PRINCIPLES OF
PROBABILITY & STATISTICS

CHAPTER 1

[redacted] Descriptive Statistics 1

- ❖ Basic Concepts and definitions
- ❖ Organization and Graphical Representation of the Data
- ❖ Measures of Central Tendency
- ❖ Measures of Dispersion
- ❖ Exercises

INTRODUCTION

Data:

A collection of information collected by means of experiments or real life events and stored in proper format degree

مجموعة المعلومات التي جمعها من التجارب أو الأحداث الحياتية وضعها في تصميم مناسب تسمى**(بيانات)**

سوف يشمل هذا الكورس نوعان من الإحصاء:

إحصاء وصفي Descriptive stat

Those statistical methods which are used for presenting and summarizing data

وفيه تستخدم الطرق الإحصائية لعرض وتلخيص البيانات في صورة جداول أو أشكال بيانية وتشمل دراسة المتوسطات والمائنيات والتشتت

Example

a) Average height, average weight

b) Human population date (gender, proportion of population)

إحصاء إستدلالي Inferential stat

Those statiscal methods wich are used for making conclusion a bout population

و هو ذلك الفرع من فروع الإحصاء الذي يهتم بعمل استدلال إحصائي للمجتمع عن طريق عينة مأخوذة من ذلك المجتمع ويشمل

- Point estimate
- interval estimat
- hypothesis testing

Example

لدراسة دخل المواطنين في بلد ما بهذه الطريقة يتم سحب عينة عشوائية ومن خلالها يتم تعميم نتائج لعينة على المجتمع

Population (المجتمع):

It's is the set of all out comes such as individuals, objects, things.

هو مجموعة كل المدخلات من أشخاص وعناصر وأشياء

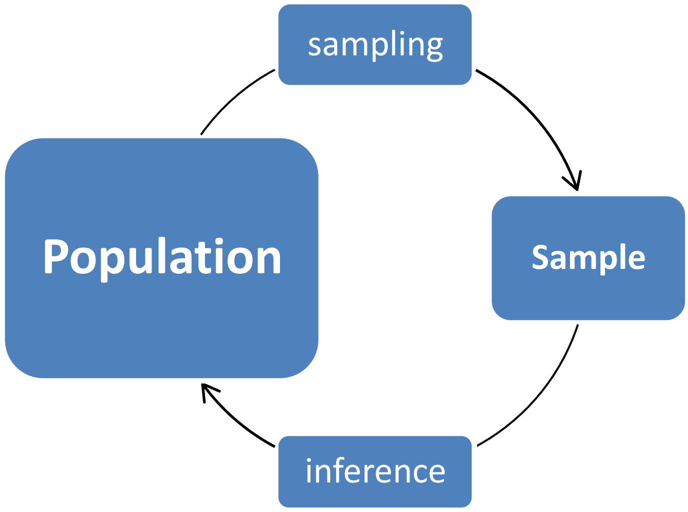
Example

- The set of all KSU students
- The set of income of all citizens of a country

Sample (العينة):

Is a subset of population

ويتم سحب العينات عن طريق العينة العشوائية البسيطة أو الطبقية أو العنقودية أو المنتظمة وتكون العلاقة بين العينة والمجتمع كما بالشكل التالي:



Parameter (المعلمة):

It's a numerical characteristics of population هي خاصية مميزة للمجتمع

Example حساب الوسط الحسابي للمجتمع – حساب الانحراف المعياري للمجتمع

Statistic (الإحصاءة):

It is a function of sample or it is numerical characteristic of the sample

هي خاصية مميزة للعينة

Example حساب الوسط الحسابي للعينة – حساب الانحراف المعياري للعينة

Variable (المتغير):

Is a characteristic, feature or factor that varies from one individual to another. هو خاصية أو سمة أو عامل يتغير من شخص إلى آخر داخل المجتمع

Example

- for countries [gross domestic product, sex ratio, birth rates]
  الناتج القومي الكلي، نسبة الجنس – معدل المواليد
- for humans [height, weight, sex, marital status, ...]
  الارتفاع والوزن والجنس والحالة الاجتماعية

Types of variable

Qualitative variable: these variable can only take values which are non-numerical (ليست أرقام) بيانات وصفية

Example

marital status, gender

Can not be ordered

No mathematical operation

Nominal scale مقياس

Nominal
إسمية لا تقبل الترتيب

- Sex
- Color
- Marital status

Ordinal
ترتيبية تقبل الترتيب

- Grade
- Cientific degree

Quantitave variable: these are variable which take numerical values بيانات كمية (أرقام)

Example

height, temperature

Can be ordered in increasing or decreasing

Can under go mathematical operations

Discrete
لا تقبل كسور

- Has gaps
- Whole number

Example

- Number of accidents
- Number of laptops
- Number of goals
- Number of children

[can assume only finite number of variables]

Continuous
تقبل كسور

- No gaps
- Interval

Example

- Age
- Weight
- Distance
- Temperature
- Sex ratio

[there is not finite number of values]

ORGANIZATION AND GRAPHICAL REPRESENTATION OF THE DATA

في هذا الباب سوف نستعرض إنشاء جداول التوزيع التكراري للبيانات بأنواعها وطريقة عرضها بيانيا

Example consider the blood groups of 40 persons below

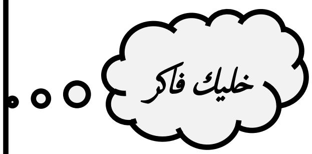
O, O, A, B, A, O, A, A, A, O, B, OB, O, O, A, O, O, A, A, A, A, AB, A, B, A, A, O, O, A, O, O, A, A, A, O, A, O, O, AB

- Construct the frequency distribution for the a bove data

Solution

Blood group	f	R.f	p.f	C.f	C.R.f
O	16	0.4	40%	16	0.4
A	18	0.45	45%	34	0.85
B	4	0.1	10%	38	0.95
AB	2	0.05	5%	40	1
Total	40	1	100%		

- The sum of all relative frequency equal one
- The sum of all percent frequency equal 100
- Sample size, n القيمة الأخيرة في التكرار التراكمي هي
- القيمة الأخيرة في التكرار النسبي التراكمي دائما واحد
- Relative frequency (R.f) = $\frac{\text{frequency}}{\text{total}}$
- Percent frequency =$\text{R.f} \times (100)$



