المملكة العربية السعودية

وزارة التعليم

MINISTRY OF EDUCATION

هام

لكل المهتمين و المهتمات

بدروس و مراجع الجامعية

مدونة المناهج السعودية eduschool40.blog

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

1

Ch. 1 - Part 1

- Introduction.
- Descriptive and inferential Statistics.
- Variables and Types of Data.
- Data Collection and Sampling Techniques.
- Observational and Experimental Studies.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 1 Part. 1

جمال السعدي

* Statistics:

العلم يربط الدراسات جمع تنظيم

Is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from the data.

تلخيص تحليل رسم النتائج

فروع علم الإحصاء
Branches of statistics

وصفي
Descriptive

Consists of:

- The collection,
- organization,
- summarization,
- Presentation of the data by the tables and graphs.

- بعض الكلمات التي تدل على هذا النوع.

Average – mean – median – mode.

- الحديث عن الماضي مثل عام 1996, 2000
- تقديم خبر – أو تقرير.

استدلالي
Inferential

Consists of:

- Generalizing from sampling to population.
- Performing estimation.
- Determining relations between variables and making prediction.

- بعض الكلمات التي تدل على هذا النوع.

Soon – Maybe – next – Can

- الحديث عن المستقبل مثل عام 2020 (التوقع).
- استخدام عينة للتعميم على المجتمع.

● **In each of these statements, tell whether descriptive or inferential statistics have been used :**

يلتحقوا

(a) In the year **2020**, 148 million Americans will be enrolled in an HMO. "العبارة تتحدث عن المستقبل" → **(Inferential)**

ضحايا العمل

(b) Nine out of ten on- the – job fatalities are men .
"العبارة بيها تقرير ناتج عن دراسة سابقة" → **(Descriptive)**

الإنفاق

(c) Expenditures for the cable industry were $ 5.66 billion in **1996**. "العبارة تتحدث عن الماضي" → **(Descriptive)**

الدخل القومي لعمال المنازل متوسط

(d) The **median** household income for people aged 25- 34 is $ 35.888. **(Descriptive)**

تم معرفة النسبة عن طريق عينة ثم تعميمها على المجتمع

(f) Drinking decaffeinated coffee **can** raise cholesterol levels by 7% . **(Inferential)**

متوسط الإنفاق على الأدوية للشخص الواحد

(g) The national **average** annual medicine expenditure per person is $ 1052 . **(Descriptive)**

خبراء الرهن العقاري

(h) Experts say that mortgage rates may **soon** hit bottom. **(Inferential)**

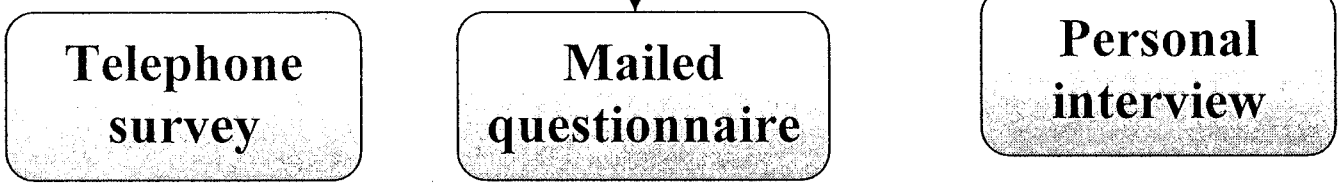
ALSAADI

● **Name and define the two areas of statistics :**

- Descriptive statistics: describe the data set. استنتاجات
- Inferential statistics: use the data to draw conclusions about the population.

● **Data Collection**

* جمع البيانات بثلاث طرق :
الهاتف
الايميل
المقابلة الشخصية



● * Population: A set consists of all subjects → المجتمع : هو المجموعة الشاملة لكل الأشياء .

Types Of Population

انواع المجتمع

مستويات الطلاب
* Example: Students grade
(A , B , C , D , F)

الأعداد الصحيحة
* Example: integer numbers
(...., -2, -1, 0, 1, 2, ...)

● • **Sample: is a group of subjects selected from population.**

• **Sample ⊂ population.**

العينة : هي مجموعة من الأشياء المختارة من المجتمع .
العينة جزء من المجتمع .

● **Identify the sample and population in each of the following statements:**

1. In order to study the response times for emergency 988 calls in Jeddah 50 calls are selected randomly over a six month period and the response times are recorded.

** Population: all calls (988).

** Sample: 50 calls.

---

2. **1500 listeners to talk radio program of various types are selected.**

** Population: all listeners to radio program.

** Sample: 1500 listeners.

● **Why we must use a sample Instead of population?**
لماذا نستخدم العينه بدلاً من المجتمع ؟

We must use a sample Instead of population because:
** نستخدم العينه بدلاً من المجتمع للأسباب الآتيه :

1. The size of population may be very large.
كبر حجم المجتمع .

2. Study the whole population may be very expensive.
دراسه المجتمع كاملاً ربما تكون أكثر تكلفه

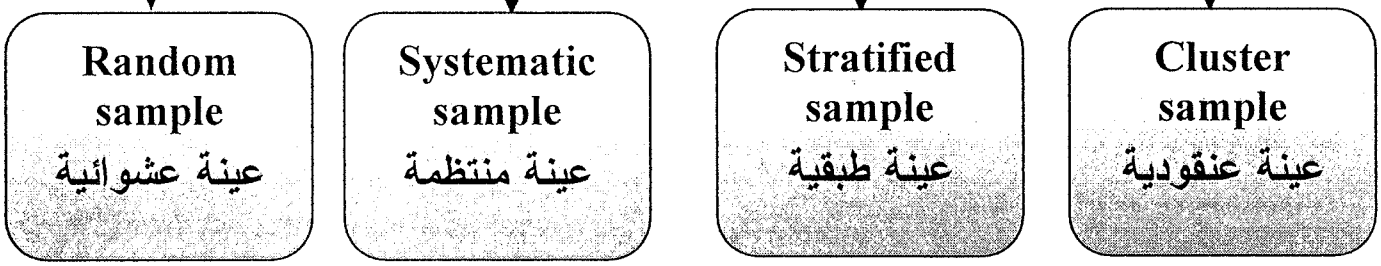
(Save Money)
توفير المال

3. Study the whole population may be need to a long time .
دراسه المجتمع ربما تحتاج لوقت طويل .

(Save Time)
توفير الوقت

4. Study the whole population may be destructive for the elements of population.
دراسه المجتمع ربما تلحق الأضرار بعناصر المجتمع
* مثال : تجربه عقار جديد
لا يصح تجربته على المجتمع بالكامل
لأنه ربما يلحق الضرر بعناصر المجتمع .
لذا يتم تجربته على عينه .

Type of Samples
انواع العينات



العينة العشوائية Random Sample:

* فيها تكون كل حالة من حالات المجتمع لها نفس فرصة الاختيار (اختيار عشوائي بدون شروط)

* All units of the population has the same chance of selecting.

العينة المنتظمة System sample:

فيها يقسم المجتمع إلى مجموعات عددها يساوي عدد أفراد العينة المختارة.

يتم اختيار حالة من المجموعة الأولى عشوائياً فإذا كان ترتيبها A مثلاً نختار الحالة التي ترتيبها A من كل المجموعات حتى نكون العينة المطلوبة.

بعض الكلمات = التي تدل عليه ← Seventh ..... , ...l**oo**th

العينة الطبقية Stratified sample:

فيها يقسم المجتمع إلى مجموعات غير متقاطعة (مستقلة) تسمى طبقات (غير متداخلة) ثم يتم اختيار عينة عشوائية بسيطة من كل طبقة.

العينة العنقودية Cluster sample:

فيها يكون المجتمع مقسم إلى مجموعات رئيسية

نختار من هذه المجموعات **بعض المجموعات** ثم نستخدم جميع عناصر المجموعات المختارة.

Classify each sample as random, systematic, stratified, or cluster.

مدرسة كبيرة مقسمة إلى مباني

(a) In a large school district, all teachers from two buildings are interviewed to determine whether they believe the students have less homework to do now than in previous years. السنوات السابقة

(Cluster)

- اختيار مبنيين من مباني المدرسة.
- ثم اختيار كل المدرسين من المبنيين.

مول تجاري

(b) Every seventh customer entering a shopping mall is asked to select her or his favorite store. المحل المفضل

(Systematic)

(c) Nursing supervisors are selected using **random** numbers in order to determine annual salaries. الرواتب السنوية

(Random)

الرقم ١٠٠ و مضاعفاته

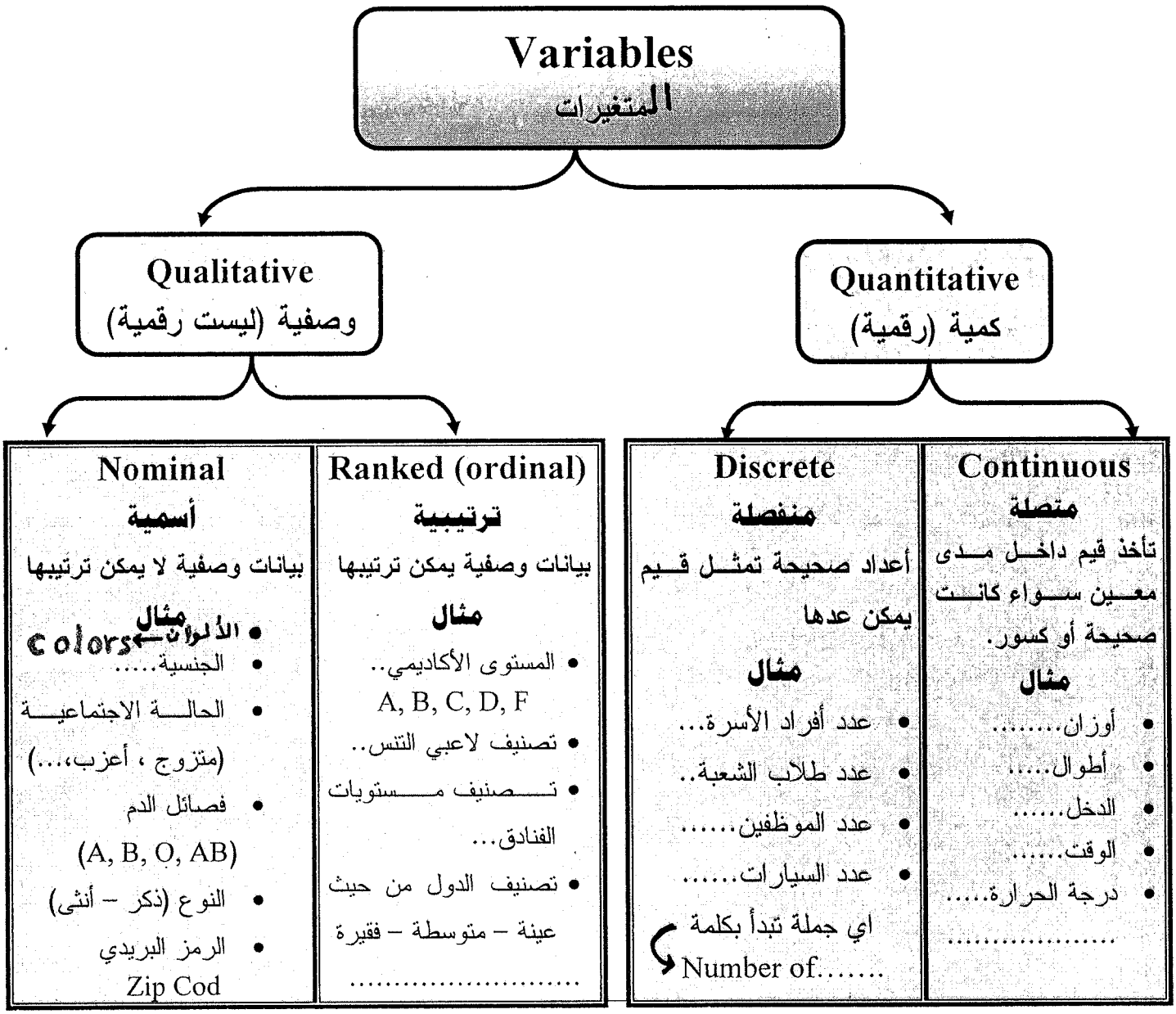
(d) Every 100^{th} hamburger **manufactured** is checked to determine its fat content. مرضية

(Systematic)

سعاة البريد

(e) Mail carriers of a large city are divided into four groups according to gender **(male or female)** and according to whether they walk يمشون or ride يركبون on their routes. Then 10 are selected from each group and interviewed مقابلتهم to determine whether they have been bitten by a dog in the last year.

(Stratified) قسموا إلى ٤ مجموعات طبقاً للنوع
تم اختيار ١٠ من كل مجموعة


Variables
المتغيرات
Qualitative
وصفية (ليست رقمية)
Quantitative
كمية (رقمية)
Nominal
أسمية
بيانات وصفية لا يمكن ترتيبها
مثال
الألوان ← Colors
الجنسية.....
الحالة الاجتماعية (متزوج ، أعزب،...)
فصائل الدم (A, B, O, AB)
النوع (ذكر – أنثى)
الرمز البريدي Zip Cod
Ranked (ordinal)
ترتيبية
بيانات وصفية يمكن ترتيبها
مثال
المستوى الأكاديمي.. A, B, C, D, F
تصنيف لاعبي التنس..
تصنيف مستويات الفنادق...
تصنيف الدول من حيث عينة – متوسطة – فقيرة
.....
Discrete
منفصلة
أعداد صحيحة تمثل قيم يمكن عدها
مثال
عدد أفراد الأسرة...
عدد طلاب الشعبة..
عدد الموظفين.....
عدد السيارات.....
اي جملة تبدأ بكلمة Number of.....
Continuous
متصلة
تأخذ قيم داخل مدى معين سواء كانت صحيحة أو كسور.
مثال
أوزان.....
أطوال.....
الدخل.....
الوقت.....
درجة الحرارة.....
.....

*** Variables** : Is quantity can taken different values. → الكميه التى تأخذ قيم مختلفه

Example : length , weight, age, coller, time,.....

فئات أصناف

*** Qualitative** : The variables expressing by categories or classes.

Example: * Gender (Male, Female) → category.

* Classes (2-5, 6-9, 10-13) → classes.

*** Qualitative variables are two types**

ⓐ **Nominal Variables:**

Gives names in which there is no order.

Example: * Types of blood.

* Name your country.

ⓑ **Ranked (ordinal)** Variables:

تصنيف أصناف ترتيبها
Classifies variables into categories that can be ranked.

Example: * Academic level. A, B, ... ← المستوى الأكاديمى

* Level of the Hotel. *, **, ...,*****

ALSAADI

المتغيرات الكمية : تأخذ قيم عددية على مقياس معين ويمكن ترتيبها .

Quantitative It is the variables which takes numerical values with measure scale and can be ordered or ranked.

Example: Age, height, weight,.....

*** Quantitative variables are two types**

ⓐ **Discrete Variables** are can be count. يمكن عدها وتأخذ قيم صحيحه .

Example: * Number of student in class.

* Number of cars in park.

ⓑ **Continuous variables:**

It is the variable which can take all possible numerical values in a given interval. تأخذ كل القيم العددية صحيحه وكسور داخل فتره معينه .

Example: * Height of a student.

- Weight of a student.

صنّف كمي وصفي

Classify each variable as qualitative or quantitative.

(a.) Number of bicycles sold in 1 year by a large sporting goods Store.

ألوان القبعات لكرة المضرب في متجر

** (b.) Colors of baseball caps in a store.

الزمن المستغرق لقطع العشب

(c.) Times it takes to cut a lawn.

السعة بالقدم المكعب لـ ٦ أسـرّة أطفال

(d.) Capacity in cubic feet of six truck beds.

رضيع

** (e.) Classification of children in a day-care center (infant – toddler, preschool).

روضة يحبو

بحيرة جورج

(f.) Weights of fish caught in Lake George.

الحالة الاجتماعية لأعضاء كلية في جامعة كبيرة

** (g.) Marital statues of faculty members in a large university.

Solution

(a) , (c) , (d) and (f) are Quantitative. كمي

(b) (e) and (g) are Qualitative. وصفي

صنف منفصل متصل

Classify each variable as discrete or continuous.

الدونات

** (a.) Number of doughnuts sold each day by doughnut Heaven.

درجة حرارة ست حمامات سباحة

(b.) Water temperatures of six swimming pools in Pittsburgh on a given day.

ملجأ للحيوانات

(c.) Weights of cats in a pet shelter.

العمر الافتراضي بالساعات لعدد ١٢ حجر بطارية

(d.) Lifetime (in hours) of 12 flashlight batteries.

** (e.) Number of cheeseburgers sold each day by a hamburger stand on a college campus.

المؤجرة

** (f.) Number of DVDs rented each day by a video store.

السعة بالجالونات لعدد ٦ خزانات ماء في مقاطعة جيفر سون

(g.) Capacity (in gallons) of six reservoirs in Jefferson County.

← مقاطعة جيفرسون

Solution

a , e , and f are Discrete. كلها جمل تبدأ بـ Number of

b , c , d and g are Continuous.

Level Measurement Of The Data

- تقاس البيانات بأربعة أنواع من المقاييس

① Nominal level data

المقياس الأسمي

هي بيانات أسمية غير رقمية (غير متقاطعة) لكل مجموعة خصائص تميزها عن غيرها.

Example: * blood types (A, B,O, AB).

* gender (male, female)

② Ordinal level data

المقياس الترتيبي

هي بيانات مقسمة إلى مستويات أو فئات يمكن ترتيبها تصاعدياً أو تنازلياً.

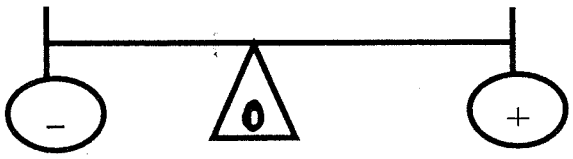
Example: * Grade of course (A, B, C, D. F)

* Rating scale (Poor, Fair, Good, Excellent).

* Ranking of tennis players. ← تصنيف لاعبي التنس

③ Interval level data

مقياس الفترة



هي بيانات رقمية تقاس بمقدار بعدها عن الصفر

Example: * Temperature

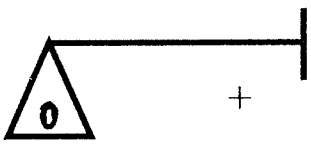
* I Q. test.

ممكن موجب أو سالب أو صفر

الصفر لا يعني انعدام الظاهرة

④ Ratio level data

مقياس النسبة



Example: * height , weight, time Salary, Age,.....

بدايته صفر: أي الصفر له معنى وهو انعدام الظاهرة ويكون موجب ولا يكون سالب

Classify each as nominal – level, ordinal - level, interval level, or ratio – level measurement.

صفحات دليل التليفونات

(a) Pages in the city of Cleveland telephone book. **(Ratio)**

صفحات دليل التليفون تبدأ من ١ إلى ....

أي ليست سالبة ولا تبدأ من Zero

(b) Rankings of tennis players. **(Ordinal)**

(c) Weights of air conditioners. **(Ratio)**

ثلاجات

(d) Temperatures inside 10 refrigerators. **(Interval)**

(e) Salaries of the top five CEO in the United States. **(Ratio)**

ترتيب مسرحيات محلية رديئة مقبولة جيدة ممتازة

(f) Ratings of eight local plays (Poor, Fair, Good, Excellent)

(Ordinal)

(g) Times required for mechanics to do a trune-up **(Ratio)**

(h) Ages of students in a classroom. **(Ratio)**

الحالة الاجتماعية المرضى المتردد على عيادة الطبيب

(i) Marital status of patients in a physician's office. **(Nominal)**

القدرة الجرار محرك

(j) Horsepower of tractor engines. **(Ratio)**

Observational and experimental studies

Observational study
دراسة مشاهدية

The researcher

Observe only and not effect

ملاحظة الحدث دون تدخل للتأثير فى النتائج

- مثال: ملاحظة عدد المرضى الذين تم شفاؤهم من مرض معين.
- كلمات تدل على هذا النوع .....Find – See .....

Experimental Study
دراسه تجريبيه

The researcher

Effect and observe.

يحدث تدخل من الباحث فى الظاهره محل الدراسه ثم ملاحظه مدى تأثير هذا التدخل على نتائج الظاهره.

* **مثال:** اعطاء المرضى عقار معين ثم ملاحظة تأثير هذا العقار على عدد من تم شفاءهم.

* كلمات تدل على هذا النوع .....Placed on - given....

ALSAADI

Identify each study as being either observational or experimental:

(a.) Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After 6 months, the numbers of respiratory tract infections each group had were compared.

تم التدخل فى الدراسه وذلك بإعطاء نوع معين من العلاج لكل مجموعه

(Experimental)

(b.) A researcher stood at a busy intersection to see if the color of the automobile that a person drives is related to running red lights.

to see يشاهد فقط بدون تدخل

(Observational)

(c.) A researcher find that people who are more hostile have higher total cholesterol levels than who are less hostile.

* find بدون تدخل

(Observational)

(d.) Subjects are randomly assigned to four groups. Each group is placed on one of four special diets – a low- fat diet, a high- fish diet, a combination of low – fat diet, and a regular diet. After 6 months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure.

تم التدخل فى دراسة " تأثير الرجيم على ضغط الدم"

Each group is placed on diets

(Experimental)

$2x + 1 = y$

*** Independent** متغير مستقل X : (مؤثر)

هو متغير محل اهتمام الباحث (موضوع الدراسه)

Another name: **Explanatory variable**

Example: * number of study hours

* Room temperature

*** Dependent variable** متغير غير مستقل (تابع) y (متأثر)

هو ناتج العمليات التى أجريت على المتغير المستقل

Another name: **Outcome variable**

Example: * Student score.

* Bacteria growth.

*** Confounding**

تعريف

متغير لا يمكن فصله عن المتغيـــر المـــستقل

ويؤثر على النتائج يتداخل مع متغيرات أخرى

Interferes With other variables

مثال : ممارسه الرياضه (متغير مستقل)

+

الامتناع عن التدخين

⇐ = تحسين الصحه (م. تابع)

ALSAADI

Identify the independent variable and the dependent: variable for each of the studies in last Exercise: من التمرين السابع

(a-) Independent var. : " type of pill " مؤثر — نوع قرص الدواء

Dependent var. : " number of infections " متأثر — عدد الاصابات

(b-) Independent var. : " color of automobile " مؤثر — لون السيارة

Dependent var. : " running red lights " متأثر — قطع الاشارة الحمراء

(c-) Independent var. : " level of hostility "مؤثر — مستوى العصبية

Dependent var. : "cholesterol level" متأثر — مستوى الكولسترول

(d-) Independent var. : " type of diet " مؤثر — نوع الحمية

Dependent var. : " blood pressure " متأثر — ضغط الدم

- Suggest possible confounding variable in last exercise: (Page16)

(A) Workplace of subjects , smoking habits, ...

(B) Gender , age, ...

(C) Diet, type of job, ...

(D) Exercise , heredity , age, ...

Summary ملخص

** The two major areas of statistics are **descriptive** and **inferential** (استدلالي وصفي)

** Descriptive statistics: includes the collection. Organization, summarization and presentation of data

** Inferential statistics: includes making inferences from samples to populations, estimations, determining relationships and making predictions. Inferential statistics: is based on probability theory. (نظرية الاحتمال)

** Since in most cases the populations under study are large, statisticians use subgroups called samples to get the necessary data for their studies. There are four basic methods used to obtain samples: **random, systematic**, **stratified** and **cluster.**

** Data can be classified as qualitative or quantitative. Quantitative data can be either discrete or continues, depending on the values they can assume. Data can also be measured by various scales (مقاييس مختلفه) the four basic levels of measurement are **nominal**, **ordinal**, **interval** and **ratio**.

** There are two basic types of statistical studies: observational studies and **experimental studies**.

when conducting **observational studies**, researchers observe what is happening or what has happened and then draw conclusions (استناداً) based on these observations. (المشاهدات)

ALSAADI

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

2

Ch. 1 - Part 2

- Chapter Quiz.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 1 Part. 2

Chapter Quiz

جمال السعدي

Determine whether each statement is true or false if the statement is false explain why.

1- Probability is used as a basis for inferential statistics (✓)

2- The height of president Lincoln is an example of variable (×)
طول الرئيس الأمريكي ثابت ← Constant

3- The highest level of measurement is the interval level (×)
← Ratio

4- When the population of college professors is divided into groups according to their rank (instructor, assistant professor. etc.) and then several are selected from each group to make up a sample, the sample is called cluster (×)
Stratified
تم التقسيم إلى طبقات مختلفة حسب الرتبة ← Stratified

5- The variable age is an example of a qualitative variable (×)
← quantitative

6- The weight of pumpkins is considered be a continuous variable (✓)
اليقطين

7- The boundary of a value such as 6 inches would be 5.9-6.1 inches (×)
5.5 – 6.5

Select the best answer.

8- The number of absences (غياب) per year that a worker has is an example of what type of data? عدد أيام غياب العامل سنوياً

a. nominal

b. Qualitative

c. Discrete

d. Continuous

مثيل متغير منفصل
لأنه عدد الأيام دائماً اعداد صحيحه
Discrete

9- **What are the boundaries of 25.6 ounces?**

a. 25-26 ounces

b. 25.55-25.65 ounces

c. 25.5 – 25.7 ounces

d. 20 – 39 ounces

25.6
25.55 – 25.65

10- **A researcher divided subjects into two groups according to gender (لنوع) and then selected members from each group for her sample. What sampling method was the researcher using?**

a. Cluster

b. Random

c. Systematic

d. Stratified

التقسيم طبقاً للنوع
∴ طبقية
Stratified

11- **Data that can be classified according to color are measured on what scale?**

a. Nominal

b. Ratio

c. Ordinal

d. Interval

طبقاً للون
∴ البيانات
اسمية Nominal

ALSAADI

يتضمن - يشمل

12- A study that involves no researcher intervention is called

يغير نتيجه

a. An experimental study

b. A noninvolvement stud

c. An observational study

d. A quasi – experimental study

13- A variable that interferes with other variables in the study is called

يتداخل

تعريف

a. An confounding variable.

b. An explanatory variable.

c. An outcome variable.

d. An interfering variable.

Use the best answer to complete these statements.

14- Two major branches of statistics are **descriptive** and **Inferential.**

15- Two uses of probability are **Gambling** and Insurance.

16- The group of all subjects under study is called a(n) **population.**

17- A group of all subjects selected from the group of all subjects under study is called (n) **sample.**

18- Three reasons why samples are used in statistics are

* ثلاث اسباب لاستخدام العينات

a. save time **b. save money** **c. when population is large.**

ALSAADI

19- The four basic sampling methods are....

a. Random **b. systematic** **c. stratified** **d. cluster**

20- A study that uses intact groups when it is not possible to (شبه) randomly assign participants to the groups is called **Quasi** (تجريبى) **Experimental** study.

21- In a research study, participants (المشاركين) should be assigned to groups **random** methods, if possible.

22- **For each statement decide whether descriptive or inferential statistics is used.**

a. **The average** (المتوقع) life expectancy in New Zealand is 78.49 years.

(descriptive).

b. A diet high in fruits and vegetables will lower blood pressure.

(Inferential)

c. the total amount of estimated (الخسارة المقدرة) losses (كمية) from hurricane Hougo (اعصار) was $ 4.2 billion.

(Descriptive)

d. Researchers stated that the shape (شكل هيئه) of a person's ear (أذن الشخص) is related (مرتبط بـ) to the person's aggression (بمدى انحراف الشخص).

(Inferential) ... يستدل بشيء على شيء آخر

e. In 2020, the number of high school graduates will be 3.2 million students. (خريجين)

(Inferential)

23- Classify each as nominal level, ordinal level, or ratio level measurement

a. Rating of movies as PG. and R **(nominal)**

b. Number of candy bars sold on a fund drive. (بالماكينة المباعة اعمدة السكر (قوالب الحلوى) **(ratio)**

c. Classification of automobiles as subcompact, compact, (سيارة صغيرة) standard, and luxury. (عادية فخمة) **(ordinal)**

d. Temperatures of hair dryers. (مجفف الشعر) **(Interval).**

e. Weights of suitcases on a commercial airline. (خطوط تجاريه) **(ratio)**

24- Classify each variable as discrete or continues.

a. Ages of people working in a large factory. (أعمار العمال) **(continuous)**

b. Number of cups of coffee served at a restaurant. (المقدمه مطعم)

(discrete)

c. The amount of drug injected into a guinea pig (المحقونه المخدر كمية)

(continuous)

d. The time it takes a student to drive to school. **(continuous)**

e. The number of gallons of milk sold each day at a grocery store.

(discrete)

ALSAADI

طريقة الحصول على Boundaries

- لابد وزن الفاصلة العشرية في class limits .
- نضع الرقم 5 في نهاية الـ upper limit.
- يطرح ١ من الرقم الأخير في lower limit ثم نضع الرقم 5.

What are the boundaries of the class limits 8.2 - 10.4?
A) 8.25 - 10.45 B) 8.15 - 10.45 C) 8.15 - 10.35 D) 8.25 - 10.35

طريقة سهلة جداً

class limits : 8.2 — 10.4

* نطرح من الرقم الأخير 1 ثم نضع 5
* نضيف 5 بعد الرقم الأخير

⇒ boundaries : 8.15 — 10.45

Find the boundaries:

6.1 - 8.32 ⟶

* أولاً وزن الفاصلة
6.10 — 8.32
* ثانياً ايجاد boundaries

6.095 — 8.325

(25) Give the boundaries of each.

a. 48 seconds ⟶ 47.5 – 48.5
b. 0.56 Centimeter ⟶ 0.555 – 0.565
c. 9.1 quarts ⟶ 9.05 – 9.15
d. 13.7 pounds ⟶ 13.65 – 13.75
e. 7 feet ⟶ 6.5 – 7.5

Which one of the following is an example of qualitative variable?

A) Grade point average (GPA) B) Temperature C) Nationality D) Age

Nationality ⟶ Qualitative variable
الجنسية متغير نوعي (وصفي)

The process of selecting every sixth customer entering a shopping center is called ...
A) systematic sampling B) cluster sampling C) random sampling D) stratified sampling

Sixth ⟶ systematic sampling

A researcher divided subjects into two groups according to nationality (Saudi and non Saudi) and then she selected members from each group for her sample. What sampling method was the researcher using

A) Cluster B) Systematic C) Convenience D) Stratified

* التقسيم تم طبقًا للجنسية وهما طبقات مختلفة

Saudi and non Saudi
طبقات مختلفة ⟶ stratified.

جزء
A portion of a population is called a ...
A) tally. B) frequency distribution. C) sample. D) random survey.

* الجزء من المجتمع يسمى العينة .

A high school counselor selected randomly a group of students from each class to form a sample for an experimental study; this is an example of a ... sample.
A) systematic B) cluster C) stratified D) random

* الاختيار عشوائي ← random

اقسام
What type of sampling is being used if a university is divided into departments and a random sample is chosen from each departments to be surveyed?
A) Stratified sampling.
B) Random sampling.
C) Systematic sampling.
D) Cluster sampling.

جامعه مقسمه إلى أقسام :
* التقسيم إلى اقسام مختلفه (طبقات)
ثم الاختيار من هذه الأقسام لتكوين العينه
↳ Stratified sampling

In a study it has been reported that a diet high in fruits and vegetables will lower blood pressure. The outcome variable is ...
A) diet. B) vegetables. C) blood pressure. D) fruits.

* المتغير المتأثر هو ضغط الدم يتأثر بنوع الحمية
↳ outcome variable is blood pressure

A researcher needs to conduct a(n) ... study to see if the number of people in a car is related to running a red lights.
A) observational B) manipulative C) experimental D) independent

To see → Observational study.

If we have measured the height of 20 students and compute their average, this will be an example of ...
A) inferential statistics.
B) descriptive statistics.
C) population statistics.
D) predictive statistics.

average height of 20 ---→ descriptive stat.

ALSAADI

Which of the following is an example of a continuous variable?

A) Students identification number.
B) Students heights.
C) Ranking of baseball teams in a league.
D) Number of students in a statistics course

أطوال الطلاب تمثل متغير متصل.

Students heights ⟶ continuous variable.

Students' age is an example of which level of measurement?

A) Nominal B) Ordinal C) Discrete D) Continuous

age ⟶ continuous متصل

Nursing Supervisors are selected using random numbers in order to determine annual salaries. This is an example of

A) random sampling B) systematic sampling C) cluster sampling D) stratified sampling

using random numbers ⟶ random sampling

A high school counselor uses a computer to generate 50 random numbers and then picks students whose names correspond to the numbers; this is an example of ... sample.

A) stratified B) cluster C) random D) systematic

* استخدام الكمبيوتر لإيجاد 50 رقم عشوائي ثم انتقاء (أختيار) الطلاب التي اسمائهم مناظره لـ 50 رقم.

هذا مثال للعينه العشوائيه ← random

What is a portion of a population called?
A) Sample B) Tally. C) Frequency distribution. D) Random survey.

* الجزء المحدود من المجتمع ماذا يسمى ؟
العينة هي جزء من المجتمع
Sample

Nationality represents a
A) continuous variable B) quantitative variable C) discrete variable D) qualitative variable

الجنسية تمثل متغير وصفي

The number of high school graduates will be 3.2 million students in the year 2020; this is an example of ... statistics.
A) descriptive B) empirical C) inferential D) traditional

تاريخ يدل على المستقبل

∴ inferential statistics.

The number of high school students will exceed 499999 students in ~~2010~~ 2012. This statement is an example of a (an) ..... statistics
A) quantitative B) qualitative C) inferential D) descriptive

will سوف (الحديث عن المستقبل) → inferential

The average life expectancy in New Zealand is 78.49 years. This statement is an example of a (an) ..... statistics
A) inferential B) qualitative C) descriptive D) quantitative

average → descriptive

Researchers used ... statistics to state that "The shape of a person's ears is related to the person's aggression."
A) qualitative B) quantitative C) descriptive D) inferential

* Inferential
aggression على The shape لأنه يستدل بـ

What type of sampling is being used if a company is divided into sections and a random sample from the sections is chosen and all the employees within the chosen sections are surveyed?

A) Cluster sampling.
B) Stratified sampling.
C) Random sampling.
D) Systematic sampling.

* العينه نختاره من اقسام مختلفه من الشركه (طبقات مختلفه)
↳ Stratified sampling

The process of counting the number of students in each room of building 44 at a specific time is called a (an) ...

A) experimental study. B) count study. C) dependent study. D) observational study.

* عمليه تحديد عدد الطلاب بدون تدخل للتأثير على هذا العدد
↳ Observational study.

A (An) ... study need to be conducted if a researcher would like to compare the effectiveness of three types of medicines.

تأثيرات مقارنه

A) observational B) manipulative C) experimental D) independent

* عمليه مقارنه تأثير انواع مختلفه من الدواء (حدث تدخل بتحديد نوع الدواء)
↳ experimental study

Temperature is an example of what level of measurement?

A) Interval. B) Ratio. C) Nominal. D) Ordinal.

Temperature ⟶ interval level of measurement.

The color of the cars that are parked along a street, is an example of which level of measurement?

A) Ordinal. B) Nominal. C) Ratio. D) Interval.

* الألوان متغير أسمى

The color of cars ⟶ Nominal level of measurement.

If a variable can take any value between 0 and 20, then this variable is ...

A) continuous and qualitative. B) discrete and quantitative. C) continuous. D) discrete.

* المتغير الذي يأخذ أي قيمه (عدد صحيح أو كسر) بين 0 ، 20
هو متغير متصل

any value ⟶ Continuous.

ALSAADI

"A variable that interferes with other variables in the study is called a(n) ...variable."
A) confounding B) interfering C) outcome D) explanatory

المتغير المتداخل : هو المتغير الذي يتداخل مع متغيرات أخرى.
تعريف (يحفظ)

The score on an IQ test is an example of which level of measurement
A) nominal B) ordinal C) interval D) ratio

IQ test اختبار الذكاء → Interval.

A researcher stood at a busy intersection to see if the color of the automobile that a person drives is related to running red lights. The type of study used here is ... study.
A) experimental B) convenience C) observational D) quasi-experimental

to see كلمة
تدل على أنه الباحث شاهد ولم يؤثر
↓
observational

السيارات تصنيف
Classification of automobiles as Nissan, Mercedes, Lexus, and Honda is an example of ... data.
A) ordinal B) continuous C) discrete D) nominal

* تصنيف السيارات :

نيسان ، مرسيدس ، لكزس ، هوندا

بيانات اسمية ← nominal

All subjects that are being studied is called ...
A) population. B) sample. C) parameter. D) statistic.

* All subjects ⟶ population .

A ... is a group of subjects that are being studied.
A) statistic B) population C) parameter D) sample

* group of subjects ⟶ sample.

طبقاً لـ
Classification of employees in a company according to their employment levels is an example of ... data.
A) ordinal B) discrete C) nominal D) continuous
تصنيف الموظفين في شركه طبقاً لمستويا تهم الوظيفيه (طبقات مرتبه) .

* employment levels ⟶ ordinal

Capacity of six dams in Saudi Arabia is an example of a(n) ... variable.
A) ordinal B) nominal C) continuous D) discrete

* Capacity هـ ⟶ continuous

Temperatures inside 9 refrigerators is an example of what level of measurement?
A) Ordinal. B) Nominal. C) Interval. D) Ratio.

* Temperatures (درجات الحراره) ⟶ Interval level of measurement.

The number of trees in a garden is an example of a(n) ... variable.
A) continuous B) discrete C) nominal D) ordinal

عدد الأشجار في الحديقه مثال للمتغير المنفصل .

* number of trees ⟶ discrete variable.

The average shoe size of Saudi's students is an example of which level of measurement?
A) Ratio B) Interval C) Nominal D) Ordinal

The average shoe size ---
is an example of Interval level measurement

يُعلم (يدرس)

Subjects were assigned randomly to two groups for an educational study. One group was taught using the traditional method and the other group using a new method. After 3 years, the students performance of the groups were compared.

What is the type of this study?
A) Experimental study.
B) Quasi-Experimental study.
C) Convoluted study.
D) Observational study.

(تحديد نوع الطريقه يعني تدخل من الباحث) * التقسيم إلى مجموعتين :
التعليم من احدهما بالطرق التقليديه و من الأخرى بالطرق الحديثه
بعد ثلاث سنين تم مقارنه مدى استفاده
الطلاب من المجموعتين ← دراسه تجريبيه .

What is the dependent variable in this study?
A) Method of teaching.
B) Period of study.
C) Students performance.
D) Number of groups.

* المتغير التابع (المتأثر) هو مدى استفاده الطلاب
Students performance.

السرعه عدد الحوادث

In the relationship between the number of accidents and speed, the speed is assumed to be ...
A) qulaitative variable B) independent variable C) dependent variable D) nominal variable

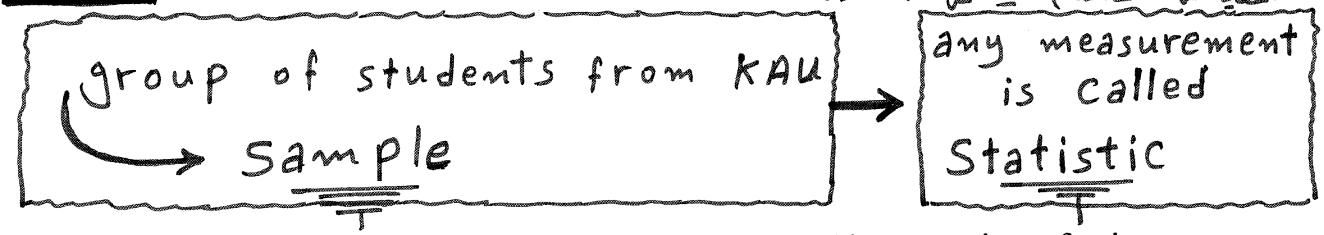
The speed (المؤثر)

is independent Variable متغير مستقل

When a study is conducted on a group of students from KAU, every measurement calculated for this study would be called a ...

A) statistic. B) sample. C) population. D) parameter.

هذه المجموعة تمثل عينة ∴ المقياس المستخدم يسمى احصاءه



If we have calculated the average height of 200 students and used it as an estimate for the average height for all students, this will be an example of ... statistics.

A) inferential B) population C) descriptive D) predictive

استخدام عينة للتعميم على المجتمع ← (احصاء استدلالي)

* استخدام عينة مكونة من أطوال 200 طالب للتنبؤ (تخمين) بمتوسط أطوال كل الطلاب ← Inferential

The average shoe size of Saudi's students is an example of which level of measurement?

A) Ratio B) Interval C) Nominal D) Ordinal

average shoe size → Interval

In a study it has been reported that a diet high in fruits and vegetables will lower blood pressure. The dependent variable is ...

A) blood pressure. B) fruits. C) vegetables. D) diet.

* المتغير التابع هو ضغط الدم dependent → blood pressure

* المتغير المستقل هو الحمية (الرجيم) independent → diet

If the values of a variable are determined randomly, then this variable is called ...

A) constants variables. B) independent variables. C) intact variables. D) random variables.

* المتغير العشوائي: هو المتغير التي تحدث قيمه عشوائيًا (بالصدفه).

The variable that is resulted (الناتج) from an experimental study is called ... variable.

A) dependent B) independent C) confounding D) explanatory

* المتغير الناتج (المحصل) من دراسه تجريبيه يسمى متغير تابع → dependent

ALSAADI

Use the best answer to complete these statements.

- Majors of high school students ➡ Nominal

* تخصصات الطلاب في المدارس الثانوية تمثل بيانات أسمية .

- Amount of fat (in grams) in 6 Cookies ➡ Ratio

* كمية الدهون بالجرامات في ....

- Rating of hotels by a number of stars ➡ Ordinal

* تصنيف الفنادق بعدد نجومها تمثل بيانات ترتيبية .

- Study that involves **No** researcher intervention (يغير النتيجة) is called ➡ Observational study

* الدراسة التي تتضمن عدم تدخل الباحث ... (دراسة شاهدية)

- One advantage of Observational study is that it occurs in a natural setting.

* احد ميزات الدراسة الشاهدية انها تتم في ظروف طبيعية

- Performing estimations and hypothesis tests are

example of : ➡ Descriptive statistics

* تقديم التقديرات واختبارات الفروض مثال للإحصاء الوصفي.

- If a variable has the possible values:

–2 , 4 , 6 , 10 and 12

Then this variable is ➡ Discrete variable

* المتغير الذي قيمه الممكنه ..... (اعداد صحيحه ليس بينها كسور)
مثال للمتغير المنفصل .

- Temperature , IQ test and average shoe size

are ➡ Interval level data

* درجه الحراره ، اختبار الذكاء ، متوسط مقاس الحذاء ...
كلها امثله لـ مقاييس الفتره .

- Height , weight , time , age , salary and number of ...

are ➡ Ratio level data

* الطول ، الوزن ، الوقت ، العمر ، الراتب ، عدد ...
كلها امثله لـ مقاييس النسبه .

