة متفقة مع النتائج المتوقعة

التكرارات المشاهدة و التكرارات المتوقعة

الأيام	مشاهد o_i	متوقع e_i
السبت	230	246
الأحد	192	246
الإثنين	290	246
الثلاثاء	250	246
الأربعاء	238	246
الخميس	257	246
الجمعة	265	246
الإجمالي	1722	1722

إحصاءة اختبار كاي مربع

H_0:	تتوزع المكالمات بانتظام على أيام الأسبوع (تتبع التوزيع المنتظم)
H_A:	توزيع المكالمات ليس منتظماً (لا تتبع التوزيع المنتظم)

- إحصاءة الاختبار هي

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \qquad (df = k - 1)$$

حيث:

k = عدد الفئات

o_i = تكرار الخلية المتوقع للفئة i

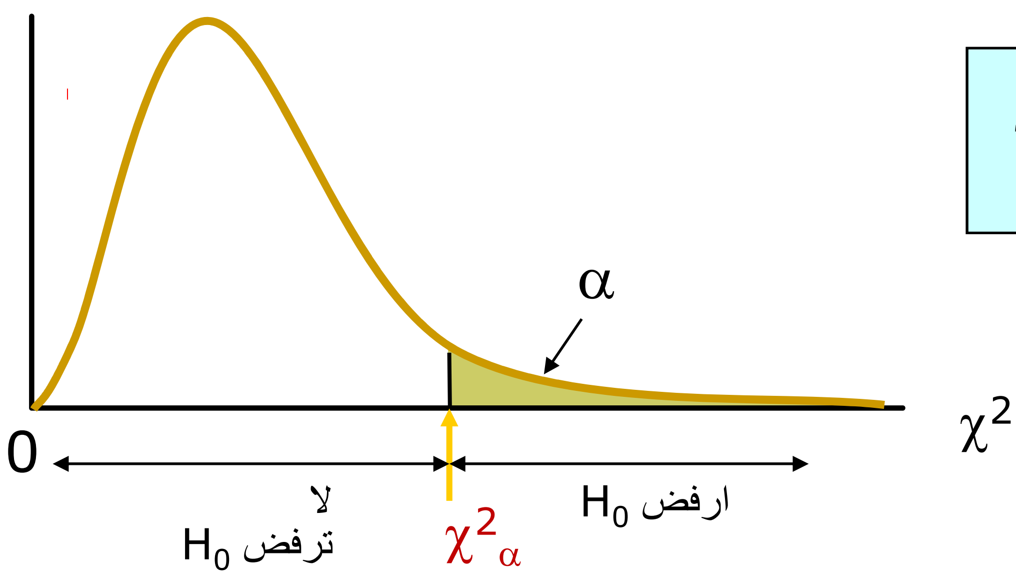
e_i = تكرار الخلية المشاهد i

منطقة الرفض

H_0: تتوزع المكالمات بانتظام على أيام الأسبوع (تتبع التوزيع المنتظم)

H_A: توزيع المكالمات ليس منتظماً (لا تتبع التوزيع المنتظم)

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$



ارفض H_0 إذا كان $\chi^2 > \chi^2_\alpha$

(مع $k - 1$ درجة حرية)