Example Consider the sample of 40 students as follows

0, 1, 2, 3, 1, 2, 2, 1, 1, 2, 0, 2, 1, 0, 1, 0, 1, 1, 2, 1, 2, 1, 3, 1, 2, 1, 1, 0, 0, 2, 1, 1, 0, 1, 2, 2, 2, 1, 0, 1

Construct the frequency table

Solution

No of subjects	f	R.f	p.f	C.f	C.R.f
0	8	0.2	20%	8	0.2
1	18	0.45	45%	26	0.65
2	12	0.3	30%	38	0.95
3	2	0.05	5%	40	1
Total	40	1	100%		

Example The following data is the price of accessories for girls

4, 1, 7, 9, 12, 16, 17, 7, 12, 19, 22, 24, 3, 2, 8, 6, 13, 24, 14, 11, 18, 16, 23, 20, 1, 2, 6, 25, 15, 7, 11, 12, 16, 17, 21, 22, 15, 17, 14, 5, 7, 8, 12, 13, 20, 23, 13, 19, 18, 12

Construct the frequency distribution table for the data by using 5 classes

Solution

البيانات السابقة هي بيانات متصلة "تقبل كسور"

لعمل جدول التوزيع التكراري في حالة البيانات المتصلة لابد من :

1. حساب المدى"Range" وهو عبارة عن أكبر قيمة ناقص أصغر قيمة

$$R = max - min = 25 - 1 = 24$$

2. إيجاد عدد الفئات "K" وغالبا ما يعطى في السؤال وهو يتراوح بين 5 -20 فئة ولكن العدد الأمثل يتراوح بين 4 -10 فئات

3. نحسب قيمة "C" حيث $C = \frac{R}{K} = \frac{24}{5} = 4.8 \approx 5$ وبالتالي فإن طول الفئة هو

$$\text{Class width} = \text{C} - 1 = 5 - 1 = 4$$

4. لحساب حدود الفئات نطرح 0.5 من الحد الدنى لكل فئة ونضيف 0.5 إلى الحد الأعلى لكل فئة

5. لحساب منتصف الفترة $\text{midpoint} = \frac{\text{upper+lower}}{2}$

6. لحساب التكرار المتجمع الصاعد ascending cumulative frequency (ACF) وتكون القيمة الأولى هي نفس قيمة التكرار

القيمة الثانية مجموع التكرار الأول +الثاني

القيمة الثالثة مجموع (الأولى + الثانية + الثالثة) ... وهكذا

- إذا كانت البيانات الخام raw data أقل من 32 في هذه الحالة ليست هناك حاجة إلى جدولة البيانات في جدول تكراري
- إذا كانت عدد الفئات غير معلوم فإنه يمكن إيجادها من العلاقة $K = [3.32 \log n]$

Where **n** is the simple size

[x] is the greatest integer less than or equal **x**

Example [3.2]=3 & [3.9]=3

Class limit	Class Boundories	Midpoint	F	R.F	LESS Than	ACF
1-5	$0.5 \to 5.5$	3	7	0.14	5.5	7
6-10	$5.5 \to 10.5$	8	9	0.18	10.5	16
11-15	$10.5 \to 15.5$	13	14	0.28	15.5	30
16-20	$15.5 \to 20.5$	18	12	0.24	20.5	42
21-25	$20.5 \to 25.5$	23	8	0.16	25.5	50
Total			**50**	**1**		

Example Consider The mileage of 40 cars per liter

12, 16, 15, 15, 12, 19,17, 18, 16, 14, 13, 12, 20, 12, 15, 16, 20, 16, 15, 12, 18, 16, 17, 19, 15, 16, 17, 15, 16, 15, 14, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20

Construct the frequency distribution table

Solution

$$Rang = max - min = 20 - 12 = 8$$

Number of classes $K = [3.32 \ \log n] = 3.32 \ log 40 \ ] = [5.3] = 5$

$$C = \frac{R}{K} = \frac{8}{5} = 1.6 \approx 2$$

Therefore length of class = C-1=2-1=1

Class limit	Class Boundaries	Midpoint	F	R.F	ACF	ACRF	ACPF
12-13	11.5 → 13.5	12.5	8	0.20	8	0.20	20%
14-15	13.5 → 15.5	14.5	10	0.25	18	0.45	45%
16-17	15.5 → 17.5	16.5	12	0.30	30	0.75	75%
18-19	17.5 → 19.5	18.5	6	0.15	36	0.90	90%
20-21	19.5 → 21.5	20.5	4	0.10	40	1	100%
Total			**40**	**1**			

GRAPHICAL REPRESENTATION

Pie-charts (القطاع الدائري):

Is a simple way of representing the proportion of each class of data on a circular disk.

Example For the frequency distribution table below

Blood group	F	R.F	P.F
O	16	0.40	**40%**
A	18	0.45	**45%**
B	4	0.10	**10%**
AB	2	0.05	**5%**
Total	40	1	**100%**

Construct the pie-chart

لرسم القطاع الدائري لابد من معرفة الزاوية المركزية لكل قطاع

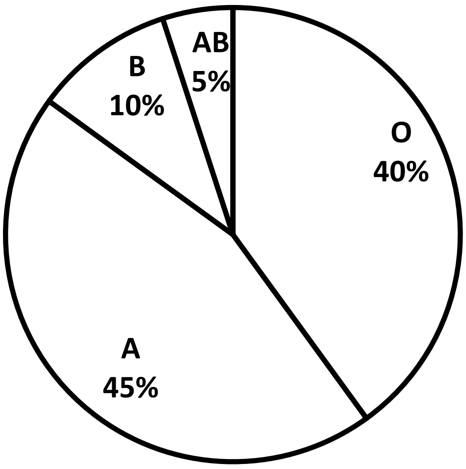
$$measure\ central\ angle = R.F\ \times 100$$

$$measure\ central\ angle\ of\ sector\ O = 0.40\ \times 360 = 144°$$

$$measure\ central\ angle\ of\ sector\ A = 0.45\ \times 360 = 162°$$

$$measure\ central\ angle\ of\ sector\ B = 0.10\ \times 360 = 36°$$

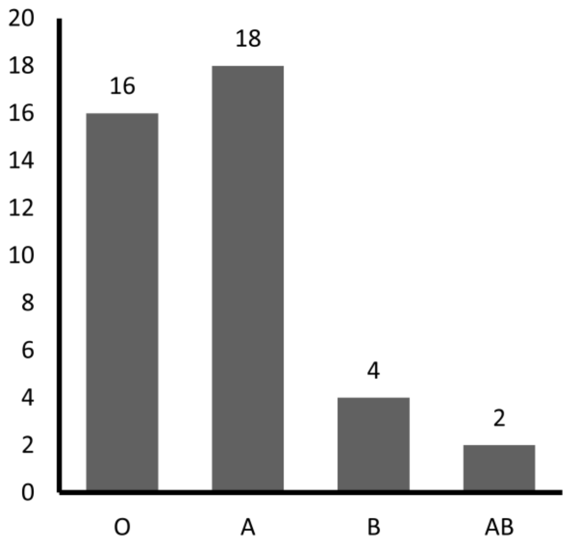
$$measure\ central\ angle\ of\ sector\ AB = 0.05\ \times 360 = 18°$$