تعريفات مهمة

- Probability is chance of an event occurring.
- Probability deals (يتعامل) more with creating models and theoretical data.
- while statistics deals more with applying models and real data.

- Data are the values of variables can assume.
- Each value in the data set is called a data value or a datum.
- A data set is a collection of data values.

Statistical studies usually include one or more independent (متغير مستقل أو أكثر) variables and one dependent (متغير واحد تابع) variable.

- Inferential statistics is based on probability theory (نظرية الاحتمالات).
- The two major areas of statistics are descriptive and inferential.

Observational study: the researcher observes what is happening or what has happened and tries to draw conclusions based on these observations.

Experimental study: the researcher manipulates (يؤثر في) one of the variables and tries to determine how that influences (تأثير) other variables.

Random samples: are selected using chance methods or random methods.

Systematic samples: numbering each subject of the populations and then selecting every $\underline{K^{th}}$ number.

Stratified samples: dividing the population into groups called strata according to some characteristic (صفه مميزه) that is important to the study, then sampling from each group or strata.

Cluster samples: intact groups called clusters. Thus, dividing the population into groups and then taking samples of the groups.

The independent variable or explanatory variable is the one that is being manipulated by the researcher. (المتغير المؤثر (المستقل) هو المتغير المؤثر به من طريقه الباحث)

When the populations to be studied are large. Statisticians use subgroups called samples. (عندما يكون المجتمع المراد دراسته كبير الاحصائين يستخدمون العينات)

In an experimental study, the subjects should be assigned to groups randomly. If this is not possible, then it is called a: quasi- experimental study.

Descriptive statistics ; consists of the collection, organization, summarization, and presentation of data.

Inferential statistics : consists of generalizing from samples to populations, performing estimations and hypothesis testing, determining relationships among variables, and making predictions.

A variable is a characteristic or attribute that can assume different values.

* المتغير : هو الخاصيه أو الصفه التي تأخذ قيم مختلفه.

Random variable have values that are determined by chance.

* المتغير العشوائي : هو المتغير الذي تحدث قيمه بالصدفه .

إنتهى Ch.1

كل التمنيات بالنجاح والتوفيق

وأخيراً
أدعو الله أن يتقبل هذا العمل
ويكون فيه النفع للجميع ....
جمال السعدي

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

3

Ch. 2 - Part 1

- Organizing Data.
- Histogram , frequency Polygons, and Ogives.

STAT.110

جمال السعدي

رياضيات - إحصاء

CH. 2 PART 1

جمال السعدي

Frequency distributions and graphs → التوزيعات التكرارية والرسومات

Organizing data: for the raw data → تنظيم البيانات للبيانات الخام

When data are collected in original form, they are called raw data.

* عندما تجمع البيانات بشكلها الأصلي تسمى البيانات الخام .

When the raw data are organized into a table which is called frequency distribution, the frequency will be the number of values in a specific class of the distribution.

* عندما تنظم البيانات في جدول يسمى التوزيع التكراري
* التكرار : هو عدد القيم داخل فئة محددة في التوزيع .

A frequency distribution is the organization of raw data in table form, using classes and frequencies.

* التوزيع التكراري : هو تنظيم البيانات الخام في جدول باستخدام الفئات والتكرار .

Types of frequency distributions are categorical frequency distribution, ungrouped frequency distribution and grouped frequency distribution.

* أنواع التوزيعات التكرارية

توزيعات تكرار نوعيه (ليست رقميه)

(1) Categorical frequency distributions:

Is used for: nominal data or ordinal data

Example:

Twenty – five army inductees were given a blood test to determine their blood type. The data set is

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

Construct a frequency distribution for the data

Solution:

Class	Tally	Frequency F	Percent P
A	𝍸	5	20
B	𝍸 //	7	28
O	𝍸 ////	9	36
AB	////	4	16
		$n = \sum F = 25$	100

$$P = \frac{F}{n} .100\%$$

- For the sample, people have type O blood more than any other type

التوزيعات التكراريه المجمعه

(2) Grouped frequency distributions (للبيانات الكميه)

Is used when the range of the data is large.

The data must be grouped into classes.

- The width of the class must be > 1
- The number of the classes should be between 5 and 20

Example:

These data represent the record high temperatures in each 50 states construct a grouped frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Solution

* **Range = H –L = 134 – 100 = 34**

* Width = $\frac{R}{\text{Number of classes}} = \frac{34}{7} = 4.9$

Rounded up = 5

Class limits	Class boundaries	Tally	Frequency	Cumulative frequency
100 -104	99.5 – 104.5	//	2	2
105 – 109	104.5 – 109.5	𝍸///	8	10
110 – 114	109.5 – 114.5	𝍸𝍸𝍸///	18	28
115 – 119	114.5 – 119.5	𝍸𝍸///	13	41
120 – 124	119.5 – 124.5	𝍸//	7	48
125 – 129	124.5 -129.5	/	1	49
130 - 134	129.5 – 134.5	/	1	50

$n = \sum f = 50$

ALSAADI

Number of classes can be found by suing the formula:

Number of classes $\approx 1 + 3.3 \times \log(n)$ ← قانون

Where n is sample size.

Example:

If sample size $n = 50$, Find the number of classes.

Solution

Number of classes $\approx 1 + 3.3 \times \log(n)$

$\approx 1 + 3.3 \times \log(50)$ بالآلة

≈ 6.6066

$\approx \boxed{7}$

(3) Ungrouped frequency distributions:

are used for data that can be enumerated and when the range of values in the data set is small (discrete data)

and the sample size (n) is large.

- Example: * number of children per family,

  * number of cars in a parking.

- Example: Number of patients in the waiting rooms of 16 clinics within a hospital at a specific.

Represent the flowing data by using:

Ungrouped frequency distributions

5	4	4	8
8	5	8	4
4	4	8	4
5	8	4	4

Solution

Class	Frequency	Cumulative Frequency	Percent
4	8	8	50%
5	3	11	19%
8	5	16	31%
Total	16	-	100%

Shown here are four frequency distributions each is incorrectly constructed state the reason why.

* توزيعات تكراريه
منشأه بطريقه خاطئه
وضح سبب الخطأ .

a.

class	Frequency	
27 -32	1	6
33 – 38	0	6
39 – 44	6	6
45 - 49	4	5
50 - 55	2	

غير متساوى

⟹ class width is not uniform

---

b.

class	Frequency
5- 9	1
9 – 13	2
13 – 17	5
17 – 20	6
20 - 24	3

متداخله

⟹ class limits overlap and class width is not uniform

---

c.

class	Frequency
123 - 127	3
128 - 132	7
138 - 142	2
143 - 147	19

فئه ناقصه

⟹ A class has been omitted

---

d.

class	Frequency	
9 - 13	1	5
14 - 19	6	6
20 – 25	2	6
26 - 28	5	3
29 - 32	9	

غير متساوى

⟹ class width is not uniform

Note

طريقة الحصول على Boundaries

لابد وزن الفاصلة العشرية.

نضع الرقم 5 في نهاية الـ upper limit.

يطرح ١ من الرقم الأخير في lower limit ثم نضع الرقم 5.

- class width = upper boundary – Lower boundary

- class midpoint x_m = $\dfrac{\text{Lower bou. + upper bou.}}{2}$

$$= \frac{\text{Lower limit + upper limit}}{2}$$

What are the boundaries of the class limits 8.2 - 10.4?
A) 8.25 - 10.45 B) 8.15 - 10.45 C) 8.15 - 10.35 D) 8.25 - 10.35

طريقة سهلة جداً

class limits : 8.2 — 10.4

* نطرح من الرقم الأخير 1 ثم نضع 5
* نضيف 5 بعد الرقم الأخير

⇒ boundaries : 8.15 — 10.45

حدود الفئة نقطة المنتصف العرض

Find: the class boundaries, Midpoints and width for each class:

(1) 12 -18 (2) 13.6 – 14.7 (3) 2.15 – 3.93

Solution

(1) 12 -18

* Class boundaries

11.5 – 18.5

* Midpoint = $\frac{12 + 18}{2} = \frac{30}{20} = 15$

* Class width = upper boundary – lower boundary = 18.5 – 11.5 = 7

(2) 13.6 – 14.7

* Class boundaries

13.55 – 14.75

* Midpoint = $\frac{13.6 + 14.7}{2} = 14.15$

* Class width = upper boundary – lower boundary = 14.75 – 13.55 = 1.2

(3) 2.15 – 3.93

* Class boundaries

2.145 – 3.935

* Midpoint = $\frac{2.15 + 3.93}{2} = 3.04$

* Class width = upper boundary – lower boundary = 3.935 – 2.145 = 1.79

Graphs الرسومات البيانيه

• For Continuous data:- للبيانات المتصله

(1) Histogram: المدرج

مدرجات على شكل مستطيلات متلاصقه رأسيه

يستخدم لوصف الكميات المتصله.

- On x – axis, put class boundaries.
- On y – axis, put frequency.

---

(2)The frequency polygon : المضلع

- On x – axis, put midpoint of classes.
- On y – axis, put frequency

---

(3) Ogive (cumulative frequency graph)

- On x – axis, class boundaries..
- On y – axis, put cumulative frequency.

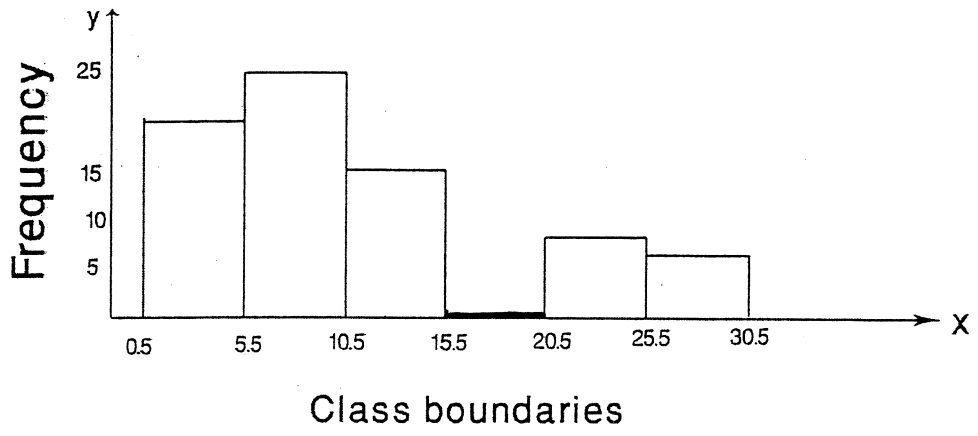
Example:

For 75 employees of a large department store, the following distribution for years of service was obtained construct a histogram. Frequency polygon and O give for the data.

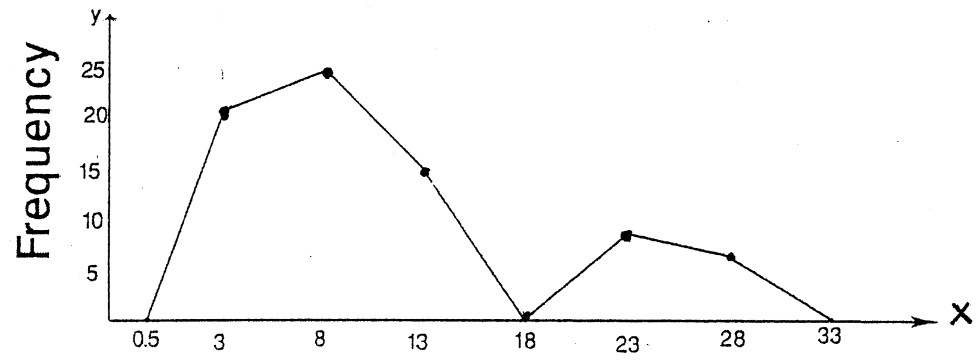
Class limits	Frequency
1-5	21
6-10	25
11-15	15
16-20	0
21-25	8
26-30	6

Solution

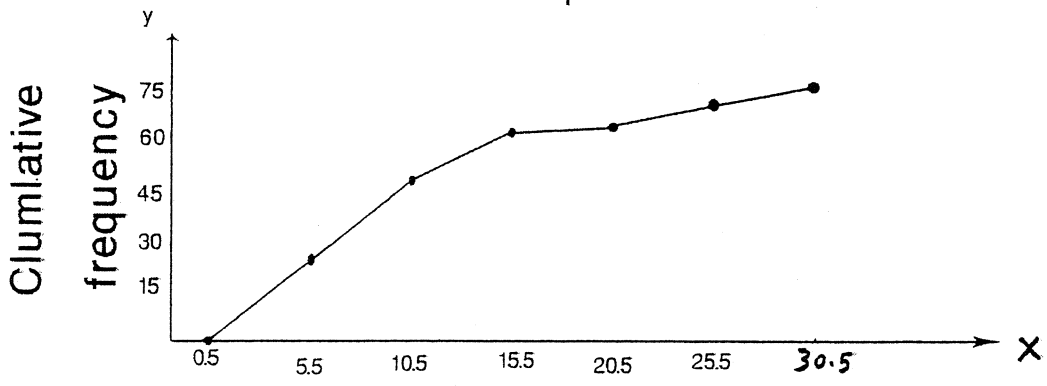
Class boundaries	Midpoint	Frequency	Cumulative freq
0.5 – 5.5	3	21	21
5.5 – 10.5	8	25	46
10.5 – 15.5	13	15	61
15.5 – 20.5	18	0	61
20.5 – 25.5	23	8	69
25.5 – 30.5	28	6	75



Class boundaries

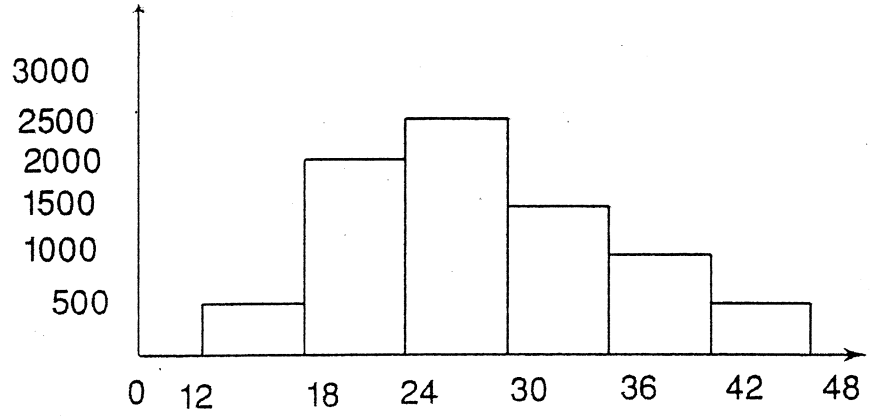


Midpoints



Class boundaries

Example:



In the above graph:

What type of this graph?

(a) Ogive (b) Pie chart (c) pereto chart (d) Histogram

---

2. The class that has highest frequency is

(a) 24-30 (b) 30-36 (c) 12-18 (d) 18-24

---

3. The class with frequency 2000 is

(a) 24-30 (b) 30 -36 (c) 12- 18 (d) 18 – 24

---

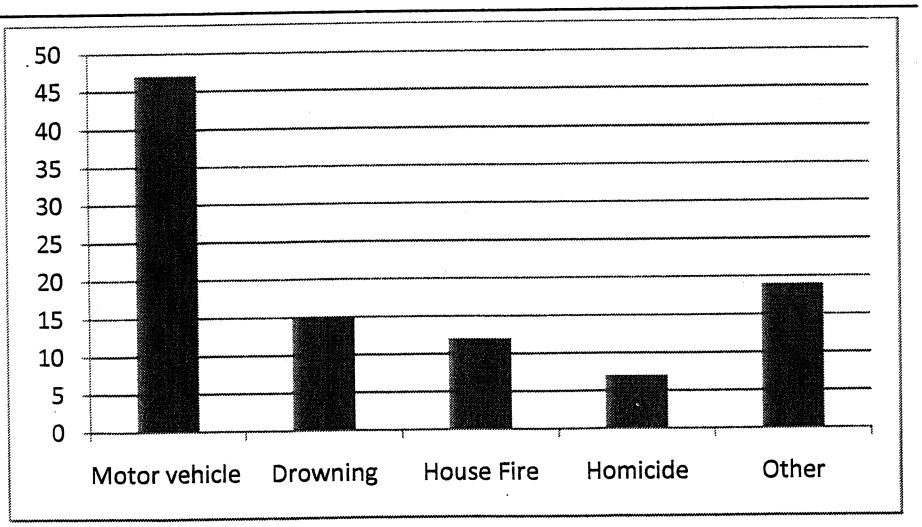
4. The total frequency of the data shown is

(a) 6000 (b) 8000 (c) 7500 (d) 1200

- **The bar charts: display the data by using vertical bars of various heights to represent the frequencies of discrete or categorical variables.**

- Example : Represent the data set by using **bar chart**:

Cause of Death(سبب الموت)	Frequency (التكرار)
Motor vehicle (حادث سيارة)	47
Drowning(الغرق)	15
House Fire(حريق منزل)	12
Homicide (القتل)	7
Other(أخرى)	19
Total	100



Notes

* Sum of relative frequencies = 1 مجموع التكرارات النسبية يساوي ١

---

* Distributions that are used when the proportions of the data is more important than the actual number of the data are known as:

Relative frequency distribution.

---

* The graphs that their distributions as proportions instead of raw data as frequencies are called:

Relative frequency graph.

---

* Histogram is suitable for representing continuous data.

* المدرج التكراري مناسب لتمثيل البيانات المتصلة .

---

* The graph that represent the cumulative frequencies for the classes in freq. distribution is called a: Ogive

Or: cumulative frequency graph.

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

4

Ch. 2 - Part 2

- Other Types of Graphs

* Pareto Chart.
* The time series graph.
* The Pie graph.

STAT.110

جمال السعدي

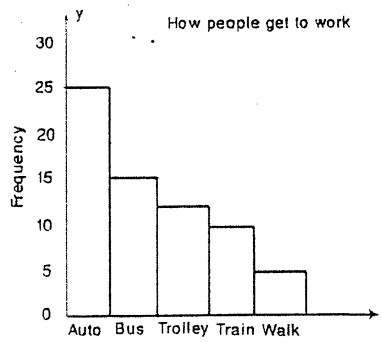
رياضيات - إحصاء

CH.2 Part 2

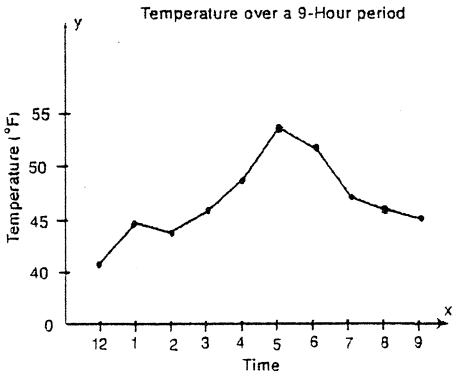
جمال السعدي

Other types of graphs:

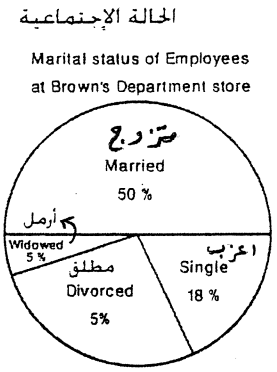
In addition to **the histogram**, **the frequency polygon**, and **the Ogive**, several other types of graphs are often used in statistics. They are **the Pareto chart**, **the time series graph**, and **the pie graph**. Figure shows an example of each type of graph.



(a) Pareto chart



(b) Time series graph



(c) Pie graph

Pareto chart

is used to represent a frequency distribution for a categorical variable and the frequencies are displayed by the heights of vertical bars, which are arranged in order from highest to lowest.

On a Pareto chart must be:

1. Make the bars the same width.
2. Arrange the data from largest to smallest according to frequency.
3. Make the units that are used for the frequency equal in size.

- **Example:**

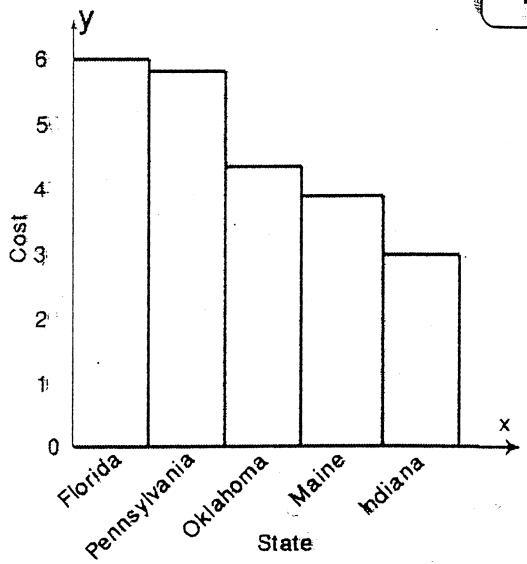
The table shown here is the average cost per mile for passenger vehicles on state turnpikes. Construct and analyze a Pareto chart for the data.

State	Number
Indiana	2.9c
Oklahoma	4.3c
Florida	6.0c
Maine	3.8c
Pennsylvania	5.8c

- **Solution:**

Arrange the data from the largest to smallest according to frequency.

State	Number
Florida	6.0c
Pennsylvania	5.8c
Oklahoma	4.3c
Maine	3.8c
Indiana	2.9c



The Pareto chart shows that Florida has the highest cost per mile: the cost is more than twice as high as the cost for Indiana.

السلاسل الزمنية

The time series Graph:

When data are collected over a period of time: they can be represented by a time series graph.

البيانات التي تحدث في فترة زمنية محددة

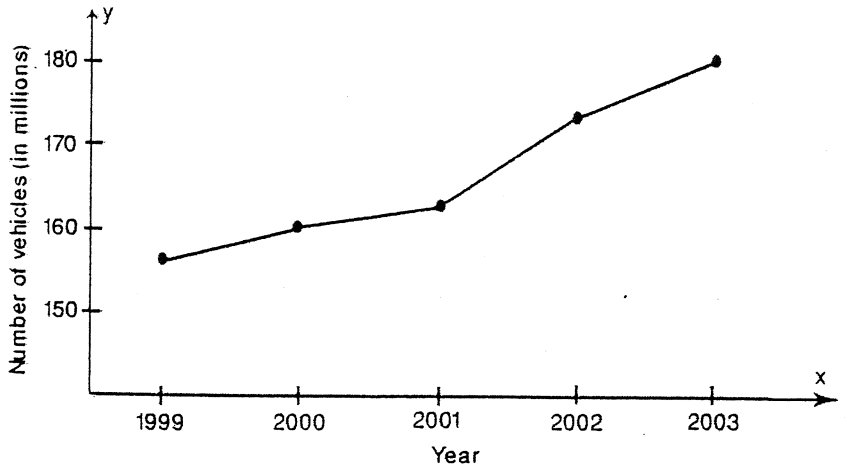
A time series graph represents data occur over a specific period of time.

Example:

المسافر

The number (in millions) of vehicles both passenger and commercial, تجاري that used the Pennsylvania طريق رئيسي Turnpike for the years 1999 through 2003 is shown. Construct and analyze a time series graph for the data.

Year	Number
1999	156.2
2000	160.1
2001	162.3
2002	172.8
2003	179.4

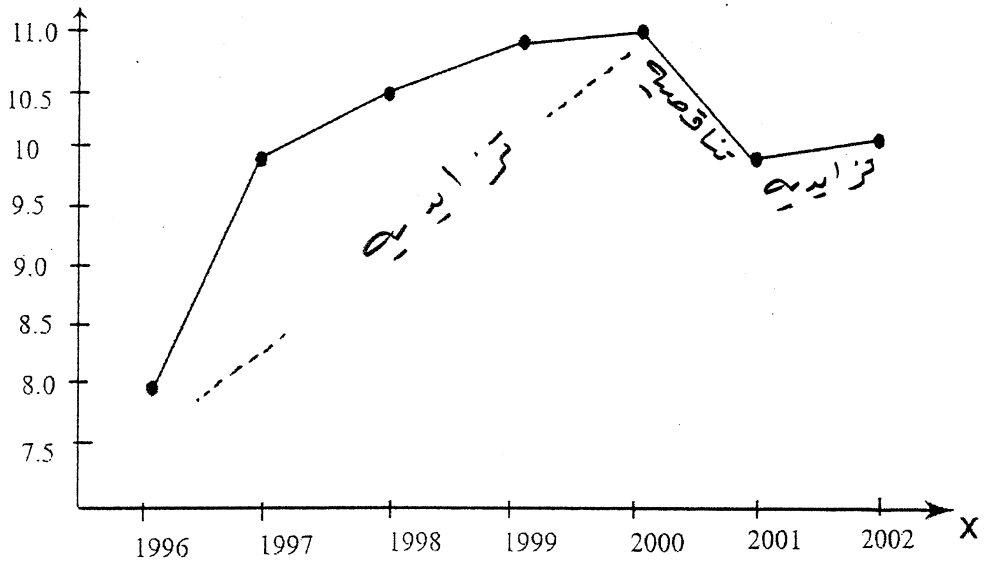


ALSAADI

Example:

Draw a time series graph to represent the data for the number of airline departure (in millions) for the given years. Over the years is the number of departure increasing, decreasing or about the same?

Year	1996	1997	1998	1999	2000	2001	2002
Number of departure	7.9	9.9	10.5	10.9	11.0	9.8	10.1



تزايدية تناقصية

Departures increased until 2000, decreased from 2000 to 2001, and increased from 2001 to 2002.

The Pie Graph:

- تعريف pie graph.....
- missing part.....
- How many degrees ..... كسر – نسبة

is a circle that is divided into sections according to the percentage of frequencies in each category of the distribution.

- * **The variable is nominal or categorical.**

Example:

Construct a pie graph showing the blood types of the army inductees described in the frequency distribution is repeated here.

Class	Frequency	Percent
A	5	20
B	7	28
O	9	36
AB	4	16
	25	100

Find the number of degrees for each class.

Solution

Using the formula

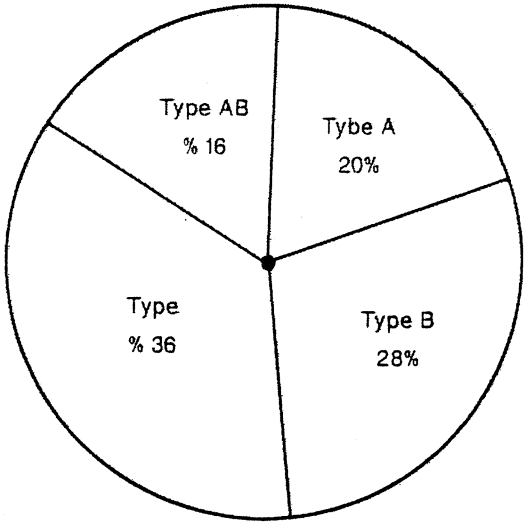
$$\text{Degrees} = \frac{F}{n} \times 360^\circ$$

$$A : \frac{5}{25} \times 360^\circ = 72^\circ$$

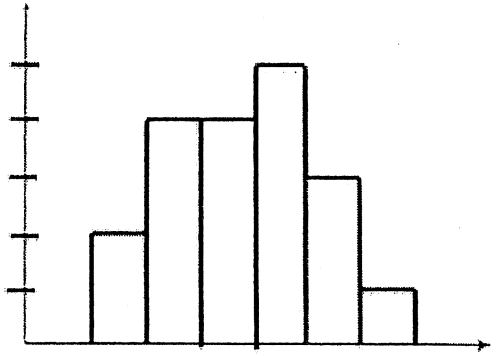
$$B : \frac{7}{25} \times 360^\circ = 100.8^\circ$$

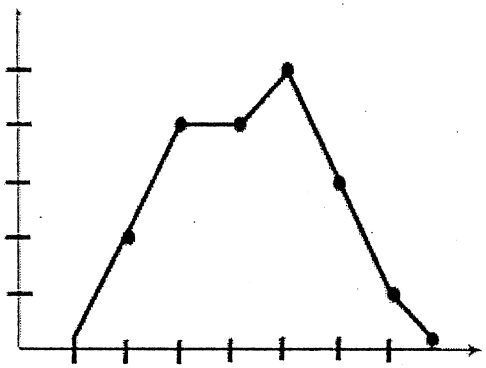
$$O : \frac{9}{25} \times 36^\circ = 129.6^\circ$$

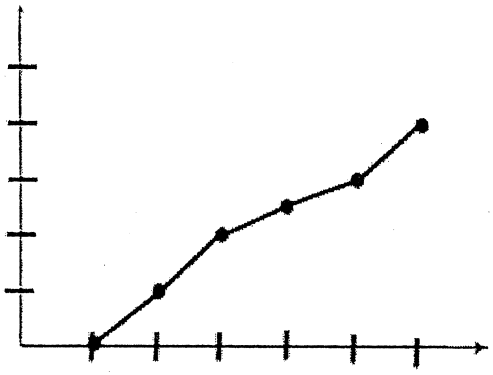
$$AB : \frac{4}{25} \times 360o = 57.6^\circ$$



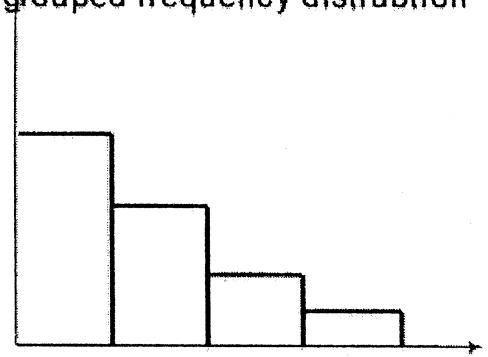
Summary of Graphs and used of Each



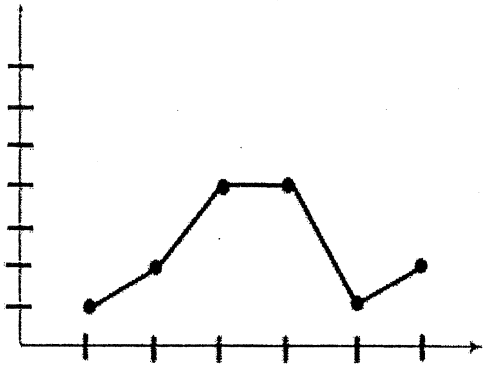




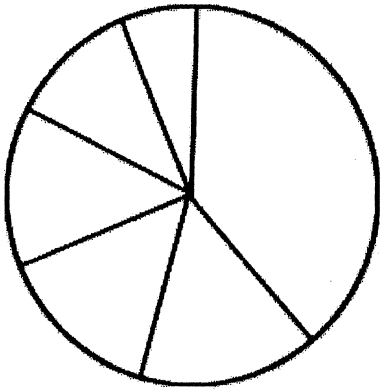
(a) Histogram, frequency polygon: ogive
Used when the data are contained in a grouped frequency distrubtion



(b) pareto chart
Used to show frequencies for nominal or qualitative variables



(c) Time series graph
Used to show a pattern or trend that occurs over a period of time



(d) Pie graph
Used to show the relationship between the parts and the whole (Most often uses percentage)

مخطط الساق والورقة

Stem and leaf plot:

is a data plot that uses part of the data value as the stem and part of the data value as the leaf.

Example:

Construct a stem and leaf plot for the data.

25	31	20	32	13
14	43	02	57	23
36	32	33	32	44
32	52	44	51	45

Solution

- *Arrange the data from L to H --- • رتب البيانات تصاعدياً

02 13 14 20 23 25 31 32 32 32
32 33 36 43 44 44 45 51 52 57

- * Separate the data according to the first digit from the left.

Stem	Leaf
0	2
1	3 4
2	0 3 5
3	1 2 2 2 2 3 6
4	3 4 4 5
5	1 2 7

Example:

The number of stories in two selected samples of tall buildings in Atlanta and Philadelphia are shown.

Construct a back-to-back stem and leaf Plot. and compare the distributions.

Atlanta					Philadelphia				
55	70	44	36	40	61	40	38	32	30
63	40	44	34	38	58	40	40	25	30
60	47	52	32	32	54	40	36	30	30
50	53	32	28	31	53	39	36	34	33
52	32	34	32	50	50	38	36	39	32
26	29								

Solution

Arrange the data from L to H

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

• Range = H − L = 61 − 25 = 36

• Range = H − L = 70 − 26 = 44

Compare the distributions. The buildings in Atlanta have a large (اكثر تشتت) variation in the number of stories per building.

ALSAADI

Example:

If the stem and leaf plot of the data is

1	5	
2	3	4
3	4	7

What is the raw data? ← ما هي البيانات الخام

Raw data is

15 , 23 , 24 , 34 , 37

On a Pareto chart, frequencies should be represented on the.....y-axis

---

The graph should be used to show the relationship between the parts and the whole is .. * الرسم البياني الذي يوضح العلاقة بين الاجزاء والكل

(a) Histogram (b) Pareto chart (c) pie graph (d) ogive

In a frequency distribution, the number of classes should between: * في التوزيع التكراري يتراوح عدد الفئات بين 5 ، 20

(a) 10 and 20 (b) 5 and 25 (c) 2 and 20 (d) 5 and 20

Another name for the Ogive is المسمى الأخر لـ ogive

(a) Histogram (b) freq. polygon (c) cumulative freq. graph

Find the boundaries for 8.6 - 8.8 ?

Lower boundary = 8.55

Upper boundary = 8.85

Chapter Quiz

Determine whether each statement is true or false. .

If the Statement is false , explain why.

1. In the construction of a frequency distribution, it is a good idea to have overlapping class limits, such as 10-20, 20-30 , 30 – 40. (×)

2. Histograms can be drown by using vertical or horizontal bars. (×)

3. It is not important to keep the width of each class the same in a frequency distribution. (×)

4. Frequency distributions can aid the researcher in drawing charts and graphs. (✓)

5. The type of graph used to represent data is determined by the type of data collected and by the researcher's purpose. (✓)

6. In construction of a frequency polygon, the class limits are used for the x - axis. (×)

---

7. Data collected over a period of time can be graphed by using a pie graph. (×)

Select the best answer:

8. What is another name for the ogive?
   a. Histogram.
   b. Frequency polygon
   c. Cumulative frequency graph
   d. Pareto chart

---

9. What are the boundaries for 8.6 - 8.87
   a. 8 – 9 b. 8.5 - 8.9 c. 8.55 – 8.85 d. 8.65 - 8.75

---

10. What graph should be used to show the relationship between the parts and the whole?
    a. Histogram b. Pie graph c. Pareto chart d. Ogive

---

11. Except for rounding errors , relative frequencies should add up to what sum?
    a. 0 b. 1 c. 50 d. 100

Complete these statements with the best answers:

(12) The three types of frequency distributions are **categorical** **ungrouped** and **grouped**.

(13) In a frequency distribution, the number of classes should be between **5** and **20** .

(14) Data such as blood types (A, B, AB, O) can be organized into a (n) **categorical**. Frequency distribution.

(15) Data collected over a period of time can be graphed using a (n) **time series** graph.

(16) A statistical device used in exploratory data analysis that is a combination of a frequency distribution and a histogram is called a (n) **Steam and leaf**.

(17) On a Pareto chart, the frequencies should be represented on the **y-** axis

● A pareto chart is useful for which of the following purposes?

A) Representing relative frequencies of categories in a specific year
B) Representing the cumulative frequencies of the data
C) Representing the frequencies of the data, sorted from largest to smallest
D) Representing the frequencies of a data category over a period of several years

In pareto chart :
the data sorted from:
Largest to smallest.

● In a categorical frequency distribution, the number of observations in a class is called a(an) ...
A) interval. B) frequency. C) midpoint. D) category.

* في التوزيع التكراري الثوي
عدد المشاهدات في الفئة يسمى التكرار ← frequency

●The graph that should be used to show the relationship between the whole and the parts is called ...
A) time series graph. B) Pareto chart. C) frequency ploygon. D) pie graph.

* الرسم البياني الذي يوضح العلاقة بين الكل والأجزاء يسمى
→ Pie graph .

● In a frequency distribution, if the relative frequencies are 0.1, Y, 0.38 and 0.22, then the relative frequency Y is ...

A) 0.30 B) 0.60 C) 0.22 D) 0.10

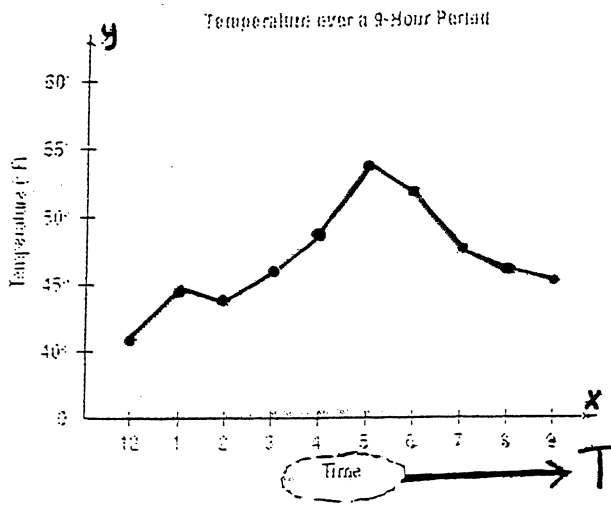
$$\Sigma \text{ relative frequencies} = 1$$

$$0.1 + y + 0.38 + 0.22 = 1$$

$$y + 0.70 = 1 \Rightarrow y = 1 - 0.70$$

$$y = 0.30$$

● This graph is an example of



Time s. graph

A) Ogive B) Time series graph C) Frequency Polygon D) Pie graph

● The data set that is collected over a period of time can be best represented by a (an) ...

A) time series graph B) histogram C) pie graph D) ogive

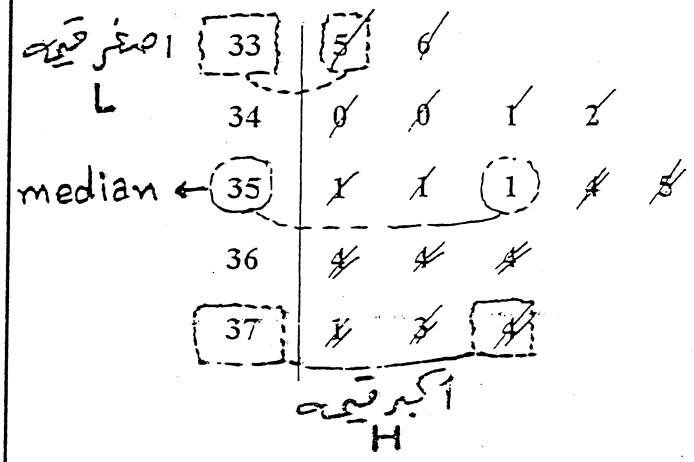
* البيانات التي تجمع خلال دوره زمنيه تمثل بيانياً بالسلاسل الزمنيه .

● The monthly incomes of eight computer operators are \$1,950, \$1,775, \$2,060, \$1,840, \$1,795, \$1,890, \$1,925 and \$1,810. What are these ungrouped values called?

A) Class limits. B) Class frequency. C) Raw data. D) Class boundaries

* الدخول الشهريه لثمان ..... ← تمثل بيانات خام
↳ Raw data

Use the following stem-leaf plot to answer the following three questions:



● The range value of the raw data set for the above stem and leaf plot is ...
A) 43 **B) 39** C) 44 D) 42

$$\text{Range} = H - L = 374 - 335 = 39$$

● The raw data set for the above stem and leaf plot is called ...
A) bimodal. B) unimodal. C) trimodal. D) multimodal.

* العددان 364 ، 351
كلاً منهما تكرر ثلاث مرات
∴ البيانات السابقه ثنائيه المنوال ← **bimodal**

● The median value of the raw data for the above stem and leaf plot is ...
A) 35 **B) 351** C) 51 D) 1

* الوسيط هو القيمه التي تتوسط البيانات

$$\therefore \text{median} = 351$$

What is the stem and leaf of 45?

* رقم الآحاد هو 5 * باقي العدد هو 4

ⓐ stem = 5 leaf = 4 ⓑ stem = 4 leaf = 5

What is the stem and leaf of 127?

ⓐ stem = 1 leaf = 27 ⓑ stem = 12 leaf=7

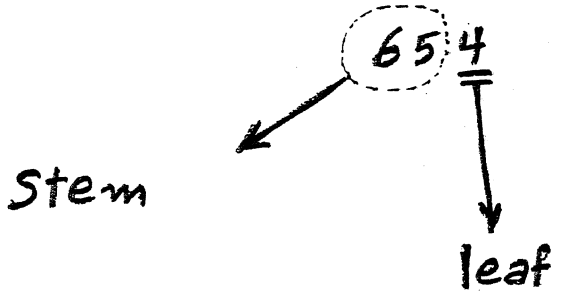
ⓒ stem = 27 leaf = 1 ⓓ stem = 7 leaf=12

The.....is a method of organizing data and is a combination of sorting and graphing?

ⓐ pareto chart ⓑ pie graph ⓒ stem and leaf plot

The stem part for the number 654 is ...

A) 6 B) 54 C) 4 D) 65



* رقم الآحاد leaf 4
* باقي العدد stem 65

In a pie graph, if the blood type A represents 9/72 of the distribution, how many degrees would be needed to represent A?

A) 8° B) 9° C) 72° D) 45°

** عدد الدرجات = الكسر × 360°

number of degrees $= \frac{9}{72} \times 360^\circ = 45^\circ$

A (An) ... is used to represent a data set that is collected over a period of time.

A) Pareto chart B) time series graph C) ogive D) pie graph

period of time → time series graph (تعريف)

ALSAADI

Use the following to answer questions

The following table shows the distribution of the blood type for 70 students:

Classes	A	B	O	AB
Frequency	15	20	17	18

- The type of the frequency distribution is ... distribution.
A) grouped frequency B) probability **C) categorical frequency** D) ungrouped frequency

The type is categorical frequency

∴ الفئات classes ليست رقمية

أي أن الفئات عبارة عن كلمات أو رموز

∴ نوع التوزيع التكراري : نوعية (وصفية)

- The probability of selecting a student with AB blood type is ...
A) 0.154 B) 0.348 **C) 0.257** D) 0.27

$$P(AB) = \frac{\text{عدد الطلاب من نوع فصيلة الدم AB}}{\text{العدد الكلي للطلاب}} = \frac{18}{70}$$

$$= \boxed{0.257}$$

- The range value ...
A) cannot be calculated B) is 3. C) is 5. D) is A - AB.

The range : can not be calculated

* لا يمكن حساب المدى لأن البيانات نوعية (وصفية)

- In a pie graph, how many degrees would be needed to represent A?
A) 72° B) 99.69° **C) 77.14°** D) 103.64°

The number of degrees for class A

$$= \frac{15}{70} \times 360 = \boxed{77.14^\circ}$$

$$\frac{\text{تكرار الفئة A}}{\text{مجموع التكرار}} \times 360^\circ$$

Use the following table to answer the following five questions:

Ages	Number of Students
20.2 – 29.3	16
29.4 – 38.5	25
38.6 – 47.7	51
47.8 – 56.9	79
57.0 – 66.1	21
66.2 – 75.3	8

* منتصف الفئه الثالثه = $\frac{\text{مجموع طرفي الفئه}}{2}$

● What is the age midpoint for the third class?