إحصاءة اختبار كاي مربع

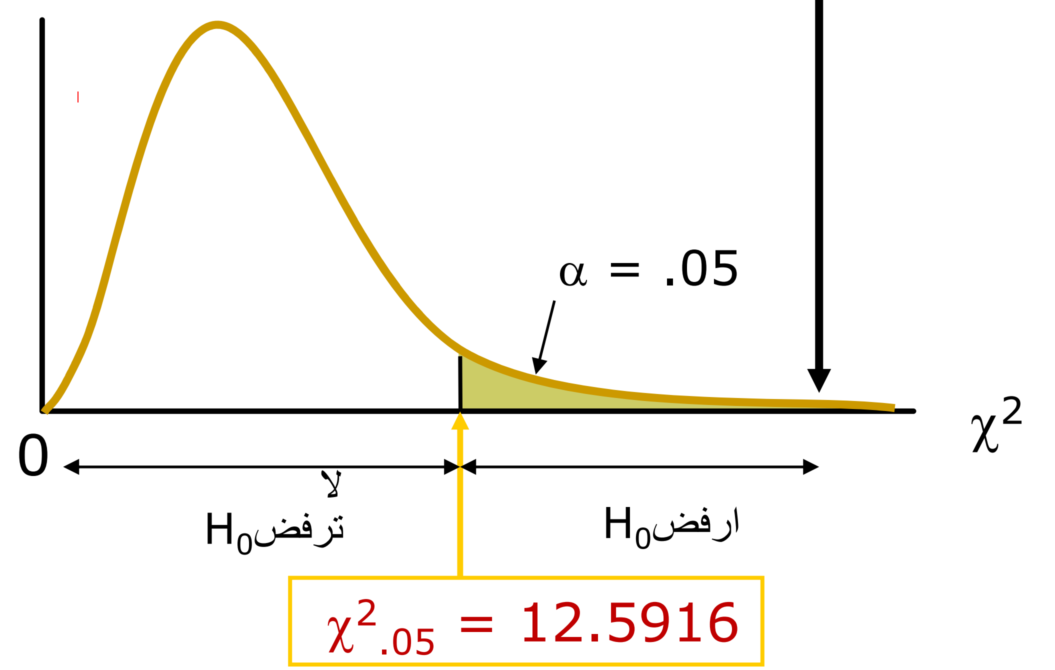
H_0: تتوزع المكالمات بانتظام على أيام الأسبوع (تتبع التوزيع المنتظم)

H_A: توزيع المكالمات ليس منتظماً (لا تتبع التوزيع المنتظم)

$$\chi^2 = \frac{(290-246)^2}{246} + \frac{(250-246)^2}{246} + \ldots + \frac{(192-246)^2}{246} = \boxed{23.05}$$

بما أن: $k - 1 = 6$ df

$\chi^2_{.05} = 12.5916$



الاستنتاج:

$\chi^2 = 23.05 > \chi^2_{\alpha} = 12.5916$

ارفض H_0 و استنتج أن التوزيع غير منتظم

مثال على التوزيع الطبيعي

- هل تتبع القياسات المأخوذة من عملية الإنتاج التوزيع الطبيعي مع $\mu = 50$ و $\sigma = 15$؟

- العملية:
  - اجمع بيانات العينة
  - قم بتجميع نتائج المعاينة في فئات (خلايا) و لاحظ أن التكرار المتوقع للخلية يجب ألا يقل عن 5 من أجل كل خلية
  - قارن التكرارات الفعلية مع التكرارات المتوقعة للخلايا

مثال على التوزيع الطبيعي

(تتمة)

- اسحب العينة و قم بتجميع البيانات في فئات:

150 قياساً من المعاينة
80
65
36
66
50
38
57
77
59
...إلخ...

الفئة	لتكرار
< 30	10
≥ 30 & < 40	21
≥ 40 & < 50	33
≥ 50 & < 60	41
≥ 60 & < 70	26
≥ 70 & < 80	10
≥ 80 & < 90	7
≥ 90	2
الإجمالي	150

مثال على التوزيع الطبيعي

(تتمة)

- ما هي **التكرارات المتوقعة** لهذه الفئات من أجل توزيع طبيعي بوسط و انحراف قدرهما $\mu = 50$ و $\sigma = 15$

الفئة	التكرار	التكرار المتوقع
< 30	10	؟
≥ 30 & < 40	21	
≥ 40 & < 50	33	
≥ 50 & < 60	41	
≥ 60 & < 70	26	
≥ 70 & < 80	10	
≥ 80 & < 90	7	
≥ 90	2	
الإجمالي	150	

التكرارات المتوقعة

القيمة	P(X < value)	التكرار المتوقع
< 30	0.09121	13.68
≥ 30 & < 40	0.16128	24.19
≥ 40 & < 50	0.24751	37.13
≥ 50 & < 60	0.24751	37.13
≥ 60 & < 70	0.16128	24.19
≥ 70 & < 80	0.06846	10.27
≥ 80 & < 90	0.01892	2.84
≥ 90	0.00383	0.57
الإجمالي	1.00000	150.00