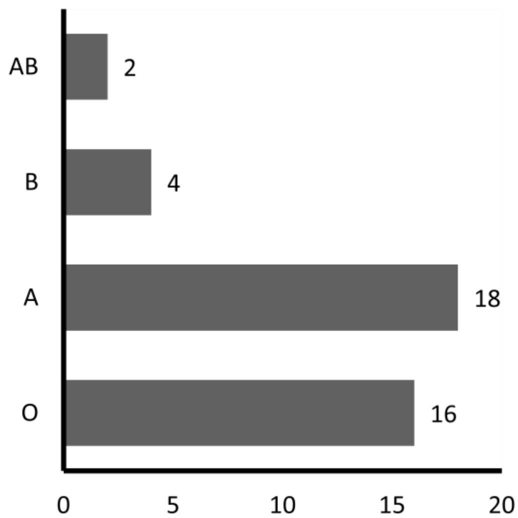
Bar charts (الأعمدة البسيطة):

The frequency of each class is represented by a bar the height of the bar corresponds to the frequency

وهناك نوعان من الأعمدة البسيطة:



vertical



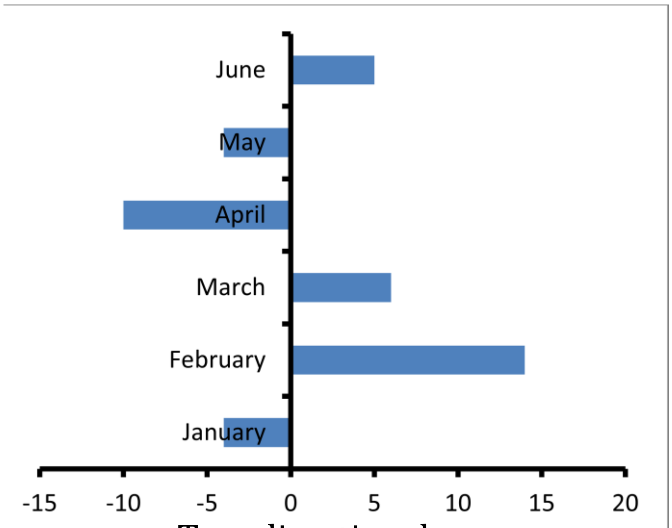
horizonal

Two directional Bar charts (الأعمدة التكرارية ذات الاتجاهين)

وتستخدم في الحالات التي تعبر عن قيم موجبة وسالبة للفئات المختلفة

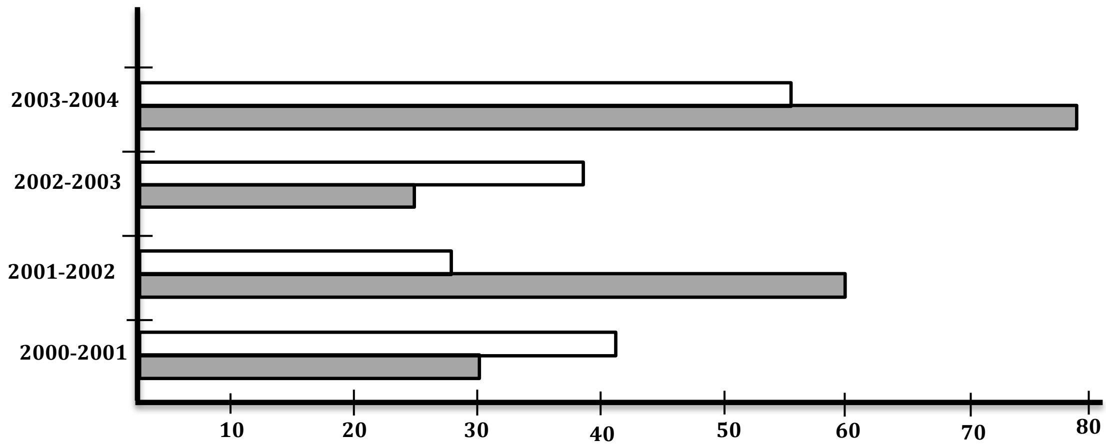
Example Consider the following changes in income of accompany

Month	Change in income
January	-4%
February	14%
March	6%
April	-10%
May	-4%
June	5%

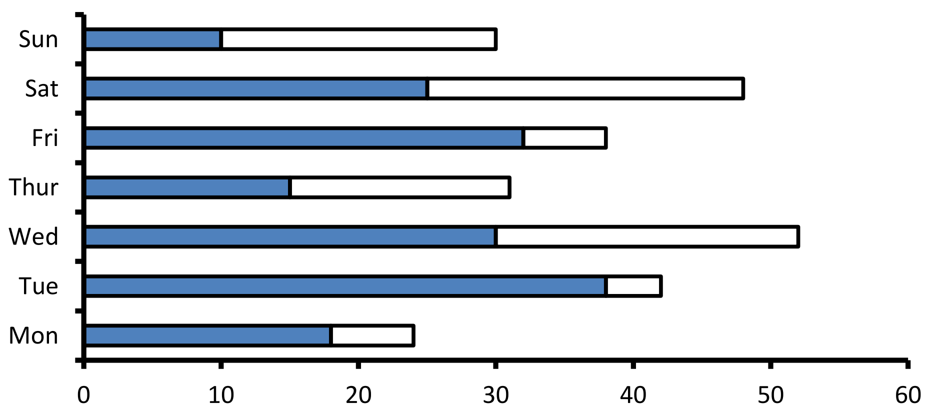


Two directional

Multiple Bar Chart



Stacked Bar Charts

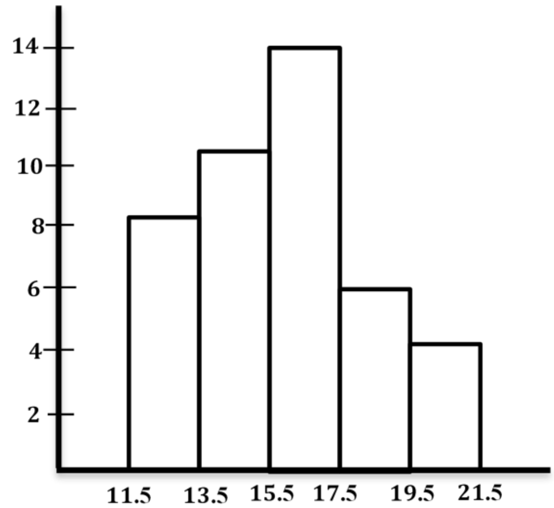


Histogram المدرج التكراري

- It is bar chart but have basic difference
- In histogram classes are adjacent
- Histograms are used to represent the quantitative continuous data
- Every class interval is called bin (خانة)