A) 33.95 B) 52.35 C) 61.55 **D) 43.15**

$$\text{midpoint for third class} = \frac{38.6 + 47.7}{2} = 43.15$$

● What is the class boundary for the fourth class limit?

A) 47.30 -- 57.40 B) 48.30 -- 56.40 **C) 47.75 -- 56.95** D) 47.85 -- 56.85

class limit = 47.8 – 56.9

class boundary = 47.75 – 56.95

الفئه الرابعه

الفاصله موزونه (بعد رقم واحد من class limit)

* نضع الرقم 5 من نهايه upper limit

* نطرح 1 من الرقم الأخير من lower limit ثم نضع الرقم 5.

أسم الجدول

- What is the name of the table?
  A) Grouped frequency distribution.
  B) Cumulative frequency distribution.
  C) Categorical frequency distribution.
  D) Ungrouped frequency distribution.

- What is the relative frequency of students whose ages within the second class?
  A) 25.5% B) 0.255 C) 12.5% D) 0.125

relative frequency
التكرار النسبي للفئة الثانية

$$= \frac{\text{تكرار الفئة}}{\text{مجموع التكرارات}} = \frac{25}{200} = 0.125$$

- The data can be represented the best using ...
  A) pie graph. B) Ogive. C) histogram. D) bar chart.

In a frequency distribution, if the relative frequencies are 0.20, 0.28, X and 0.16, then the relative frequency X is
A) 0.63 B) 0.36 C) 0.46 D) 0.64

* مجموع $P(x)$ لا بد أنه يكون 1

$$\Sigma P(x) = 1$$

$$0.20 + 0.28 + X + 0.16 = 1$$

$$X = 1 - 0.20 - 0.28 - 0.16$$

$$X = 0.36$$

The following table represents the favorite car make for a group of students

Classes	Frequency
Toyota	10
Nissan	11
Chevrolet	9
GMC	13
Honda	14

الفئة الثالثة

- The cumulative frequency for the third class is ...
  A) 30 B) 48 C) 13 D) 43

* التكرار المتجمع الناظر للفئة الثالثة :

$= 10 + 11 + 9 = \underline{\underline{30}}$

- The sample size is ...
  A) 5 B) 60 C) 57 D) 18

Sample size: $n = \Sigma f = 10 + 11 + 9 + 13 + 14$
$= \underline{\underline{57}}$

- The percentage of students who like Chevrolet is ...
  A) 19.23% B) 83.33% C) 25% D) 15.79%

$P = \frac{9}{57} \times 100\% = 15.79\%$

- If a pie graph is used to represent the data, the degree for the Honda brand would be ...
  A) 55.38 B) 15.38 C) 0.15 D) 88.42

Number of degree $= \frac{14}{57} \times 360^\circ = \underline{\underline{88.42}}$

- The type of data is ...
  A) nominal. B) ordinal. C) discrete. D) continuous.

* نوع البيانات في الجدول أسمية
وهي عبارة عن ماركات السيارات .

- The most appropriate measure of central tendency for this data is the ...
  A) mean B) mode C) midrange D) range

* المقياس المناسب
من مقاييس النزعة المركزية لهذه البيانات الأسمية هو المنوال

- The continuous data can be organized into a table that is called ...
  A) grouped frequency distribution.
  B) ungrouped frequency distribution.
  C) categorical frequency distribution.
  D) ordinal frequency distribution.

- The class width for the class limit 4.1 - 9 is ... → 4.1 - 9.0 وزن الفاصلة *
  A) 6.55 B) 5.9 C) 4.9 D) 5

class boundaries : 4.05 - 9.05

class width = 9.05 - 4.05 = 5

- What is the most appropriate measure of central tendency for the following data set?
  Ali, Saeed, Ahmed, Saeed, Ali, Ali
  A) Mean. B) Mode. C) Median. D) Midrange.

* المقياس المناسب من مقاييس النزعة المركزية
للبيانات الاسمية هو المنوال ← Mode

- The most appropriate graph for categorical data is ...
  A) pie graph. B) histogram. C) time series graph. D) stem and leaf plot.

* التمثيل البياني المناسب للبيانات النوعية (الوصفية)
هو القطاعات الدائرية ← pie graph

- The nominal data can be organized into a table that is called ...
  A) ungrouped frequency distribution.
  B) nominal frequency distribution.
  C) grouped frequency distribution.
  D) categorical frequency distribution.

* البيانات الاسمية تنظم بجدول يسمى التوزيع التكراري النوعي

The number of patients in the waiting rooms within a hospital at a specific time are given by the following frequency distribution.

Number of patient	1	2	3	4	5	Total
Frequency	4	5	?	8	4	30

Answer the following five questions

- The percentage of the number waiting rooms that have 2 patients is ...
  A) 16.67% B) 12.50% C) 36.67% D) 83.33%

$$P = \frac{5}{30} \times 100\% = 16.67\%$$

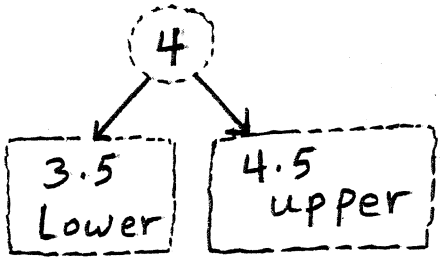
- The data can be best represented graphically by ...
  A) bar chart B) frequency polygon C) time series graph D) histogram

* الفئات عباره عن ارقام منفصله 1, 2, 3, 4, 5
∴ أفضل طريقه لتمثيل البيانات بيانياً
هي طريقه الأعمده ← bar chart

- The missing frequency for the third class is ... .
  A) 7 B) 6 C) 3 D) 9

The missing f. $= 30 - (4+5+8+4) = 9$

- The lower class boundary for the fourth class is ...
  A) 5.5 B) 4.5 C) 2.5 D) 3.5



- The sample size is ...
  A) 30 B) 5 C) 20 D) 32

* حجم العينه هو مجموع التكرارات .

Sample size $= \Sigma f = 30$

تعريفات مهمة

- The pie graph: is a circle that is divided into sections according to the percentage of frequencies in each category of the distribution.

بيانات خام
- When data are collected in the original form, they called raw data.

- Ungrouped frequency distributions: is used for discrete data.

- Grouped frequency distribution: is used for continuous data.

- Histogram, frequency polygon or ogive graphs are used to represent continuous data graphically.

- Bar graph: are used mostly to represent discrete and ordinal data graphically.

* طريقه الأعمده :
تستخدم لتمثيل البيانات المنفصله والترتيبيه بيانيًا .

- Pie graph and Pareto chart are used mostly to represent nominal data graphically.

* طريقه الـ : باى ، باريتو
تستخدم لتمثيل البيانات الأسميه بيانيًا .

- Stem and leave plot: is a combination of sorting and graphing. It retains the actual data while showing them graphically.

- The histogram displays the continuous data that are organized in a grouped frequency distribution by using vertical bars of various heights to represent the frequencies.

- The frequency polygon displays the continuous data that are organized in a grouped frequency distribution by using lines that connect points plotted for the frequencies at the midpoints of the classes.

- The cumulative frequency graph or ogive represents the cumulative frequencies for the classes in a grouped frequency distribution.

- The bar charts displays the data by using vertical bars of various heights to represent the frequencies of discrete or categorical variables.

- Categorical frequency distributions are used for data that can be placed in specific categories, such as nominal or ordinal level data.

Given the following distribution:

Class boundaries Ages	Frequencies number of students
13.5 – 18.5	4
18.5 – 23.5	9
23.5 – 28.5	12
28.5 – 33.5	15
33.5 – 38.9	17

1. Number of students where age is less than 23.5 is $= 4+9$

(a) 4 (b) 9 (c) 13 (d) 5

---

2. Number of students where age is less than 33.5 is: $= 4+9+12+15$

(a) 15 (b) 57 (c) 40 (d) 25

---

3. Except for rounding errors, relative frequencies should add up to what sum?

* باستبعاد اخطاء التقريب
يكون مجموع التكرار النسبي = 1

(a) 0 (b) 1 (c) 50 (d) 100

---

4. If class limits 23.4 – 28.4 the class width is..... $28.45 - 23.35 = 5.1$

(a) 5 (b) 2.5 (c) 5.1 (d) 6

---

5. In a pie graph if the blood type O was 36% of the distribution.

How many degrees would be needed to represent type O?

(a) 129° (b) 29.6° (c) 129.6° (d) 360°

Solution

$$\text{Degrees} = \frac{F}{n} \times 360^\circ$$

$$= \frac{36}{100} \times 360^\circ$$

$$= 129.6^\circ$$

ALSAADI

إنتهى Ch. 2

كل التمنيات بالنجاح والتوفيق

وأخيراً
أدعو الله أن يتقبل هذا العمل
ويكون فيه النفع للجميع ....
جمال السعدي

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

5

Ch. 3 - Part 1

- Measures of Central Tendency for Ungrouped data.
- Measures of Variation for ungrouped data.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch.3 Part 1

جمال السعدي

Measures of central tendency

Mean, Median and Mode.

- statistic: is measures for the sample.
- parameter: is measures for the population.

Individual data:

The mean (average)

* الوسط الحسابي = مجموع القيم / عددها

Is the sum of values, divided by total number of values.

$$\bar{x} = \frac{\sum x}{n}$$ For the sample

$$\mu = \frac{\sum x}{N}$$ For the population

The median: (MD)

$$\begin{cases} X_{\frac{n+1}{2}} \ldots\ldots\ldots n \text{ is odd} \\ [X_{\frac{n}{2}} + (X_{\frac{n}{2}} + 1)] \div 2 \ldots\ldots n \text{ is even} \end{cases}$$

* الوسيط: هو نقطة منتصف البيانات بعد ترتيبها تصاعدياً أو تنازلياً.

- Is the midpoint of the data.
- Arrange the data from low to high and select the middle point.

The Mode: ⟶ (the most typical case)
يحدث كثيراً

❖ Is the value that occurs most often in the data.
المنوال : هو القيمة الأكثر تكراراً من البيانات

❖ The model class is the class with the largest frequency.
الفئة المنوالية : هي الفئة المناظرة لأعلى تكرار .

Mode may be

- There is no mode... لا يوجد منوال
- The mode is unique . المنوال وحيد
- Bimodal ..... ثنائية المنوال
- Multimodal..... عديدة المنوال

→ For Individual data

Note

- The mean is unique
- The median is unique

- **The Midrange:** تخمين تقريبي
is a rough estimate of the middle.

$$\text{Midrange} = \frac{H + L}{2}$$

- **The range** = Highest value - lowest value
المدى
= H - L

- **The variance:** التباين
is the average of the squares of the distance each value is from the mean.

- **The standard deviation:** الانحراف المعياري
is the square root of the variance.

* الانحراف المعياري هو الجذر التربيعي للتباين .

Important Formulas

	Sample	Population
• Mean	$\overline{X} = \frac{\sum X}{n}$	$\mu = \frac{\sum X}{N}$
• Variance	$S^2 = \frac{\sum X^2}{n} - \overline{X}^2$ OR $S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{2}}{n-1}$ اذا كثر استخداماً	$\sigma^2 = \frac{(X-\mu)^2}{N}$
• Standard deviation	$S = \sqrt{S^2}$	$\sigma = \sqrt{\sigma^2}$
• Coefficient of variation	$C.Var = \frac{S}{\overline{X}} .100\%$	$C.Var = \frac{\sigma}{\mu} .100\%$

Note

Sample: $\overline{X}$, S^2 , S	**Population:** μ, σ^2 , σ

Note

Measures of variation:

Range, variance and standard deviation

Note

When: C. var for sample 1 > C.var for sample 2

Then: sample 1 is more than variable sample 2

Example:

Twelve major earthquakes had Richter magnitudes shown here

7.0 , 6.2 , 7.7 , 8.0 , 6.4 , 6.2

7.2 , 5.4 , 6.4 , 6.5 , 7.2 , 5.4

Find ① Mean ② Median ③ Mode

④ Midrange ⑤ Range ⑥ Variance

⑦ Stander deviation ⑧ Coefficient of variation

Solution

Arrange the data:

5.4 , 5.4 , 6.2 , 6.2 , 6.4 , 6.4 , 6.5 , 7 , 7.2 , 7.2 , 7.7 , 8

① Mean : $\overline{X} = \frac{\sum x}{n} = \frac{79.6}{12} = 6.63$

② Median: $\overline{X} = \frac{X_6 + X_7}{2} = \frac{6.4 + 6.5}{2} = 6.45$

③ Mode: D = 5.4 , 6.2 , 6.4 , 7.2

④ Midrange $= \frac{H + L}{2} = \frac{8 + 5.4}{2} = 6.7$

⑤ Range $= H - L = 8 - 5.4 = 2.6$

⑥ Variance: $S^2 = \frac{\sum x^2}{n} - \overline{X}^2 = \frac{535.34}{12} - (6.63)^2 \cong 0.65$

⑦ Stander deviation: $S = \sqrt{S^2} = \sqrt{0.65} \cong 0.81$

⑧ Coefficient of variation

$$C.\,var = \frac{S}{\overline{X}} \times 100\ \%$$

$$= \frac{0.81}{6.63} \times 100\ \%$$

$$= 12.22\ \%$$

ALSAADI

●● لاحظ الكلمه المميزه في كل جمله .

Exercises: * حدد المقياس الأنسب لكل موقف

Describe which measure of Central tendency: Mean, Median, and Mode was probably used in each situation.

(a) **One half** of the factory workers make more than $ 5.37 per hour, and **one half** make less than $ 5.37 per hour.

- One half = Median **(Median)**

(b) The **average** number of children per family is 1.8.

- Average + عدد = Mean **(Mean)**

(c) **Most** people prefer red convertibles over any other color.

- Most = Mode **(Mode)**

(d) The average person cuts the lawn **once a week**.

- Average + once a week = Mode
  Twice **(Mode)**

(e) the **most** common fear today is fear of speaking in public.

- Most = Mode **(Mode)**

(f) The **average** age of college professor is **42.3** years.

- Average + عدد = Mean **(Mean)**

ALSAADI

- The weighed mean is used when the values in a data set are not equally represented.

Example:

In a survey of third – grade students, this distribution was obtained for the number of "best friends" each had.

Number of students	Number of best friends
8	1
6	2
5	3
3	0

Find the average number of best friends for the class use the weighted mean.

Solution

$$\overline{X}_{\omega} = \frac{\sum w \times x}{\sum w} = \frac{8\times 1 + 6\times 2 + 5\times 3 + 3\times 0}{8+6+5+5} = \frac{35}{22} = 1.6$$

Example:

- The average score on an English final exam was 85,

  With a standard deviation of 5.

- The average score on a history final exam was 110.

  With a standard deviation of 8 which class was more variable?

Solution

We find coefficient of variation

- For English exam:

$$C.Var = \frac{S}{\bar{X}} \cdot 100\% = \frac{5}{85} \cdot 100\% = 5.8\%$$

- For history exam:

$$C.Var = \frac{S}{\bar{X}} \cdot 100\% = \frac{8}{110} \cdot 100\% = 7.2\%$$

∴ History exam is more variable.

Grouping data

Important formulas

* صيغ هامة : تُعطى في ورقة الأسئلة

- Width = $\dfrac{\text{Rang}}{\text{number of class}}$

- Range = Highest value − lowest value
  = H − L

- Mean $\bar{X} = \dfrac{\sum X.F}{\sum F}$

- Variance $S^2 = \dfrac{\sum F.X^2 - [(\sum F.X)^2 / n]}{n-1}$

- Standard deviation $S = \sqrt{s^2}$

- Coefficient of variation $C\,var = \dfrac{S}{\bar{X}}.100\%$

- Z score or standard score: * الدرجة المعيارية (حفظ)

  Tells how many standard deviation of the data above or below the mean.

  For sample $Z = \dfrac{X - \bar{X}}{S}$ For population $Z = \dfrac{x - \mu}{\sigma}$

ALSAADI

Example:

This distribution represents the data for weights of fifth – grade boys.

جميع طرق الفئة / 2

Find: ① Mean ② standard deviation

Example		Solution		
Weight	Frequency f	X	X.f	$X^2.f$
52.5 – 55.5	9	54	486	26244
55.5 – 58.5	12	57	684	38988
58.5 – 61.5	17	60	1020	61200
61.5 – 64.5	22	63	1386	87318
64.5 – 67.5	15	66	990	65340
	$n = \sum f = 75$		$\sum X.f = 4566$	$\sum X^2 f = 279090$

① Mean: $\bar{X} = \frac{\sum X.f}{\sum f} = \frac{4566}{75} = 60.88$

② Variance: $S^2 = \frac{\sum x^2 f - \frac{(\sum X.f)^2}{n}}{n-1} = \frac{279090 - \frac{(4566)^2}{75}}{75-1}$

$= 15.02594595$

⇒ Standard deviation: $S = \sqrt{S^2}$ $= 3.876$

ALSAADI

Example:

A final exam has a mean of 84 and a standard deviation of 4.
Find: the corresponding Z score for each raw score.

(a) 87 (b) 79 (c) 93

Solution

$$Z = \frac{X - \mu}{\sigma} \qquad \mu = 84 \qquad \sigma = 4$$

(a) $Z = \frac{87 - 84}{4} = 0.75$

(b) $Z = \frac{79 - 84}{4} = -1.25$

(c) $Z = \frac{93 - 84}{4} = 2.25$

Example:

A student scores 60 on a math test that has mean of 54 a standard dev. Of 3 and she scores 80 on a history test.

With a mean 78 and a standard dev. Of 2

On which test did she perform better?

Solution

$$Z = \frac{X - \bar{X}}{S}$$

$$Z_1 = \frac{60 - 54}{3} = 2 \qquad Z_2 = \frac{80 - 78}{2} = 1$$

$$\longrightarrow Z_1 > Z_2$$

Then:

Score on math is better relative position.

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

6

Ch. 3 - Part 2

* Quartiles.
* Five number summary.
* Skew.
* Outliers.
* Box plot.

STAT.110

جمال السعدي

رياضيات - إحصاء

CH.3 Part 2

جمال السعدي

$$\text{Percentile} = \frac{(\text{number of values below x}) + 0.5}{\text{Total number of values}} \times 100\%$$

ALSAADI

Example 1:

A teacher gives a 20 – point test to 10 students. The scores are:

18, 15, 12, 6, 8, 2, 3, 5, 20, 10

Find: the percentile rank of score of 12.

Solution

Arrange the data from low to high

$$2 , 3 , 5 , 6 , 8 , 10 , \overset{\times}{\boxed{12}} , 15 , 18 , 20$$

$$\text{Percentile} = \frac{6 + 0.5}{10} \times 100\% = 65\%$$

∴ A student whose score was 12 better than 65% of the class.

* Arrange the value corresponding to a given percentile

ترتيب القيمة المقابلة لنسبة مئوية معينة $C = \frac{n \cdot p}{100}$

Where: n is total number of values

: P is percentile.

- إذا كانت C عدد عشري نأخذ العدد الصحيح التالي له ثم توجد القيمة المناظرة للنسبة P.
- إذا كانت C عدد صحيح تأخذ القيمة المناظرة للعدد C والقيمة التالية وتوجد وسطها الحسابي فيكون هو القيمة المناظرة للنسبة P.

From example 1:

1. Find the value corresponding to percentile 25%

Arrange the data.....

2 , 3 , 5 , 6 , 8 , 10 , 12 , 15 , 18 , 20

ترتيب القيمة $C = \frac{n \times p}{100} = \frac{10 \times 25}{100} = 2.5$

- If: C is not whole number

Round it up to the next whole number

يقرب للعدد الصحيح التالي له أي إلى 3

$\therefore C = 3$ (Third)

$\therefore$ The value 5 corresponding to 25%

2. Find the value corresponding to percentile 60 %

2 , 3 , 5 , 6 , 8 , [10] , [12] , 15 , 18 , 20

ترتيب القيمة $C = \frac{n \times p}{100} = \frac{10 \times 60}{100} = 6$

- If: C is whole number

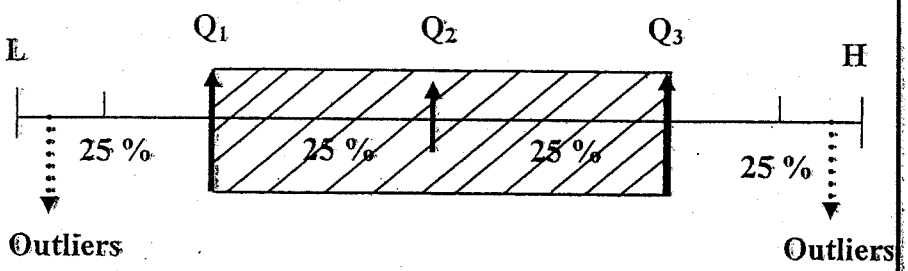
Use the value halfway between C and C + 1

$\frac{10 + 12}{2} = 11$ $\therefore$ corresponding to 60%

$\therefore$ Anyone scoring 11 is better than 60% of the class.

ALSAADI

Quartiles



- Quartiles divide the distribution into four groups separated by Q_1 , Q_2, Q_3.
  - Q_2 is the median
  - Inter quartile range: IQR = $Q_3 - Q_1$

- Skew : S.K = $\dfrac{\overline{X} - D}{S}$ OR $S.K = \dfrac{3(\overline{X} - median)}{S}$

Mean, Mode, Standard Deviation

القانون الأكثر استخداماً

Where $\overline{X} = \dfrac{\sum x}{n}$ and $S = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n-1}}$

- Mid Q = $\dfrac{Q_1 + Q_3}{2}$

Note

Five number summary for the data set are:

- **low value**
- **Q_1**
- **Q_2**
- **Q_3**
- **High value**

ALSAADI

Example:

For the values

33 , 27 , 51 , 30 , 31 , 23 , 38 , 40 , 42 , 43 , 29

Find: Q_1 , Q_2 , Q_3 , IQR , mid Q and Skew.

Solution

* Arrange the data from low to high: * ترتيب البيانات تصاعدياً

23 , 27 , 29 , 30 , 31 , $\underset{Q_2}{\boxed{33}}$, 38 , 40 , 42 , 43 , 51

* Q_2 (Median) = 33

* Q1 = 29 * Q_3 = 42

* IQR = $Q_3 - Q_1$ = 42 – 29 = 13

$$* \text{ Mid Q} = \frac{Q1 + Q3}{2} = \frac{29 + 42}{2} = 35.5$$

* To find skew : we find $\bar{X} = \frac{\sum x}{n} = \frac{387}{11} = 35.181$

$$\text{and } S = \sqrt{\frac{\sum x^2 - \frac{(\sum x^2)}{n}}{n-1}} = \sqrt{\frac{14307 - \frac{(387)^2}{11}}{11-1}} = 8.32$$

$$\therefore \text{ S.K} = \frac{3(\bar{x} - \text{med.})}{S} = \frac{3(35.1818 - 33)}{8.32} = 0.79 > 0$$

موجب

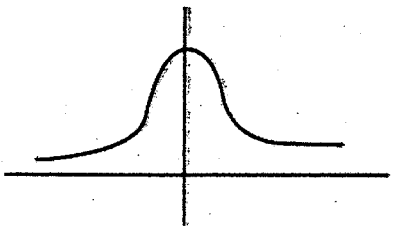
∴ Skew to the right.

Note

Skew :	$S.K = \dfrac{3(\bar{x} - \text{median})}{S}$

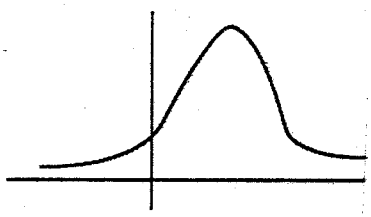
1. Symmetric : if S.K = 0

   Mean = median = mode



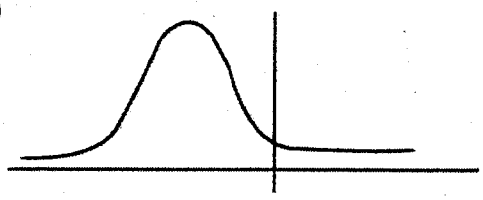
2. Skew to **the right**: if S.K > 0

   Mode < median < mean



3. Skew to **the left**: if S.K <0

   Mean < median < mode



Example:

- If the mean = 4 and mode = 2 ⟹ Mode < Mean

  then the distribution is : right skew.

- If the mode = 8 and median = 5 ⟹ Median < Mode

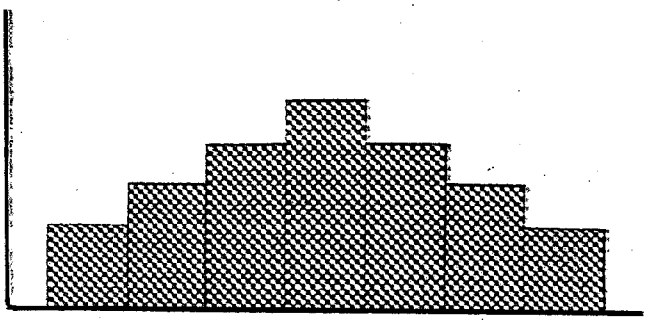
  then the distribution is : left skew.

- If the mean, median and mode are equal * الثلاثة متساويين *

  the distribution is: Symmetric.

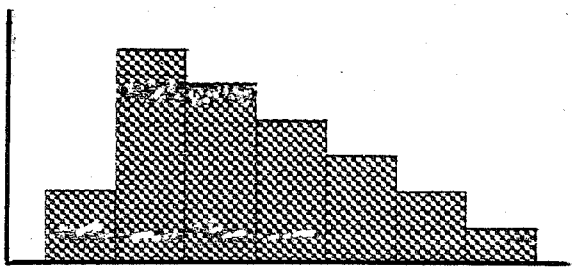
- In a *symmetrical distribution*, the data values are evenly distributed on both sides of the mean.

mean=median=mode



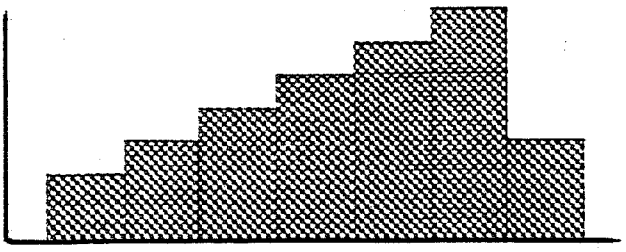
- In a *positively skewed* or *right skewed distribution*, the majority of the data values fall to the left of the mean and cluster at the lower end of the distribution.

mode<median<mean

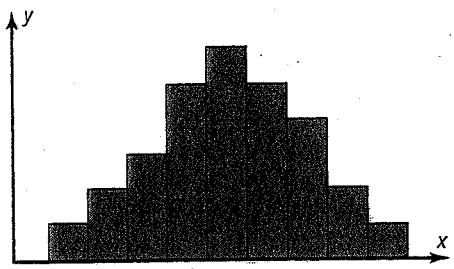


- In a *negatively skewed* or *left skewed distribution*, the majority of the data values fall to the right of the mean and cluster at the upper end of the distribution

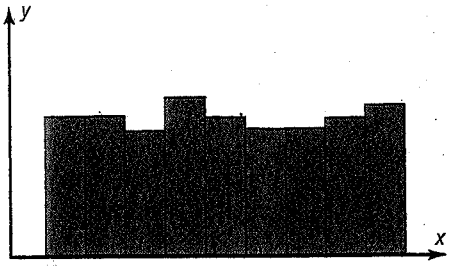
mean<median<mode



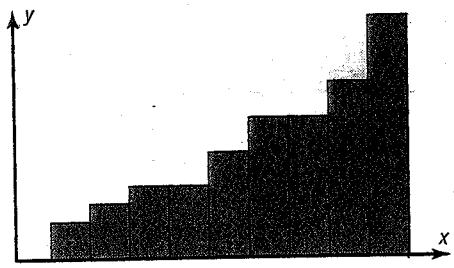
Distribution Shapes



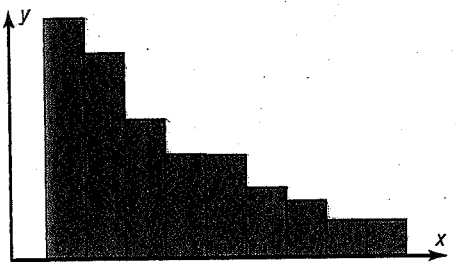
(a) Bell-shaped



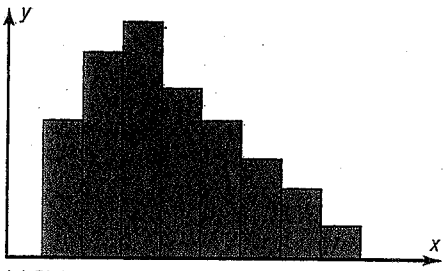
(b) Uniform



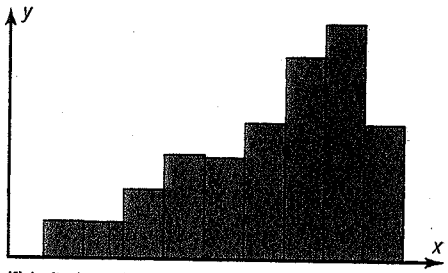
(c) J-shaped



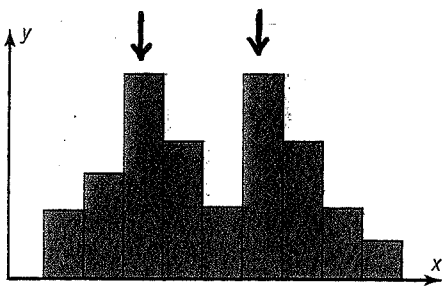
(d) Reverse J-shaped



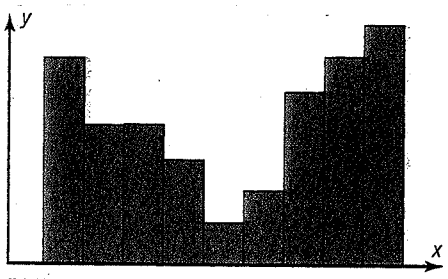
(e) Right-skewed



(f) Left-skewed



(g) Bimodal



(h) U-shaped

Exercises:

Find: Q_1 , Q_2 , Q_3 , IQR , mid Q and Skew.

For the value.

22 , 50 , 15 , 18 , 6 , 13 , 12 , 5

If: the number of data set is 6

and the mean is 8

* Find the sum of the data values.

For the following data set:

79 , 83 , 86 , 90 , 96 , 100

Find the value corresponding to the 60^{th} percentile

Find the five number summary for the following data

10 , 6 , 12 , 2 , 16 , 20 , 14

If the variance is 36.

Find the standard deviation.

Example:

Check the following data set for outliers.

5 , 6 , 12 , 13 , 15 , 18 , 22 , 50

Solution

* Q_2 (median) $= \dfrac{13 + 15}{2} = 14$

* $Q_1 = \dfrac{6 + 12}{2} = 9$

* $Q_3 = \dfrac{18 + 22}{2} = 20$

* IQR = Q3 – Q1 = 20 – 9 = 11

طريقة تحديد Outliers

١ نوجد Q_3 , Q_1 , Q_2

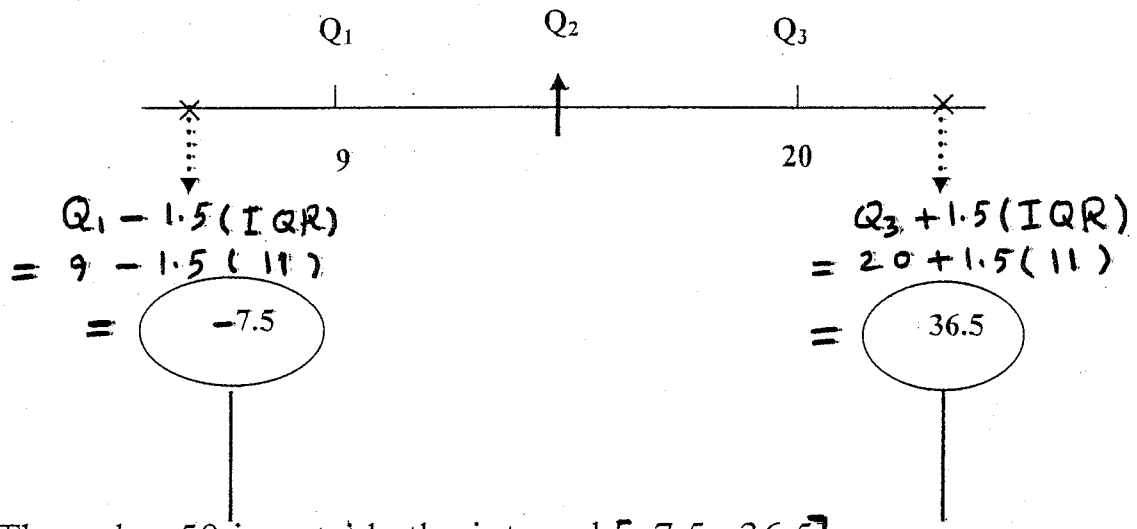
٢ نوجد $IQR = Q_3 - Q_1$

٣ نوجد $Q_1 - 1.5\,(IQR)$, $Q_3 + 1.5\,(IQR)$

٤ القيمة الأقل من $Q_1 - 1.5\,(IQR)$

القيمة الأكبر من $Q_3 + 1.5\,(IQR)$

Outliers يطلق عليها



The value 50 is outside the interval [-7.5 , 36.5]

∴ 50 it can be considered an outlier.

Pox Plot

Consists of:

- **Five – number summary of the data set:**

(1) The lowest value of the data set.

(2) The first quartile Q_1.

(3) The median Q_2.

(4) The third quartile Q_3.

(5) The highest value of the data set.

Example:

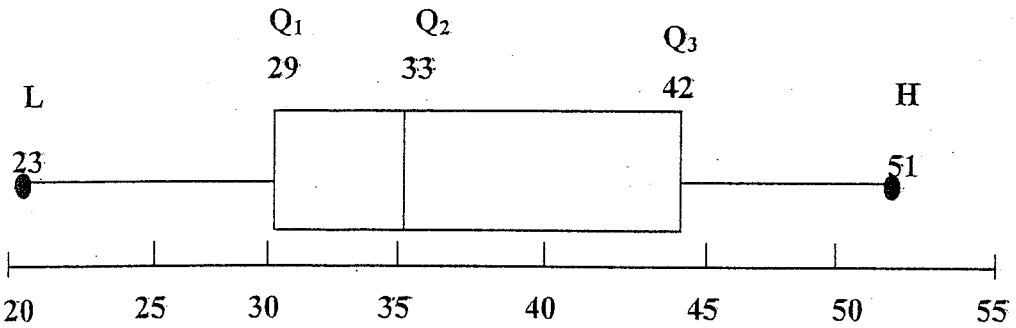
Construct a box Plot for the data.

33 , 38 , 43 , 30 , 29 , 40 , 51 , 27 , 23 , 31

Solution:

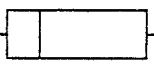
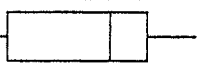
Arrange the data from low to high:

23 , 27 , (29), 30, 31 , [33], 38 , 40 , (42), 43 , 51



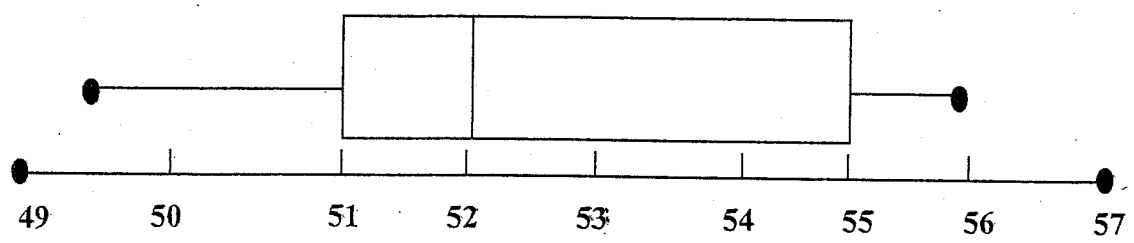
Note

- The distribution is right skew.

- The distribution is symmetric  إذا كان الوسيط Q_2 في منتصف الـ Box
- The distribution is right skew  إذا كانت المساحة الأكبر داخل الـ Box جهة اليمين
- The distribution is left skew  إذا كانت المساحة الأكبر داخل الـ Box جهة اليسار

Example:

Use the box plot to find:



1. Q_1
2. Q_3
3. MD
4. IQR
5. Min. value
6. Max .value
7. This distribution is positive or negative skew.

Solution

1. $Q_1 = 51$
2. $Q_3 = 55$
3. Median: $Q_2 = 52$
4. $IQR = Q_3 - Q_1 = 55 - 51 = 4$
5. Min. Value = 49.
6. Max. Value = 57.
7. The distribution is : right skew

السعدي ALSAADI
نسخة جديدة منقحة
1433/32

7

Ch. 3 - Part 3

Chapter Quiz.

- measures of Central tendency.
- measures of variation.
- measures of position.

STAT.110

جمال السعدي
رياضيات - إحصاء

CH.3 Part 3

جمال السعدي

Chapter quiz:

Determine whether each statement is true of false.

If the statement is false, explain why.

(1) When the mean is computed for individual data.

All values in the data set are used. (√)

---

(2) The mean can not be found for grouped data.

When there is an open class. (√)

---

(3) A Single, extremely large value can affect.

The median more than the mean. العكس صح (x)

---

(4) One – half of all the data values will fall above.

The mode and one- half will fall below the mode. (x)
median ✓ median ✓

---

(5) In a data set, the mode will always be unique (x)

May be * there is no mode (لا يوجد منوال / ربما)
* Bimodal (ثنائي)
* unique mode (منوال وحيد)
* Multimodal (عديد المنوال)

---

(6) The range and midrange are both measures of variation (x)

variation ليس ضمن مقاييس الـ

(7) One disadvantage of the median is that it is not unique. (x)

عيوب — ليس وحيد

Because: the median is unique. وحيد

(8) The mode and midrange are both measures of variation (x)

× ×

(9) If a person's score on an exam corresponds to 75th percentile, then that person obtained 75 correct answer out of 100 questions (x)

May be 3 correct answers from 4 questions.

Note

- **Mean – Median – Mode :**
  are measures of central tendency.
- **Range – Variance – Standard deviation**
  are measures of variation.
- **Percentiles – quartiles – deciles:**
  are measures of position.

- **The coefficient of variation CVar :**

العلاقه

Show relation between mean and standard deviation.

Select the best answer:

(10) what is the value of the mode when all values in the data set are different ? * اذا كانت كل البيانات مختلفه ← لايوجد منوال

(a) 0 (b) 1 (c) there is no mode

---

(11) When data are categorized as, for example, places of residence (Rural ريفي , Suburban ضاحية , Urban متسوب للمدينة) , the most appropriate المناسب - الملائم measure of central tendency مركزية is the تزعة

(a) Mean (b) Median (c) Mode (d) Midrange

---

(12) P_{50} Corresponds to يماثل - يقابل $P_{50} = Q_2 = D_5$

(a) Q_2 (b) D_5 (c) IQR (d) a and b are correct

---

(13) Which is not part of the five – number summary ?

(a) Q_1 and Q_3 (b) the mean (c) the median

(d) the smallest and largest data values.

---

(14) A statistic that tells the number of standard devotions a data value is above or below the mean is called

(a) A quartile (b) A Percentile (c) A Z-score ← تعريف

(d) A coefficient of variation

ALSAADI

Complete these statements with the best answer

16 A measure obtained from sample data is called a **(statistic)**

الحروف μ , σ^2 , σ

17 Generally, Greek letters are used to represent **(parameters)**

الحروف $\bar{X}$, S^2 , S

,and Roman letters are used to represent **(statistic)**.

---

18 The positive square root of the variance is called :

* الجذر التربيعي الموجب للتباين يسمى الانحراف المعياري .

(The standard deviation).

---

19 The symbol for the population standard deviation is (σ).

---

20 When the sum of the lowest data value and the highest data value is divided by 2 , the measure is called **(Midrange)** $= \frac{L+H}{2}$

---

21 If the mode is to the left of the medium and the mean is to the right of the medium then the distribution is **(Positive)** skewed.

---

22 An extremely high or extremely low data value is called **(an Outlier).**

ALSAADI

Home work

(23) For the values:

12 , 15, 13 , 14 , 15 , 16 , 17 , 16 , 17 , 18

Find each of these:

1. mean **2. median** **3. Mode** **4. Midrange**

5. Range **6. variance** **7. standard devation**

Exercises:

For these situations, state which measure of central tendency – mean, median , or mode – should be used.

a) The most typical case is desired. **(Mode)**

b) The distribution is open- ended. **(Median)**

c) There is an extreme value in the data set. **(Median)**

d) The data are categorical. **(Mode)**

e) Further statistical Calculation will be needed. **(Mean)**

f) The values are to be divided into two approximately equal groups, one group containing the larger values and one containing the smaller values. **(Median)**

Home work

(24) the distribution of the number of errors that 10 students made on a typing test is shown

Errors	Frequency
0-2	1
3-5	3
6-8	4
9-11	1
12-14	1

Find each of these:

1. Mean.
2. Model class.
3. Variance.
4. Standard deviation.

Note

- The median is used for an open – ended distribution.
- The mode is the only measure of central tendency that can be used when the data are nominal or categorical.
- When the distribution is extremely skewed the median rather than, the mean as measure of central tendency.
- In box plot, if the median is near the center of the box, the distribution is approximately symmetric.
- When all the data transformed into z – scores the resulting distribution will have

  Mean = 0 and standard deviation = 1

Exercises:

1. What is a z score or standard score?

- A Z-score tells how many standard deviations the data value is above or below the mean. تعريف

---

2. Define: Percentile rank, percentiles and Deciles.

- Percentile rank: $\frac{\text{(number of values below x) + 0.5}}{\text{Total number of values}}$ 100%
- Percentiles: divide the data set into 100 equal groups.
- Deciles : divide the data set into 10 equal groups

---

3. What is the difference between a percentage and a percentile?

مقياس نسبي

- A percentile: is a relative measurement of position.
- A percentage: is an absolute measure of the part to the total.

---

4. Define quartile.

Position in fourths that a data value holds in the distribution Q.

---

5. what is the relationship between quartiles and percentiles ?

$Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$

6. What is a decile ?

Position in tenths that a data value holds in the distribution D.

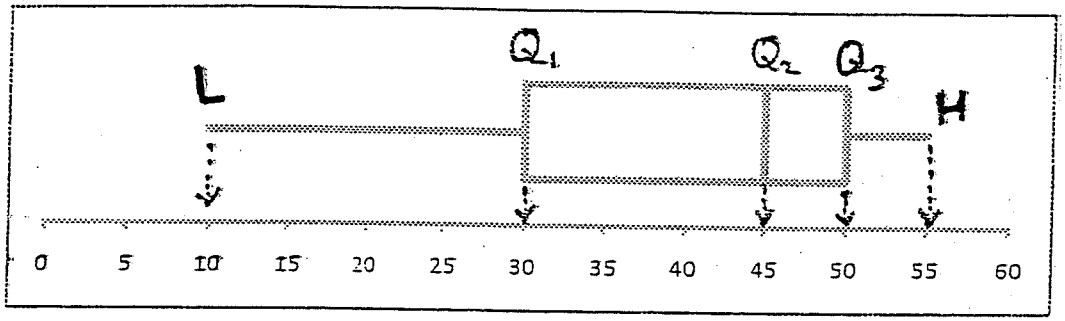
7. How are deciles related to percentiles ?