التكرارات المتوقعة في عينة من الحجم n=150 ، من توزيع طبيعي ذي μ=50 و σ=15

مثال:

$$P(x < 30) = P\left(z < \frac{30-50}{15}\right)$$

$$= P(z < -1.3333)$$

$$= .0912$$

$$(.0912)(150) = 13.68$$

إحصاءة الاختبار

الفئة	التكرار (المشاهد، o_i)	التكرار المتوقع، e_i
< 30	10	13.68
≥ 30 & < 40	21	24.19
≥ 40 & < 50	33	37.13
≥ 50 & < 60	41	37.13
≥ 60 & < 70	26	24.19
≥ 70 & < 80	10	10.27
≥ 80 & < 90	7	2.84
≥ 90	2	0.57
الإجمالي	150	150.00

إحصاءة الاختبار هي

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

ارفض H_0 إذا كان:

$$\chi^2 > \chi^2_\alpha$$

مع $(k - 1)$ درجة حرية

منطقة الرفض

H_0: $\mu = 50$ و $\sigma = 15$ القيم تتبع التوزيع الطبيعي مع

H_A: توزيع القيم لا يتبع هذا التوزيع

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} = \frac{(10 - 13.68)^2}{13.68} + \ldots + \frac{(2 - 0.57)^2}{0.57} = \boxed{12.097}$$

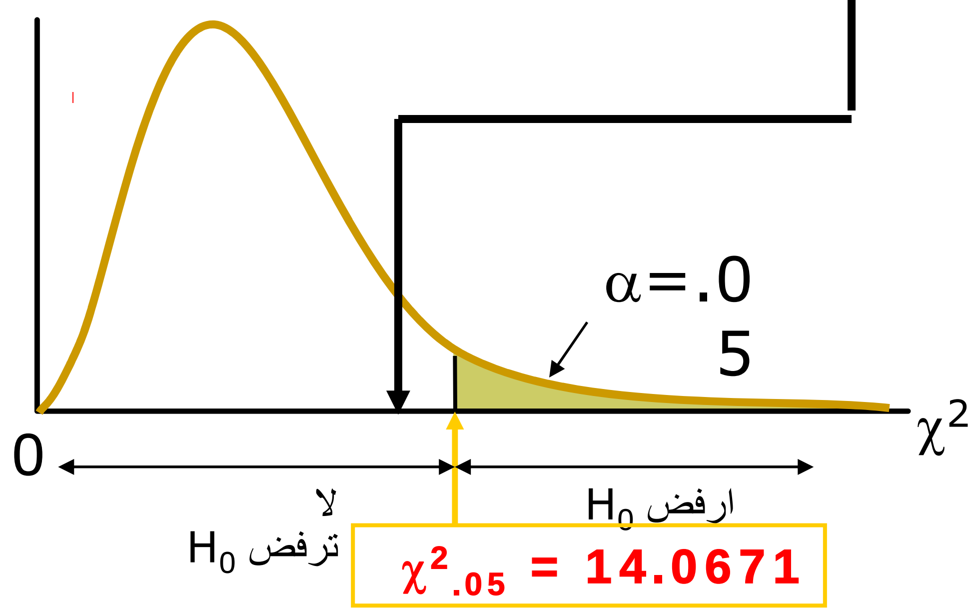
8 فئات و لذلك نستعمل 7 درجات حرية

$\chi^2_{.05} = 14.0671$

الاستنتاج:

$\chi^2 = 12.097 < \chi^2_\alpha = 14.0671$

لذلك لا ترفض H_0



جداول الاستقلال

جداول الاستقلال أو Contingency Tables

- حالات تنطوي على نسب مجتمعية متعددة.
- يستخدم لتصنيف مشاهدات العينة تبعاً لاثنين أو أكثر من الصفات
- و يسمى أيضاً بالجدولة التقاطعية Crosstabulation

مثال على جداول الاستقلال

خاصية اليد اليسارية إزاء الجنس

- اليد الأساسية : يسار مقابل يمين
- الجنس : ذكر مقابل أنثى

H_0: كون اليد الأساسية يمين أو يسار مسألة مستقلة عن الجنس

H_A: مسألة تفضيل إحدى اليدين على الأخرى ليست مستقلة عن الجنس

مثال على جداول الاستقلال

(تتمة)

نتائج المعاينة بعد ترتيبها في جدول استقلال:

حجم العينة: $n = 300$

120 أنثى منهن 12 يساريات،

و 180 ذكر منهم 24 يساريين.