Example Consider the data

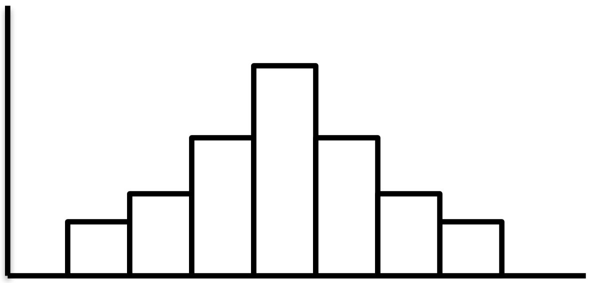
Classe boundaries	Frequency
$11.5 \to 13.5$	**8**
$13.5 \to 15.5$	**10**
$15.5 \to 17.5$	**12**
$17.5 \to 19.5$	**6**
$19.5 \to 21.5$	**4**
Total	**40**



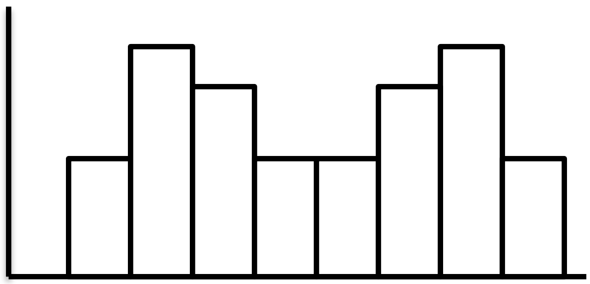
TYPES OF HISTOGRAMS

❶ *Symmetric histogram (المدرجات المتماثلة)*

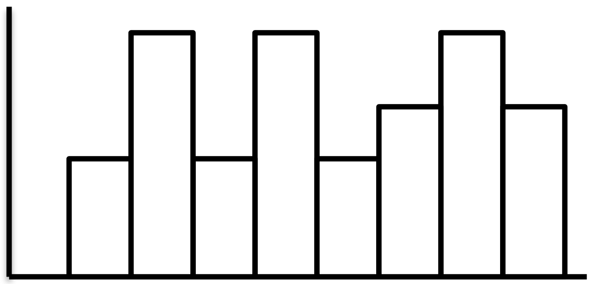
Unimodal:(one peak) قمة واحدة



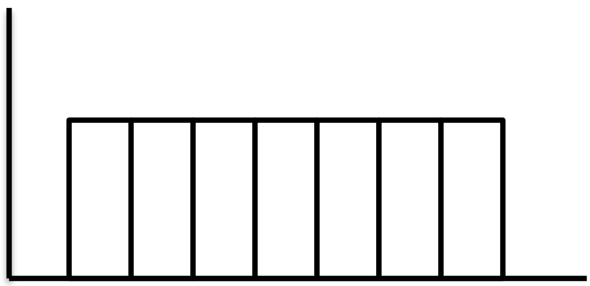
Bimodel:(two peaks) قمتان



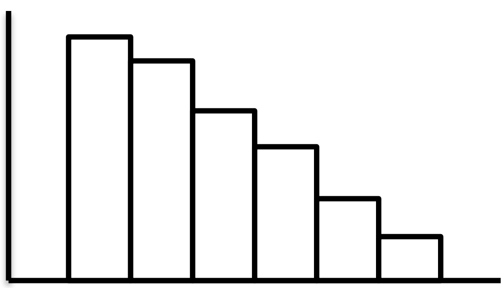
Multimodel (more than two) أكثر من قمة



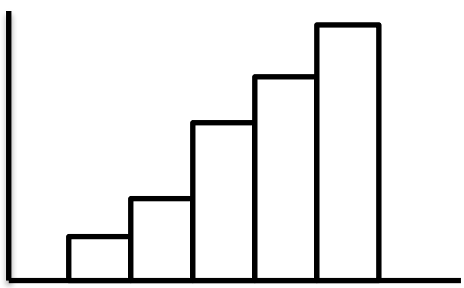
Uniform (no peaks) منتظم



❷ ***Symmetric skewed Histogram*** (مدرجات ملتوية متماثلة)



Right skewed



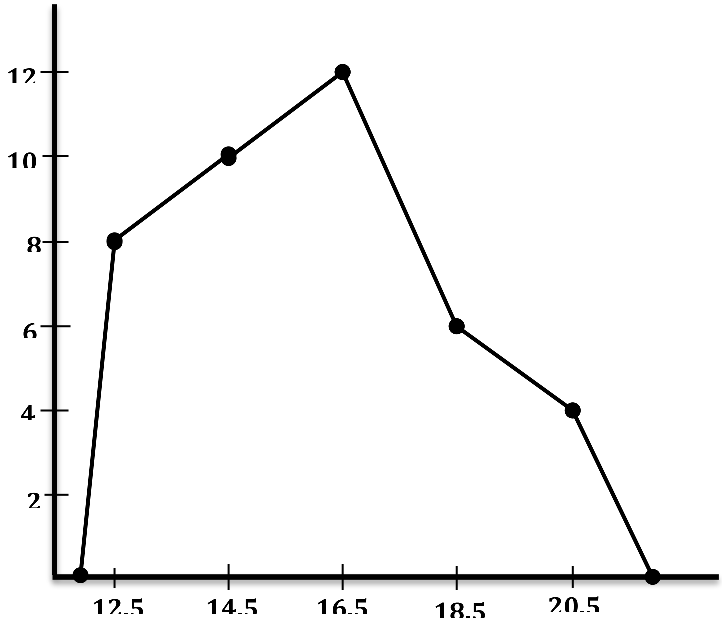
left skewed

POLYGON (المضلع التكراري)