$D_1 = P_{10}$, $D_2 = P_{20}$, $D_3 = P_{30}$, ..., $D_{10} = P_{100}$

8. To which Percentile, Quartile and decile does the median correspond ?

Median = $P_{50} = Q_2 = D_5$

Use the following boxplot to answer the following four questions:



The midrange value of the raw data for the above boxplot is ...
A) 42.5 B) 47.5 C) 32.5 D) 27.5

$$\text{Midrange} = \frac{L + H}{2} = \frac{10 + 55}{2} = 32.5$$

The mode value of the raw data for the above boxplot ...
A) is 30. B) is 50. C) can't be determined. D) is 45.

لا يمكن تعيين الـ mode من boxplot

The IQR value of the raw data for the above boxplot is ...
A) 40 B) 30 C) 20 D) 15

$$IQR = Q_3 - Q_1 = 50 - 30 = 20$$

What is the relationship among the mean, median and mode for the above boxplot?
A) The mean is the smallest value.
B) Can't be determined.
C) They are all equal.
D) The mean is the largest value.

* Left skew منحرف لليسار
∴ أصغر المقاييس هو mean وأكبرها mode

Use the following to answer questions

The following table shows the distribution of the blood type for 32 students:

Classes	A	B	O	AB
Frequency	5	10	8	9

- The mean value
  A) is B B) is 8 C) is 2 **D) cannot be calculated**

The mean value can not be calculated

because: The data set are nominal.

* لا يمكن حساب قيمة الوسط الحسابي لأن البيانات أسمية

- The mode value
  A) is AB B) is 10 **C) is B** D) cannot be calculated

المنوال
The mode : is B

هي الفئة المتاظرة لأكبر تكرار.

- Which measures of central tendency will always have unique values?
  A) Mode and weighted mean.
  B) Mode and median.
  C) Mean and median.
  D) Mean and mode.

* مقاييس النزعة المركزية التي لها قيمة وحيدة دائماً هي:

Mean and Median

* أما اختيارية Mode لا يصلح لأن الـ Mode ربما يكون

may be →
- no mode
- unique
- bimodal
- multimodal

Use the following to answer questions

The following table shows the distribution of the blood type for 24 students:

Classes	A	B	O	AB
Frequency	8	3	6	7

- The midrange value ...
  A) is 5 B) is B C) cannot be calculated D) is 1

midrange : cannot be calculated لا يمكن حسابه
لأنه البيانات وصفيه (الفئات ليست اعداد)

- The mode value ...
  A) is 8 B) is 6 C) cannot be calculated D) is A

المنوال هو الفئه ذات التكرار
mode = A

- In a pie graph, how many degrees would be needed to represent the blood type A?
  A) 120° B) 28° C) 105° D) 90°

The number of degrees

$$= \frac{f}{n} \times 360^\circ$$

$$= \frac{8}{24} \times 360^\circ = 120^\circ$$

● Which is not part of a five-number summary?

A) The mean B) Q_1 and Q_3 C) The median D) The smallest and the largest data values

Five - number summary are

Smallest data , Q_1 , Q_2 , Q_3 , Largest data.

The mean: is not part of five - number summary.

● Which is not part of a five-number summary?

A) Q_2 B) The midrange C) The smallest and the largest data values D) Q_1 and Q_3

The midrange: is not part of five - number summary

● Which measures are mostly affected by outliers?

A) Mean and median
B) Mean and IQR
C) Mode and median
D) Mean and standard deviation

Mean and standard deviation

are affected by outliers

الوسط الحسابي والانحراف المعياري أكثر تأثرًا بالقيم الشاذه .

● Which measures are mostly affected by outliers?

A) Mode and median. ×
B) Midrange and range. $\frac{H+L}{2}$, $H-L$
C) Mode and midrange. ×
D) Mean and mode. ×

Midrange and range
are affected by outliers

● Which measures are mostly affected by outliers?

A) Mean and mode ×
B) Mode and midrange ×
C) Mean and midrange
D) Mean and median ×

* ملحوظه : الأكثر تأثرًا بالقيم الشاذه هي المقاييس التي يتم عليها عمليات حسابيه .

● Which measures of central tendency are not affected largely by outliers?

لا يتأثر بالقيم الشاذه

A) Mean and mode.
B) Weighted mean and mean.
C) Mean and midrange.
D) Mode and median.

* أي من مقاييس النزعه المركزيه لا يتأثر بالقيم الشاذه ؟

→ Mode and median

المنوال والوسيط لا يتأثران بالقيم الشاذه لأنه لا يتم عليهما اي عمليات حسابيه .

- Approximately what percentage of normally distributed data values will fall within 1 standard deviation above or below the mean.

A) 68% B) 95% C) 99.7% D) 13.5%

حفظ

* لو كان 1 standard dev. يكون الناتج 68%
* // // // 2 // // // 95%
* // // // 3 // // // 99.7%

- Approximately what percentage of normally distributed data values will fall within 3 standard deviation above or below the mean

A) 95% B) 99.7% C) 68% D) 34.1%

A) → 2 S B) → 3 S C) → 1 S

حفظ

3 standard deviation ⇒ 99.7 %

- Approximately what percentage of the standard normal distribution data values will fall between -2 and 2.

A) 99.7% B) 13.5% C) 95% D) 68%

حل السؤال * Data values fall between -2 and 2 is 95 %

حالات أخرى:

Data values fall between -3 and 3 is 99.7%

Data values fall between -1 and 1 is 68 %

* الحالات الثلاثة السابقة حفظ.

- Which one of the following is referred to as a statistic?
  A) The sample mode B) The sample data C) The population mean D) The population data

- Which one of the following is referred to as a parameter?
  A) Population mean B) Population C) Sample mode D) Sample

Parameter ⟶ Population mean

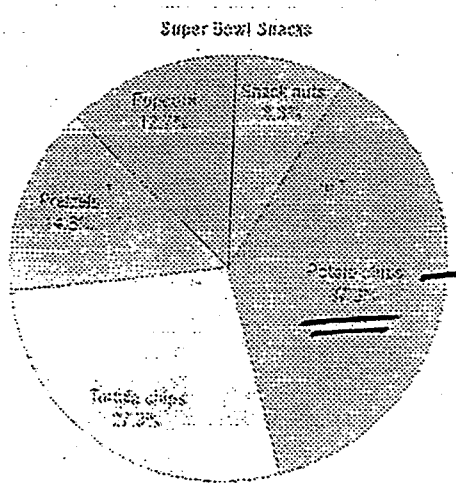
- Which one of the following is referred to as a parameter?
  A) The population data.
  B) The sample variance.
  C) The sample data.
  D) The population variance.

* **Parameter** (المعلمة)
لـ عباره عن مقياس يخص المجتمع Population
وينقسم إلى
μ : mean
σ^2 : variance (إجابة السؤال)
σ : standard deviation

* **Statistic** (الاحصاءة)
لـ عباره عن مقياس يخص العينة sample
وينقسم إلى
$\bar{X}$: mean
S^2 : variance
S : standard deviation

← معلومات اضافيه.

● This Pie graph represents snacks people eat at a sport game. What is the mode for this data



المنوال هو القطاع الذي له أكبر نسبة

∴ The mode is potato chips

A) Potato chips B) Tortilla chips C) Pretzels D) Popcorn

● What is the mean, median, mode of the following numbers? 1, 3, 6, 8, 12

A)	mean= 6	median= 6	mode= no mode
B)	mean= 5	median= 6	mode=1
C)	mean= 4	median= 6.5	mode= 3
D)	mean= 20	median= 3	mode=12

$$* \text{mean} = \frac{\Sigma x}{n} = \frac{1+3+6+8+12}{5} = 6$$

* median = 6 * القيمة التي تتوسط القيم بعد ترتيبهم تصاعدياً

* mode : no mode ← لأن القيم مختلفة *

● For a data set, half of the observations are always greater than the ...
A) mean. B) weighted mean. C) mode. D) median.

* لمجموعة البيانات :
نصف عدد المشاهدات دائماً
أكبر من الوسيط ← median .
والنصف الآخر من المشاهدات أقل من الوسيط

● If the standard deviation of a data set is 8.00, and a value X=12.00 has a z-score of 0.50, then the mean value

A) is 8 B) is -8 C) is 8.5 D) cannot be determined from the data given

$\sigma = 8$
$X = 12$
$z = 0.50$

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

$$0.50 = \frac{12 - \mu}{8}$$

$$(0.50)(8) = 12 - \mu$$
$$4 = 12 - \mu \Rightarrow \mu = 8$$

● Find the average grade points for a student who has the following results:

Course	STAT	CHIM	BIO	ENG	ARAB
Grade	88	70	75	80	90
Credit	3	3	4	3	2

A) 73.33 B) 76.80 C) 80.40 D) 79.60

Average grade point

$$= \frac{88\times3 + 70\times3 + 75\times4 + 80\times3 + 90\times2}{3 + 3 + 4 + 3 + 2} = 79.6$$

ALSAADI

Use the following to answer questions

The weights (in grams) of the contents of several small bottles are 4, 2, 5, 4, 5, 2 and 4. Use this information to answer the following three questions:

- What is the value of the mean?
  A) 5.50 B) 4.57 **C) 3.71** D) 4.00

أوزان محتويات عدد زجاجات صغيره هي
4 , 2 , 5 , 4 , 5 , 2 , 4

$$\text{mean: } \bar{x} = \frac{\Sigma x}{n} = \frac{26}{7} = 3.71$$

- What is the value of the standard deviation?
  A) 1.57 B) 0.80 **C) 1.25** D) 0.89

$$S = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{106 - \frac{(26)^2}{7}}{7-1}} = 1.25$$

- What is the value of the coefficient of variation?
  A) 42.32% B) 20.00% **C) 33.69%** D) 22.25%

Coefficient of Variation

$$CVar = \frac{S}{\bar{x}} \cdot 100\%$$

$$= \frac{1.25}{3.71} \cdot 100\%$$

$$= 33.69\ \%$$

- A statistic that tells the number of standard deviations a data value is above or below the mean is called a ....
  A) coefficient of variation. (B) z score. C) percentile. D) quartile.

z score ⟶ {تعريف احفظ}

: tells the number of standard deviations above or below the mean.

- If the value X=6 has a z-score of -0.50 and standard deviation 6 in a data set, then the mean value ...
  (A) is 9. B) is 12. C) cannot be determined. D) is -9.

$X = 6 \quad , \quad z\text{-score} = -0.50 \quad , \quad \sigma = 6$

The mean : μ ?

$\because z = \frac{x - \mu}{\sigma}$ {قانون} $\Rightarrow \quad -0.50 = \frac{6 - \mu}{6}$

$\Rightarrow (-0.50)\cdot(6) = 6 - \mu$

$-3 = 6 - \mu \quad \Rightarrow \mu = 6 + 3 \Rightarrow \boxed{\mu = 9}$

- If the value X=6 has a z-score of -0.50 in a data set, then the mean
  A) is 5.50 B) is 6.00 C) is 6.50 (D) cannot be determined from the given data

$Z\text{-score} = \frac{X - \mu}{\sigma}$

{لا يمكن ايجاد ال mean μ لأنه المعادلة تحتوي على مجهولين: μ, σ}

$-0.50 = \frac{6 - \mu}{\sigma}$ ∴ mean can not be determined From the given data.

● If the mean of a set of data equals 18.00 and a value X=22.50 has a z-score of 0.65, then the variance must be:

(A) 47.93 B) 73.94 C) 6.92 D) 3.24

$$Z\text{-score} = \frac{X-\mu}{\sigma}$$

$$0.65 = \frac{22.5-18}{\sigma} \Rightarrow \sigma = \frac{22.5-18}{0.65}$$

$$\Rightarrow \text{Variance}: \sigma^2 = \left(\frac{22.5-18}{0.65}\right)^2 = 47.928$$

● If a student scored X points on a test where the mean score was 82.4, the variance was 9, and the student's Z- score was 2.7 then X must be ...

A) 74.3 B) 10.4 C) 47.6 (D) 90.5

$$Z\text{-score} = \frac{X-\mu}{\sigma}$$

$$2.7 = \frac{X-82.4}{3}$$

$X = ?$
$\mu = 82.4$
$\sigma^2 = 9 \Rightarrow \sigma = 3$
$z\text{-score} = 2.7$

$$\Rightarrow X - 82.4 = (3)(2.7)$$

$$\Rightarrow X = (3)(2.7) + 82.4 \Rightarrow X = 90.5$$

● What is the most appropriate measure of central tendency for the following data set?

A, C, B, B, C, A, B

A) Median. B) Midrange. C) Mean. (D) Mode.

* البيانات ليست رقمية (وصفية)

∴ المقياس المناسب لها هو ← Mode

انتشار

●The ... can be used to determine the spread of a data set.

A) mode and the standard deviation
B) midrange and the variance
C) standard deviation and the mean
D) variance and the range

spread ≡ variation * تقايس

Range – variance – standard deviation

●The measures that can be used to determine the spread of a data set are the ...

A) variance and the mean
B) midrange and the standard deviation.
C) variance and the mode.
D) range and the standard deviation.

ليست ضمن

●If a data value is not within the range $[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$, then this value is called ...

A) an outlier B) the median C) the third quartile D) the first quartile

$[Q_1 - 1.5(IQR), Q_3 + 1.5(IQR)]$ * القيم التي ليست ضمن الفترة تعرف بالقيم الشاذة

outlier

●If a data value is smaller than $Q_1 - 1.5(IQR)$, this value is considered to be ...

A) the range. B) the minimum. C) a z-score. D) an outlier.

* $Q_1 - 1.5(IQR)$ القيم التي أصغر من
* $Q_3 + 1.5(IQR)$ أو أكبر من
كلها تعرف بالقيم الشاذة

outlier

●A ... is referred to every measurement calculated for a study that is conducted on all students of KAU.

A) statistic. B) population. C) sample. D) parameter.

المجتمع كاملاً

parameter * المقياس الناتج من دراسة المجتمع كاملاً يسمى

●A ... is referred to every measurement calculated for a study that is conducted on a group of students from KAU.

A) sample. B) population. C) statistic. D) parameter.

عينة من المجتمع

Statistic * المقياس الناتج من دراسة عينة من المجتمع يسمى

ALSAADI

تعريفات مهمة

- Measures of average are also called measures of central tendency and include: the mean, median, mode, midrange, and weighted mean.

- Measures of variation such as the range, variance and standard deviation are used to describe the spread of data.

- The values that are smaller than $Q_1 - 1.5$ (IQR)
  or larger than $Q_3 + 1.5$ (IQR)
  are called outliers.

- The weighed mean is used when the values in a data set are not equally represented.

- A statistic is a characteristic or measure calculated using the data values of a sample.

- A parameter is a characteristic or measure calculated using all the data values of a specific population.

- Variances and standard deviations can be:
  used to determine the spread of the data.
  If the variance or standard deviation is large, the data are more dispersed.

Summarize data using measures of central tendency, such as the mean, median, mode, and midrange.

Describe data using the measures of variation, such as the range, variance, and standard deviation.

Identify the position of a data value in a data set using various measures of position, such as standard scores and quartiles.

Measures of average are also called measures of central tendency and include the mean, median, mode, midrange, and weighted mean.

When all the values in a data set occur with the same frequency is said to have no mode.

The midrange (MR) is a rough estimate of the middle and defined as the sum of the lowest and highest values in a data set divided by 2.

The mean cannot be computed for an open – ended frequency distribution.

The mean is affected by extremely high or low values and may not be the appropriate average.

The median is used to find the average of an open-ended distribution.

The median is affected less than the mean by extremely high or extremely low values.

The midrange is easy to compute.

The midrange gives the midpoint.

The midrange is affected by extremely high or low values in a data set.

Large coefficient of variation means large variability.

A standard score or z score is used when direct comparison of raw scores is impossible.

Boxplots are graphical representations of a five-number summary of a data set. Data set. The five specific values that make up five-number summary are minimum, Q_1, Q_2, Q_3 and maximum.

The range is the distance between highest value and lowest value.

The variance is the average of the squares of the distance between the mean and each value in a data set.

The standard deviation is the square root of the variance.

وأخيراً
أدعو الله أن يتقبل هذا العمل
ويكون فيه النفع للجميع ....
جمال السعدي

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

9

Ch. 4 - Part 1

- Introduction.
- Sampel Spaces and Probability.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 4 Part. 1

• sample space	فراغ العينه	• Product	حاصل ضرب
• Experiment	تجربه	• Certain occur	حادثة مؤكدة
• probability	احتمال	• Never occur	حادثة مستحيلة
• Toss	القاء	• Complement	حادثة مكملة
• coin	قطعه نقود	• Outcomes	نواتج (عناصر)
• Roll	يتدحرج (القاء)	• Empirical	تجريبي (مبني على الملاحظة
• Dice	حجر نرد	• Subjective	وهمي (مبني على التخمين)
• Tree diagram	الشجرة البيانية	• Urn	صندوق
• Head	وجه القطعه النقدية	• Drawn	يسحب
• Tail	خلفيه القطعه النقدية	• Mutually exclusive	متنافية
• Event	حادثه	• Recent Study	دراسة حديثة
• Simple event	حادثه بسيطه	• Common	مشترك
• Odd number	عدد فردى	• Exactly	بالضبط
• Prime number	عدد أولى	• Contain	يحتوى على
• Even number	عدد زوجى	○ Consists of	يتكون من
• Compound event	حدث مركب	• Select	يختار
• Random	عشوائى	• At least	على الأقل
• Gender	نوع	• At most	غلى الأكثر

Sample Spaces and Probability

A probability experiment

A chance leads to defined results called outcomes

An outcome

Is the result of a single trial of a probability experiment.

A sample space

Is the set of all possible outcomes.

An event

Consists of a set of outcomes of a probability experiment.

Equally likely events

Are events that have the same probability of occurring.

Classical Probability

The probability of any event E

$$\text{Is } P(E) = \frac{\text{Number of outcomes in E}}{\text{Total number of outcomes in the sample space}} \quad P(E) = \frac{n(E)}{n(S)}$$

ALSAADI

Empirical probability

Based on observation

$$P(E) = \frac{\text{Frequency for the class}}{\text{Total frequency in the distribution}}$$

Subjective probability

Based on estimate and inexact information

Simple event

Is an event contain **one** outcome.

Compound event

Is an event contain **more than one** outcome

Coin

- **The number of outcomes in the sample space:**

$n(S) = 2^N$ ⟶ عدد الرميات أو عدد القطع

Find the sample space for:

(1) Toss one coin → S = { H , T} ⟶ $N(s) = 2^1 = 2$

Head Tail

Tree diagram

H

T

(2) Toss two coins → S = {HH, HT, TH, TT} ⟶ $N(s) = 2^2 = 2 = 4$

Tree diagram

H H

T

T H

T

H. W

(3) Toss 3 Coins

ALSAADI

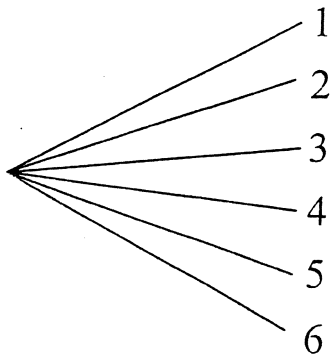
Dice

The number of outcomes in the sample space.

$n(S) = 6^N$ ⟶ عدد الرميات أو عدد القطع

- Roll a dice ➔ S = {1, 2, 3, 4, 5, 6} ⟶ $N(s) = 6^1 = 6$

Tree diagram



A is even number = {2, 4, 6}

B is odd number = {1, 3, 5}

C is prime number = {2, 3, 5}

D = {4} is simple event

A and B are mutually exclusive: where $A \cap B = \phi$

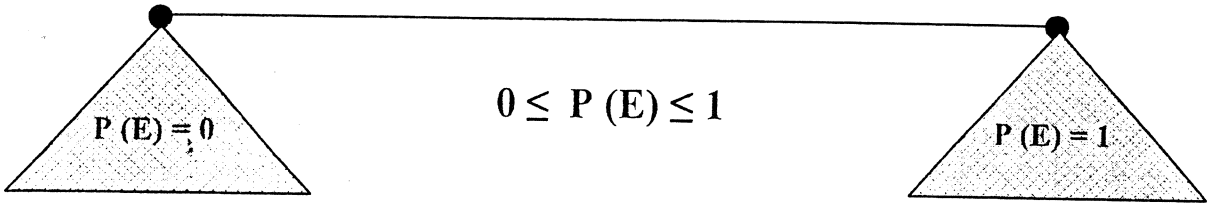
H. W

- Roll two dice → S = ... , n (s) =.....

  Tree diagram.....

Probability Rules

For any event E



- The Range of the values of the probability = [0 , 1]******
- P (E) = 0 where E can never occur.
- P (E) = 1 where E certain occur.
- $\sum_{i=1}^{n} P\ (a\,i) = 1$

The sum of the probabilities of all the outcomes in the sample space equal 1.

- $P(E') = 1 - P(E) \rightarrow P(E) + P(E') = 1$

Where E' is the complement of E

If s = {1, 2, 3 , 4, 5, 6}

A = {1, 2, 3, 4} → A′ = {5 , 6 }

$P(A) = \frac{4}{6} \rightarrow P(A') = \frac{2}{6}$

ALSAADI

11.Classify each statement as an example of classical probability, empirical probability, or subjective probability.

a) The probability that a person will watch the 6 o' clock evening news is 0.15 → **(Empirical)**

b) The probability of winning at a Chuck-a-Luck game is $\frac{5}{36}$ → **(Classical)**

c) The probability that a bus will be in an accident on a specific run is about 6% → **(Empirical)**

d) The probability of getting a royal flush when five cards are selected at random is $\frac{1}{649.740}$ → **(Classical)**

e) The probability that a student will get a C or better in a statistics course is about 70% → **(Empirical)**

f) The probability that a new fast-food restaurant will be a success in Chicago is 35% → **(Empirical)**

g) The probability that interest rates will rise in the next 6 months is 0.50 → **(Subjective)**

Note

Empirical	Classical	Subjective
• نسبة مئوية • فاصلة عشرية	• كسور	• يبنى على التخمين ويشير إلى المستقبل Next.....

10. A probability experiment is conducted. Which of these cannot be considered a probability of an outcome?

a. $\frac{1}{3}$	**b.** $-\frac{1}{5}$	**c. 0.80**
d. -0.59	**e. 0**	**f. 1.45**
g. 1	**h. 33%**	**i. 112%**

Solution

(b) $-\frac{1}{5}$

(d) -0.59

(f) 1.45

(i) 112% = 1.12

Are can not be considered a probability of an outcome.

Where $0 \leq P(E) \leq 1$

12. If a die is rolled one time, find these probabilities.

a. Of getting a 4.

b. Of getting an even number.

c. Of getting a number greater than 4.

d. Of getting a number less than 7.

e. Of getting a number greater than 0.

f. Of getting a number greater than 3 or an odd number.

g. Of getting a number greater than 3 and an odd number.

Solution

$S = \{1, 2, 3, 4, 5, 6\}$

(a) $a = \{4\} \rightarrow P(a) = \frac{1}{6}$

(b) $b = \{2, 4, 6\} \rightarrow P(b) = \frac{3}{6} = \frac{1}{2}$

(c) $c = \{5, 6\} \rightarrow P(c) = \frac{2}{6} = \frac{1}{3}$

(d) $d = \{1, 2, 3, 4, 5, 6\} \rightarrow P(d) = \frac{6}{6} = 1 \longrightarrow$ certain occur

(e) $e = \{1, 2, 3, 4, 5, 6\} \rightarrow P(e) = \frac{6}{6} = 1 \longrightarrow$ certain occur

(f) $f = \{4, 5, 6, 1, 3\} \rightarrow P(f) = \frac{5}{6}$

(g) $g = \{5\} \rightarrow p(g) = \frac{1}{6}$

Example:

A couple has three children find each probability:

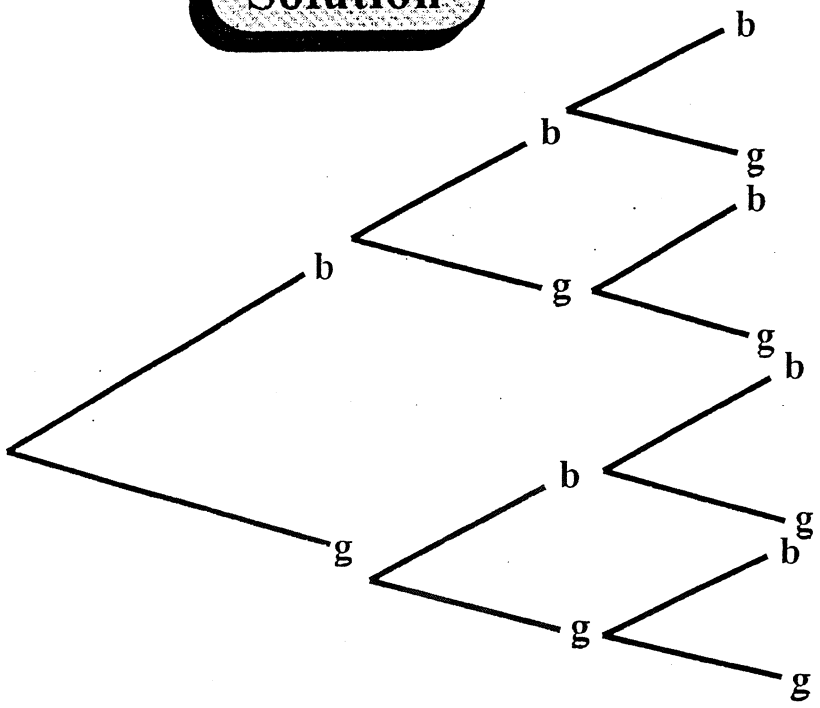
a) All boys

b) All girls or all boys

c) Exactly two boys or two girls.

D) At least one child of each gender.

Solution



S = { bbb , bbg , bgb , bgg , gbb, gbg , ggb , ggg}

(a) P (all boys) = $\frac{1}{8}$

(b) P (all girls or all boy) = $\frac{2}{8} = \frac{1}{4}$

(c) P (Exactly two boys or two girls) = $\frac{6}{8} = \frac{3}{4}$

(d) P (at least one child of each gender) = $\frac{6}{8} = \frac{3}{4}$

13. if two dice are rolled one time, find the probability of getting these results.

a. A sum of 6.

b. Doubles.

c. A sum of 7 or 11.

d. A sum greater than 9.

e. A sum less than or equal to 4.

Solution

S = {(1,1) , (1,2) ,....., (1,6)

(2,1) , (2,2),....., (2,6)

(6,1) , (6,2),....., (6,6) }

(a) a = {(1,5) , (2,4) , (3,3) (4,2) , (5,1)}

$$P(a) = \frac{5}{36}$$

(b) b = { (1,1) , (2,2) , (3,3) , (4,4) , (5,5) , (6,6)}

$$P(b) = \frac{6}{36} = \frac{1}{6}$$

(c) c = {(1,6) , (2,5) , (3,4) , (4,3) , (5,2) , (6,1) , (5,6) , (6,5)}

$$P(c) = \frac{8}{36} = \frac{2}{9}$$

(d) d = {(4,6) , (5,5) , (5,6) , (6,4) , (6,5) , (6,6) }

$$P(d) = \frac{6}{36} = \frac{1}{6}$$

(c) c= {(1,1) , (1,2) , (1,3) , (2,1) , (2,2) , (3,1)}

$$P(c) = \frac{6}{36} = \frac{1}{6}$$

Roll two dice and multiply the number together.

a. Write out the sample space.

b. What is the probability that the product is a multiple of 6 ?

c. What is the probability that the product is less than 10?

Solution

a) S = {(1,1) , (1,2) , (1,3) , (1,4) , (1,5) , (1,6),
(2,1) , (2,2) , (2,3) , (2,4) , (2,5), (2,6),
(3,1) , (3,2) , (3,3) , (3,4) , (3,5), (3,6),
(4,1) , (4,2) , (4,3) , (4,4) , (4,5), (4,6),
(5,1) , (5,2) , (5,3) , (5,4) , (5,5), (5,6),
(6,1) , (6,2) , (6,3) , (6,4) , (6,5), (6,6)}

b) P (Product is multiple of 6) = $\frac{15}{36}$

c) P (Product is less than 10) = $\frac{17}{36}$

19. For a recent year, 51% of the families in the United Slates had no children under the age of 18; 20% had one children 19% had two children; 7% had three children; and 3% had four or more children. If a family is selected at random, find the probability that the family has:

a. Two or three children
b. More than one child
c. Less than three children
d. Based on the answers to parts a, b, and c, which is most likely to occur? Explain why.

Solution

Information's:

* P (0 children) = 0.51 * P (1 children) = 0.20

* P (2 children) = 0.19 * P (3 children) = 0.07

* P (4 children or more) = 0.03

(a) P (2 or 3 children)

= P (2) + P (3)

= 0.19 + 0.07

= **0.26**

(b) P (more than one children)

= P (2) + P (3) + P (4 children or more)

= 0.19 + 0.07 + 0.03

= **0.29**

(c) P (less than three children)

= P (0) + P (1) + P (2)

= 0.51 + 0.20 + 0.19

= 0.90

(d) In part c the event is most likely to occur.

Because the probability is greeter than any one.

الحادثه وبكملتها

An event and its complement are ...

A) mutually exclusive B) not mutually exclusive C) independent D) dependent

الحادثه ومكملتها دائماً متنافيتان

mutually exclusive

Two events are ... if they cannot occur at the same time.

A) not mutually exclusive
B) independent events
C) dependent events
D) mutually exclusive ⇒ تعريف الحادثتان المتنافيتان

Two events that can occur at the same time are called ...

A) mutually exclusive. B) not mutually exclusive. C) independent. D) dependent.

لاحظ أن :

* Can not occur at the same time ⇒ mutually exclu.

* can " " " " " ⇒ not mutually exclu.

• أي وجود not في احد طرفي الجملة وعدم وجودها في الطرف الآخر .

تكملة

The complement of guessing 10 incorrect answers on a 10-question true/false exam is

A) guessing 10 incorrect answers
B) guessing at least 1 incorrect answer
C) guessing at least 1 correct answer
D) guessing 10 correct answers

Complement: guessing 10 incorrect answers

is ⇒ guessing at least 1 correct

When 10 adults are tested for high blood pressure, the complement of at least one of the results are positive.

A) None of the adults have high blood pressure.
B) Nine of the adults have high blood pressure.
C) All of the adults have high blood pressure.
D) One of the adults have high blood pressure.

تكملة واحد على الأقل ايجابي

Complement of at least one are positive

لا أحد ايجابي

None of the adults have high blood Pressure

يتأثر ب
Two events are said to be ... if the probability of one event occurring is affected by the occurrence of the other event.
A) not mutually exclusive B) independent C) dependent D) mutually exclusive

تعريف: اذا كان احتمال حدوث حادثة يتأثر ب حدوث حادثة أخرى.

فإنه يقال أن الحادثتان مرتبطتان (أي غير مستقلتان)

↳ dependent.

(84) ..... probability uses a frequency distribution to compute probabilities.
A) Classical B) Empirical C) Subjective D) continuous

التوزيع التكراري
(use frequency distribution)

↳ Empirical تجريبي

Consider this table to answer the following two questions:

Cookie Types	Number Sold
Chocolate Chip	20
Peanut Butter	15
Oatmeal	30
Sugar	10
	$\Sigma = 75$

243

What is the level of measurement for the cookie type

نوع الكيك
مقياس أسمى .

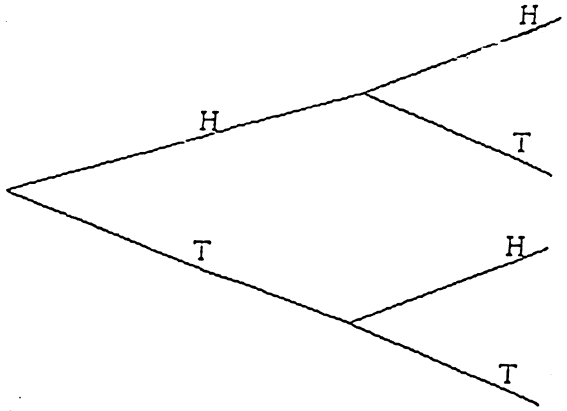
A) Discrete B) Ordinal C) Nominal D) Continuous

What is the probability of selecting a chocolate chip cookie?

A) 4/15 B) 1/2 C) 1/15 D) 1/4

$$P(\text{chocolate chip}) = \frac{20}{75} = \frac{4}{15}.$$

How many times was the coin tossed in the figure below?



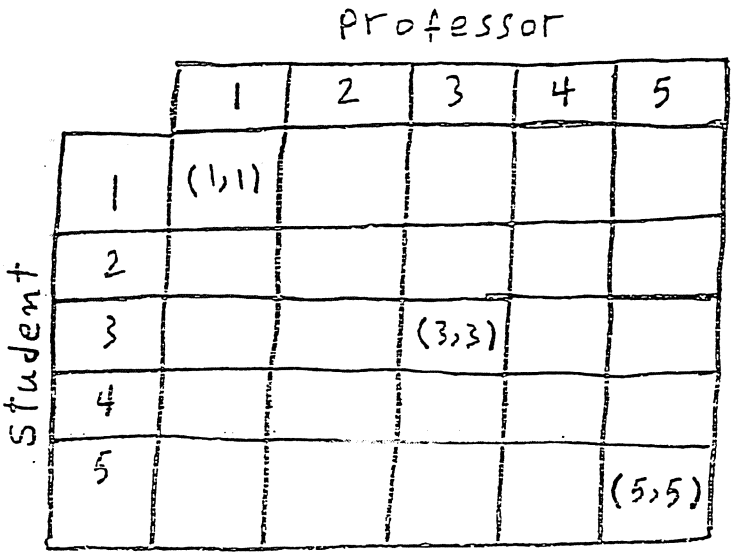
A) 3 B) 6 C) 4 D) 2

من الشكل
كم عدد مرات
القاء قطعه النقود

" مرتين "

A student and a professor each choose a number between 1 and 5 (1 and 5 are both possible choices). What is the probability that the two choose the same odd number?

A) 0.04 B) 0.08 C) 0.12 D) 0.16



اختيار نفس العدد الفردي

$$P = \frac{n(A)}{n(S)} = \frac{3}{25} = 0.12$$

عنصر واحد

Consider the experiment of selecting one item at random from a box containing an equal number of defective (D) and non-defective(N) items. The sample space for this experiment is

A) 2 B) S={D,N} C) 4 D) S={DD,DN,ND,NN}

defective (D) معيبة صندوق يحتوي على

non-defective (N) غير معيبة

حب عنصر واحد فقط اما N أو D

$S = \{D, N\}$

ALSAADI

The complement of guessing at least 5 correct answers on a 6-question true/false exam is
A) guessing 6 incorrect answers
B) guessing at least 2 incorrect answer
C) guessing at least 2 correct answer
D) guessing 6 correct answers

الحادثه المكمله لتخمين 5 أجابات صحيحه على الأقل من ضمن 6 أسئله .

guessing at least 2 incorrect هي على الأقل 2 خطأ

الحادثه: at least 5 correct

المكمله: at least 2 incorrect

$S = \{ 6\checkmark , (5\checkmark, 1X) , (4\checkmark, 2X), (3\checkmark, 3X), (2\checkmark, 4X), (1\checkmark, 5X), 6X \}$

The complement of guessing 4 correct answers on a 4-question true/false exam is
A) guessing 4 incorrect answers
B) guessing at least 1 incorrect answer
C) guessing at least 1 correct answer
D) guessing no incorrect answers

Complement " All correct"

= At least 1 incorrect

S={(4 correct), (3 correct and 1 incorrect), (2 correct and 2 incorrect), (1 corrent and 3 inco.), (4 incorrect)}

Use the following to answer questions

Two dice are rolled. Let X represents the summation of the two faces that will appear.

Die 2 \ Die 1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The probability of $X \geq 10$ is ...
A) 0.167 B) 0 C) 0.028 D) 0.083

X : is sum. of two faces

$X \geq 10 = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\}$

$P(X \geq 10) = \frac{\text{عدد الازواج المرتبة السابقة}}{\text{العدد الكلي للأزواج المرتبة}} = \frac{6}{36} = 0.166 \simeq \boxed{0.167}$

The probability of $X \leq 12$ is ...
A) 0 B) 0.083 C) 0.056 D) 1

$X \leq 12$ مجموع الوجهين أقل من أو يساوي 12

عدد الأزواج المرتبة التي تحقق هذا الشرط 36

$\therefore P(X \leq 12) = \frac{36}{36} = \boxed{1}$ ⟵ certain occur حادثة مؤكدة

Use the following to answer questions

مجموع الوجهين.

If two dice are rolled and X is the random variable represents the summation of two faces that will appear

Die 2 \ Die 1	1	2	3	4	5	6
1	2	3	(4)	5	6	7
2	3	(4)	5	6	7	8
3	(4)	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	(12)

7) The probability of $X = 4$ is

A) 0.083 B) 0.139 C) 0.861 D) 0.917

$$P(X=4) = \frac{3}{36} = 0.083$$

مجموع الوجهين 4

8) The probability of $X = 12$ is

A) 0.972 B) 0.917 C) 0.028 D) 0

$$P(X=12) = \frac{1}{36} = 0.028$$

مجموع الوجهين 12

When 2 dice (6-sided each) are rolled , find the probability of getting a sum of 13

A) 0.25 B) 0.72 C) 0.17 D) 0

مجموع الوجهين 13

صادث مستحيله ∅

لأنه أكبر مجموع ممكن الحصول عليه هو 12 من ← (6,6)

$$P(\text{sum of } 13) = P(\emptyset) = 0$$

عدد زوجي

If a die were rolled, the event of getting an even number would be called a ....
A) compound event B) simple event C) sample space D) sample size

even number = {2, 4, 6} حادثة الحصول على عدد زوجي
تتكون من أكثر من عنصر
← Compound event
حادثة مركبة .

If there is a 20% chance that it will rain tomorrow, what is the probability that it will not rain tomorrow?
تمطر غداً — لا تمطر

A) 0.08 B) 0.20 C) 0.80 D) 0

P(rain) = 20% = 0.20

P(not rain) = 1 − P(rain) = 1 − 0.20

متماثلة

If there are 20 equally likely events, then the probability of the second one occurring is ....
A) 1 B) 1/10 C) 1/20 D) 0

يوجد 20 حادثة متماثلة (لها نفس احتمالية الحدوث)

∴ احتمال حدوث اي حادثة منهم

$\frac{1}{20}$

* لو كان عدد الحوادث 50 كانت الاجابة $\frac{1}{50}$

ALSAADI

The number of outcomes in a compound event can be ...
A) E={2, 4, 6} B) 3 C) 1 D) E={3}

عدد عناصر الحادثة المركبة (أكثر من عنصر
ولكن ليس على شكل مجموعة { }

What type of probability uses sample spaces to determine the numerical probability that an event occurs?
A) Empirical probability
B) Classical probability
C) Subjective probability
D) Conditional probability

تعريف

Ueses sample spaces
to determine numerical probability
↳ Classical probability

A child gets 22 heads out of 30 tosses of a coin. This would be an example of ... type of probability.
A) empirical B) subjective C) classical D) empirical and classical

عدد محاولات التجربة 30
حصل منهم الطفل على 22 صورة
∴ احتمال تجريبي

empirical

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

10

Ch. 4 - Part 2

- The Addition Rules for Probability.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 4 Part. 2

The addition rules for probability

Two events are **mutually exclusive** events if they cannot occur at the same time they have no outcomes in common.

Addition Rule 1

When two events A and B are mutually exclusive, the probability that A or B will occur is

$P(A \text{ or } B) = P(A) + P(B)$

Addition Rule 2

If A and B are not mutually exclusive, then

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

In a hospital unit there are 8 nurses (ممرض) and 5 physicians (طبيب); 7 nurses and 3 physicians are females. If a staff person is selected, find the probability that the subject (شخص) is a.

(1) Nurse or male. (2) Physicians or females.

Solution

Staff	Females	Males	Total
Nurses	7	1	8
Physicians	3	2	5
Total	10	3	13

(1) P (Nurse or male)

= P (Nurse) + P (male) – P (nurse and male)

$$= \frac{8}{13} + \frac{3}{13} - \frac{1}{13} = \frac{10}{13}$$

(2) P (Physician or female)

= P (Physician) + P (female) – P (Physician and female)

$$= \frac{5}{13} + \frac{10}{13} - \frac{3}{13} = \frac{12}{13}$$

2. Determine whether these events are mutually exclusive:

a. Roll a die: Get an even number, and get a number less than 3.

b. Roll a die: Get a prime number, and get an odd number.

c. Roll a die: Get a number greater than 3,
and get a number less than 3.

أشقر

d. Select a student in your class: The student has blond hair,
and the student has blue eyes.

e. Select a student in your college: the student is a
طالب في السنة الثانية تجارة تخصص
sophomore, and the student is a business major.

f. Select any course: it is a calculus course,
and it is an English course.

الناخب جمهوري

g. Select a registered voter: the voter is a Republican,
and the voter is a democrat.

Solution

(a) $A = \{2,4,6\}$ $B = \{1,2\}$

$A \cap B = \{2\}$ → not mutually exclusive.

(b) $A = \{2,3,5\}$ $B = \{1,3,5\}$

$A \cap B = \{3,5\}$ → not mutually exclusive.

(c) $A = \{4,5,6\}$ $B = \{1,2\}$

$A \cap B = \phi$ → A and B mutually exclusive.

(d) A: blond hair B = blue eyes

$A \cap B \neq \phi$ → not mutually exclusive.

(e) A: sophomore B: business major

$A \cap B \neq \phi$ → not mutually exclusive.

(f) A: calculus course B: English course

$A \cap B = \phi \rightarrow$ A and B mutually exclusive

(g) A: Republican B: Democrat

$A \cap B = \phi \rightarrow$ A and B mutually exclusive

Example:

At a convention there are 7 mathematics instructors, 5 computer science instructors, 3 statistics instructors, and 4 science instructors. If an instructor is selected, find the probability of getting a science instructor or a math instructor.

Solution

Total instructors = 7 + 5 + 3 + 4 = 19

P (science instructor or math instructor)

= P (science instructor) + P (math instructor)

$$= \frac{4}{19} + \frac{7}{19} = \frac{11}{19}$$

7. A recent study of 200 nurses found that of 125 female nurses, 56 had bachelor's degrees; and of 75 male nurses, 34 had bachelor's degrees. If a nurse is selected at random, find the probability that the nurse is

a. A female nurse with a bachelor's degree.

b. A male nurse.

c. A male nurse with a bachelor's degree.

d. Based on your answers to parts a, b, and c,
Explain which is most likely to occur. Explain why.