الجنس	اليد المفضلة		
	يمين	يسار	
أنثى	108	12	120
ذكر	156	24	180
	264	36	300

منطق الاختبار

H_0: كون اليد الأساسية يمين أو يسار مسألة مستقلة عن الجنس

H_A: مسألة تفضيل إحدى اليدين على الأخرى ليست مستقلة عن الجنس

- إذا كانت فرضية العدم صحيحة فإن نسبة الإناث اليساريات يجب أن تكون مماثلة لنسبة الذكور اليساريين.

- يجب أن تكون النسبتان أعلاه مماثلتان لنسبة اليساريين عند الناس عامةً.

إيجاد التكرارات المتوقعة

120 أنثى F منهن 12 يساريات LH

180 ذكر M منهم 24 يساريين LH

بالإجمال:

$P(LH) = 36/300 = .12$

في حالة الاستقلال:

$P(LH|\ F) = P(LH|\ M) = .12$

و لذلك فإننا نتوقع أن 12% من 120 أنثى و 12% من 180 ذكر سيكونون يساريين / يساريات

$(120)(.12) = 14.4$ **و عليه نتوقع وجود 14.4 أنثى يسارية**

$(180)(.12) = 21.6$ **و عليه، نتوقع وجود 21.6 ذكراً يسارياً**

التكرارات المتوقعة للخلية

(تتمة)

- توقع تكرارات الخلية (التكرارات المتوقعة)

$$e_{ij} = \frac{(i^{th}\ \text{Row total})(j^{th}\ \text{Column total})}{\text{Total sample size}}$$

- أي جداء إجمالي قيم السطر بإجمالي قيم العمود مقسوماً على الحجم الإجمالي للعينة

مثال:

$$e_{11} = \frac{(120)(36)}{300} = 14.4$$

التكرارات المتوقعة و التكرارات المشاهدة

التكرارات المشاهدة O مقابل التكرارات المتوقعة E:

الجنس	اليد المفضلة		
	يسار	يمين	
أنثى	O = 12	O = 108	120
	E = 14.4	E = 105.6	
ذكر	O = 24	O = 156	180
	E= 21.6	E = 158.4	
	36	264	300

إحصاءة اختبار كاي مربع (اختبار الاستقلال)

إن إحصاءة كاي مربع الخاصة بالاستقلال هي:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(o_{ij}-e_{ij})^2}{e_{ij}}$$

$$d.f. = (r-1)(c-1)$$

- حيث:

$o_{ij} = (i,\ j)$ التكرار المشاهد في الخلية

$e_{ij} = (i,\ j)$ التكرار المتوقع للخلية

$r =$ عدد الأسطر

$c =$ عدد الأعمدة

التكرارات المتوقعة و التكرارات المشاهدة

الجنس	يسار	يمين	
	اليد المفضلة		
أنثى	O = 12 E = 14.4	O = 108 E = 105.6	120
ذكر	O = 24 E = 21.6	O = 156 E = 158.4	180
	36	264	300

$$\chi^2 = \frac{(12-14.4)^2}{14.4} + \frac{(108-105.6)^2}{105.6} + \frac{(24-21.6)^2}{21.6} + \frac{(156-158.4)^2}{158.4} = 0.6848$$