- Depend on two points $(x_i\ ,\ f_i)$ Where x_i is midpoint and f_i is frequency
- It is useful to compare two or more variables

Example Consider the frequency table below

Class Boundaries	Frequency	Midpoint
$11.5 \rightarrow 13.5$	8	**12.5**
$13.5 \rightarrow 15.5$	10	**14.5**
$15.5 \rightarrow 17.5$	12	**16.5**
$17.5 \rightarrow 19.5$	6	**18.5**
$19.5 \rightarrow 21.5$	4	**20.5**
Total	**40**	

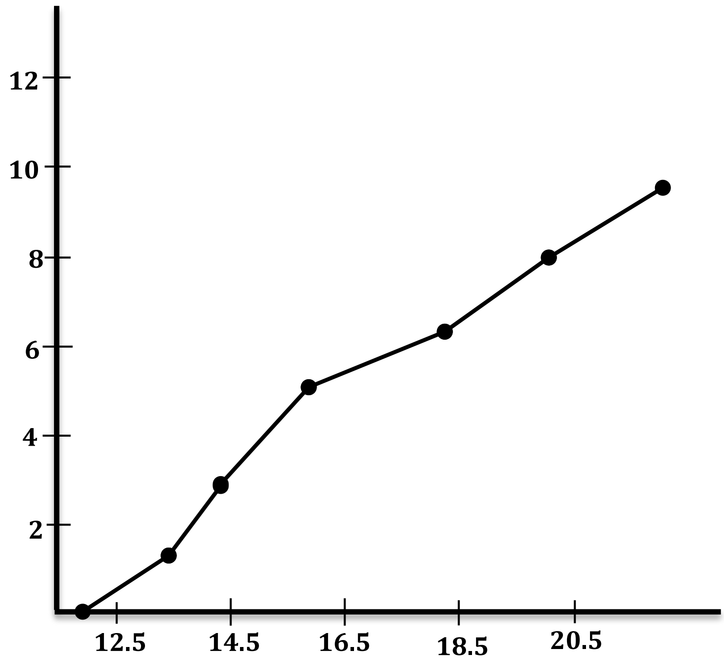


polygon

Ogive (المنحنى المتجمع الصاعد)

Depend on two point $(b_i\ , f_i)$,where b_i is the upper bound and f_i is cumulative frequency

Less than ogive



The measure of central tendency

- mean
- median
- mode
- percentiles
- deciles
- quartiles

1 mean $\bar{X}$ (الوسط الحسابي)

mean is the sum of values divided by the number of values, so $\bar{X} = \frac{\sum x}{n}$

Example

Calculate the mean for the data **20, 18, 15, 15, 14, 12, 11, 9, 7, 6, 4, 1**

Solution

$$\bar{X} = \frac{\sum x}{n} = \frac{20+18+15+15+14+12+11+9+7+6+4+1}{12}$$

$$= \frac{132}{12} = 11$$

الحاسبة: يمكن حساب الوسط الحسابي للبيانات السابقة بالخطوات التالية

لإدخال البيانات $\boldsymbol{mod} \rightarrow \mathbf{3} \rightarrow \mathbf{1}$ (إدخال البيانات) $\rightarrow \boldsymbol{AC}$

للنواتج $\boldsymbol{Shift} \rightarrow \mathbf{1} \rightarrow \mathbf{4} \rightarrow \mathbf{2}(\bar{x})$

البيانات المجدولة يمكن إيجاد الوسط الحسابي لها من القانون

$$\bar{X} = \frac{\sum Fx}{\sum F}$$

Example find the mean for the data below

No of student	0	1	2	3	Total
Frequency	8	18	12	2	40

Solution

No of student	F	Fx
0	8	**0**
1	18	**18**
2	12	**24**
3	2	**6**
Total	$\sum F = 40$	$\sum Fx = 48$

$$\bar{X} = \frac{\sum Fx}{\sum F} = \frac{48}{40} = 1.2$$

إذا كانت البيانات متصلة يمكن إيجاد الوسط الحسابي لها بالقانون التالي :

$$\bar{X} = \frac{\sum FM}{\sum F}$$

حيث M تمثل منتصف الفترة midpoint

Example calculate the mean for the mileage of the cars

Class boundaries	$11.5 \to 13.5$	$13.5 \to 15.5$	$15.5 \to 17.5$	$17.5 \to 19.5$	$19.5 \to 21.5$	Total
Frequency	**8**	**10**	**12**	**6**	**4**	**40**

Solution

Class boundary	Midpoint M	Frequency F	F.M
$11.5 \to 13.5$	12.5	8	**100**
$13.5 \to 15.5$	14.5	10	**145**
$15.5 \to 17.5$	16.5	12	**198**
$17.5 \to 19.5$	18.5	6	**111**
$19.5 \to 21.5$	20.5	4	**82**
Total		40	**636**

$$\bar{X} = \frac{\sum FM}{\sum F} = \frac{636}{40} = 15.9$$

الحاسبة: يمكن إيجاد الوسط الحسابي للبيانات المتصلة كالتالي:

تفعيل الـ Frequency بالحاسبة

$$shift \rightarrow mod \rightarrow 4(stat) \rightarrow 1(on)$$

إدخال القيم في الحاسبة وتكون قيمة x هي الـ midpoint

$$mod \rightarrow 3 \rightarrow 1\ (\text{القيم}) \rightarrow AC$$

للنواتج $shift \rightarrow 1 \rightarrow 4 \rightarrow 2(\bar{x})$

Advantages

- quick and easy
- all values are considered
- unique

Disadvantages

- can not found for qualitattive
- highly of fected by extreme value
- not applicable if a data is lost

Weighted mean $\bar{X}_w$ (المتوسط المرجح)

$$\bar{X}_w = \frac{\sum Wx}{\sum W}$$

Example a person wants to decide which car better, where look 20%, mileage 30%, engine 50%

Car A: 7 for look, 6 for mileage and 8 for engine

Car B: 6 for look, 4 for mileage and 9 for engine

Solution

$$\bar{X}_A = \frac{(0.2 \times 7) + (0.3 \times 6) + (0.5 \times 8)}{0.2 + 0.3 + 0.5} = 7.2$$

$$\bar{X}_B = \frac{(0.2 \times 6) + (0.3 \times 4) + (0.5 \times 9)}{0.2 + 0.3 + 0.5} = 6.9$$