Solution

	Male	Female
Bachelor's degree	34	56
Without bachelor degree	75-34 = 41	125-56 = 69
Total	75	125

(a) $P(A) = \frac{56}{200} = 0.28$

(b) $P(B) = \frac{75}{200} = 0.38$

(c) $P(C) = \frac{34}{200} = 0.17$

(d) Event B is most likely to occur

Because the probability is greeter than any one.

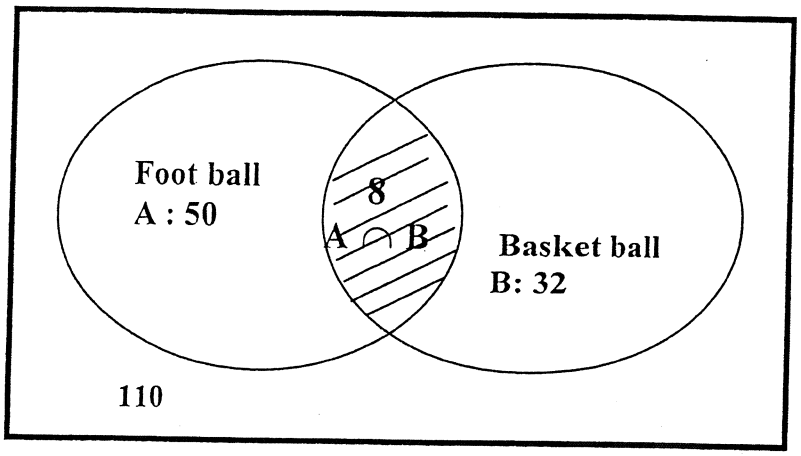
9. At a particular (خاصة) school (مدرسة) with 200 male students, 58 play football, 40 play basketball, and 8 play both.

What is the probability that a randomly selected male student?

a. Plays neither sport.

b. Plays (Football or Basketball).

c. Plays (Football and Basketball).



a. P (Play Neither sport) $= \frac{110}{200} = 0.55$

b. P (Play: Football or basketball)

$= P(A \text{ or } B)$

$= P(A) + P(B) - P(A \cap B)$

$= \frac{58}{200} + \frac{40}{200} - \frac{8}{200} = \frac{90}{200} = 0.45$

c. P (Play: Football and basketball)

$= \frac{8}{200} = 0.04$

مركز تجاري لمقايضة السيارات المخزن

13. The Bargain Auto Mall has these cars in stock.

	SUV	سيارة صغيرة Compact	متوسطة الحجم Mid – Sized
Foreign أجنبي	20	50	20
Domestic داخلي - وطني	65	100	45

If a car is selected at random, find the probability that it is:

a. Domestic

b. Foreign and mid – sized

c. Domestic or an SUV.

Solution

Total cars = 20 + 50 + 20 + 65 + 100 + 45 = 300

(a) $P(\text{Domestic}) = \frac{65 + 100\ 45}{300} = 0.7$

(b) $P(\text{Foreign and mid – sized}) = \frac{20}{300} = 0.07$

(c) P (Domestic or SUV)

= P (Domestic) + P (SUV) – P (Domestic and SUV)

$$= \frac{65 + 100 + 45}{300} + \frac{20 + 65}{300} - \frac{65}{300} = 0.77$$

25. An urn contains 6 red balls, 2 green balls, 1 blue ball and 1 white ball. If a ball is drawn, find the probability of getting:

(a) Red or green.

(b) blue or white

(c) not green

(d) blue and white

Solution

Total balls = 6 + 2 + 1 + 1 = 10

(a) P (red or green) = P (red) + P (green) $= \frac{6}{10} + \frac{2}{10} = \underline{\underline{0.8}}$

(b) P (blue or white) = P (blue) + P (white) $= \frac{1}{10} + \frac{1}{10} = \underline{\underline{0.2}}$

(c) P (not green) = 1 – P (green) $= 1 - \frac{2}{10} = \underline{\underline{0.8}}$

(d) P (blue and white) = P (ϕ) $= \underline{\underline{0}}$

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

11

Ch. 4 - Part 3

- The Multiplication Rules and Conditional Probability.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 4 Part. 3

The Multiplication Rules and Conditional Probability

Two events A and B are independent events if :

A occurs does not affect the probability of B occurring.

Multiplication Rule 1

When two events are independent, the probability of both occurring is P(A and B) = P(A) × P(B)

Example:

An urn contains 3 red balls, 2 blue balls, and 5 white balls.
A ball is selected and its color noted. Then it is replaced. A second ball is selected and its color noted. Find the probability of each of these.

a. Selecting 2 blue balls.

b. Selecting 1 blue ball and then 1 white ball.

c. Selecting 1 red ball and then 1 blue ball.

Solution

a. P (blue and blue) = P(blue) × P(blue) = $\frac{2}{10} \times \frac{2}{10} = \frac{4}{100} = \frac{1}{25}$

b. P (blue and white) = P(blue) × P(white) = $\frac{2}{10} \times \frac{5}{10} = \frac{10}{100} = \frac{1}{10}$

c. P (red and blue) = P(red) × P(blue) = $\frac{3}{10} \times \frac{2}{10} = \frac{6}{100} = \frac{3}{50}$

Example:

Approximately 9% of men have a type of color blindness that prevents them from distinguishing between red and green. If 3 men are selected at random, find the probability that all of them will have this type of red-green color blindness.

Solution

Let C denote red-green color blindness. Then

$P(\text{C and C and C}) = P(C) \times P(C) \times P(C)$

$= (0.09)(0.09)(0.09)$

$= 0.000729$

Example:

A coin is flipped and a die is rolled. Find the probability of getting a head on the coin and a 4 on the die.

Solution

$$P(\text{head and 4}) = P(\text{head}) \times P(4) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$$

Note

- **IF $P(A) < 0.5$**

A is unlikely to occur

- **If $P(A) \geq 0.5$**

A is likely to occur

- **If $P(A) = L$**

➔ **$P(\text{none A}) = 1 - L$**

Example:

If 28% of U.S. medical degrees are conferred to women, find the probability that 3 randomly selected medical school graduates are men. Would you consider this event likely or unlikely to occur? Explain your answer.

Solution

- P (W) = 0.28
- P (M) = 1 – P (W)

  = 1 – 0.28 = 0.72

  P (3M) = P (M) . P (M) , P (M)

  = (0.72) (0.72) (0.72)

  = 0.373

The event is unlikely to occur because P (3M) < 0.5

Example:

Eighty-eight percent of U.S. children are covered by some type of health insurance. If 4 children are selected at random, what is the probability that none are covered?

Solution

P(covered) = 0.88

P(non covered) = 1 – 0.88 = 0.12

P(4 children are non covered) = (0.12) (0.12) (0.12) (0.12) = 0.0002

Multiplication Rule 2

When two events are dependent, the probability of both occurring is

$P(A \text{ and } B) = P(A) \times P(B/A)$

Example:

A person owns a collection of 30 CDs, of which 5 are country music. If 2 CDs are selected at random, find the probability that both are country music.

Solution

Since the events are dependent,

$$P(C_1 \text{ and } C_2) = P(C_1) \times P(C_2 \mid C_1) = \frac{5}{30} \times \frac{4}{29} = \frac{20}{870} = \frac{2}{87}$$

Example:

In a civic organization, there are 38 members; 15 are men and 23 are women. If 3 members are selected to plan the July 4th parade, find the probability that all 3 are women. Would you consider this event likely or unlikely to occur? Explain your answer.

Solution

The total members

= 15 men + 23 women = 38

$$P(3 \text{ women}) = \frac{23}{38} \times \frac{22}{37} \times \frac{21}{36} = 0.21 \quad < 0.5$$

There for: This event unlikely to occur.

Conditional probability

- P(A/B)

Probability that A occur
After B already occurred

$$P(A/B) = \frac{P(A \text{ and } B)}{P(B)} \rightarrow P(A \text{ and } B) = P(B) \times P(A/B)$$

$$P(B/A) = \frac{P(A \text{ and } B)}{P(A)} \rightarrow P(A \text{ and } B) = P(A) \times P(B/A)$$

Example:

A recent survey asked 100 people if they thought women in the armed forces should be permitted to participate in combat. The results of the survey are shown.

Gender	Yes	No	Total
Male	32	18	50
Female	8	42	50
Total	40	60	100

Find these probabilities.

a. The respondent answered yes, given that the respondent was a female.

b. The respondent was a male, given that the respondent answered no.

Solution

Let: M = Male F = Female
Y = yes N = No

$$\text{a. } P(Y/F) = \frac{P(Y \text{ and } F)}{P(F)} = \frac{\frac{8}{100}}{\frac{50}{100}} = \underline{\underline{0.16}}$$

$$\text{b. } P(M/N) = \frac{P(M \text{ and } N)}{P(N)} = \frac{\frac{18}{100}}{\frac{60}{100}} = \underline{\underline{0.3}}$$

Example:

An insurance company classifies drivers as low-risk, medium-risk, and high-risk. Of those insured, 60% are low-risk, 30% are medium-risk, and 10% are high-risk. After a study, the company finds that during a 1-year period, 1% of the low-risk drivers had an accident, 5% of the medium-risk drivers had an accident, and 9% of the high-risk drivers had an accident. If a driver is selected at random, find the probability that the driver will have an accident during the year.

Low
P(A) × P(B/A)
(0.60) × (0.01)

Medium
P(A) × P(B/A)
(0.30) × (0.05)

High
P(A) × P(B/A)
(0.10) × (0.09)

Solution

P(have an accident)

= P(low – risk and have an accident) ➔ (0.6) (0.01)

+ P(medium – risk and have an accident) ➔ (0.3) (0.05)

+ P(high – risk and have an accident) ➔ (0.1) (0.09)

= 0.03

حالة الـ Coin

At least one

Find the probability of getting at least one .....

- (1) A coin is tossed 3 times:
  $N(s) = 2^3 = 8$

  $\therefore P(\text{at least one tail}) = \frac{N(s) - 1}{N(s)} = \frac{8-1}{8} = \frac{7}{8}$

- (2) A coin is tossed 5 times :
  $N(s) = 2^5 = 32$

  $\therefore P(\text{at least one head}) = \frac{N(s) - 1}{N(s)} = \frac{32-1}{32} = \frac{31}{32}$

حالة النسب المئوية

Rule

- **A: at least one .....**
- **A′: no ..... ≡ .....**
- **P(A′) = ( ) ( ) ( ).....**
- **P(A) = 1 - P(A′)**

Example:

It has been found that 6% of all automobiles on the road have defective brakes. If 5 automobiles are stopped and checked by the state police, find the probability that at least one will have defective brakes.

Solution

A = at least one have defective brakes
A′ = no have defective brakes
P(defective) = 0.06 P(undefective) = 1 – 0.06 = 0.94
P(A′) = (0.94) (0.94) (0.94) (0.94) (0.94) = **0.7339**
P(A) = 1 – P(A′) = 1 – 0.7339 = **0.266**

ALSAADI

Use the following to answer questions

An apartment building has the following apartments:

	1st Floor	2nd Floor	3rd Floor	Total
2 Bedroom	3	1	2	6
3 Bedroom	1	3	2	6
Total	4	4	4	12

If an apartment is selected at random,

what is the probability that it is on the 2nd floor (A) or (∪) has 3 bedrooms (B)?

A) 4/12 B) 6/12 C) 7/12 D) 3/12

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

$$= \frac{4}{12} + \frac{6}{12} - \frac{3}{12} = \boxed{\frac{7}{12}}$$

what is the probability that it is a 3 bedroom apartment (B) given that it is on the 3rd floor (A)?

A) 2/12 B) 6/12 C) 4/12 D) 1/12

** given that كلمة
تدل على الاحتمال المشروط
A قبل given that مثل
B " " " " بعد

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{12}}{\frac{4}{12}} =$$

$$= \frac{2}{4} = \frac{1}{2} = \boxed{\frac{6}{12}}$$

ALSAADI

Use the following to answer questions

In a recent study, the following data was obtained in response to the question, "Do you favor recycling in your neighborhood? "

	Yes	No	No opinion	Total
Males	23	17	8	48
Females	7	8	12	27
Total	30	25	20	75

If a person is selected at random, use the above table to answer the following questions.

The probability that a person is a female given that she answered yes regarding recycling is:
A) 0.68 B) 0.32 C) 0.767 D) 0.233

$$P(F/y) = \frac{P(F \cap y)}{P(y)} = \frac{7/75}{30/75} = \frac{7}{30} \approx \boxed{0.233}$$

What is the probability that a person has no opinion regarding recycling?
A) 0.267 B) 0.333 C) 0.4 D) 0.64

$$P(\text{no opinion}) = \frac{20}{75} \approx \boxed{0.267}$$

What is the probability that a person is a male and he answered no regarding recycling?
A) 0.107 B) 0.227 C) 0.093 D) 0.16

$$P(m \cap No) = \frac{17}{75} \approx \boxed{0.227}$$

The probability that a person is a male or he has no opinion regarding recycling is:
A) 0.467 B) 0.8 C) 0.747 D) 0.587

$$P(m \cup \text{no opinion})$$
$$= P(m) + P(\text{no opinion}) - P(m \cap \text{no opinion})$$
$$= \frac{48}{75} + \frac{20}{75} - \frac{8}{75} = \boxed{0.8}$$

The manager of a bank recorded the amount of time each customer spent waiting in line during peak (الذروة) business hours one Monday. The frequency table below summarizes the results.

Waiting Time (minutes)	Number of Customers
0-3	14
4-7	9
8-11	11
12-15	6
16-19	7
20-23	3
24-27	2

$n = \Sigma f = 52$

If we randomly select one of the customers represented in the table, what is the probability that the waiting time is at least 12 minutes or between 8 and 15 minutes?

A) 0.519 B) 0.63 C) 0.558 D) 0.2

$P(\text{at least 12 minutes } \underset{\cup}{\text{or}} \text{ between 8 and 15 minutes})$

A — B

$= P(A \cup B)$

$= P(A) + P(B) - P(A \cap B)$

$= \frac{18}{52} + \frac{17}{52} - \frac{6}{52}$ بالرد $= \frac{29}{52} \simeq \boxed{0.558}$

The probability that a student has a computer is 0.91 and the probability that he has a car is 0.49 while the probability that he has both is 0.46. Find the probability that the student has a computer given that he has a car.

A) 0.82 B) 0.51 C) 0.94 D) 0.65

$P(\text{has computer}) = 0.91$ ، $P(\text{has car}) = 0.49$ ، $P(\text{has both}) = 0.46$

(has computer given that has car)

$= P(\text{has computer} \mid \text{has car})$

given that كلمة تعني احتمال مشروط

$= \frac{P(\text{has comp. and has car})}{P(\text{has car})}$

$= \frac{P(\text{both})}{P(\text{car})} = \frac{0.46}{0.49} = 0.938 \simeq \boxed{0.94}$

Use the following to answer questions

The table below shows the number of earned degrees in the year 2008 in a university by level and gender. A person who earned a degree in the year 2008 from this university is randomly selected. Find the probability of selecting someone who

		Male	Female
Level of Degree	Bachelor's	240	180
	Master's	35	15
	PhD's	25	5

is a female given that the person earned a bachelor's degree.
A) 0.83 B) 0.57 C) 0.17 D) 0.43

* مجموع الدرجات العليا

$240 + 180 + 35 + 15 + 25 + 5 = 500$

* كلمة given that تعني احتمال مشروط Conditional probability

$P(\text{Female} / \text{bach.})$

$= \dfrac{P(\text{Female} \cap \text{bach.})}{P(\text{bach.})}$

* ليجاد التقاطع خط افقي مع جذر رأسي

$= \dfrac{\frac{180}{500}}{\frac{240+180}{500}} = \dfrac{180}{240+180} = 0.428 \simeq \boxed{0.43}$

earned a master's degree or is a female.
A) 0.47 B) 0.63 C) 0.45 D) 0.61

$P(\text{master} \cup \text{Female})$

$= P(\text{mas.}) + P(\text{F.}) - P(\text{mas.} \cap \text{F.})$ (خط افقي مع جذر رأسي)

or $\longrightarrow \cup$
and $\longrightarrow \cap$
given that $\longrightarrow /$

$= \dfrac{50}{500} + \dfrac{200}{500} - \dfrac{15}{500}$

$= \dfrac{235}{500} = \boxed{0.47}$

The probability that a student has a computer (A) is 0.82 and the probability that he has a car (B) is 0.48 while the probability that he has either a computer or a car ($A \cup B$) is 0.86. Find the probability that the student has both.

A) 0.38 B) 0.44 C) 0.34 D) 0.52

* تطبيق قانون ثم حل معادلة

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

$A \cap B$: has both

$$0.68 = 0.82 + 0.48 - P(A \cap B)$$

$$P(A \cap B) = 0.82 + 0.48 - 0.68 = \boxed{0.44}$$

Use the following to answer questions

A supermarket employs cashiers, managers and cleaner. The distribution of employees according to marital status is shown here.

	Cashier	Manager	Cleaner	Total
Married	9	12	3	24
Not married	5	15	2	22
Total	14	27	5	46

Find the probability that ...

$P(A \cup B)$

the employee is a manager or married

A) 24/46 B) 12/46 C) 27/46 D) 39/46

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

$$= \frac{27}{46} + \frac{24}{46} - \frac{12}{46} = \frac{39}{46}$$

$P(A / B)$

the employee is a cashier given that he is married.

A) 9/24 B) 9/22 C) 9/46 D) 9/14

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{9}{46}}{\frac{24}{46}} = \frac{9}{24}$$

Use the following to answer questions

Two dice are rolled. Let X represents the summation of the two faces that will appear.

المتغير X يمثل
مجموع الوجهين الظاهرين

Die 2 \ Die 1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The probability of X= 4 is
A) 0.083 B) 0.833 C) 0 D) 0.028

X : sum 4 $= \{ (1,3) , (2,2) , (3,1) \}$

$$P(X=4) = \frac{3}{36} = 0.083$$

The probability of X=15 is
A) 0.056 B) 0.028 C) 0.083 D) 0

$P(X=15)$

لا يوجد وجهين زهرة النرد مجموعهما 15

$= P(\phi) = \text{Zero}$

ALSAADI

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

12

Ch. 4 - Part 4

- Counting Rules.
- Probabiltiy and Counting Rules.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 4 Part. 4

Counting Rules

- Fundamental Counting rule:

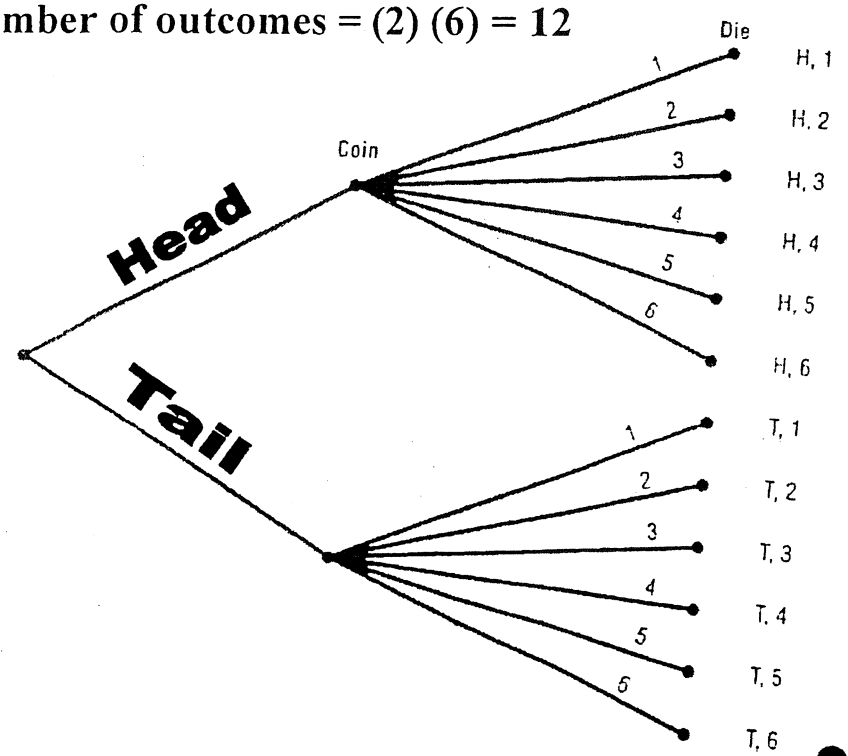
In a sequence of n events in which the first one has k_1 possibilities and the second event has k_2 and the third has k_3, and so forth, the total number of possibilities of the sequence will be $K_1 \times K_2 \times K_3 \times \ldots\ldots\ldots \times K_n$

Example:

A coin is tossed and a die is rolled. Find the number of outcomes for the sequence of events.

Solution

The number of outcomes = (2) (6) = 12



Example:

The digits 0, 1,2, 3, and 4 are to be used in a **four-digit** ID card. How many different cards are possible if repetitions are permitted?

5	5	5	5

Solution

Since there are 4 spaces to fill and 5 choices for each space,

The number of cards $= 5 \times 5 \times 5 \times 5 = 5^4 = 625$

Permutations

A permutation is an arrangement of n objects in a specific order.

Factorial Formulas

For any counting n

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

$5! = (5)(4)(3)(2)(1) = 120$

$0! = 1$

Permutation Rule

The arrangement of n objects in a specific order using r objects at a time is called a permutation of n objects taking r objects at a time. It is written as ${}_nP_r$, and the formula is

$${}_nP_r = \frac{n!}{(n-r)!}$$

- order is important

$${}_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1} = 60$$

ممكن بالآلة بدون الفك

$${}_5P_5 = \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{1} = 120$$

${}_5P_5 = 5!$ **$0! = 1$** **${}_nP_n = n!$**

ALSAADI

Example:

How many different ways can a chairperson and an assistant chairperson be selected for a research project if there are seven scientists available?

Solution

$${}_7P_2 = \frac{7!}{(7-2)!} = \frac{7!}{5!} = 42$$

Example:

A store manager wishes to display 8 different brands of shampoo in a row. How many ways can this be done?

Solution

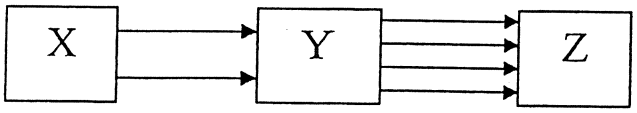
Numbers of ways

$= 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 8! = 40320$

Example:

There are 2 major roads from city X to city Y and 4 major roads from city Y to city Z. How many different trips can be made from city X to city Z passing through city Y?

Solution



Numbers of ways $= 2 \times 4 = 8$

Example:

If 50 tickets are sold and 2 prizes are to be awarded.

Find the probability that one person will win 2 prizes if that person buys 2 tickets.

Solution

التذكرتان x , y

(x ثم y) أو (y ثم x)

$$P(2\text{ prizes}) = \left(\frac{1}{50} \times \frac{1}{49}\right) + \left(\frac{1}{50} \times \frac{1}{49}\right) = \frac{1}{1225}$$

Combination Rule

The number of combinations of r objects selected from n objects is denoted by ${}_nC_r$ and is given by the formula: $${}_nC_r = \frac{n!}{(n-r)!r!}$$

Example:

In a club there are 7'women and 5 men. A committee of 3 women and 2 men is to be chosen.

How many different possibilities are there?

Solution

$${}_7C_3 \times {}_5C_2 = \frac{7!}{(7-3)!3!} \times \frac{5!}{(5-2)!2!} = 350$$

How many different tests can be made from a test bank of 20 questions if the test consists of 5 questions?

Solution

Number of tests = $20\ C_5 = 15504$

Example:

There are 7 women and 5 men in a department.

1. How many ways can a committee of 4 people be selected?
2. How many ways can this committee be selected if there must be 2 men and 2 women on the committee?
3. How many ways can this committee be selected if there must be at least 2 women on the committee?

Solution

1- number of committee = $12\ C_4 = 495$

2- number of committee = $7\ C_2 \times 5C_2 = 210$

3- number of committee where at least 2 women

(2 w and 2 m) or (3 w and 1 m) or 4 w)

$= (7\ C_2 \times 5C_2) + (7\ C_3 \times 5C_1) + (7\ C_4)$

$= (21 \times 10) + (35 \times 5) + 35$

$= 420$

Example:

How many ways can a dinner patron select 3 appetizers and 2 vegetables if there are 6 appetizers and 5 vegetables on the menu?

Solution

appetizers vegetables

Number of ways = $6C_3 \times 5C_2$

$= 20 \times 10 = 200$

Example:

How many different ways can an instructor select 2 textbooks from a possible 17?

Solution

Number of ways = $17C_2 = 136$

Example:

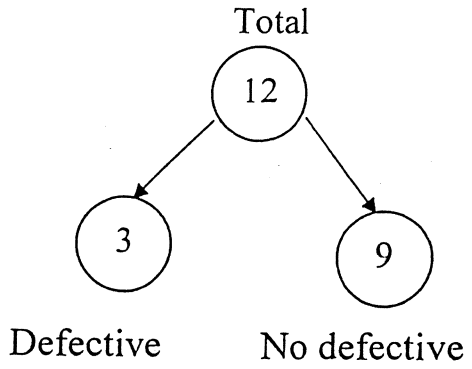
A package contains 12 resistors, 3 of which are defective.
If 4 are selected, find the probability of getting
a. No defective resistors
b. 1 defective resistor
c. 3 defective resistors

Solution

a. $P(\text{No defective}) = \dfrac{9C_4}{12C_4} = \underline{0.255}$

b. $P(1 \text{ defective}) = \dfrac{3C_1 \times 9C_3}{12C_4} = \underline{0.509}$

c. $P(3 \text{ defective}) = \dfrac{3C_3 \times 9C_1}{12C_4} = \underline{0.018}$



Example:

A box contains 24 transistors, 4 of which are defective. If 4 are sold at random, find the following probabilities.

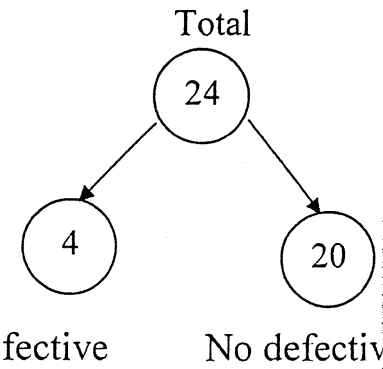
a. Exactly 2 are defective.

b. None is defective.

c. All are defective,

d. At least 1 is defective.

Solution



a.

$$P(\text{exactly 2 defectives}) = \frac{{}_4C_2 \times {}_{20}C_2}{{}_{24}C_4} = \frac{1140}{10626} = \frac{190}{1771}$$

b.

$$P(\text{no defectives}) = \frac{{}_{20}C_4}{{}_{24}C_4} = \frac{4845}{10626} = \frac{1615}{3542}$$

c.

$$P(\text{all defective}) = \frac{{}_4C_4}{{}_{24}C_4} = \frac{1}{10626}$$

d.

$$P(\text{at least 1 defective}) = 1 - P(\text{no defectives})$$

$$= 1 - \frac{{}_{20}C_4}{{}_{24}C_4} = 1 - \frac{1615}{3542} = \frac{1927}{3542}$$

ALSAADI

Example:

A store has 6 TV Graphic magazines and 8 News time magazines on the counter. If two customers purchased a magazine, find the probability that one of each magazine was purchased.

Solution

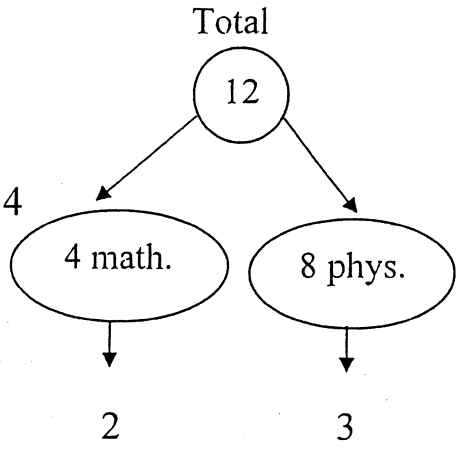
$$P(\text{1 TV Graphic and 1 News time}) = \frac{{}_6C_1 \times {}_8C_1}{{}_{14}C_2} = \frac{6 \times 8}{91} = \frac{48}{91}$$

Example:

Find the probability of randomly selecting 2 mathematics books and 3 physics books from a box containing 4 mathematics books and 8 physics books.

Solution

$$P(\text{2 math and 3 phys.}) = \frac{{}_4C_2 \times {}_8C_3}{{}_{12}C_5} = 0.424$$



Example:

An insurance sales representative select 3 policies to review. The group of policies she can select from contains 8 life policies, 5 automobile policies, and 2 homeowner policies.

Find the probability of selecting

a. All life policies

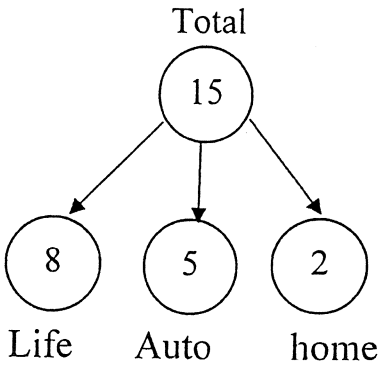
b. Both homeowner policies

c. All automobile policies

d. 1 of each policy

e. 2 life policies and 1 automobile policy

Solution



(a) $P(\text{All life}) = \frac{8C_3}{15C_3} = 0.123$

(b) $P(\text{Both homeowner}) = \frac{2C_2 \times 13C_1}{15\,C_3} = 0.029$

(c) $P(3\ \text{Auto}) = \frac{5C_3}{15C_3} = 0.022$

(d) $P(1\ \text{of each policy}) = \frac{8C_1 \times 5C_1 \times 2C_1}{15\,C_3} = 0.176$

(e) $P(2\ \text{life and}\ 1\ \text{Auto}) = \frac{8C_2 \times 5\,C_1}{15\,C_3} = 0.308$

Example:

There are **2 math** – students and **5 stat** – students in a class

How many ways can a group of **3 students be selected** if there must be **at least one math** – student on this group ?

Solution

2 math
5 stat
} Select 3 students

At least 1 math – student

= (1 math and 2 stat) or (2 math and 1 stat)

= ($2\ c_1$ × $5\ c_2$) + ($2\ C_2$ × $5\ C_1$)

= 25

There are 3 different mathematics courses, 2 different science courses, and one history course. If a student must take one of each, how many different ways can this be done?

A) 1 B) 6 C) 9 D) 3

When objects are arranged in a specific order, the arrangement is called .... تباديل
A) a combination B) with replacement C) without replacement D) a permutation

تعريف

A store manager wants to display 5 different brands of toothpaste in a row. How many ways can this be done?

A) 5 B) 20 C) 24 D) 120

number of ways = 5 x 4 x 3 x 2 x 1

= 120

In one town, 70% of adults بالغ سن الرشد have health insurance صحي تأمين. What is the probability that 6 adults selected at random from the town all have health insurance?
A) 0.086 B) 4.2 C) 0.118 D) 0.7

P(All 6 have health insurance) = $(0.70)^6$

$= 0.117649$

= 0.118

It is known that 9% of men have a type of color blindness that prevents them from distinguishing between red and green. If 3 men are selected at random, find the probability that all of them will have this type of red-green color blindness.

التمييز صن يمنعهم عمى الوان

A) 0.27 B) 0.07 C) 0.09 D) 0.0007

P(blindness) = 9% = 0.09

P(3 men have blindness) = $(0.09)^3$ = 0.0007

How many different ways can 3 tickets be selected from 30 tickets if each ticket wins a different prize?
A) 6840 B) 4060 C) 1140 D) 24360

تباديل لأنها تذاكر مختلفة السعر
different prize

$30P_3 = \boxed{24360}$

بالآلة الحاسبة: 991 ES → 30 sHift X 3 = 24360

How many different ways can be 3 cars selected from 12 cars?
A) 36 B) 6 C) 220 D) 1360

الاختيار بدون قيود (أي أن الترتيب غير مهم)

∴ الآلة توافيقه

number of diff. ways = $12C_3$

$= \boxed{220}$

مستطلع رأي الجمهور

A pollster wants to minimize the effect the order of the questions has on a person's response to a survey. How many different surveys are required to cover all possible arrangements if there are 6 questions on the survey?
A) 120 B) 720 C) 6 D) 36

The number of different ---

$= 6 \times 5 \times 4 \times 3 \times 2 \times 1 = \boxed{720}$

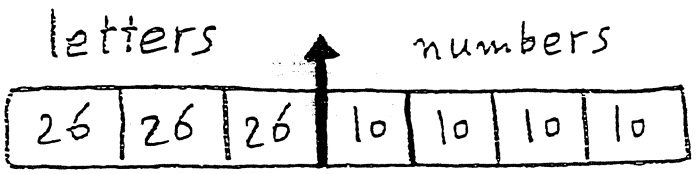
ALSAADI

لوحة السيارة

If the 26 alphabetic letters and the numbers from 0 to 9 is to be used to form a car plate that consists of 3 letters and 4 numbers, how many different plates are possible if repetitions are permitted?

A) 611531 B) 447174 C) 786240 D) 175760000

مع التكرار مسموح



➡ number of plates

$= 26 \times 26 \times 26 \times 10 \times 10 \times 10 \times 10$ بالآلة

$= \boxed{175760000}$

How many different ways can 4 tickets be selected from 10 tickets if each ticket wins a different prize?

A) 270 B) 720 C) 5040 D) 120

The number of different ways

$= {}_{10}P_4 = 5040$

بالآلة 10 shift x 4 = 5040

Model fx-991 ES ← نوع الآلة المستخدمة

كاسيو

6 خانات

If the letters A,B,C,D,E, and F are to be used in a letter code consists of six digits, how many different codes are possible if the first letter must be A and repetitions are not permitted?

A) 720 B) 1440 C) 46656 D) 120 التكرار غير مسموح

A B C D E F

الحرف الأول يجب أن يكون A

∴ عدد طرق اختيار الحرف الأول = 1

1	5	4	3	2	1

The number of diff. Codes

$= 1 \times 5 \times 4 \times 3 \times 2 \times 1 = 120$

Determine the number of all possible outcomes of guessing the last two digits in a telephone number if repetition of digits is allowed.

A) 30 B) 100 C) 1000 D) 20

The number of all possible outcomes

$= 10 \times 10 = 100$

لأنه التكرار مسموح repetition is allowe

لو كان التكرار غير مسموح repetition is not allowed

$10 \times 9 = 90$ يكون الحل هو

How many different 4-letter permutations can be formed from the letters in the word *orange*?
A) 360 B) 840 C) 120 D) 420

Orange

كم عدد التباديل لأربعة أحرف
ممكن تكوينها من كلمة orange التي تحتوي طبعاً على 6 أحرف

$= {}_6P_4$ بالآلة $= 6 \times 5 \times 4 \times 3 = \boxed{360}$

How many different ways can 3 tickets be selected from 20 tickets if each ticket wins a different prize?
A) 6840 B) 8000 C) 1140 D) 60

اختيار 3 تذاكر من ضمن 20 تذكرة بشرط أنها مختلفة
↓
تباديل

The number of ways $= {}_{20}P_3$ بالآلة $= \boxed{6840}$

How many different tests can be made from a test bank of 10 questions if the test consists of 3 questions?
A) 240 B) 720 C) 120 D) 360

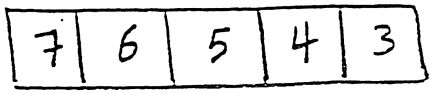
* لم يذكر قيود على الاختيار (أي لا أهمية للترتيب)
∴ الحالة توافيق

number of tests $= {}_{10}C_3$ بالآلة $= \boxed{120}$

التكرار

How many 5-digit numbers can be formed using the digits 1, 2, 3, 4, 5, 6, 7 if repetition of digits is not allowed?

A) 120
B) 16.807
C) 2520
D) 119



number of digits

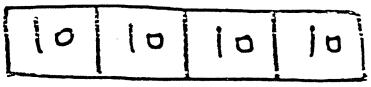
$= 7 \times 6 \times 5 \times 4 \times 3 = 2520$

Determine the number of all possible outcomes of guessing the last four digits in a telephone number if repetition of digits is allowed.

A) 10000 B) 24 C) 5040 D) 210

repetition is allowed

التكرار مسموح

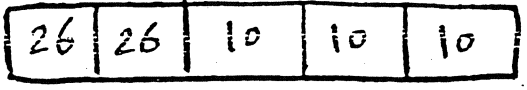


$= 10 \times 10 \times 10 \times 10 = 10000$

An ID card consists of 2 letters followed by 3 digits. How many different ID cards can be made if repetitions are allowed?

A) 650000 B) 486720 C) 468000 D) 676000

الحروف letters الأرقام digits



$= 26 \times 26 \times 10 \times 10 \times 10 = 676000$

A sample of 4 different calculators is randomly selected from a group containing 12 that are defective and 40 that have no defects. What is the probability that at least one of the calculators is defective?

A) 0.338 B) 0.002 C) 0.662 D) 0.998

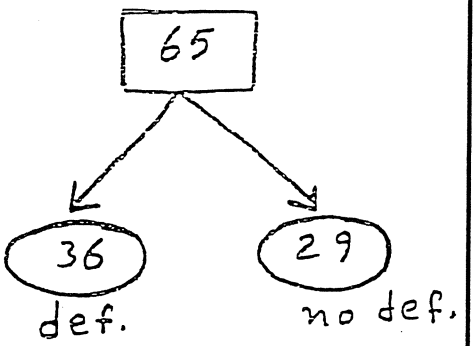
$P(\text{at least one is defective})$

$= 1 - P(\text{All no def.})$

$= 1 - \frac{{}_{40}C_4}{{}_{52}C_4} = \frac{2759}{4165} \simeq \boxed{0.662}$

A sample of 4 different calculators is randomly selected from a group containing 36 that are defective and 29 that have no defects. What is the probability that at least one of the calculators is not defective?

A) 0.965 B) 0.087 C) 0.913 D) 0.035



$P(\text{at least one no def.})$

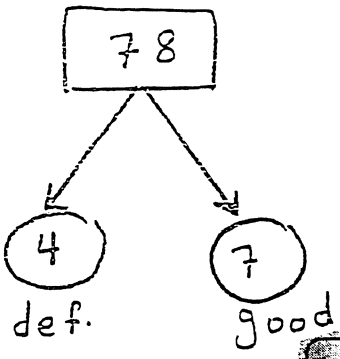
$= 1 - P(\text{All 4 def.})$

$= 1 - \frac{{}_{36}C_4}{{}_{65}C_4} = \frac{5887}{6448} \simeq \boxed{0.913}$

مصباح كهربي صندوق

A bin contains 78 light bulbs of which 4 are defective. If 5 light bulbs are randomly selected from the bin, find the probability that all the bulbs selected are good ones.

A) 0.763 B) 0.769 C) 0.051 D) 0.949



العدد المختار 5

$P(\text{all good}) = \frac{{}_{74}C_5}{{}_{78}C_5}$ بالآلة

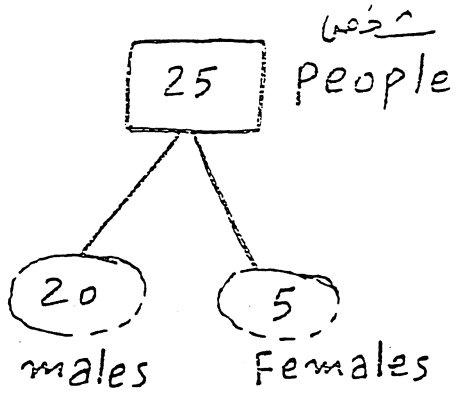
$= 0.763$

لجنة اربعة اشخاص

A committee consisting of 4 people is to be formed from 20 males and 5 females. Find the probability that the committee will consist of males only. ذكور فقط

A) 0.150 B) 0.791 C) 0.383 D) 0.451

اللجنة تتكون من أربعة اشخاص كلهم ذكور



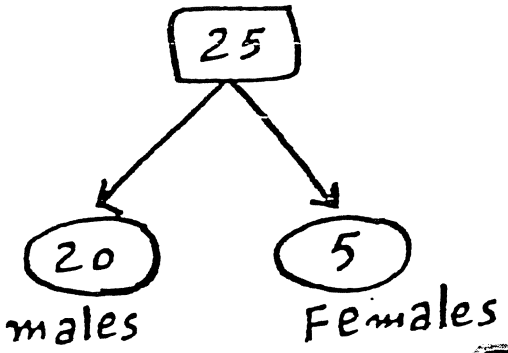
$\therefore P = \frac{{}_{20}C_4}{{}_{25}C_4}$ بالآلة

$= 0.383$

A committee consisting of 8 people is to be formed from 20 males and 5 females. Find the probability that the committee will consist of males only.

A) 0.116 B) 0.292 C) 0.161 D) 0.219

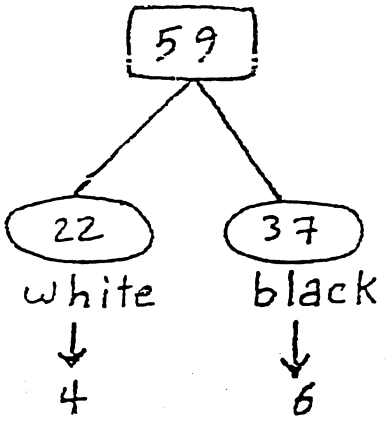
8 العدد المختار



$$P(\text{All 8 males only}) = \frac{{}_{20}C_8}{{}_{25}C_8} \simeq 0.116$$

A bag contains 22 white marbles and 37 black marbles. The probability P(X) of selecting without replacement a sample of size 10 consists of 4 white marbles and 6 black marbles is

A) $\{{}_{22}C_4 \times {}_{37}C_6\} / {}_{59}C_{10}$ B) $\{{}_{6}C_4 \times {}_{37}C_{22}\} / {}_{59}C_{10}$ C) ${}_{59}C_{10} / \{{}_{22}C_4 \times {}_{37}C_6\}$ D) ${}_{59}C_{10}$



$$P(X) = \frac{{}_{22}C_4 \times {}_{37}C_6}{{}_{59}C_{10}}$$

ALSAADI

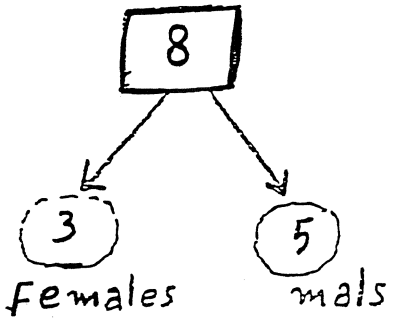
Given eight students, three of which are females. If two students are selected at random, what is the probability that both students are female?

A) 9/56 B) 3/32 C) 3/28 D) 3

اختيار اثنين

$$P = \frac{3C_2}{8C_2}$$ بالآلة

$$= \boxed{\frac{3}{28}}$$



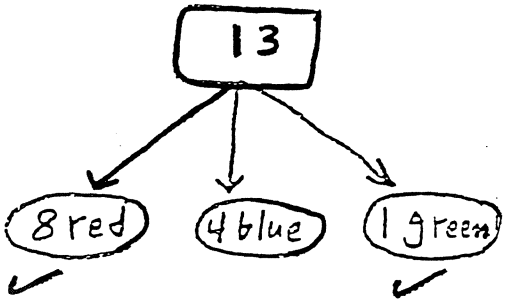
A bag contains 8 red marbles, 4 blue marbles, and 1 green marble. Find P(not blue).