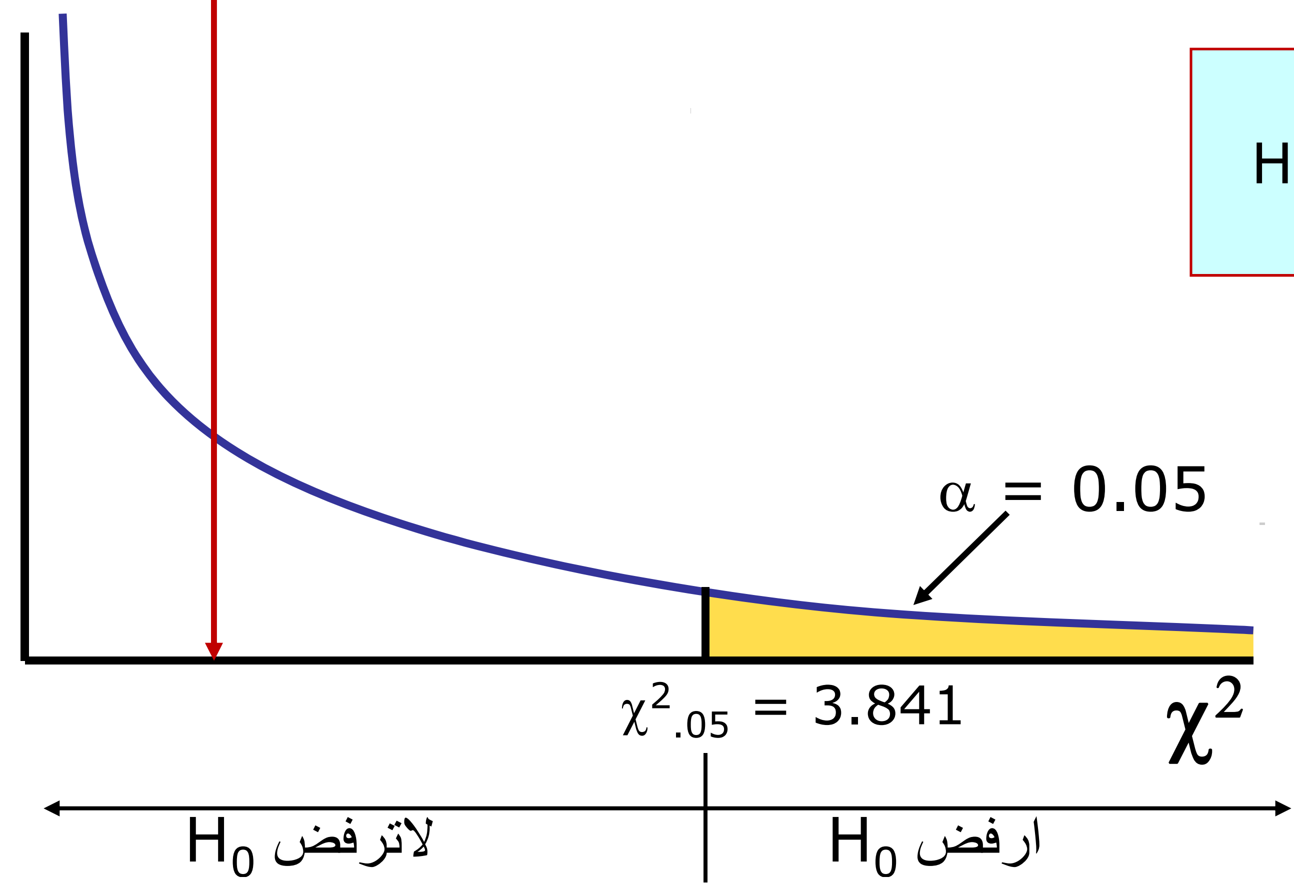
تحليل الاستقلال

$$\chi^2 = \boxed{0.6848} \quad ; \quad d.f. = (r-1)(c-1) = (1)(1) = 1$$

قاعدة اتخاذ القرار:

إذا كان $\chi^2 > 3.841$ ارفض H_0

فإن لم يكن: لا ترفض H_0



لدينا هنا:

$\chi^2 = 0.6848 < 3.841$

ولذلك لا نرفض H_0

و نستنتج أن:

الجنس مستقل عن نوع اليد المفضلة