Car A is better

2 Median (الوسيط)

Is that value which divides the data in half after ordering them

هو القيمة التي تتوسط البيانات بعد ترتيبها تصاعديا أو تنازليا

Example calculate the median

First sample: 28, 22, 26, 29, 21, 23, 24

Second sample: 28, 22, 26, 29, 21, 23, 24, 35

Solution

لإيجاد الوسيط: 1- ترتيب البيانات

2- إيجاد القيمة التي تقسم البيانات إلى 50% قبلها و50% بعدها

First sample (odd) we ordered data as ~~21~~, ~~22~~, ~~23~~, 24, ~~26~~, ~~28~~, ~~29~~

median

Second sample (even) ordered data as ~~21~~, ~~22~~, ~~23~~, 24, 26, ~~28~~, ~~29~~, ~~35~~

$$\boldsymbol{median = \frac{24+26}{2} = 25}$$

median for frequency table ⟶ discrete

- حساب التكرار التراكمي
- نوجد أصغر تكرار تراكمي أكبر من $\frac{n}{2}$
- القيمة التي تناظر هذا التكرار تكون هي الوسيط

Example consider the number of student in the following

No. student	Frequency	ACF
0	8	**8**
1	18	**8+18 = 26**
2	12	**26+12 = 38**
3	2	**38+2 =40**
Total	40	

Solution

$n = 40$, so $\frac{n}{2} = \frac{40}{2} = 20$

Then, the smallest cumulative frequency greater than 20 is 26

The corresponding value to 26 is 1

Hence the median is 1

median for frequency distribution → continuous

أولا نوجد الفئة الوسيطية وهي الفئة التي تكرارها التراكمي أكبر من $\frac{\sum F}{2}$ من خلال المعادلة

$$\tilde{X} = L + \frac{\frac{\sum f}{2} - (F - f)}{f} \times C$$

حيث

الحد الأدنى للفئة الوسيطية L

التكرار التراكمي للفئة الوسيطية F

تكرار الفئة الوسيطية f

طول الفئة C

Note that:

The median class is the first class whose cumulative frequency is greater than or equal $\frac{\sum f}{2}$

Example

Class boundary	Midpoint	Frequency	Relative frequency	ACF
11.5 → 13.5	12.5	8	0.2	8
13.5 → 15.5	14.5	10	0.25	18
15.5 → 17.5	16.5	12	0.30	30
17.5 → 19.5	18.5	6	0.15	36
19.5 → 21.5	20.5	4	0.10	40
Total		**40**	**1**	

Solution

$$\frac{\sum f}{2} = \frac{40}{2} = 20 \Longrightarrow median\ class\ is\ 15.5 \rightarrow 17.5$$

$$L = 15.5 \quad \& \quad F = 30 \quad \& \quad f = 12 \quad \& \quad C = 2$$

$$\tilde{X} = L + \frac{\frac{\sum f}{2} - (F - f)}{f} \times C = 15.5 + \frac{20 - (30 - 12)}{12} \times 2 = 15.83$$

3 Mode (المنوال)

The value which has occurred maximum number

خليك فاكر

ꝏ يمكن إيجاد المنوال لجميع أنواع البيانات لذلك هو المقياس الأوسع انتشارا

ꝏ إذا تكررت كل البيانات بنفس المقدار فلا يوجد منوال

ꝏ يمكن للبيانات أن تحتوى أكثر من منوال

21	**22**	**24**	**23**	**29**	**No mode**

21	**28**	**24**	**28**	**28**	**Mode is 28**

Blood group	**O**	**A**	**B**	**AB**	**Mode is O&B**
Frequency	16	8	16	4	

21	**22**	**21**	**22**	**23**	**23**	**No mode**

mode for frequency distribution table → continuous

$$\hat{X} = L + \frac{d_1}{d_1+d_2} \times C$$

L الحد الأدنى للفئة المنوالية

f_m تكرار الفئة المنوالية

f_p التكرار السابق & f_n التكرار اللاحق

$d_1 = f_m - f_p$ & $d_2 = f_m - f_n$

Example

Seconds20	Frequency
$50.5 \to 55.5$	**2**
$55.5 \to 60.5$	**$7 \to f_p$**
$60.5 \to 65.5$	**$8 \to f_m$**
$65.5 \to 70.5$	**$4 \to f_n$**
Total	**21**

Solution

$L = 60.5$ $d_1 = 8 - 7 = 1$ $d_2 = 8 - 4 = 4$ $C = 5$

$$\hat{X} = 60.5 + \left(\frac{1}{1+4}\right) \times 5 = 61.5$$

العلاقة بين الوسط والوسيط والمنوال

⌗ في حالة البيانات المتماثلة يمكن استخدام المقاييس الثلاثة وتقع في منتصف البيانات في هذه الحالة.

i.e (for symmetric distribution mean=mode=median)

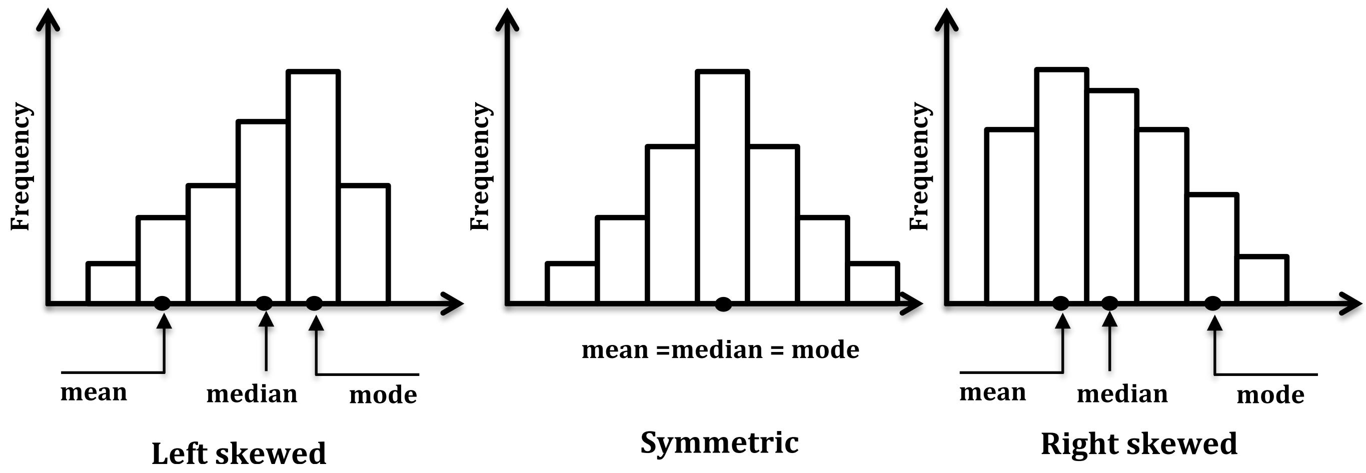
⌗ في حالة البيانات الملتوية يسار يكون المنوال اكبر الوسيط اكبر من الوسط لان وجود القيمة الشاذة (extreme value) يزيح الوسط ناحية اليسار وبالتالي يكون الوسط هو القيمة الأصغر في حالة البيانات ملتوية لليسار.