بليه

A) $\frac{9}{13}$ B) 9 C) $\frac{13}{9}$ D) $\frac{4}{13}$

$$P(\text{not blue}) = \frac{9C_1}{13C_1} =$$ ليست زرقاء

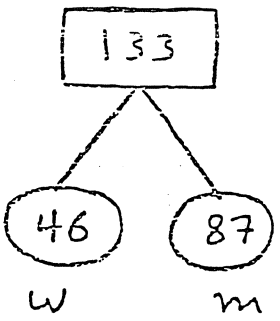
$$= \boxed{\frac{9}{13}}$$



A class consists of 46 women and 87 men. If a student is randomly selected, what is the probability that the student is a woman?

A) $\frac{1}{133}$ B) $\frac{46}{133}$ C) $\frac{87}{133}$ D) $\frac{46}{87}$

$$P(w) = \frac{46C_1}{133C_1} = \boxed{\frac{46}{133}}$$



ALSAADI

Use the following to answer questions

الدرجات المتحققه

The table below shows the number of earned degrees in the year 2008 in a university by level and gender. A person who earned a degree in the year 2008 from this university is randomly selected. Find the probability of selecting someone who

		Male	Female
Level of Degree	Bachelor's	300	200
	Master's	35	15

earned a master's degree or is a female

A) 0.7 B) 0.45 C) 0.64 D) 0.48

$$P(\text{master's or Female}) = P(\text{master's}) + P(\text{Female}) - P(\text{master's and Femal})$$

$$= \frac{50}{550} + \frac{215}{550} - \frac{15}{550} = 0.45\ 45\ 45\ 45\ 45 = 0.45$$

T

is a female given that the person earned a bachelor's degree.

A) 0.4 B) 0.67 C) 0.36 D) 0.6

$$P(F/\text{bach.}) = \frac{P(F \cap \text{bach.})}{P(\text{bach.})} = \frac{\frac{200}{550}}{\frac{500}{550}} = \frac{200}{500} = 0.4$$

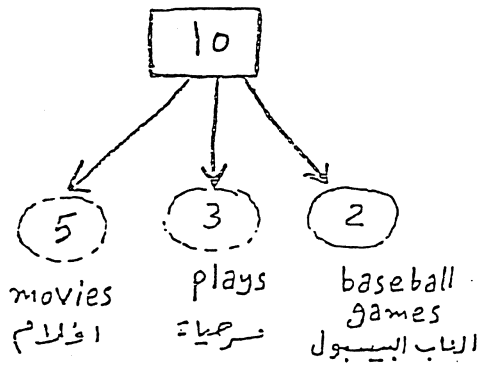
T

A newspaper advertises 5 different movies, 3 plays, and 2 baseball games for the weekend. If a couple selects 3 activities, find the probability that they attend 2 plays and 1 movie.

A) 0.20 B) 0.021 C) 0.083 D) 0.125

مسرحيتين — فيلم

العدد المختار 3



$$P(\text{2 plays and 1 movie}) = \frac{{}_3C_2 \times {}_5C_1}{{}_{10}C_3}$$

بالآله

$$= \boxed{0.125}$$

It has been found that 6% of all automobiles on the road have defective brakes. If 5 automobiles are stopped and checked by the state police, find the probability that at least one will have defective brakes.

A) 0.0127 B) 0.266 C) 0.734 D) .9872

*** هذه الانواع من التمارين التي تشمل على

نسبة مئوية + عدد + at least one

تحل بقانون واحد هو

$$P = 1 - (\text{تكملة النسبة})^{\text{العدد}}$$

$$= 1 - (0.94)^5 = 0.266$$

In a batch of 8,000 clock radios 2% are defective. A sample of 12 clock radios is randomly selected without replacement from the 8,000 and tested. The entire batch will be rejected if at least one of those tested is defective. What is the probability that the entire batch will be rejected?

A) 0.785 B) 0.0833 C) 0.0200 D) 0.215

*** كل المسائل من هذه النوعية :

يبها علاقة (أو) وهي وجود

نسبة مئوية 2% + عدد 12 + عبارة at least one

تحل بهذا القانون

$$P = 1 - (\text{طرح النسبة من 1})^{\text{العدد}}$$

(تكملة النسبة)

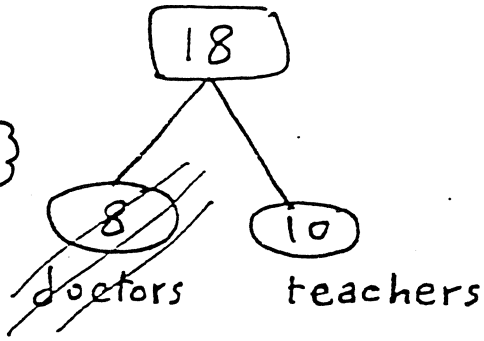
$$= 1 - (0.98)^{12} = 0.215$$

A committee of 4 people is to be formed from 8 doctors and 10 teachers. Find the probability that the committee will consist of at least one doctor

A) 0.93 B) 0.07 C) 0.02 D) 0.98

$P(\text{at least one doctor})$ نأخذ 4 من الفرخ الآخر

$= 1 - \frac{10C_4}{18C_4}$ ← قانونه المكمل

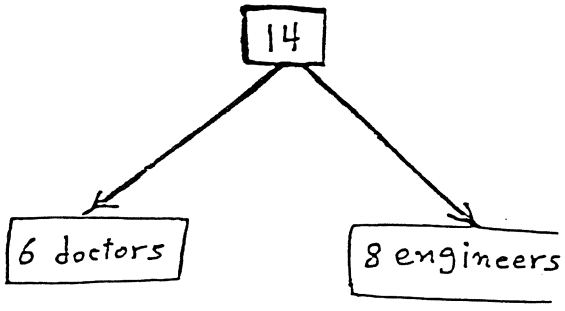


$= 0.93$

A committee of 4 people is to be formed from 6 doctors and 8 engineers. Find the probability that the committee will consist of at least two doctors.

A) 0.83 B) 0.17 C) 0.41 D) 0.59

لابد كل اختيار يتكون من أربعة أشخاص
4 people حيث أنه الكيف اقتصر على



(At least two doctors)

اثنين أطباء أو أكثر

(2 d. and 2 e.) or (3 d. and 1 e.) or (4 d.)

أربعة أربعة أربعة

$$\therefore P = \frac{6C_2 \times 8C_2}{14C_4} + \frac{6C_3 \times 8C_1}{14C_4} + \frac{6C_4}{14C_4}$$

$= 0.5944 = 0.59$

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. Subjective probability has little use in the real world. (×)
2. Classical probability uses a frequency distribution to compute probabilities. (×)
3. In classical probability, all outcomes in the sample space are equally likely. (√)
4. When two events are not mutually exclusive, P (A or B) = P (A) + P (B) (×)
5. If two events are dependent, they must have the same probability of occurring. (×)
6. An event and its complement can occur at the same time. (×)
7. The arrangement ABC is the same as BAC for combinations. (√)
8. When objects are arranged in a specific order. the arrangement is called a combination. (×)

Select the best answer.

9. The probability that an event happens is 0.42. What is the probability that the event won't happen?
a. -0.42 c. 0
b. 0.58 d. 1

10. When a meteorologist says that there is a 30% chance of showers, what type of probability is the person using?
a. Classical c. Subjective
b. Empirical **d.** b and c are correct

11. The sample space for tossing 3 coins consists of how many outcomes?
a. 2 c. 6
b. 4 **d.** 8

12. The complement of guessing 5 correct answers on a 5-question true/false exam is
a. Guessing 5 incorrect answers
b. Guessing at least 1 incorrect answer
S = {(5 inc.), (1c, 4 inc), (2c, 3 inc), (3c, 2 inc), (4c, inc), (5c)}
A: 5 correct answers
A': at least 1 incorrect answers
c. Guessing at least 1 correct answer
d. Guessing no incorrect answers

13. When two dice are rolled, the sample space consists of how many events?
a. 6 **c.** 36
b. 12 d. 54

14. What is nP_0?
a. 0 c. n
b. 1 d. It cannot be determined.

15. What is the number of permutations of 6 different objects taken all together?
a. 0 c. 36
b. 1 **d.** 720

16. What is 0!?
a. 0 c. Undefined
b. 1 d. 10

17. What is ${}_nC_n$?
a. 0 c. n
b. 1 d. It cannot be determined.

Complete the following statements with the best answer.

18. The set of all-possible outcomes of a probability experiment is called the ***sample space***
19. The probability of an event can be any number between and including ***0*** and ***1***
20. If an event cannot occur, its probability is ***0***
21. The sum of the probabilities of the events in the sample space is ***1***
22. When two events cannot occur at the same time, they are said to be ***mutually exclusive***

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

13

Ch. 5 - Part 1

- Introduction.
- Probability Distributions.
- Mean , Variance , Standard Deviation and Expectation.

STAT.110

جمال السعدي
رياضيات - إحصاء

Ch. 5 Part. 1

Discrete Probability Distributions

Probability Distributions

- **A random variable is a variable whose values are determined by chance.**
- Variables that can assume all values in the interval between any two given values are called **continuous variables**. For example, if the temperature goes from 60° to 70°.
- If a variable can assume only a specific number of values, such as the outcomes for the roll of a die or the outcomes for the toss of a coin, then the variable is called a **discrete variable**.
- **For these Exercises, state whether the variable is discrete or continuous.**

1. The speed of a jet airplane. **(Continuous)**

2. The number of cheeseburgers a fast-food restaurant serves each day. **(Discrete)**

3. The number of people who play the state lottery each day. **(Discrete)**

4. The weight of a Siberian tiger. **(continuous)**

5. The time it takes to complete a marathon. **(continuous)**

6. The number of mathematics majors in your school. **(Discrete)**

7. The blood pressures of all patients admitted to a hospital on a specific day. **(Discrete)**

Example:

Construct a probability for rolling a single die.

Solution

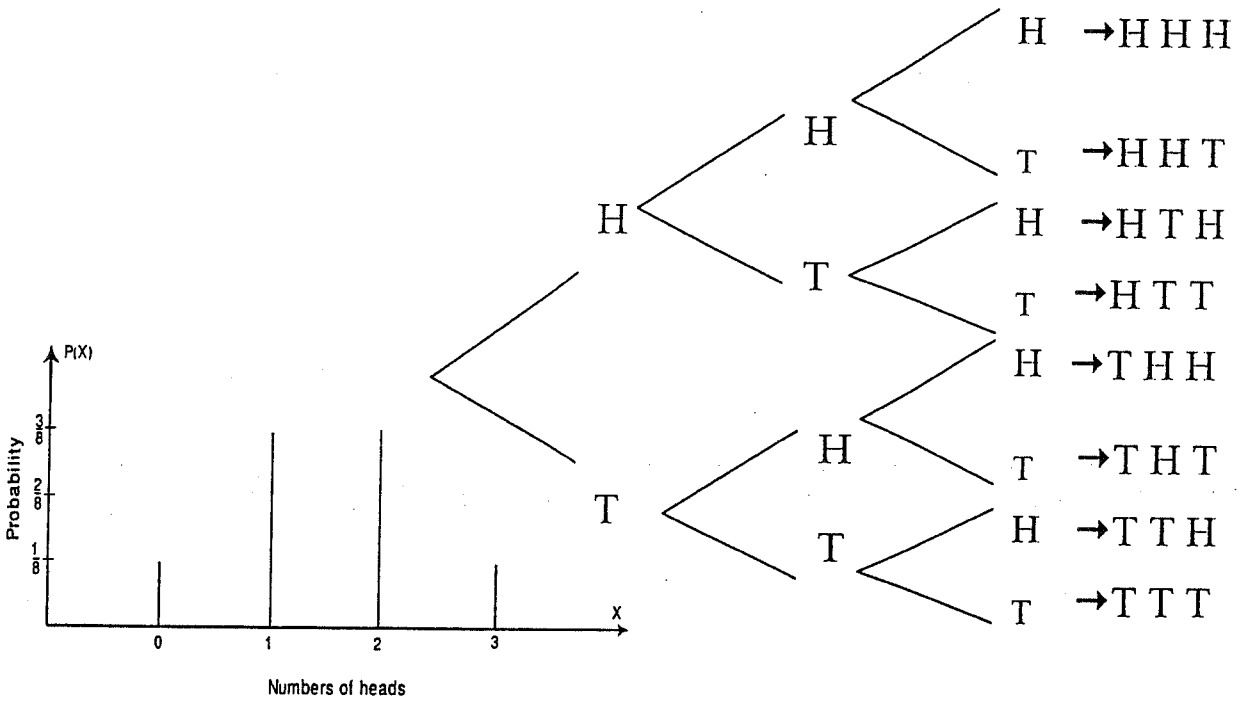
Outcome x	1	2	3	4	5	6
Probability: P(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Example:

Represent graphically the probability distribution for the sample space for tossing three coins.

Number of heads x	0	1	2	3
Probability: P (x)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Solution



شرطين أساسين
Two Requirements For a Probability Distribution

1. The Sum of the probabilities of all the events in the sample space must be equal 1 $\sum P(X) = 1$
2. The probability of each event in the sample space must be between or equal to 0 and 1. $0 \leq P(X) \leq 1.$

Example:

Determine whether each distribution is a probability distribution.

(a)

X	0	5	10	15	20
P (X)	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

Yes, it is a probability distribution.

(b)

X	0	2	4	6
P (X)	- 1.0	1.5	0.3	0.2

No. It is not a probability distribution, since P (x) cannot be 1.5 or – 1.0

(c)

X	1	2	3	4
P (X)	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{9}{16}$

Yes, it is a probability distribution.

(d)

X	2	3	7
P (X)	0.5	0.3	0.4

No, it is not, since $\sum$ p (X) = 1.2

ALSAADI

Mean, Variance, Standard Deviation, and Expectation

- **Formula for the mean of a probability distribution**

The mean of a random variable with a discrete probability distribution

$$\mu = X_1 \cdot P(X_1) + X_2 \cdot P(X_2) + X_3 \cdot P(X_3) + \ldots\ldots\ldots\ldots + X_n \cdot P(X_n)$$
$$\mu = \sum X \cdot P(X)$$

- **Formula for the variance of a probability distribution**

$$\sigma^2 = \sum [X^2 \cdot P(X)] - \mu^2$$

- **The standard deviation of a probability distribution is**

$$\sigma = \sqrt{\sigma^2} \quad \text{or} \quad \sqrt{\sum [X^2 \cdot P(X)] - \mu^2}$$

- **The expected value:**

$$\mu = E(X) = \sum X \cdot P(X)$$

- **Remember that variance and standard deviation cannot be negative.**

If X is a discrete random variable with $\sum[X^2 P(X)] = 6$ and $E(X) = 2$. The variance for the probability distribution of X is

A) 1.732 B) 2 C) 4 D) 1.141

$E(x) = 2 \Rightarrow \mu = 2$

$$\sigma^2 = \sum x^2 \cdot p(x) - \mu^2$$

$$= 6 - (2)^2 = 6 - 4 = \boxed{2}$$

Find the mean of the distribution shown.

x	1	2
P(x)	0.40	0.60

A) 1.60 B) 0.87 C) 1.09 D) 1.33

$$\mu = \sum x \cdot p(x) = (1)(0.40) + (2)(0.60) = \boxed{1.6}$$

In a frequency distribution, if the percentages are 20%, 38%, X and 16%, then the percentage X is ...

A) 26% B) 11% C) 16% D) 21%

In frequency distribution: $\sum p(x) = 1$ ← كأعداد عشرية

$\sum p(x) = 100\%$ ← كنسب

$$20\% + 38\% + X + 16\% = 100\%$$

$$74\% + X = 100\%$$

$$X = 100\% - 74\% \Rightarrow \boxed{X = 26\%}$$

Example:

A pizza shop owner determines the number of pizza that are delivered each day. Find the mean variance, and standard deviation for the distribution shown. If the manager stated that 45 pizzas were delivered on one day. Do you think that this is a believable claim? يستحق تصديقه

Number of deliveries X	35	36	37	38	39
Probability: P (X)	0.1	0.2	0.3	0.3	0.1

Solution

X	P (x)	X. P (x)	X^2 . P (x)
35	0.1	3.5	122.5
36	0.2	7.2	259.2
37	0.3	11.1	410.7
38	0.3	11.4	433.2
39	0.1	3.9	152.1
		$\sum$ x. P (x) = 37.1	$\sum$ x^2. p (x) = 1377.7

- Mean: $\mu = \sum x \cdot p(x) = 20.8$
- Variance: $\sigma^2 = \sum x^2 \cdot p(x) - \mu^2$

  $= 1377.7 - (37.1)^2$

  $= 1.29$
- Standard deviation: $\sigma = \sqrt{\sigma} = \sqrt{1.29} = 1.1$

Example: متجر بيع بالتجزئة

The number of suits sold per day at a retail store is shown in the table, with the corresponding probabilities. Find the mean, variance, and standard deviation of the distribution.

Number of suits sold X	19	20	21	22	23
Probability P (X)	0.2	0.2	0.3	0.2	0.1

If the manager of the retail store wants to be sure that he has enough suits for the next 5 days, how many should the manager purchase ?

Solution

X	P (x)	X. P (x)	X^2 . P (x)
19	0.2	3.8	72.2
20	0.2	4	80
21	0.3	6.3	132.3
22	0.2	4.4	96.8
23	0.1	2.3	52.9
		$\sum$ x. P (x) = 20.8	$\sum$ x^2. P (x) = 434.2

- Mean. $\mu = \sum x \,.\, p(x) = 37.1$
- Variance: $\sigma^2 = \sum x^2 \,.\, p(x) - \mu^2$

  $= 434.2 - (20.8)^2$

  $= 1.56$
- Standard deviation: $\sigma = \sqrt{\sigma^2} = \sqrt{1.56} = 1.2$
- The number of suits = (20.8) × (5) = 104 suits

ALSAADI

Example:

From past experience, a company has found that in cartons of transistors, 92 % contain no defective transistors, 3% contain one defective transistor, 3% contain two defective transistors, and 2% contain three defective transistors. Find the mean, variance, and standard deviation. For the defective transistors. About how many extra transistors per day would the company need to replace the defective ones if it used 10 cartons per day?

Solution

X	P (x)	X. P (x)	X^2 . P (x)
0	0.92	0	0
1	0.03	0.03	0.03
2	0.03	0.06	0.12
3	0.02	0.06	0.18
		$\sum$ x. P (x) = 0.15	$\sum$ x^2. P (x) = 0.33

- Mean. $\mu = \sum x.p(x) = 0.15$
- Variance: $\sigma^2 = \sum x^2 . p(x) - \mu^2$

  $= 0.33 - (0.15)^2$

  $= 0.3075$
- Standard deviation: $\sigma = \sqrt{\sigma^2} = \sqrt{0.3075} = 0.555$
- Number of extra transistors = (0.15) . (10) = 1.5 is $\cong$ 2.

What is the sample size for the following probability distribution?

X	1	3	5	7	9
$P(X)$	1/7	1/7	2/7	2/7	1/7

A) It cannot be determined B) 25 C) 5 D) 1

* لا يمكن أن يستدل على حجم العينة sample size من جدول التوزيع الاحتمالي .

A box contains 4 red balls and 7 black balls. 5 balls are selected with replacement. The standard deviation of the number of red balls that will be obtained is ...

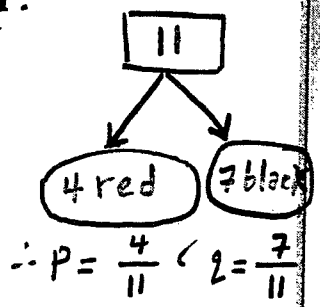
A) 0.968. B) 1.08 C) 0.938 D) 1.16

4 red , 7 black , 5 are selected $\therefore n = 5$

To find standard deviation for red.

$$\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{5\left(\frac{4}{11}\right)\left(\frac{7}{11}\right)}$$

$$= 1.075 \simeq \boxed{1.08}$$



A box contains 3 red balls and 5 black balls. 4 balls are selected with replacement. The standerd deviation of the number of red balls that will be obtained is

A) 0.938 B) 5 C) 4 D) 0.968

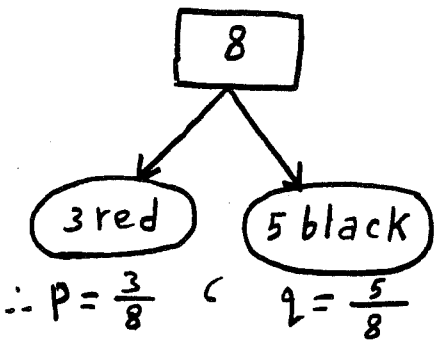
3 red , 5 black , 4 are slected $\therefore n = 4$

To find standard deviation for red balls.

$$\sigma = \sqrt{n \cdot p \cdot q}$$

$$= \sqrt{4 \cdot \left(\frac{3}{8}\right) \cdot \left(\frac{5}{8}\right)}$$

$$= 0.968$$



Example:

A person decides to invest $ 50.000 in a gas well. Based on history, the Probabilities of the outcomes are as follows.

Outcome x	P (x)
$ 80.000 (Highly successful)	0.2
$ 40.000 (Moderately successful	0.7
- $ 50.000 (Dry well) خسارة كبيرة	0.1

استثمار

- Find the expected value of the investment.

  Would you consider this a good investment?

Solution

$$E(x) = \sum x \cdot P(x)$$

$$= (80000)(0.2) + (40000)(0.7) + (-50000)(0.1)$$

$$= \$ 39000$$

This a good investment.

If X is a discrete random variable with $\sum[X^2 P(X)] = 4$ and $E(X) = -2$. The standard deviation for the probability distribution of X is

A) 8 B) 1.41 C) 2.828 D) 0

$\Sigma X^2 P(X) = 4$ ، $\mu = E(X) = -2$ ، σ ؟

$\sigma^2 = \Sigma X^2 P(X) - \mu^2$

$= 4 - (-2)^2 = 4 - 4 = 0$

$\therefore \sigma^2 = 0 \Rightarrow \sigma = \sqrt{\sigma^2} = \sqrt{0} = \boxed{0}$

If X is a discrete random variable with $\sum[X^2 P(X)] = 7$ and $\sigma^2 = 2$, then the mean for the probability distribution of X is ...

A) 2.24. B) 5 C) 1.141 D) 2

$\sigma^2 = \Sigma X^2 \cdot P(X) - \mu^2$ ← قانون

$2 = 7 - \mu^2$ حل معادلة

$\mu^2 = 7 - 2 = 5$

$\therefore \mu = \sqrt{5} = 2.236 \simeq \boxed{2.24}$

ALSAADI

السعدي ALBAADI

نسخة جديدة منقحة

1433/32

14

Ch. 5 - Part 2

- The Binomial Distribution.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 5 Part. 2

The Binomial Distribution

A binomial experiment is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial has only two outcomes: success or fail.
3. The outcomes of each trial must be independent of each other.
4. The probability of a success must remain the same for each trial.

Mean, Variance, and standard deviation for the binomial distribution

The mean, variance, and standard deviation of a variable that has the binomial distribution can be found by using the following formulas.

- **Mean:** $\mu = n \cdot p$
- **Variance:** $\sigma^2 = n \cdot p \cdot q$
- **Standard deviation:** $\sigma = \sqrt{n \cdot p \cdot q}$

Example:
A dice is rolled 480 times. Find the mean, variance, and standard deviation of the number of 2s that will be rolled.

Solution

Getting a 2 is a success and not getting a 2 is a failure:

- $n = 480$, $P = \frac{1}{6}$, and $q = \frac{5}{6}$
- $\mu = n.p = 480.\frac{1}{6} = 80$
- $\sigma^2 = n.p.q = 480.\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = 66.7$
- $\sigma = \sqrt{n.p.q} = \sqrt{66.7} = 8.2$

Example:

A coin is tossed 4 times. Find the mean, variance, and standard deviation of the number of heads that will be obtained.

Solution

The binomial distribution and

$n = 4$, $p = \frac{1}{2}$ and $q = \frac{1}{2}$

$\mu = n \cdot p = 4.\frac{1}{2} = 2$

$\sigma^2 = n \cdot p \cdot q = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1$

$\sigma = \sqrt{1} = 1$

Example:

If 3% of calculators are defective, find the mean, variance, and standard deviation of a lot of 300 calculators.

Solution

$n = 300$ $p = 0.03$ $q = 0.97$

- $\mu = n.p = (300)(0.03) = 9$
- $\sigma^2 = n.p.q = (300)(0.03)(0.97) = 8.7$
- $\sigma = \sqrt{\sigma^2} = \sqrt{8.73} = 2.9 \cong 3$

Example:

In a restaurant, a study found that 42% of all patrons smoked. If the seating capacity of the restaurant is 80 people, find the mean, variance, and standard deviation of the number of smokers. About how many seats should be available. For smoking customers?

Solution

$n = 80$ $p = 0.42$ $q = 0.58$

- $\mu = n.p = (80)(0.42) = 33.6$
- $\sigma^2 = n.p.q = (80)(0.42)(0.58) = 19.5$
- $\sigma = \sqrt{\sigma^2} = \sqrt{19.5} \cong 4.4$

Note

- **two outcomes:** yes or no → **(binomial)**
- **more than two outcomes → (not binomial)**

Which of the following are binomial experiments or can be reduced to binomial experiments?

صابون كثير الرغوة

(a.) Surveying 100 People to determine if they like sudsy soap.

(Binomial)

(b.) Tossing a coin 100 times to see how many heads occur

(Binomial)

(c.) Asking 1000 people which brand of cigarettes they smoke.

(Not binomial)

(d.) Testing one brand of aspirin by using 10 people to determine whether it is effective **(Binomial)**

(f.) Asking 100 people if they smoke **(Binomial)**

(g.) Checking 1000 applicants to see whether they were admitted to white Oak college. **(Binomial)**

(h.) Surveying 300 prisoners to see how many different crimes they were convicted of. (أدينو بها) **(Not binomial)**

(i.) Surveying 300 prisoners to see whether this is their first offense.

(Binomial)

Binomial Probability Formula

In a binomial experiment, the probability of exactly X successes in n trials is

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

Example: يخمن

A student takes a 20 – question, true/ false exam and guesses on each question. Find the probability of passing اختيار if the lowest passing grade is 15 correct out of 20. Would you consider this event likely to occur? Explain your answer.

Solution

$n = 20$ $\quad p = \frac{1}{2}$ $\quad q = \frac{1}{2}$

$$p(\text{passing}) = p(x \geq 15)$$

$$= p(x=15) + p(x=16) + p(x=17) + p(x=18) + p(x=19) + p(x=20)$$

$$= 20\,C_{15}\left(\frac{1}{2}\right)^{15}\left(\frac{1}{2}\right)^{5} + 20C_{16}\left(\frac{1}{2}\right)^{16}\left(\frac{1}{2}\right)^{4}$$

$$+ 20\,C_{17}\left(\frac{1}{2}\right)^{17}\left(\frac{1}{2}\right)^{3} + 20\,C_{18}\left(\frac{1}{2}\right)^{18}\left(\frac{1}{2}\right)^{2}$$

$$+ 20C_{19}\left(\frac{1}{2}\right)^{19}\left(\frac{1}{2}\right)^{1} + 20C_{20}\left(\frac{1}{2}\right)^{20}\left(\frac{1}{2}\right)^{0}$$

$$= 0.015 + 0.005 + 0.001 + \ldots\ldots = 0.021 < 0.5$$

There for P (passing) unlikely to occur.

Example:

ضحية
A survey found that 86% of Americans have never been a victim
جريمة العنف
of violent crime. If a sample of 12 Americans is selected at
معقول يبدو
random, find the probability that 10 or more have never been victims of violent crime. Does it seem reasonable that 10 or more have never been victims of violent crime?

Solution

$n = 12 \qquad p = 0.86 \qquad q = 0.14$

$p(x \geq 10) = p(x = 10) + p(x = 11) + p(x=12)$

$= {}^{12}C_{10} (0.86)^{10} (0.14)^{2} + {}^{12}C_{11} (0.86)^{11} (0.14)^{1} + {}^{12}C_{12} (0.86)^{12} (0.14)^{0}$

$= 0.77 > 0.5$

Yes: it seem reasonable.....

Example: مجتمع خدمة

If 80% of the people in a community have internet access from their homes, find these probabilities for a sample of 10 people.

a. At most 6 have internet access. على الأكثر

b. Exactly 6 have internet access.

c. At least 6 have internet access.

d. Which event a, b, or c is most likely to occur? Explain why?

Solution

$n = 10$ $p = 0.8$ $q = 0.2$

(a) P (at most 6) = $p(x \leq 6)$

$= p(x=6) + p(x=5) + p(x=4) + p(x=3) + p(x=2) + p(x=1) + p(x=0)$

$= 10C_6 (0.8)^6 (0.2)^4 + 10C_5 (0.8)^5 (0.2)^5 + 10C_4 (0.8)^4 (0.2)^6$

$+ 10C_3 (0.8)^3 (0.2)^7 + 10C_2 (0.8)^2 (0.2)^8 + 10C_1 (0.8)^1 (0.2)^9$

$+ 10C_0 (0.8)^0 (0.2)^{10} = 0.121$

(b) $P(x=6) = 10C_6 (0.8)^6 (0.2)^4 = 0.088$

(c) p (at least 6) = $p(x \geq 6)$ =.....= 0.967.

(d) Event c is most likely to occur because it's > 0.5

A die is rolled 5 times. the probability of getting a number 4 one time only is
A) 0.402 B) 0.167 C) 0.015 D) 0.386

$n = 5 \qquad p = \frac{1}{6} \qquad q = \frac{5}{6}$

$P(X = ?) = {}_nC_x \cdot p^x \cdot q^{n-x}$ ← قانون

$P(X = 1) = {}_5C_1 \cdot \left(\frac{1}{6}\right)^1 \cdot \left(\frac{5}{6}\right)^{5-1}$

$= 0.4018 \simeq 0.402$

T

A die is rolled 5 times, the probability of getting a number 5 exactly two times only is
A) 0.598 B) 0.161 C) 0.839 D) 0.402

$x = 2$

$n = 5 \qquad p = \frac{1}{6} \Rightarrow q = \frac{5}{6}$

$P(X) = {}_nC_x \, p^x q^{n-x}$ قانون

$P(X=2) = {}_5C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{5-2}$ بالآلة $\simeq 0.161$

$n = 7$ $\quad P = \frac{1}{4}$ $\quad \therefore q = \frac{3}{4}$

A student takes a 7 question multiple choice quiz with 4 choices for each question. If the student guesses at random on each question, what is the probability that the student gets exactly 3 questions correct?

$?$ $\quad X = 3$

A) 0.130 B) 0.346 C) 0.173 D) 0.043

$$P(x) = {}_nC_x \, p^x \, q^{n-x}$$

$$P(X=3) = {}_7C_3 \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^{7-3}$$ بالآلة $= \boxed{0.173}$

The outcomes of each trial in a binomial experiment
A) are unlimited B) are independent C) are dependent D) must be fixed

(حفظ) من ضمن الخواص الاربعه لـ binomial

The number of trials in a binomial experiment ... من ضمن خواص binomial
A) are independent. B) are dependent. C) must be fixed. D) are unlimited

A study shows that 70% of drivers consider themselves above average in driving ability. If 10 drivers at random are chosen, what is the mean and variance of the number of drivers who consider themselves above average?

A) mean = 7 and variance = 7.
B) mean = 10 and variance = 1.45
C) mean = 7 and variance = 2.1.
D) mean = 10 and variance = 10.

$p = 0.70 = 0.7 \Rightarrow q = 0.3$, $n = 10$

mean: $\mu = n \cdot p = (10)(0.7) = \boxed{7}$

Variance: $\sigma^2 = n \cdot p \cdot q = (10)(0.7)(0.3) = \boxed{2.1}$

"A distribution using the means computed from all possible random samples of a specific size taken form a population." The previous statement is the definition of
A) central limit theorem B) sampling distribution C) sampling error D) empirical distribution

Chapter Quiz

Determine whether each statement is true or false. If the statement is false explain why.

1. The expected value of a random variable can be thought of as a long – run average. (فترة طويلة) (✓)

2. The number of courses a students is taking this semester is an example of a continuous random variable. (x)

3. when the multinomial distribution is used, the outcomes must be dependent. (x)

4. A binomial experiment has a fixed number of trials. (✓)

Complete these statements with the best answer:

5. Random variable values are determined by **chance.**

6. The mean for a binomial variable can be found be found by using the formula **$\mu = n \cdot p$.**

7. One requirement for a probability distribution is that the sum of all the events in the sample space must equal **1.**

** A probability distribution can be graphed using bar char.

Select the best answer:

8. **What is the sum of the probabilities of all outcomes in a probability distribution?**

a. 0 c. 1

b. 1/2 d. It cannot be determined.

9. How many outcomes are there in a binomial experiment?

a. 0 c. 2

b. 1 d. It varies

For questions 11 through 14, determine if the distribution represents a probability distribution. If not, state why.

11

X	1	2	3	4	5
P (X)	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

→ No where $\sum P(x) > 1$

12

X	3	6	9	12	15
P (X)	0.3	0.5	0.1	0.08	0.02

→ yes

13

X	50	75	100
P (X)	0.5	0.2	0.3

→ yes

14

X	4	8	12	16
P (X)	$\frac{1}{6}$	$\frac{3}{12}$	$\frac{1}{2}$	$\frac{1}{12}$

→ yes

إنتهى Ch. 5

كل التمنيات بالنجاح والتوفيق

السعدي

ALSAADI

15

السعدي ALSAADI

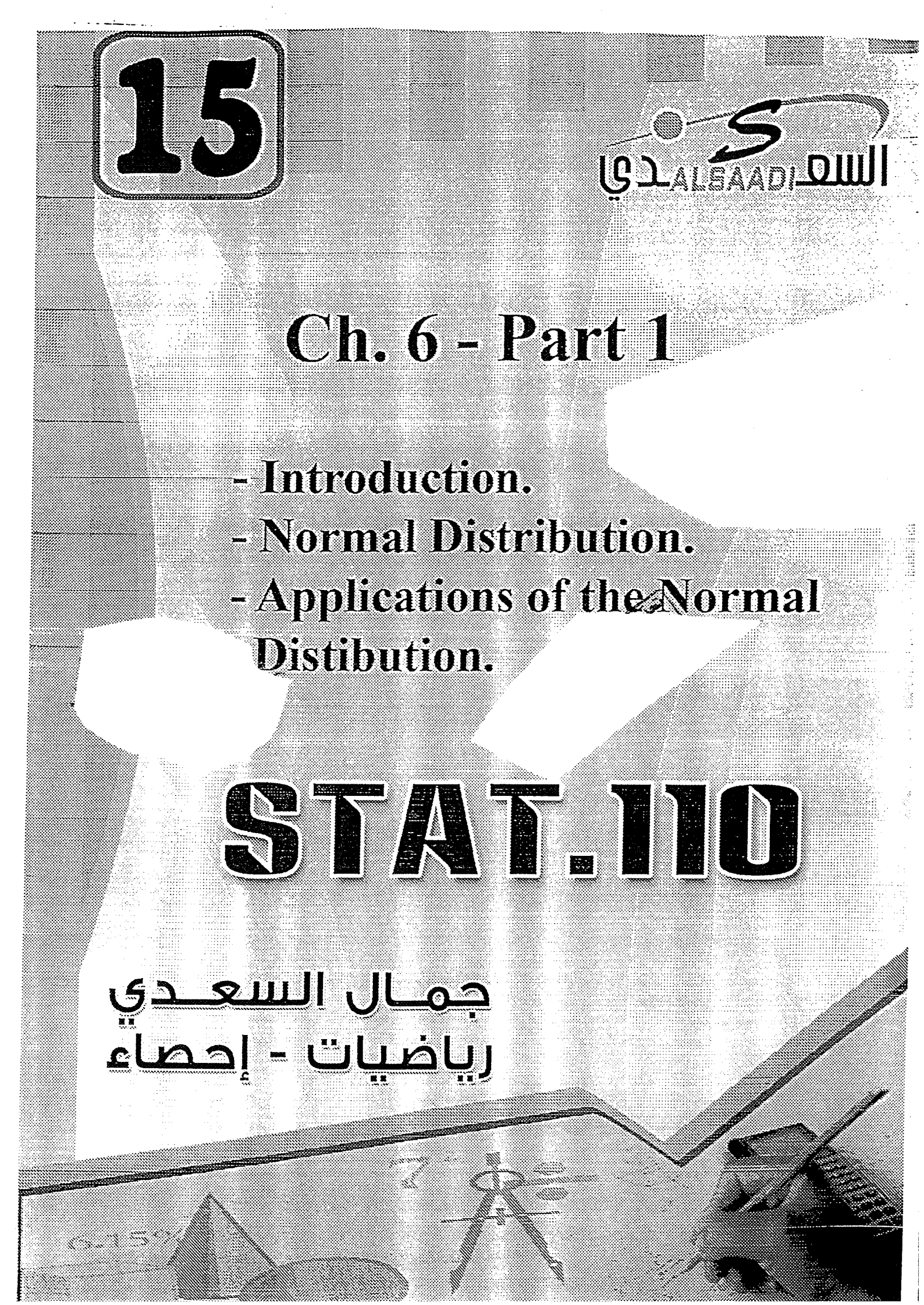
Ch. 6 - Part 1

- Introduction.
- Normal Distribution.
- Applications of the Normal Distibution.

STAT.110

جمال السعدي

رياضيات - إحصاء

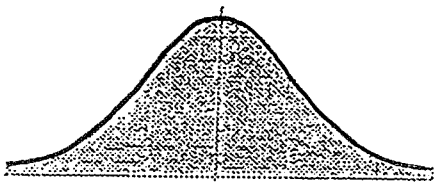


Ch. 6 Part. 1

التوزيع الطبيعي
The Normal Distribution

When the data values are evenly distributed about the mean, a distribution is said to be a symmetric distribution. (A normal distribution is symmetric.).

* تماثل

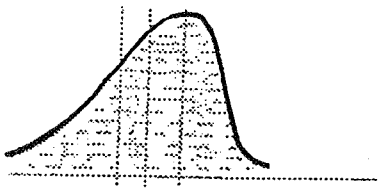


Mean = Median = Mode

Normal

When the majority of the data values fall to the right of the mean, the distribution is said to be a negatively or left-skewed distribution.

* منحرف يسار

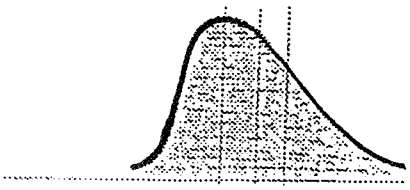


Mean Median Mode
Negatively skewed

Mean < Median < Mode

When the majority of the data values fall to the left of the mean, a distribution is said to be a positively or right-skewed distribution.

* منحرف يمين



Mode Median Mean
Positively skewed

Mode < Median < Mean

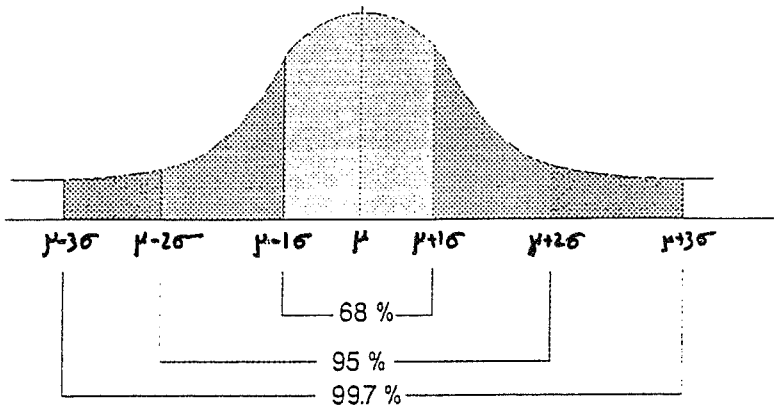
التوزيع الطبيعي

A normal distribution is a continuous, symmetric, bell-shaped distribution of a variable.

ملخص الخواص النظرية للتوزيع الطبيعي

Summary of the Properties of the Theoretical Normal Distribution

1. A normal distribution curve is bell-shaped.
2. The mean, median, and mode are equal and are located at the center of the distribution.
3. A normal distribution curve is unimodal (it has only one mode).
4. The curve is symmetric about the mean
5. The curve is continuous, that is, there are no gaps or holes.
6. The curve never touches the x axis.
7. The total area under a normal distribution curve is equal to 1.00, or 100%.
8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%.