⌗ في حالة البيانات الملتوية يمين يكون الوسط هو القيمة الأكبر ثم الوسيط وأخيرا المنوال وهنا وجود القيم المتطرفة يزيح الوسط ناحية اليمين وبالتالي يكون الوسط هو القيمة الأكبر هنا

خليك فاكر

- الوسط الحسابي هو أدق مقاييس النزعة المركزية
- الوسط الحسابي يتأثر بكل العمليات التي تجرى عليه "جمع – طرح –ضرب – قسمة"
- الوسط الحسابي يمكن أن يكون قيمة سالبة
- أكثر المقاييس تأثرا بالقيم الشاذة هو الوسط الحسابي

The relationship among the three measures of central tendency



4 Percentiles (المئينيات)

The percentiles of a variable divide the value into 100 equal parts and we denote by P_1, P_2 and P_{99}

خليك فاكر

- P_{50} is the median which divides the data into two equal part
- P_1 divides the data into 1% from bottom and 99% from top
- $p_r = \frac{r(n+1)}{100}$ where r=1, 2, ... 99 n is the sample size
- $P_r = X_K + S(X_{K+1} - X_K)$ K is the integer , S the reset of p_r

Example Calculate the 35th percentile of the data

40, 51, 92, 10, 36, 60, 70, 36, 36, 40, 80, 39, 53, 56, 60, 60, 70, 72, 88, 92, 50, 92, 20, 70, 38, 95, 56, 60, 88, 70

Solution

نقوم بترتيب البيانات تصاعديا

10, 20, 36, 36, 36, 38, 39, 40, 40, 50, 51, 53, 56, 56, 60, 60, 60, 60, 70, 70, 70, 70, 72, 80, 88, 88, 92, 92, 92, 95

$$p_r = \frac{r(n+1)}{100} \quad \text{where} \quad r = 35 \quad n = 30$$

$$P_{35} = \frac{35(30+1)}{100} = 10.85$$

أي أن القيمة التي تقسم البيانات إلى 35% قبلها و65% بعدها تقع بين القيمة العاشرة والحادية عشر ويمكن ايجاد قيمتها

$$P_r = X_K + S(X_{K+1} - X_K) \quad \text{where } K = 10 \;\; S = 0.85$$

$$P_{35} = X_{10} + 0.85(X_{11} - X_{10})$$

$$= 50 + 0.85(51 - 50) = 50.85$$

5 Deciles (العشيرات)

The deciles divide the data into 10 equal parts

خليك فاكر

- D_1 is the 10th percentile, D_2 is the 20th percentiles
- $dr = \frac{r(n+1)}{10}$ where r=1,2,…9
- $D_r = X_K + S(X_{K+1} - X_K)$ $\quad$ *where K is the integer S is the rest*

Example For the above example calculate the sixth decile

Solution

$d_r = \frac{r(n+1)}{10} \qquad r = 6, \qquad n = 30$

$$d_6 = \frac{6(30+1)}{10} = 18.6$$

أن موقع القيمة التي تقسم البيانات إلى 0.6 قبلها و0.4 بعدها تقع بين القيمة الثامنة عشر والتاسعة عشر في البيانات ويمكين إيجاد قيمتها تحديدا من خلال القانون

$$Dr = X_K + S(X_{K+1} - X_K)$$

$$D_6 = 60 + 0.6(70 - 60) = 66$$

6 Quartiles (الربيعات)

The quartiles divide the data into 4 equal parts

خليك فاكر

- Q_1 is 25^{th} percentile, Q_2 is the median and Q_3 is the 75^{th} percentile
- $q_r = \frac{r(n+1)}{4}$ where $r = 1, 2, 3$
- $Q_r = X_K + S(X_{K+1} - K_K)$

Example For the above example find the first quartile

Solution

$$q_r = \frac{r(n+1)}{4} \quad where \quad r = 1 \ , \ n = 30$$

$$q_1 = \frac{1\times(30+1)}{4} = 7.75$$

$$K = 7 \qquad S = 0.75$$

$$Q_1 = X_K + S(X_{K+1} - X_K)$$

$$= 39 + 0.75 \times (40 - 39) = 39.75$$

Example Find the quartiles of the data **28, 22, 26, 29, 21, 23, 24**

Solution

$$r = 1 \qquad n = 7$$

$$q_1 = \frac{1\times(7+1)}{4} = 2 \Rightarrow Q_1 = 22 + 0 \times (23 - 22) = 22$$

$$q_2 = \frac{3\times(7+1)}{4} = 4 \Rightarrow Q_2 = 24 \qquad which\ is\ median$$

$$q_3 = \frac{3\times(7+1)}{4} = 6 \Rightarrow Q_3 = 28$$

Example Find the quartiles of the data **7, 8, 15, 36, 39, 40, 41, 56**

Solution

$q_1 = 2.25 \implies Q_1 = 8 + 0.25 \times (15 - 8) = 9.75$

$q_2 = 4.5 \implies Q_2 = 36 + 0.5 \times (39 - 36) = 37.5$

$q_3 = 6.75 \implies Q_3 = 40 + 0.75 \times (41 - 40) = 40.75$

Extreme values (القيم المتطرفة)

The value of x is said to be extreme if

$$X < Q_1 - 1.5(Q_3 - Q_1)$$

or

$$X > Q_3 + 1.5(Q_3 - Q_1)$$

For the preceding example we have

- $Q_1 - 1.5(Q_3 - Q_1) = 9.75 - 1.5(40.75 - 9.75) = -36.75$
- $Q_3 + 1.5(Q_3 - Q_1) = 40.75 + 1.5(40.75 - 9.75) = 87.25$