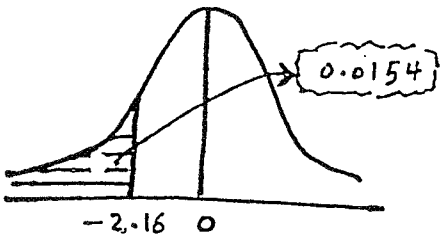
* الجدول الجديد : الكشف عن الموجب والسالب

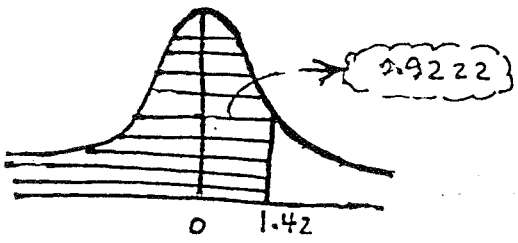
* طريقة استخدام جدول Table I : $\Phi(z)$

الذي يحتوي على z_+ / z_-

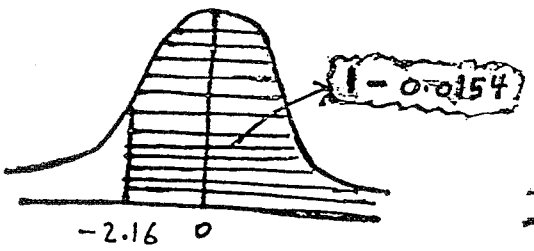
أولاً : اذا علم الـ z والمطلوب حساب الاحتمال P

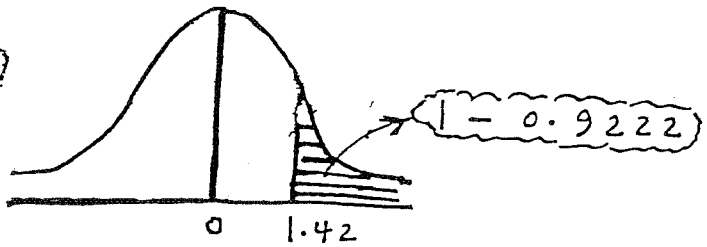
الحالة الأولى : أقل من (يسار) "P = الكشف المباشر"



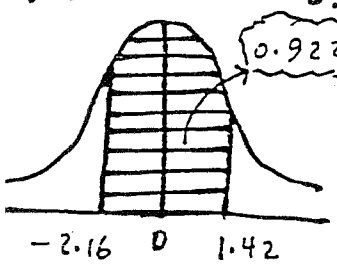


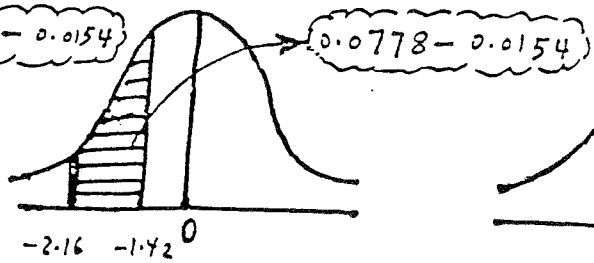
الحالة الثانية : أكبر من (يمين) "P = 1 - ناتج الكشف" ⇐

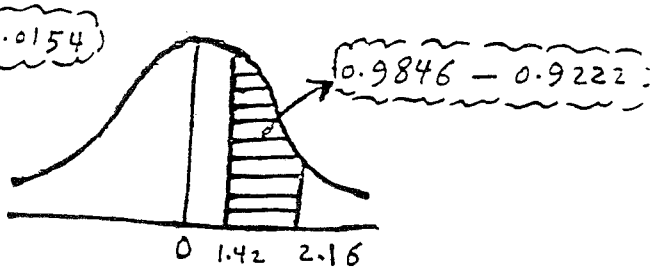




الحالة الثالثة : المساحة المحصورة بين عددين " الكشف عن العددين ثم نطرح ناتج الكشف الأصغر ⇐ ناتج الكشف الأكبر ناقص



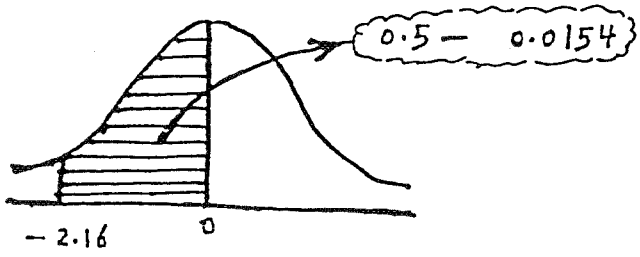




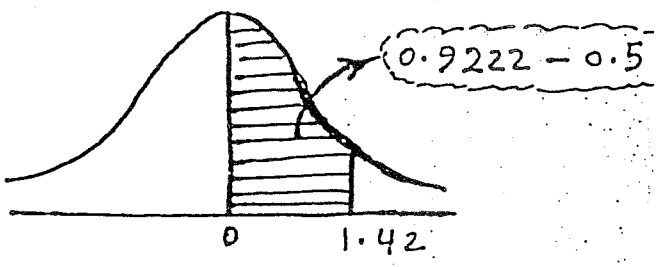
جمال السعدي
استاذ الرياضيات والإحصاء للمرحلة الجامعية
٠٥٦٦٦٦٤٧٩٠

الحاله الرابعه : من z إلى Zero

ناتج الكشف عن z − 0.5

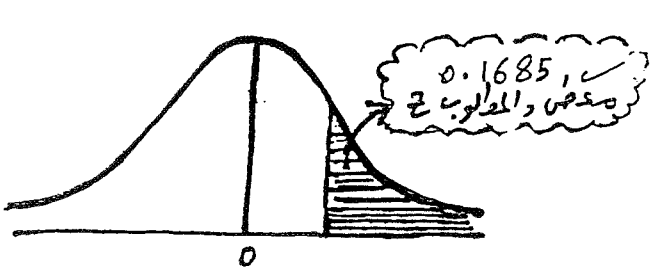


0.5 − ناتج الكشف عن z



ثانياً : اذا علم الاحتمال (المساحه) P والمطلوب ايجاد z

كشف عكسي من الداخل إلى الخارج ←



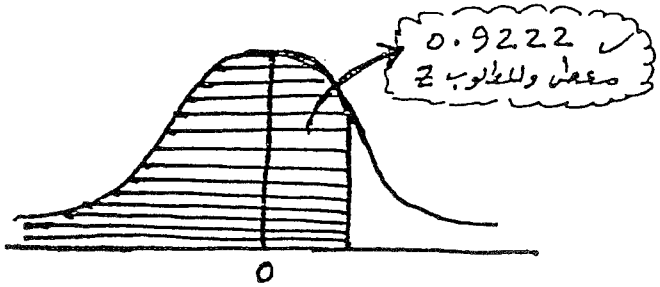
المساحه المطلوبه على اليمين

أولاً : تطرح المساحه المعلومه من 1

ثانياً : نكشف عن الناتج كشف عكسي

ثالثاً : ويكون الناتج + اذا z على اليمين

" " − " " z " اليسار



المساحه المعلومه على اليسار

أولاً : يكون الكشف عكسي مباشر .

ثانياً : الناتج + اذا z على اليمين .

" − " z " اليسار .

اللهم لا سهل إلا ما جعلته سهلاً وأنت تجعل الحزن إن شئت سهلاً

مع تمنياتي لكم بالتوفيق ،،

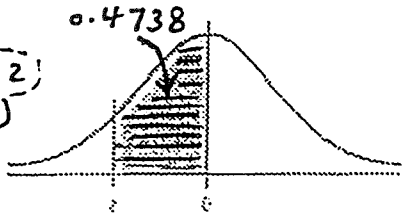
السعدي

جمال السعدي
أستاذ الرياضيات والإحصاء للمرحلة الجامعية
٠٥٦٦٦٦٤٧٩٠

Note: إذا علمت المساحة وطلب قيمة Z نكشف كشف عكسي في الجدول

For Exercises 40 through 45, find the z value that corresponds to the given area

40



تطرح المساحة المعطاة من 0.5

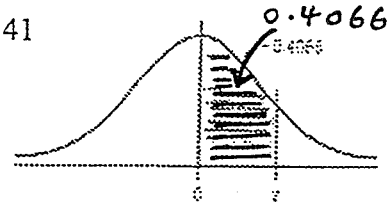
$0.5 - 0.4738 = 0.0262$

نكشف عن هذا الناتج كشف عكسي

$z = -1.94$ فتكون قيمته

$Z = -1.94$

41



يضاف للمساحة 0.4066

0.5

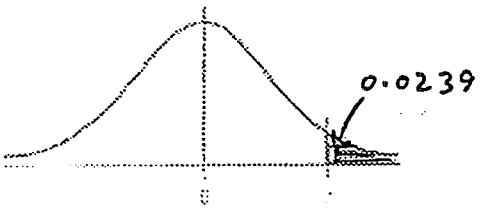
0.9066 فيكون الناتج

نكشف عن هذا الناتج كشف عكسي

$z = 1.32$ فتكون قيمته

$Z = 1.32$

42



∴ التظليل يمين

∴ تطرح المساحة المعطاة من 1

$1 - 0.0239$

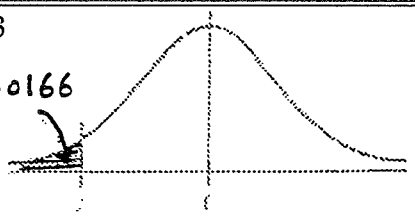
$= 0.9761$

نكشف عن هذا الناتج كشف عكسي

$z = 1.98$ تكون قيمته

$Z = 1.98$

43



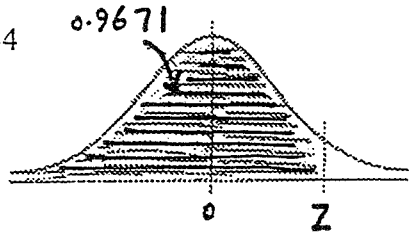
كشف عكسي مباشر

0.0166 عن

$z = -2.13$ تكون قيمته

$Z = -2.13$

44



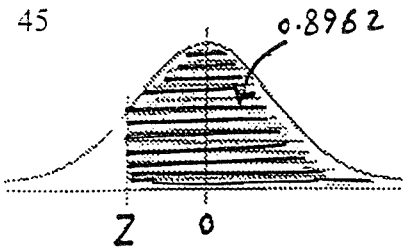
كشف عكسي مباشر

0.9671 عن

$z = 1.84$ تكون قيمته

$Z = 1.84$

45



∴ التظليل يمين

∴ تطرح المساحة المعطاة من 1

$1 - 0.8962$

$= 0.1038$

نكشف عن هذا الناتج كشف عكسي

$z = -1.26$ تكون قيمته

$Z = -1.26$

Applications of the Normal Distribution

The standard normal distribution

Is normal distribution with $\mu = 0$ *and* $\sigma = 1$

التوزيع الطبيعي المعياري

To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the formula

- $z = \dfrac{\text{value} - \text{mean}}{\text{standard deviation}}$ or $z = \dfrac{x - \mu}{\sigma}$
- $P(X > X_0)$

$$= P\left(Z > \frac{x_0 - \mu}{\sigma}\right)$$

ملحوظة

عند حساب الاحتمال حول المتغير x الذي يتبع توزيع طبيعي. يحول إلى توزيع طبيعي معياري Z

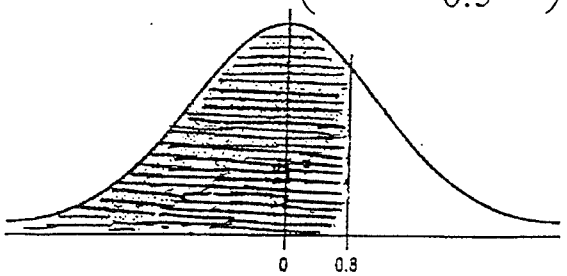
$$z = \frac{x - \mu}{\sigma}$$

Example:

The mean number of hours an American worker spends on the computer is 3.1 hours per workday. Assume the standard deviation is 0.5 hour. Find the percentage of workers who spend less than 3.5 hours on the computer. Assume the variable is normally distributed

Solution

$$P(x < 3.5) = p\left(z < \frac{3.5 - 3.1}{0.5}\right) = p(z < 0.8)$$



للتحويل إلى نسبة مئوية نضرب في ١٠٠ %

$= 0.7881 \rightarrow$

Therefore, 78.81 % of the workers spend less than 3.5 hours per workday on the computer

Example:

أفران الميكروويف

A survey found that people keep their microwave ovens an average of 3.2 years. The standard deviation is 0.56 year. If a person decides to buy a new microwave oven.

يملك

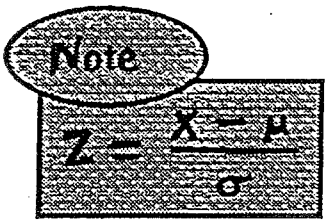
Find the probability that he or she has owned the old oven for the following amount of time. Assume the variable is normally distributed:

a. Less than 1.5 year's

b. Between 2 and 3 years

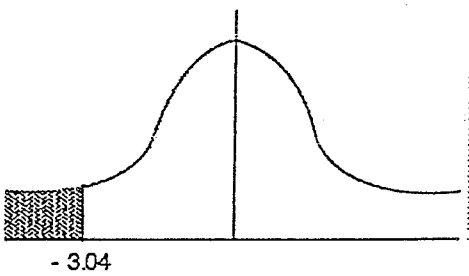
c. More than 3.2 years

d. What percent of microwave ovens would be replaced if a warranty (الضمان) of 18 months were given?



Solution

$\mu = 3.2$ $\sigma = 0.56$



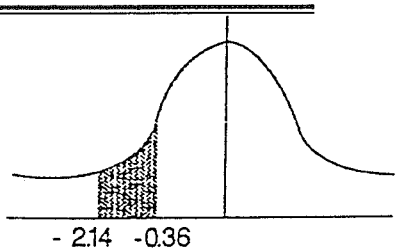
$$(a)\ P(x < 1.5) = p\left(z < \frac{1.5 - 3.2}{0.56}\right)$$

$$= P(z < -3.04)$$

$$= 0.0012$$

$$(b)\ P(2 < x < 3) = p\left(\frac{2-3.2}{0.56} < z < \frac{3-3.2}{0.56}\right)$$

$$= p(-2.14 < z < -0.36)$$



الحدان في جهه واحدة (طرح) بعد الكشف في الجدول (الأكبر – الأصغر)

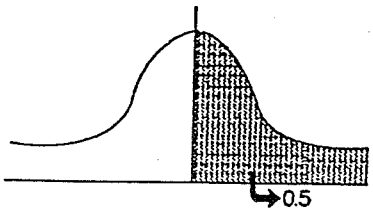
$0.3594 - 0.0162$

$= 0.3432$

(c) $p(x > 3.2) = p\left(z > \frac{3.2 - 3.2}{0.56}\right)$

$= p(z > 0)$

$= 0.5$



(d) 18 months $= \frac{18}{12} = 1.5$ years

Were μ and σ by years.

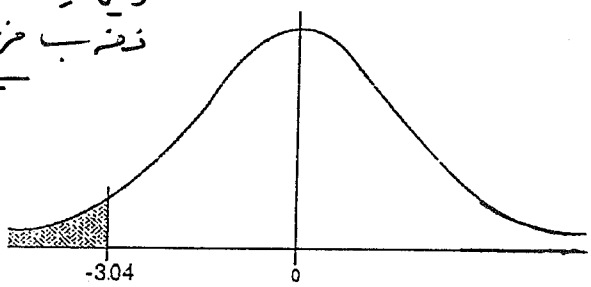
$p(x < 1.5) = p\left(z < \frac{1.5 - 3.2}{0.56}\right)$

$= p(z < -3.04)$

$= 0.0012$

$0.0012 \times 100\% = 0.12\%$ of the ovens must be replaced.

للتحويل إلى نسبة مئوية نضرب في ١٠٠ %



Example:

The average time for a mail carrier to cover his route (طريق) is 380 minutes, and the standard deviation is 16 minutes. If one of these trips is selected at random, find the probability that the carrier will have the following route time. Assume the variable is normally distributed.

a. At least 350 minutes

b. At most 395 minutes

Solution

$\mu = 380$ $\sigma = 16$

(a) At least (على الأقل) 350 minutes

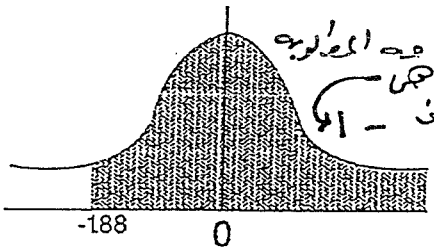
$$P(x \geq 350) = p\left(Z \geq \frac{350 - 380}{16}\right)$$

$$= p(z \geq -1.88)$$

$$= 1 - 0.0301$$

$$= 0.9699$$

التحليل يصير
المساحة المطلوبة = 1 - ناتج الكشف



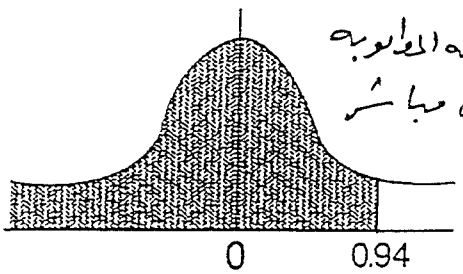
(b) At most (على الأكثر) 395 minutes

$$P(x \leq 395) = p\left(z \leq \frac{395 - 380}{16}\right)$$

$$= p(z \leq 0.94)$$

$$= 0.8264$$

التحليل يسار
المساحة المطلوبة
كشف مباشر



ALSAADI

Example:

ساعة يد

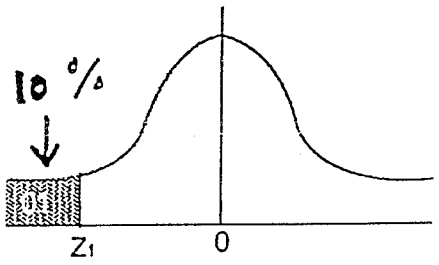
The mean lifetime of a wristwatch is 25 months, with a standard deviation of 5 months. If the distribution is normal.

For how many months should a guarantee ضمان be made if the manufacturer الصانع (صاحب المصنع) does not want to exchange more than 10% of the watches? Assume the variable is normally distributed.

Solution

$\mu = 25$ $\sigma = 5$

$P(z < z_1) = 0.1$



Therefore $z_1 = -1.28$ & $z_1 = \frac{x - 25}{\sigma}$

$$\frac{x - 25}{5} = -1.28$$

$$X - 25 = (-1.28)(5)$$

$$X = (-1.28)(5) + 25 = 18.6 \text{ month}$$

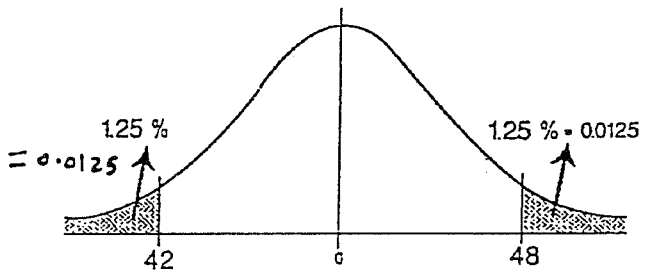
Example:

In a certain normal distribution, 1.25% of the area lies to the left of 42, and 1.25% of the area lies to the right of 48.

Find μ and σ.

Solution

$$\mu = \frac{42 + 48}{2} = \frac{90}{2} = 45$$



$$P(Z < z_1) = 0.0125$$

$$Z_1 = -2.24$$

$$\text{There for } z = \frac{x - \mu}{\sigma}$$

$$-2.24 = \frac{42 - 45}{\sigma}$$

$$\sigma = \frac{42 - 45}{-2.24} = 1.34$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

16

Ch. 6 - Part 2

- The Central Limit Theorem (Distribution of Sample Means)

STAT.110

جمال السعدي
رياضيات - إحصاء

Best Wishes And Good Luck.

Ch. 6 Part. 2

The Central Limit Theorem نظرية النهاية المركزية

•في حالة : مجتمع أخذت منه عينة حجمها n. **n: sample size taken from population**

A sampling distribution of sample means is a distribution using the means computed from all possible random samples of a specific size taken from a population Properties of the Distribution of Sample Means

1. The mean of the sample means will be the same as the population mean.
2. The standard deviation of the sample means will be smaller than the standard deviation of the population, and it will be equal to the population standard deviation divided by the square root of the sample size.

1. $z = \frac{x - \mu}{\sigma}$ في حالة فرد	Used to gain information about an individual data value when the variable is normally distributed.
2. $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ في حالة عينه حجمها n	Used to gain information when applying the central limit theorem about a sample mean when the variable is normally distributed

Example:

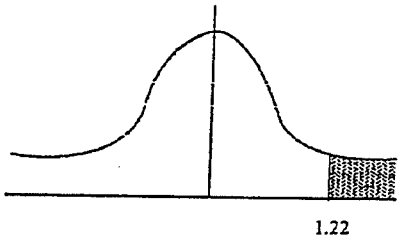
The mean weight of 15-year-old males is 142 pounds, and the standard deviation is 12.3 pounds. If a sample of thirty-six 15-year-old males is selected, find the probability that the mean of the sample will be greater than 144.5 pounds. Assume the variable is normally distributed. Based on your answer, would you consider the group overweight?

Solution

$\mu = 142$, $\sigma = 12.3$ and $n = 36$

$P(x' > 144.5)$

$$= P\left(z > \frac{144 - 142}{\frac{12.3}{\sqrt{36}}}\right)$$



$= P(z > 1.22)$

$= 1 - 0.8888$

$= 0.1112$

$= 11.12\ \%$

التحليل الجيبي
∴ المساحة المطلوبة (الاحتمال)
= 1 − ناتج الكشف في الجدول عند 1.22
الجدول موجود page 24

- No: since the average weight is within 2 standard deviation of the mean.

Example:

The average age of chemical engineers is 37 years a standard deviation of 4 years. If an engineering (شركة هندسية) firm employs 25 chemical engineers, find the probability that the average age of the group is greater than 38.2 years old. If this is the case, would it be safe to assume that the engineers in this group are generally much older than average?

Solution

$\mu = 37$, $\sigma = 4$ and $n = 25$

$P(x' > 38.2)$

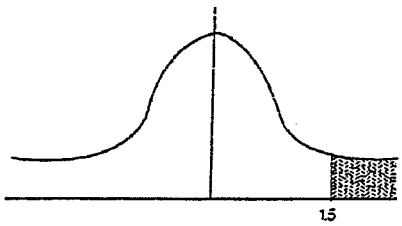
$$= p\left(z > \frac{38.2 - 37}{\frac{4}{\sqrt{25}}}\right)$$

$= p(z > 1.5)$

$= 1 - 0.9332$

$= 0.0668$

$= 6.68\%$



التحليل بسيط
∴ المساحة المطلوبة (الاحتمال)
= 1 − ناتج الكشف في الجدول عند 1.5

Example:

The average annual salary in Pennsylvania was \$24,393 in 1992. Assume that salaries were normally distributed for a certain group of wage earners, and the standard deviation of this group was \$4362.

a. Find the probability that a randomly selected individual (أشخاص) earned (مكسبه) less than \$26,000.

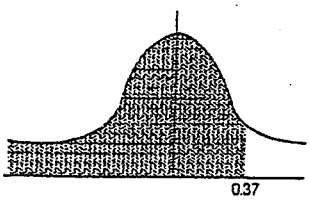
b. Find the probability that, for a randomly selected sample of 25 individuals, the mean salary was less than \$26,000.

c. Why is the probability for part b higher than the probability for part a.

Solution

$\mu = 24393$, $\sigma = 4362$

(a) $p(x < 26000) = p\left(z < \frac{26000 - 24393}{4362}\right)$



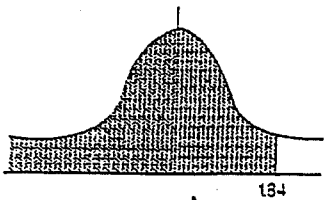
التظليل يسار
∴ كشف مباشر من الجدول

$= p(z < 0.37)$

$= 0.6443$

(b) $\mu = 24393$, $\sigma = 4362$ and $n = 25$

$P(x' < 26000) = p\left(z < \frac{26000 - 24393}{\frac{4362}{\sqrt{25}}}\right)$



التظليل يسار
∴ كشف مباشر من الجدول

$= p(z < 1.84)$

$= 0.9671$

(c) Sample means are less variable than individual data.

Example:

الانقباض ضغط الدم

Assume that the mean systolic blood pressure of normal adults بالغ سن الرشد is 120 millimeters of mercury (mm Hg) and the standard deviation is 5.6. Assume the variable is normally distributed.

(a.) If an individual is selected, find the probability that the individual's pressure will be between 120 and 121.8 mm Hg.

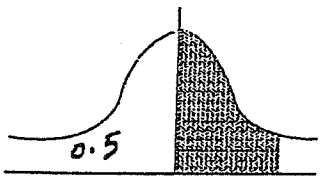
(b.) If a sample of 30 adults is randomly selected, find the probability that the sample mean will be between 120 and 121.8 mm Hg.

(c.) Why is the answer to part a so much smaller than the answer to part b?

Solution

(a) $p(120 < x < 121.8) = p\left(\frac{120-120}{5.6} < z < \frac{121.8-120}{5.6}\right)$

$= 0.6255 - 0.5 = 0.1255$



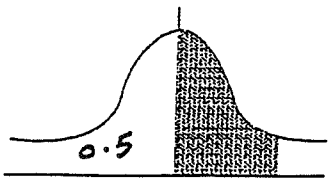
نكشف عن 0.32 من الجدول ثم نطرح من ناتج الكشف 0.5

(b) $\mu = 120$, $\sigma = 5.6$ and $n = 30$

$$P(120 < x' < 121.8) = p\left(\frac{120-120}{\frac{5.6}{\sqrt{30}}} < z < \frac{121.8-120}{\frac{5.6}{\sqrt{30}}}\right)$$

$= p(0 < z < 1.76) = 0.9608 - 0.5$

$= 0.4608$



نكشف عن 1.76 من الجدول ثم نطرح من ناتج الكشف 0.5

(c) Sample means are less variable than individual data .

A survey found that the microwave ovens have an average life of 3 years with a standard deviation of 0.5 year. Assume the variable is normally distributed. What percent of microwave ovens would be replaced if a warranty of 18 months were given?

A) 0.13% B) 2.28% C) 10.56% D) 0.52%

* average $\mu = 3$

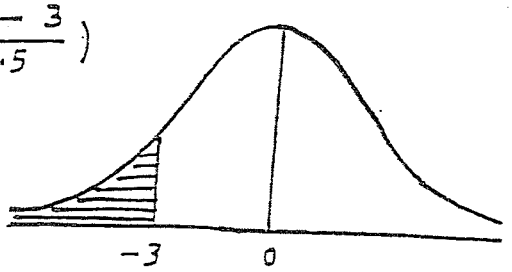
* $\sigma = 0.5$

18 months $= \frac{18}{12} = 1.5$ years

where μ and σ by years

$P(X < 1.5) = P(z < \frac{1.5 - 3}{0.5})$

$= P(z < -3)$



$= 0.0013$

لإيجاد ال Percent نضرب في 100%

$= 0.0013 \times 100\%$

$= 0.13\%$

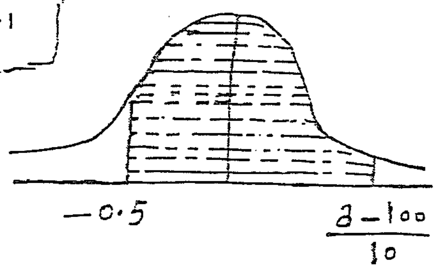
$P(95 < X < a) = 0.6309$ $\mu = 100$
A) 115.5 B) 101.5 C) 84.5 D) 15.5 $\sigma = 10$

$$P(95 < X < a) = 0.6309$$

$$= P\left(\frac{95-100}{10} < Z < \frac{a-100}{10}\right) = 0.6309$$

$$= P\left(-0.5 < Z < \frac{a-100}{10}\right) = 0.6309$$

لو كان الحدان في جهة واحدة تكون المساحة بينهما أقل من 0.5
∴ المساحة 0.6309 أكبر من 0.5
∴ الحدان في جهتين مختلفتين من خط المنتصف ،
أي أن $\frac{a-100}{10}$ يقع في الجهة اليمنى
لأن -0.5 يقع في الجهة اليسرى



$$P\left(Z < \frac{a-100}{10}\right) - P(Z < -0.5) = 0.6309$$

نكشف في الجدول ← بالموجب

$$P\left(Z < \frac{a-100}{10}\right) = 0.6309 + 0.3085$$

$$P\left(Z < \frac{a-100}{10}\right) = 0.9394$$ ← نكشف عكسي عنه

نحصل على $z = 1.55$

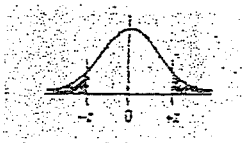
$$\frac{a-100}{10} = 1.55$$

$$a - 100 = (1.55)(10)$$

$$a = (1.55)(10) + 100 \Longrightarrow a = 115.5$$

T

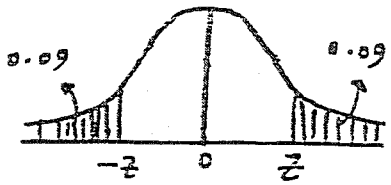
Find the value of z such that the shaded tail areas equals 18%



مجموع المساحتين المظللتين على الأطراف = 0.18

∴ المساحة الواحدة = $\frac{0.18}{2}$ = 0.09

A) 1.4 B) 1.34 C) 1.22 D) 1.47



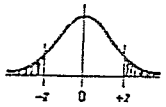
∴ نكشف عن 0.09 كشف نكسي

نجد أنها أقرب إلى 0.0901

فتكون z-المقابلة -1.34

$\Rightarrow \boxed{z = 1.34}$

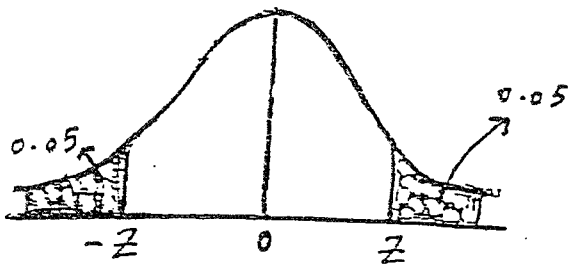
Find the value of z such that the shaded tail areas equals 0.10



مجموع المساحتين المظللتين = 0.10

∴ المساحة الواحدة = $\frac{0.10}{2}$ = 0.05

A) 1.64 B) 1.28 C) -1.96 D) 0.25



∴ نكشف عن 0.05 كشف نكسي

نجد أنها أقرب إلى 0.0505

فتكون z-المقابلة -1.64

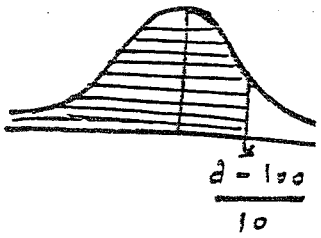
$\Rightarrow z = 1.64$

$P(X < a) = 0.8665$

A) 101.1 B) 113.3 C) 122.2 D) 111.1

$\mu = 100 \qquad \sigma = 10$

$P\left(Z < \frac{a-100}{10}\right) = 0.8665$



نكشف في الجدول كشف بكم عسى 0.8665
تكون قيمة $\frac{a-100}{10}$ هي قيمة الناتج

اما 1.1 تحت 0.1
∴ الناتج 1.11

$\therefore \frac{a-100}{10} = 1.11$

$\Rightarrow a - 100 = (1.11)(10)$

$\Rightarrow a = (1.11)(10) + 100$

$\Rightarrow a = 111.1$

Use the following to answer questions

Let X be a normally distributed random variable with mean 100 and a standard deviation 10. Use this information to find the value of a such that

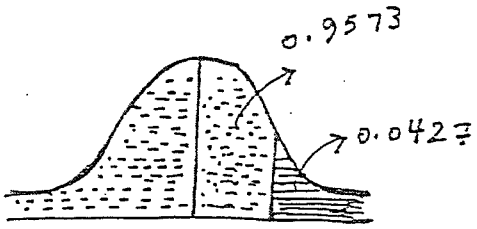
$P(X > a) = 0.0427$

A) 101.72 B) 17.2 C) 117.2 D) 82.8

$$P(X > a) = 0.0427$$

$$= P\left(Z > \frac{a-100}{10}\right) = 0.0427$$

$$\therefore P\left(Z < \frac{a-100}{10}\right) = 0.9573$$



نبحث كشف عكسي عن 0.9573 (من داخل الجدول) نجده أمام 1.7 تحت 0.02

$$\therefore Z = 1.72 \Longrightarrow Z = \frac{a-\mu}{\sigma}$$

$$1.72 = \frac{a-100}{10}$$

$$a - 100 = 17.2$$

$$a = 17.2 + 100$$

$$\Longrightarrow a = 117.2$$

Use the following to answer questions

The time T_1 to travel from A to B through city center (road R_1) is normally distributed with mean 10 minutes (μ_1) and standard deviation 2 minutes (σ_1).
The time T_2 to travel from A to B through a new ring road (road R_2) is normally distributed with mean 15 minutes (μ_2) and standard deviation 3 minutes (σ_2).
You have 12 minutes to travel from A to B on an important appointment. Use these information to solve the following

$\mu_1 = 10$, $\sigma_1 = 2$ $\quad$ $\mu_2 = 15$, $\sigma_2 = 3$

$P(T_2 > 12)$

A) 0.6587 B) 0.8413 C) 0.3413 D) 0.1587

أكبر من ٠.٥ إذاً ناقص الكشف من الجدول

$$P(T_2 > 12) = P\left(Z_2 > \frac{12-15}{3}\right) = P(Z_2 > -1) = 1 - 0.1587 = 0.8413$$

$P(T_1 > 12)$

A) 0.1587 B) 0.6587 C) 0.3413 D) 0.8413

أكبر من ٠.٥ إذاً ناقص الكشف من الجدول

$$P(T_1 > 12) = P\left(Z_1 > \frac{12-10}{2}\right) = P(Z_1 > 1) = 1 - 0.8413 = 0.1587$$

Your correct decision is

A) R_1 is better than R_2
B) R_1 is not as good as R_2
C) Both R_1 and R_2 are the same
D) R_2 is better than R_1

R_1 is better than R_2 أفضل من

((لأنه وقت الوصول عن طريقه R_1 أقل))

If the scores for a test have a mean of 70 and a standard deviation of 12, find the percentage of scores that will fall below 50. Assume the test scores are normally distributed.

كم نسبة مئوية ؟

A) 35.54% B) 4.75% C) 42.07% D) 45.54%

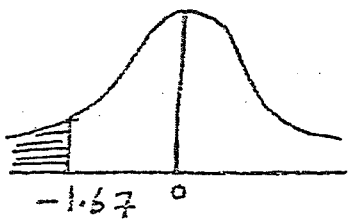
$\mu = 70$, $\sigma = 12$

$P(X < 50)$

$$= P\left(Z < \frac{50-70}{12}\right) = P(Z < -1.67)$$

كشف مباشر من جدول الموجب والسالب

$= 0.0475$



The percentage $= 0.0475 \times 100\% = 4.75\%$

Use the following to answer questions

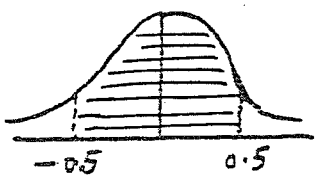
Suppose that the monthly allowance, X, of a student in a given school is normally distributed with mean S300 and standard deviation S50. $\mu = 300 \quad \sigma = 50$

The probability that the monthly allowance of a student selected at random is between S275 and S 325 is

A) 0.383 B) 0.617 C) 0.8085 D) 0.1915

$$P(275 < X < 325) = P\left(\frac{275-300}{50} < z < \frac{325-300}{50}\right)$$

$$= P(-0.5 < z < 0.5)$$



$$= P(z < 0.5) - P(z < -0.5)$$

$$= 0.6915 - 0.3085 = 0.383$$

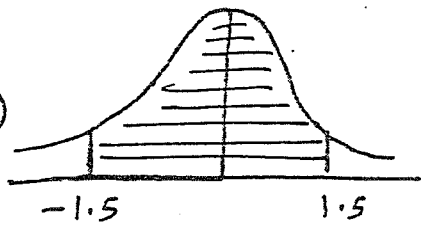
If a random sample of 9 students is selected randomly, find the probability that the mean allowance of the sample is between S275 and S325

A) 0.134 B) 0.4332 C) 0.567 D) 0.8664

$$(\text{sample}) \longrightarrow n = 9$$

$$P(275 < \bar{x} < 325) = P\left(\frac{275-300}{\frac{50}{\sqrt{9}}} < z < \frac{325-300}{\frac{50}{\sqrt{9}}}\right)$$

$$= P(-1.5 < z < 1.5)$$



$$= P(z < 1.5) - P(z < -1.5)$$

$$= 0.9332 - 0.0668$$

$$= 0.8664$$

Use the following to answer questions

The monthly income, X, of a family in a given city is normally distributed with mean $3000 and standard deviation $500.

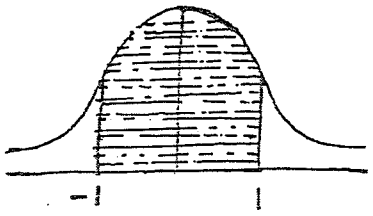
The probability that a person selected at random earns a monthly income between $2500 and $3500
A) 0.3413 B) 0.6826 C) 0.1587 D) 0.3174

$$P(2500 < X < 3500)$$

$$= P\left(\frac{2500-3000}{500} < Z < \frac{3500-3000}{500}\right)$$

$$= P(-1 < Z < 1)$$

الكشف عن العددين 1 ، 1-
من الجدول ثم طرح الناتجين الأكبر ناقص الأصغر



$$= 0.8413 - 0.1587 = 0.6826$$

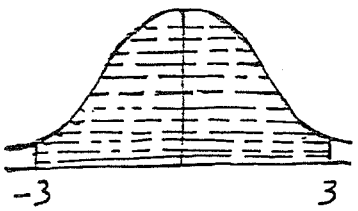
If a random sample of size 9 is selected at random, find the probability that the mean income of the sample is between $2500 and $3500
A) 0.0013 B) 0.4987 C) 0.9974 D) 0.5601

$$P(2500 < \bar{X} < 3500)$$

$$= P\left(\frac{2500-3000}{\frac{500}{\sqrt{9}}} < Z < \frac{3500-3000}{\frac{500}{\sqrt{9}}}\right)$$

$$= P(-3 < Z < 3)$$

الكشف عن العددين 3 ، 3-
من الجدول ثم طرح الناتجين الأكبر ناقص الأصغر



$$= 0.9987 - 0.0013 = 0.9974$$

Use the following to answer questions

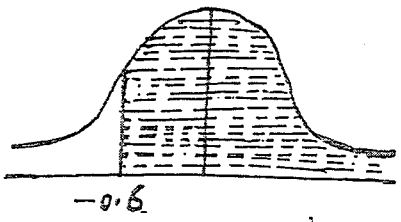
The time T_1 to travel from A to B through city center (road R_1) is normally distributed with mean 20 minutes and standard deviation 5 minutes.
The time T_2 to travel from A to B through a new ring road (road R_2) is normally distributed with mean 15 minutes and standard deviation 8 minutes.
You have 17 minutes to travel from A to B on an important appointment. Using this information, solve the following

$P(T_1 > 17)$
A) 0.2743 B) 0.7257 C) 0.2257 D) 0.7743

$$P(T_1 > 17)$$

$$= P\left(Z_1 > \frac{17 - 20}{5}\right)$$

$$= P(Z_1 > -0.6)$$



∴ التظليل يمين
∴ الاحتمال المطلوب
هو 1 ناقص الكشف فرالجدول عم 0.6-

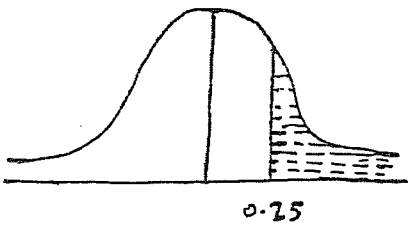
$$= 1 - 0.2743$$
$$= \boxed{0.7257}$$

$P(T_2 > 17)$
A) 0.5987 B) 0.0987 C) 0.4013 D) 0.9013

$$P(T_2 > 17)$$

$$= P\left(Z_2 > \frac{17 - 15}{8}\right)$$

$$= P(Z_2 > 0.25)$$
$$= 1 - 0.5987$$
$$= \boxed{0.4013}$$



∴ التظليل يمين
∴ الاحتمال المطلوب
هو 1 ناقص الكشف فرالجدول عم 0.25

Your correct decision is
A) R_2 is better than R_1
B) R_1 is better than R_2
C) Both R_1 and R_2 are the same
D) R_2 is not as good as R_1

Use the following to answer questions

Let X be a normally distributed random variable with mean 100 and a standard deviation 10. Use this information to find the value of a such that

$P(a < X < 111) = 0.6223$

A) 97 B) 95 C) 98 D) 93

$\mu = 100 \qquad \sigma = 10$

$P(a < x < 111) = 0.6223$

$P\left(\frac{a-100}{10} < z < \frac{111-100}{10}\right) = 0.6223$

$P\left(\frac{a-100}{10} < z < 1.1\right) = 0.6223$

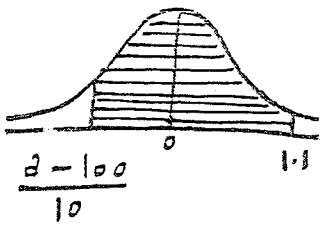
لو كان المساحة في جهة واحدة لكانت المساحة بينهما أقل من 0.5

بما المساحة 0.6223 أكبر من 0.5

∴ المساحة في جهتين مختلفتين من خط المنتصف.

أما أن $\frac{a-100}{10}$ يقع في الجهة اليسرى

و أن 1.1 يقع في الجهة اليمنى.



$P(z < 1.1) - P\left(z < \frac{a-100}{10}\right) = 0.6223$

$0.8643 - P\left(z < \frac{a-100}{10}\right) = 0.6223$

$0.8643 - 0.6223 = P\left(z < \frac{a-100}{10}\right)$

$\therefore P\left(z < \frac{a-100}{10}\right) = 0.242$ ← كشف كس عم

$z = -0.7$ بكوم قيمة

$\Rightarrow \frac{a-100}{10} = -0.7$

$\Rightarrow a - 100 = (-0.7)(10)$

$a = (-0.7)(10) + 100 \Rightarrow \boxed{a = 93}$

ALSAADI

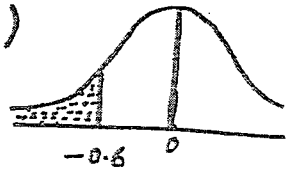
Use the following to answer questions

Let T_1, the time to travel from A to B through an old road, be normally distributed with mean 22 minutes and standard deviation 5 minutes.
Let T_2, the time to travel from A to B through a new road, be normally distributed with mean 16 minutes and standard deviation 7 minutes.
You have 19 minutes to travel from A to B on an important appointment. Using this information, solve the following

Your correct decision is that the new road is ... the old road.
A) same as B) not as good as C) worse than D) better than

** old road $\mu = 22$, $\sigma = 5$

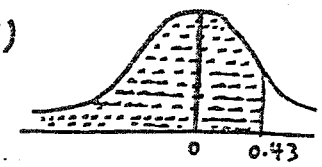
$$P(T_1 < 19) = P\left(z < \frac{19-22}{5}\right) = P(z < -0.6)$$



$\therefore P(T_1 < 19) = \boxed{0.2743}$

** New road $\mu = 16$, $\sigma = 7$

$$P(T_2 < 19) = P\left(z < \frac{19-16}{7}\right) = P(z < 0.43)$$



$\therefore P(T_2 < 19) = \boxed{0.6664}$

$\because P(T_2) > P(T_1)$

∴ احتمال أن يصل أسرع (في وقت أقل) عندما يستخدم الطريق الجديد New road

∴ New road is better than old road

$P(T_1 < 19)$ الاحتمال أقل
A) 0.7743 B) 0.2743 ↑ C) 0.7257 D) 0.2257

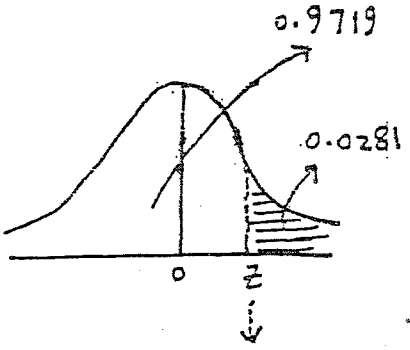
$P(T_2 < 19)$ الاحتمال أكبر
A) 0.4013 B) 0.9013 C) 0.0987 D) 0.6664 ↑

Use the following to answer questions

Let X be a normally distributed random variable with mean 100 and a standard deviation 10. Use this information to find the value of a such that

$P(X > a) = 0.0281$

A) 1.91 B) 119.1 C) 80.9 D) 19.1



$\mu = 100 \quad , \quad \sigma = 10$

$P(X > a) = 0.0281$

$P\left(Z > \frac{a-100}{10}\right) = 0.0281$

$z = \frac{a-100}{10}$

نكشف كشف عكسي من الداخل للخارج

$1 - 0.0281 = 0.9719$ ← نحسم

$z = 1.91 \longrightarrow 1.91 = \frac{a-100}{10}$

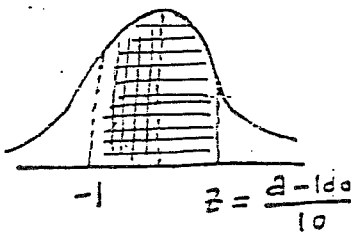
$\Rightarrow a - 100 = (1.91)(10)$

$a = (1.91)(10) + 100$

$a = 119.1$

$P(90 < X < a) = 0.6519$

A) 90.0 B) 108.8 C) 110.0 D) 91.2



$P(90 < x < a) = 0.6519$

$P\left(\frac{90-100}{10} < z < \frac{a-100}{10}\right) = 0.6519$

$P\left(-1 < z < \frac{a-100}{10}\right) = 0.6519$

$P\left(z < \frac{a-100}{10}\right) - P(z < -1) = 0.6519$

$P(z < -1) = 0.1587$ نكشف عنه من الجدول

$\therefore P\left(z < \frac{a-100}{10}\right) = 0.8106$ ← بالموجب نكشف عكسي

$\Rightarrow z = 0.88$ ← نكون قيمة z هنا

$0.88 = \frac{a-100}{10}$

$a - 100 = (0.88)(10)$

$a = (0.88)(10) + 100$

$a = 108.8$

Use the following to answer questions

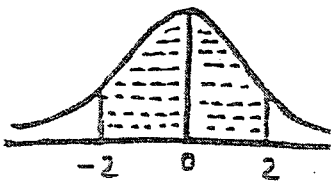
The weekly income, X, of a family in a given city is normally distributed with mean $200 and standard deviation $25.

The probability that a person selected at random earns a weekly income between $150 and $250
A) 0.3174 B) 0.9542 C) 0.6826 D) 0.1587

$\mu = 200 \qquad \sigma = 25$

$P(150 < X < 250)$

$= P\left(\frac{150-200}{25} < Z < \frac{250-200}{25}\right)$



$= P(-2 < Z < 2)$

الكشف عن العددين 2 ، 2- من الجدول
ثم طرح الأكبر ناقص الأصغر.

$= 0.9772 - 0.0228$

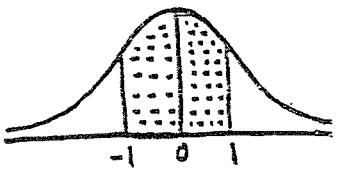
$= \boxed{0.954}$

If a random sample of size 4 is selected at random, find the probability that the mean income of the sample is between $187.5 and $212.5
A) 0.4987 B) 0.6826 C) 0.0013 D) 0.9974

Sample size : $n = 4$

$P(187.5 < X < 212.5)$

$= P\left(\frac{187.5-200}{\frac{25}{\sqrt{4}}} < Z < \frac{212.5-200}{\frac{25}{\sqrt{4}}}\right)$



$= P(-1 < Z < 1)$

الكشف عن العددين 1 ، 1- من الجدول
ثم طرح الأكبر ناقص الأصغر

$= 0.8413 - 0.1587 = \boxed{0.6826}$

To qualify for the medical school, the student must score in the top 10% on a general test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume test scores are normally distributed.

A) 276 B) 1.28 C) 25.6 D) 226

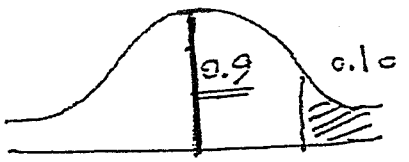
$P(X > a) = 0.10$

$P(z > \frac{a-200}{20}) = 0.1$

$P(z < \frac{a-200}{20}) = 0.9$

نكشف كشف عكسي عن 0.9
نجد أنها اقرب إلى 0.8997
التي تقابل z

$z = 1.28$



$z = \frac{a-200}{20}$

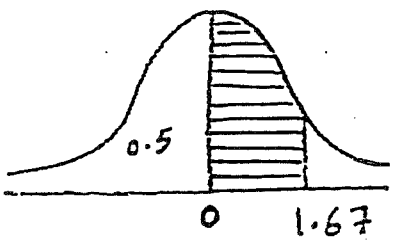
$1.28 = \frac{a-200}{20}$

$a - 200 = 25.6$

$a = 225.6$

Find the probability for $P(0 < z < 1.67)$

A) 0.4525 B) 0.4207 C) 0.3554 D) 0.4554



نكشف عن 1.67 في جدول الموجب والسالب
ثم نطرح من الناتج 0.5

$= 0.9525 - 0.5 = 0.4525$

When the distribution is **positively skewed**, the relationship of the mean, median, and mode will be ...

A) mode > median > mean.
B) median < mode < mean.
C) mean < mode < median
D) mean > median > mode.

-positive skew
mean > median > mode → اجابة السؤال

-Negative skew
mean < median < mode — شرح زيادة

Symmetric skew
mean = median = mode — شرح زيادة

A distribution of the means that are computed from all possible random samples of a specific size taken with replacement from a population." The previous statement is the definition of ...
A) central limit theorem B) empirical distribution C) sampling distribution D) sampling error

تعريف
↳Sampling distribution .

تعريف

- Sampling error is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

The standard deviation of a distribution is 20. If a sample of 225 is selected, what is the standard error of the mean?

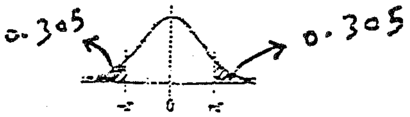
A) $\frac{4}{45}$ B) $\frac{7}{5}$ C) $\frac{4}{3}$ D) $\frac{3}{4}$

$\sigma = 20$

$n = 225$

$$\text{standard error of the mean} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{225}} = \frac{4}{3}$$

Find the value of z such that the shaded tail areas equals 0.61



A) 0.11 B) 0.305 C) 0.195 D) 0.51

المساحتين المظللتين
متساويتين مجموعهما 0.61
مساحة المنطقة الواحدة
$= \frac{0.61}{2} = 0.305$

0.3050

نكشف عنها كشف عكسي
من داخل الجدول فتكون قيمة Z المقابلة هي 0.51-

$\therefore Z = 0.51$

ALSAADI

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. The total area under a normal distribution is <u>infinite.</u> (x)
2. The standard normal distribution is a continuous distribution. (√)
3. All variables that are approximately normally distributed can be transformed to standard normal variables. (√)
4. The z value corresponding to a number <u>below</u> the mean is always negative. (√)
5. <u>The area</u> under the standard normal distribution to the left of $z = 0$ is negative. (x)
6. The central limit theorem applies to means of samples selected from <u>different</u> populations. (x)

<u>Select the best answer.</u>

7. The mean of the standard normal distribution is

a. 0 b. 1 c. 100 d. variable

8. Approximately what percentage of normally distributed data values will fall within 1 standard deviation above or below the mean?

a. 68% b. 95% c. 99.7% d. Variable

9. Which is not a property of the standard normal distribution?

a. It's symmetric about the mean. b. It's uniform.

c. It's bell-shaped. d. It's unimodal.

(10) When a distribution is positively skewed, the relationship of the mean, median, and mode from left to right will be .

a. Mean, median, mode

b. Mode, median, mean

c. Median, mode, mean

d. Mean, mode, median

---

(11) The standard deviation of all possible sample means equals

a. The population standard deviation.

b. The population standard deviation divided by the population mean.

c. The population standard deviation divided by the square root of the sample size.

d. The square root of the population standard deviation.

Complete the following statements with the best answer.

---

(12) When one is using the standard normal distribution, $P(z < 0) = \underline{0.5}$.

---

(13) The difference between a sample mean and a population mean is due to Sampling error.

---

(14) The mean of the sample means equals Population mean.

---

(15) The standard deviation of all possible sample means is called Standard error of the mean.

السعدي

ALSAADI

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

السعدي ALSAADI

نسخة جديدة منقحة

1433/32

8

Ch. 10

- Scatter Plots and Correlation.
- Regression.

Ch. 13 - Sec. 6

- The Spearman Rank Correlation Coefficient.

STAT.110

جمال السعدي

رياضيات - إحصاء

Ch. 10 Part. 1

Correlation → الارتباط

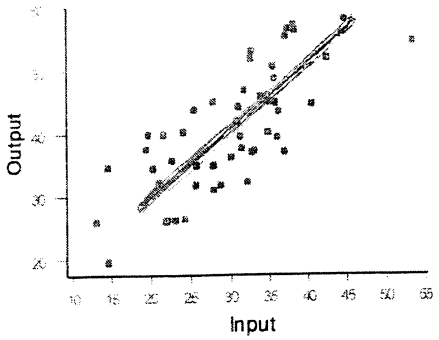
Is a statistical method to determine whether a relationship between variables exists.

Scatter plots: is show the relationship between the independent and dependent variables.

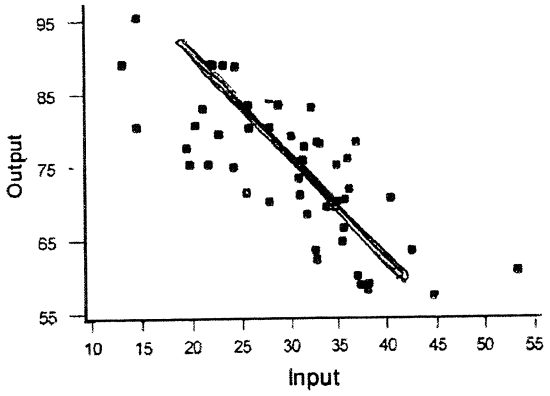
← اشكال الانتشار

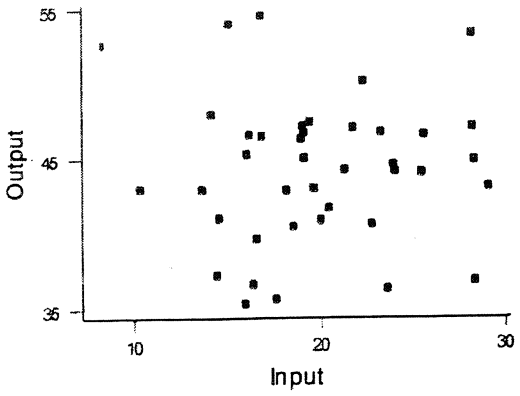
Types of correlation → انواع الارتباط

Positive correlation
As x increases, y also increases



Negative correlation
As x increases, y decreases





الرسم يوضح نوع ومدى العلاقة بين متغيرين X,Y (ظاهرتين) من حيث

نوع العلاقة	مدى العلاقة
positive طردية	Strong قوية
negative عكسية	Weak ضعيفة
no relation لا توجد علاقة	Zero منعدمة

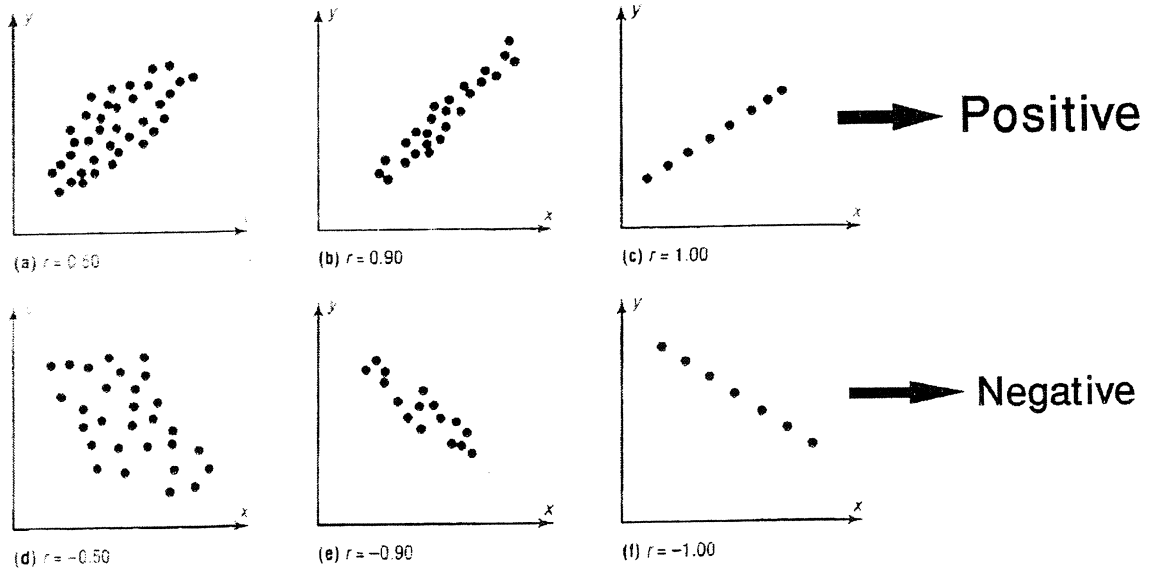
ALSAADI

معامل الارتباط

The correction coefficient computed from the sample data measures the **strength** and **direction** of linear relationship between tow variables. The symbol for the sample correlation coefficient is r. the symbol for the population correlation coefficient is ρ.

The range of the correlation coefficient is [- 1, 1].

- If r = -1: There is a perfect(complete) negative linear relationship.
- If r = 1 : There is a perfect(complete) positive linear relationship.
- If r = 0: There is no linear relationship (Does not exist).



weak ضعيف

strong قوي

Perfect (complete) تام

* The Pearson Correlation coefficient r : معامل ارتباط بيرسون

$$r = \frac{n.\sum xy - \sum x \ . \ \sum y}{\sqrt{[n.\sum x^2 - (\sum x)^2] \times [n.\sum y^2 - (\sum y)^2]}}$$

Note

مدى العلاقة	نوع العلاقة
• r = 0	no relation
• 0.01 ≤ r ≤ 0.29	Very weak ضعيف جدًا
• 0.30 ≤ r ≤ 0.49	Weak ضعيف
• 0.50 ≤ r ≤ 069	Moderate متوسط
• 0.70 ≤ r ≤ 0.89	Strong قوي
• 0.90 ≤ r ≤ 0.99	Very strong قوي جدًا
• r = 1	Perfect (complete)

- n is the number of data pairs. (عدد خانات الجدول)
- Round the value of r to two decimal places

Example

Compute the value of the correlation coefficient for the data obtained in the study of the number of absences and the final grade of the seven students

Sample **Solution**

Student	Number of Absences x	Final grad y (%)	x . y	x^2	y^2
A	6	82	492	36	6724
B	2	86	172	4	7396
C	15	43	645	225	1849
D	9	74	666	81	5476
E	12	58	696	144	3364
F	5	90	450	25	8100
G	8	78	624	64	6084
	$\sum x = 57$	$\sum y = 511$	$\sum xy = 3745$	$\sum x^2 = 579$	$\sum y^2 = 38993$

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

$$= \frac{(7)(3745) - (57)(511)}{\sqrt{[(7)(579) - (57)^2][(7)(38993) - (511)^2]}} = -0.94$$

- **Strong negative relationship between a student's Final grade and the number of absences a student has.**

ALSAADI

Example

Compute the value of the correlation coefficient for the data obtained in the study of age and blood pressure.

Sample **Solution**

Subject	Age x	Pressure y	x . y	x^2	y^2
A	43	128	5504	1849	16384
B	48	120	5760	2304	14400
C	56	135	7560	3136	18225
D	61	143	8723	3721	20449
E	67	141	9447	4489	19881
F	70	152	10640	4900	23104
	$\sum x=345$	$\sum y=819$	$\sum xy=47634$	$\sum x^2=20399$	$\sum y^2=112443$

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

$$= \frac{(6)(47634) - (345)(819)}{\sqrt{[(6)(20339) - (345)^2][(6)(1124433) - (819)^2]}} = 0.89$$

The correlation coefficient is a strong positive relationship between age and blood pressure.

Use the following to answer questions

In the study of relationship between the number of absences X and the final grade Y of 8 students in the statistic class, the data are shown as follows

$n = 8$

$\sum X = 42, \sum Y = 470, \sum XY = 3143, \sum X^2 = 354$ and $\sum Y^2 = 37358$

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]\cdot[n\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{8(3143) - (42)(470)}{\sqrt{[8(354) - (42)^2][8(37358) - (470)^2]}}$$

$$= \boxed{0.592}$$

The value of the Pearson correlation coefficient means that there is a ... linear relationship between the number of absences and the final grade.

A) strong negative B) strong positive C) moderate negative D) moderate positive

$\because r = 0.592$

موجب

$0.69 < 0.5$ وحصورة بين

$\therefore$ moderate positive

The range of the Pearson correlation coefficient value (r) for the positive linear relationship is ...

A) $0 < r \leq 1$ B) $0 \leq r \leq 1$ C) $0 \leq r < 1$ D) $-1 \leq r \leq 1$

موجب

The range of positive r is $0 < r \leq 1$

If the value of the correlation coefficient equals -0.95, then the type of the relationship is ...
A) weak negative B) strong negative C) strong positive D) weak positive

عندما تكون $r = -0.95$
فإنها تكون strong negative

If the value of the correlation coefficient equals -0.19, then the type of the relationship is ...
A) weak negative B) strong negative C) strong positive D) weak positive

---

If Pearson correlation coefficient (r_p) equals 0.45, then the relationship can be described as

A) positive, strong and non linear
B) weak and linear
C) positive, moderate and non linear
D) positive, moderate and linear

$r_p = 0.45$ is weak and linear

- A positive relationship exists when both variables increase or decrease at the same time.

- A negative relationship exists when one variable increases and the other variable decreases.

If there is a strong positive linear relationship between the variables, the value of r will be close to +1.

If there is a strong negative linear relationship between the variables, the value of r will be close to – 1.

كمية الدهون الوزن

The correlation coefficient between the amount of fats which a person eats and his or her weight may be

A) close to -1 B) close to 2 C) close to 1 D) 0

* العلاقة بين كمية الدهون التي يتناولها الشخص ووزنه الشخصي
علاقة طردية قوية ⇐ تقترب من +1
⇒ close to 1

الارتباط الدخل السنوي الوزن

If the correlation between body weight and annual income were high and positive, we could conclude that high income people

A) are eating more food
B) are eating less food
C) are loosing weight
D) are gaining weight

* اذا كانت العلاقة بين الوزن والدخل السنوي
high and positive علاقة قوية موجبة (طردية)

∴ اصحاب الدخل العالي يزدادون من الوزن
أي كلما زاد الدخل كلما زاد الوزن .
⇐ الاختيار الصحيح هو D

A negative relationship between two variables means that for the most part, as the X variable decreases, the Y variable

A) stays the same B) increases C) decreases D) equals X

Negative relationship

= X decreases ⟹ y increases

عكس بعضهم y ∝ x

CH. 10 Part 2

Regression → الانحدار

Regression is a statistical method used to describe the nature of the relationship between variables.

In a simple relationship, there are only two types of variables under study; an independent variable (explanatory variable) or predictor variable and a dependent variable (outcome variable) or response variable

---

Equation of regression Line is: $\hat{y} = a + b\,x$

- **Another name is line of best fit.**

Where b: is the slope of the line. a: is y – Intercept.

$$b = \frac{n\sum xy - \sum x.\sum y}{n\sum x^2 - (\sum x)^2}$$

$$a = \frac{\sum y.\sum x^2 - \sum y - \sum xy}{n\sum x^2 - (\sum x)^2}$$

- round the values of a and b to three decimal places

Example

Instructor is interested in finding the relationship between the final exam grades of students enrolled in calculus I (x)

And calculus II (y). use the following in formation to answer the following question:

$n = 5$, $\sum x = 400$, $\sum y = 380$, $\sum x^2 = 32656$

$\sum y^2 = 29546$, $\sum xy = 31023$, slope (b) = 0.95

(1)The Pearson correlation coefficient is :

(a) 0.94 (b) 0.81 (c) 0.68 (d) 0.34

Solution

$$r = \frac{n.\sum xy - \sum x \,.\, \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n.\sum y^2 - (\sum y)^2]}}$$

$$= \frac{5.(31023) - (400)\,(380)}{\sqrt{[5.(32656) - (400)^2][5.\,(29546) - (380)^2]}} = 0.94$$

(2) Interpret the value of the person correlation coefficient (r).

a. Weak positive linear correlation.

b. Weak negative linear correlation.

c. Strong positive linear correlation.

d. Strong negative linear correlation.

(3) The equation of the regression line is:

a. $y' = 0.154 + 1.48x$ b. $y' = 0.924 + 0.925x$

c. $y' = 0.024 + 0.95x$ d. $y' = 0.95 + 0.024x$

Solution

- $b = 0.95$ (the slope)
- $a = \dfrac{\sum y.\sum x^2 - \sum x.\sum xy}{n.\sum x^2 - (\sum x)^2}$

$$= \frac{(380)(32656) - (400)(31023)}{5(32656) - (400)^2} = 0.024$$

Equation of the regression line

$y' = a + b\,x$

there for $y' = 0.024 + 0.95x$

(4) predict a calculus II exam score for student who get 80 in calculus I.

a. 85 b. 80 c. 70 d. 76

Solution

$\therefore \hat{y} = 0.024 + 0.95\,x$

$= 0.024 + 0.95\,(80) = 76.024 \cong 76$

Example:

Age and sick days

Age x	18	26	39	48	53	58
Days y	16	12	9	5	6	2

- Find y' when x = 47 years

Solution

$n = 6$, $\sum x = 242$, $\sum y = 50$

$\sum x^2 = 10998$, $\sum y^2 = 546$, $\sum xy = 1625$

$$b = \frac{n\sum xy - \sum x.\sum y}{n\sum x^2 - (\sum x)^2} = \frac{6.(1625) - (242)\,.\,(50)}{6.\,(10998) - (242)^2} = -0.317$$

$$a = \frac{n\sum y.\sum x^2 - \sum y\,.\sum xy}{n\sum x^2 - (\sum x)^2} = \frac{(50)\,(10998) - (242)\,(1625)}{6.\,(10998) - (242)^2} = 21.1$$

Equation of regression line:

$\acute{y} = a + b\,x$

$\acute{y} = 21.1 + (-0.317)\,x$

when $x = 47$

$\acute{y} = 21.1 + (-0.317)\,(47) = 6.201 \cong 6$ days

ALSAADI

الميل
The slope of the regression line $y' = -2x + 5$ is:

A) 2 B) 5 C) -5 D) -2

$y' = -2x + 5$

-2 ← الميل slope

5 ← الجزء المقطوع من محور y y-intercept

$\therefore$ slope $= \frac{-2}{\text{T}}$

x عمر الرجل ضعف y عمر الزوجه

Suppose that every man 's age(x) is exactly twice his wife's age(y). Then the Pearson correlation coefficient(r) between x and y is

A) 0.5 B) -0.5 C) 1 D) -1

$X = 2y$

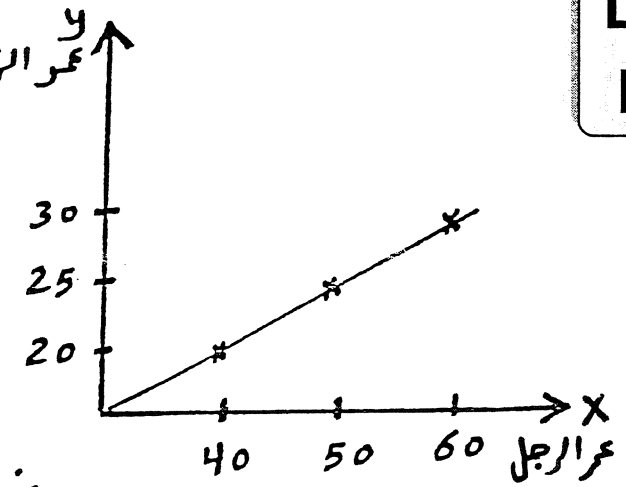
$y = \frac{1}{2} X$

على فرضه أنه:

عمر الرجل x

ضعف

عمر الزوجه y



$\therefore$ جميع النقاط تقع على استقامه واحده

وبالتالي يكون معامل الارتباط $r = 1$

Use the following to answer questions

عدد ساعات ممارسة الرياضة — عمر الشخص

The equation of the regression line between a person's age in years (x) and the number of hours he exercises per week (y) is given by $\hat{y} = 10.944 - 0.18x$

العلاقة — تمثل — الجملة الصحيحة

The correct statement that represents the relationship between (x) and (y) is:

A) When the number of hours he exercises increases by 1 hour, his age increases by 10.944 on average.

B) When the number of hours he exercises decreases by 1 hour, his age decreases by 10.944 on average.

C) When a person's age increases by 1 year, the number of hours he exercises decreases by 0.18 on average.

D) When a person's age increases by 1 year, the number of hours he exercises increases by 0.18 on average.

العلاقة بين العمر X ، وعدد ساعات التمارين y (ممارسة الرياضة)

ممثلة بالمعادلة

$\hat{y} = 10.944 - 0.18x$

* العلاقة عكسية بدليل وجود اشارة (سالب) أمام X

* كلما زاد العمر X بمقدار 1 سنة

كلما قل عدد ساعات التمارين الرياضية y بمقدار معامل X اي 0.18 ساعة

∴ الاجابة الصحيحة هي [C]

توقع — y؟

Predict the number of hours a person exercises per week when his age is 50 years. — X — عندما x = 50 توقع قيمة y؟

A) 1.49 B) 10.04 C) 19.9 D) 1.944

∵ $\hat{y} = 10.944 - 0.18x$

$\hat{y} = 10.944 - 0.18(50)$ بالآلة

$\hat{y} = 1.944$

∴ نضع x=50 من المعادلة ونوجد قيمة y

There is an approximate linear relationship between the height (y in cm) and age (x from 5 to 18 years) described by y' = 50 + 6 * x. Which of the following is NOT CORRECT?

A) Children reach the height of 50 cm when they are 50/6=8.33 years old.
B) Children's heights increase by 6 cm for each year they grow older
C) The estimated height of a 10 years old child is 110 cm
D) My son is 7 years old and is 110 cm tall. He is taller than average

اي الجمل الآتية ليست صحيحة.

العلاقة الخطية بين age (x) و height (y)

$y' = 50 + 6x$

A) $\frac{50}{6}$ يكون الطول 50 عندما يكون العمر

$x = \frac{50}{6}$ نعوضهم عن العمر

فنجد أن قيمة y هي

$y' = 50 + 6\left(\frac{50}{6}\right) = 100$

∴ الطول 100 وليس 50 ∴ الجملة A خاطئة

Not correct

When the correlation coefficient (r) equals zero, the linear relationship between the variables .....
A) is strong B) is weak C) is moderate D) does not exist ← غير موجودة

اذا كانت قيمة معامل الارتباط $r = 0$
معنا معناه لا توجد علاقة بين المتغيرين

Use the following to answer questions

The following data is recorded to determine the relationship between the number of hours a person goes without sleeping (x) and the number of mistakes he makes on a simple test (y):

$$n=10, \sum x=46, \sum y=60, \sum xy=303, \sum x^2=238, \sum y^2=616$$

The Pearson correlation coefficient (r) is

A) 0.24 B) 0.33 C) -0.24 D) -0.33

$$r=\frac{n\Sigma xy-\Sigma x\,\Sigma y}{\sqrt{[n\Sigma x^2-(\Sigma x)^2][n\Sigma y^2-(\Sigma y)^2]}}$$

قانونه

يعطى في نهاية ورقة الأسئلة.

تعويض مباشر من المعطيات

$$=\frac{10(303)-(46)(60)}{\sqrt{[10(238)-(46)^2][10(616)-(60)^2]}}$$

بالآلة $\approx$ 0.33

The value of the Pearson correlation coefficient (r) means that there is ....

A) strong negative linear relationship
B) moderate negative linear relationship
C) strong positive linear relationship
D) weak positive linear relationship

The equation of the regression line is

A) $y'=1.02-1.3x$ B) $y'=1.3-1.02x$ C) $y'=1.02+1.3x$ D) $y'=1.3+1.02x$

* لابد من حساب a و b لكتابة eq. of regression line.

$$*b=\frac{n\Sigma xy-\Sigma x\,\Sigma y}{n\Sigma x^2-(\Sigma x)^2}=\frac{10(303)-(46)(60)}{10(238)-(46)^2}=1.02$$

$$*a=\frac{(\Sigma y)(\Sigma x^2)-(\Sigma x)(\Sigma xy)}{n(\Sigma x^2)-(\Sigma x)^2}=\frac{(60)(238)-(46)(303)}{10(238)-(46)^2}=1.3$$

$$\hat{y}=a+bx$$

$$\hat{y}=1.3+1.02x$$

The correct statement that represents the relationship between (x) and (y) is:

A) When the number of hours a person goes without sleeping increases by 1 hour, his number of mistakes increases by 1.02 on average.

B) When the number of hours a person goes without sleeping increases by 1 hour, his number of mistakes decreases by 1.02 on average.

C) When the number of mistakes increases by 1, the number of hours a person goes without sleeping decreases by 1.3 on average.

D) When the number of mistakes increases by 1 year, the number of hours a person goes without sleeping increases by 1.3 on average.

من السؤال السابق

$\hat{y} = 1.3 + 1.02x$

number of hours without sleeping [x] increasing by 1

" " mistakes [y] increasing by (1.02 معامل x) .

توقع y ؟ x

Predict the number of mistakes for a person who goes 9.51 hours without sleeping.

A) -8 B) 11 C) 13 D) -11

$x = 9.51$ عندما y ؟ توقع قيمة

$x = 9.51$ نعوض من المعادلة

$\hat{y} = 1.3 + 1.02x$

$\hat{y} = 1.3 + 1.02(9.51)$

$\hat{y} \approx \boxed{11}$

ALSAADI

Use the following to answer questions

In the study of relationship between the number of absences X and the final grade Y of 6 students in the statistic class, the data are shown as follows

$\sum X = 42, \sum Y = 470, \sum XY = 3143, \sum X^2 = 354$ and $\sum Y^2 = 37358$

$n = 6$

The correlation coefficient is
A) 0.82 B) 0.92 C) -0.82 D) 1

by substitute in the formula : بالتعويض في الصيغة

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{6(3143) - (42)(470)}{\sqrt{[6(354) - (42)^2][6(37358) - (470)^2]}} = -0.815 \simeq -0.82$$

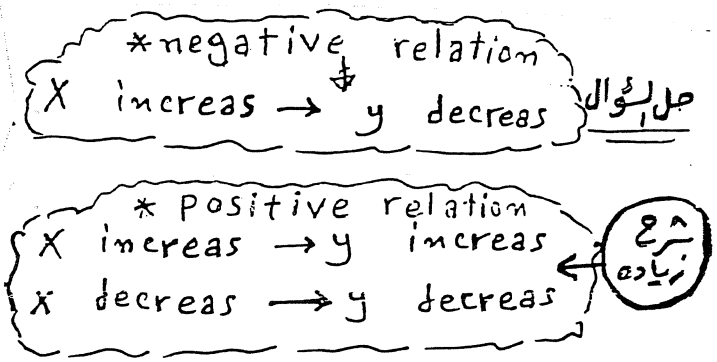
The slope of the regression line is
A) 2.45 B) -3.45 C) 3.45 D) -2.45

The slope of regression line is ميل خط الانحدار

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

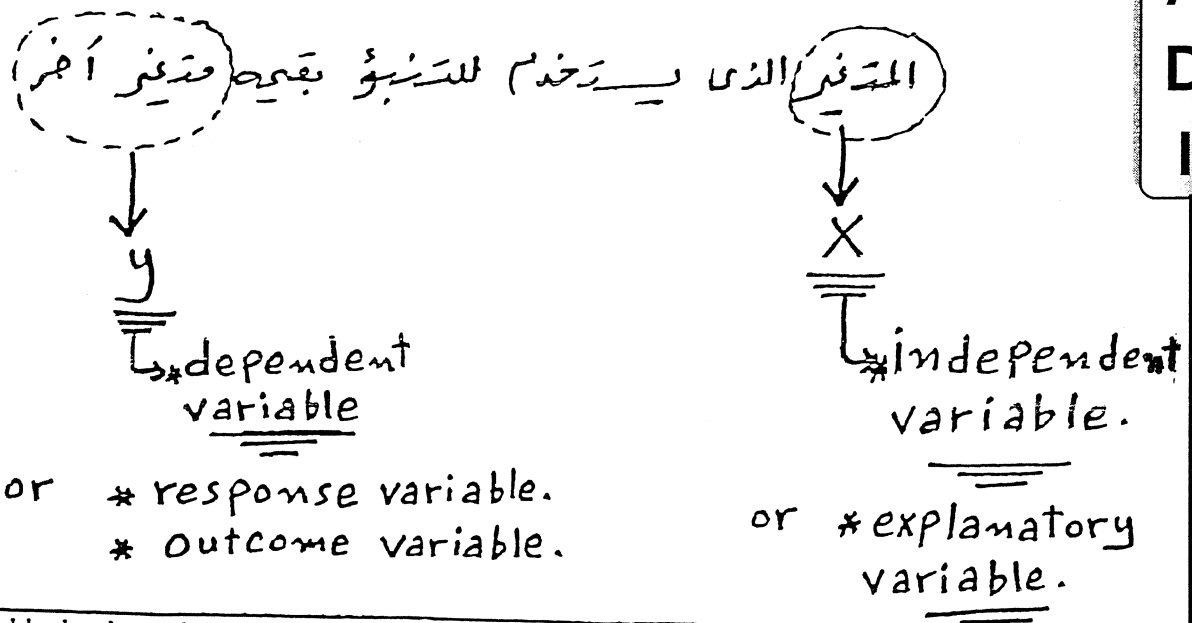
$$= \frac{6(3143) - (42)(470)}{6(354) - (42)^2} = -2.45$$

A negative relationship between two variables means that for the most part ,as the X variable increases, the Y variable
A) decreases B) increases C) equals X D) remains the same



In the relationship between the number of studying hours and an exam grade, the number of studying hours is assumed to be
A) continuous variable B) dependent variable C) nominal variable D) independent variable

The variable that is used to predict the value of another is called ...
A) independent variable.
B) dependent variable.
C) correlation variable.
D) variable of determination.



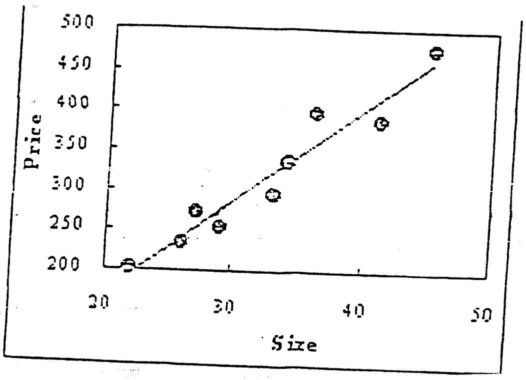
The variable that is used to predict the value of another is called ...
A) response variable.
B) explanatory variable.
C) correlation variable.
D) variable of determination.

A home owner wanted to determine if there was a relationship between the size (in 100 square feet) of a new home and the price (in thousands of dollars) of the home. He found the equation of the regression line and

graph the line on a scatter plot as presented below. $y' = -58.767 + 11.535\,x$

Use this graph to answer the following four questions

Price Size



What would be the predicted home price y if x=40

A) 461.4 B) 402.63 C) 350 D) 3419.15

عوض عن $X = 40$ في المعادلة المعطاه

$y' = -58.767 + 11.535(40) = 402.63$

When the size of the home increases one unit (100 square feet), the expected change in home price is

A) 11.535 B) -58.767 C) 0 D) 58.767

عندما تزداد X بمقدار one unit

فإن y تزداد بمقدار معامل X

اي تزداد y بمقدار 11.535

The size of the home is called

A) Explanatory variable
B) Response variable
C) Outcome variable
D) Confounding variable

Size of home (x) is called:

→ Explanatory variable

or → independent variable.

متغير مستقل

From the graph, the relationship between size and prize is

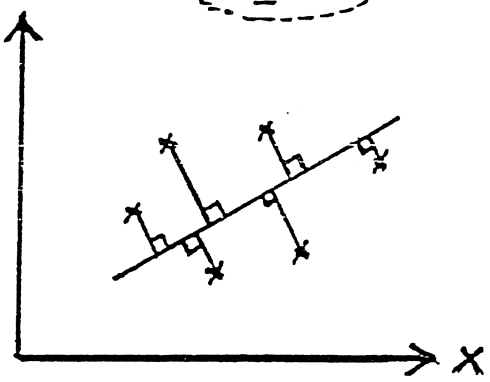
A) Strong negative B) Weak positive C) Weak negative D) Strong positive

The line of best fit means the line that the sum of the .... distances from each point to this line is at a minimum.

A) squares of the vertical B) vertical C) squares of the horizontal D) horizontal

تعريف

مجموع مربعات المسافات العمودية بين النقاط والخط (بُعد النقاط عن الخط) يكون أقل ما يمكن.



A graph of the independent variable, X, and the dependent variable, Y, is called ...

A) scatter plot. B) pie graph. C) histogram. D) frequency polygon.

تعريف

حفظ

In the $y' = a + bx$, what is y'?

A) Slope of the regression line.
B) Intercept of the regression line.
C) Predicted value of y, given a specific x value.
D) Value of y when $x = 0$.

$y' = a + bx$ في معادلة خط الانحدار

eq. of regression line

y' تعني: توقع قيمة y عند قيمة محددة لـ x

$y = a - bx$

* في حالة b السالبة المتغيرين عكس بعضهم
احدهم يزداد الاخر ينقص.

$y = a + bx$

* في حالة b الموجبة المتغيرين مثل بعضهم
سواء بالزيادة أو النقصان.

Complete the following statements with the best answer.

1. The regression line is called the **line of best fit**

2. If all the points fall on a straight line the value of r will be 1 or -1.

3. The strength and direction of the relationship between two variables is determined by the value of the **correlation coefficient**.

4. The equation of the regression line used in statistics is $y' = a + bx$

5. The sign of r and b will always be the same.

6. A statistical graph of two variables is called **scatter plot**.

7. The range of the correlation coefficient is from -1 to 1.

إنتهى Ch.10

كل التمنيات بالنجاح والتوفيق

السعدي

CH. 13 sec. 6

Section 13-6

The spearman rank correlation coefficient

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Where: n is number of data pairs.

: d is difference in ranks.

Note

- The range of r_s = [-1 , 1] or $-1 \le r_s \le 1$
- d = zero if both sets of data have the same ranks $\mathbf{r_s = 1}$.
- If the sets of data are ranked in the opposite way $\mathbf{r_s = -1}$.
- If there is no relationship between the two variables $\mathbf{r_s = 0}$.

Example:

A statistics instructor wishes to see whether is a relationship between the number of homework exercise a student completes and her or his exam score. The data are shown here.

Using the table to find spearman rank correlation coefficient.

Homework problems x	63	55	58	87	89	52	46	75	105
Exam score y	85	71	75	98	93	63	72	89	100

Solution

Arrange X and y From Low to high

H.W. Problems x	Exam score y	Rank x	Rank y	$d = R_x - R_y$	d^2
63	85	5	5	0	0
55	71	3	2	1	1
58	75	4	4	0	0
87	98	7	8	-1	1
89	93	8	7	1	1
52	63	2	1	1	1
46	72	1	3	-2	4
75	89	6	6	0	0
105	100	9	9	0	0
					$\sum d^2 = 8$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(8)}{9(81-1)} = 0.93$$

Strong Positive

Example:

For this data:

X	50	60	24	30	25	35	44	56	37	30
Y	40	37	20	25	19	25	25	42	30	20

Find the spearman rank correlation coefficient.

Solution

Arrange x and y from high to low.....

X	Y	R_x	R_y	$d = R_x - R_y$	d^2
50	40	3	2	1	1
60	37	1	3	-2	4
24	20	10	8.5	1.5	2.25
30	25	7.5	6	1.5	2.25
25	19	9	10	-1	1
35	25	6	6	0	0
44	25	4	6	-2	4
56	42	2	1	1	1
37	30	5	4	1	1
30	20	7.5	8.5	-1	1
					$\sum d^2 = 17.5$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(17.5)}{10(100-1)} = 0.89$$

Strong Positive

Use the following to answer questions

If the differences between the ranks of two variables are (-1, -3, 1, -2, 1, 3, 2, -1) then answer the following three questions:

The sample size is (حجم العينة)

A) 0 B) 30 C) 16 D) 8

Sample size حجم العينة = 8 → عدد القيم

The value of the correlation coefficient is

A) 0.357 B) -0.357 C) -0.643 D) 0.643

d: -1 , -3 , 1 , -2 , 1 , 3 , 2 , -1

d^2: 1 , 9 , 1 , 4 , 1 , 9 , 4 , 1

$\Sigma d^2 = 30$

* Correlation coefficient

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

قانون سبيرمان يدخل في نهاية ورقة الأسئلة

$$= 1 - \frac{6(30)}{8(64-1)} \simeq \boxed{0.643}$$

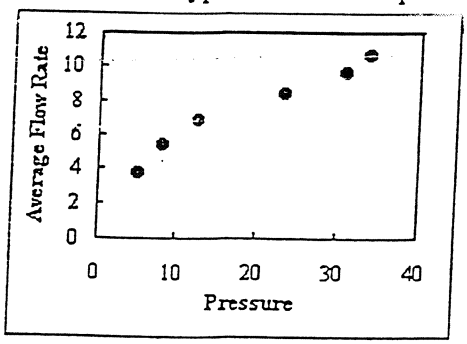
The correlation coefficient value means that there is (متوسط)

A) very weak linear relationship
B) very strong linear relationship
C) moderate linear relationship
D) no linear relationship

$r \simeq 0.64$ is moderate.

- The regression line can be used to predict a value for the dependent variable (y) for a given value of the independent variable (X).

Determine the type of relationship shown in the figure below.



A) there is no relationship B) positive C) negative D) curvilinear

●● واضح أنه العلاقة طردية (موجبة)

حيث أنه

* كلما زاد Pressure زاد Average Flow Rate

If the Spearman rank correlation coefficient (r_s) equals 0.6, then the relationship can be described as

A) positive, strong and non linear
B) weak and linear
C) positive, moderate and linear
D) moderate and non linear

Use the following to answer questions

X	1	2	3	4
Y	2	5	1	3

The sample size is ...
A) 4 B) 8 C) 0 D) 18

حجم العينه
sample size = 4 (عدد الخانات بالجدول)

The Spearman rank correlation coefficient (r_s) is ...
A) 1 B) 0 C) -1 D) -0.98

$n = 4$

X	y	R_x	R_y	d	d^2
1	2	4	3	1	1
2	5	3	1	2	4
3	1	2	4	-2	4
4	3	1	2	-1	1
					$\Sigma d^2 = 10$

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6(10)}{4(16-1)} = 0$$

The value of the Spearman rank correlation coefficient (r_s) means that there is ....
A) strong negative linear relationship
B) moderate linear relationship
C) strong positive linear relationship
D) no linear relationship

$\therefore r_s = \text{zero}$ $\therefore$ no linear relationship.
لا توجد علاقة بين المتغيرين x, y

كل التمنيات بالنجاح والتوفيق