وبالتالي فإن أي قيمة أقل من 36.75- أو أكبر من 87.25 تعتبر قيمة شاذة ولذلك لا توجد قيم شاذة في المثال أعلاه

خليك فاكر

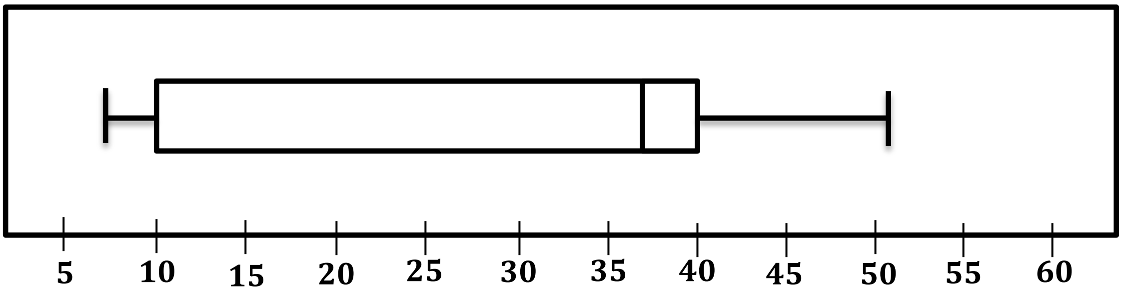
Five Numbers are:

Smallest value , Q_1, Q_2, Q_3 , largest value for the above example, the five numbers are 7, 9.75, 37.5, 40.75, 56

Example Draw the box plot for the data

7, 9.75, 37.5, 40.75 and 56

Solution



خليك فاكر

- إذا لم تحتوي البيانات قيمة متطرفة كبيرة نرسم الذراع على أكبر عدد بالبيانات
- إذا وجدت قيمة متطرفة نرسم الذراع على الحد الأعلى ناحية اليمين للسور وكذلك ناحية اليسار.

Measure of dispersion

- variance
- standard deviation
- Range
- Coefficient of variation

1 Variance (التباين)

وهو يعبر عن مدى تباعد أو تقارب القيمة الكبرى والصغرى عن مركز البيانات $\bar{X}$ ويرمز له بالرمز S^2

$$S^2 = \frac{\sum (X - \bar{X})^2}{n-1}$$

or

$$S^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}$$

خليك فاكر

- The variance always more than or equal zero
- أي أن التباين وكذلك قيم التشتت لا يمكن أن تكون قيم سالبة
- The variance unit is square unit
- وحدة التباين دائما وحدة مربعة ومعامل الاختلاف ليس له وحدة أما باقي المقاييس لها نفس الوحدة محل الدراسة
- التباين والانحراف المعياري لا يتأثر بإضافة أو طرح أي قيمة على البيانات يتأثر فقط بالضرب أو القسمة
- الوسط الحسابي يتأثر بجميع العمليات (ضرب ، قسمة ، جمع ، طرح)

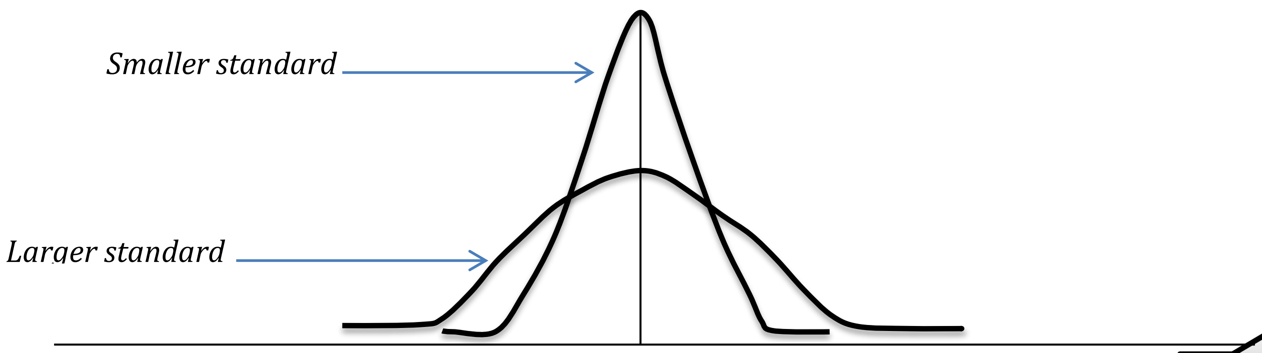
2 Standard deviation (الانحراف المعياري)

The standard deviation is the square roote of the variance and denoted by (S)

خليك فاكر

- يعتبر مقياس الانحراف المعياري هو أفضل مقاييس التشتت في حالة عدم وجود قيمة شاذة و وهو أفضل من التباين لأنه له نفس وحدة محل الدراسة
- الوسط الحسابي هو أفضل مقاييس النزعة المركزية "شرط عدم وجود قيمة شاذة"

➤ A small value of standard deviation indicates that the value of the variable tend to be close to the center where a large value indicates that they tend to be far from the mean (center)



الحاسبة

يمكن إيجاد الانحراف المعياري والتباين كما يلي

$mod \to 3 \to 1 \to$ (إدخال القيمة) $\to AC$

$shift \to 1 \to 4 \to 4(S)$ الانحراف المعياري

وبتربيع القيمة الأخير نحصل على التباين

Example let **2, 3, 6, 8, 10, 13 and 14** calculate the variance and standard deviation

Solution

في هذا المثال سوف نقوم بثلاث طرق:

الأولى

$$S=\sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}=\sqrt{\frac{130}{6}}=4.65$$

$$S^2=21.7$$

X	X^2	$\bar{X}$	$(X-\bar{X})^2$
2	4	8	**36**
3	9	8	**25**
6	36	8	**4**
8	64	8	**0**
10	100	8	**4**
13	169	8	**25**
14	196	8	**36**
$\sum X=56$	$\sum X^2=578$		$\sum(X-\bar{X})^2=130$

الثانية

$$S=\sqrt{\frac{\sum X^2-\frac{(\sum X)^2}{n}}{n-1}}=\sqrt{\frac{578-\frac{56^2}{7}}{6}}=4.65$$

وبتربيع القيمة الأخيرة نحصل $S^2=21.7$

الثالثة

$mod\rightarrow 3\rightarrow 1$

$2=3=6=8=10=13=14=$ AC

$shift\rightarrow 1\rightarrow 4\rightarrow 4(S)=4.65$

Example find the variance and the standard deviation for the data

6, 6, 6, 6, 6, 6, 6

Solution

باستخدام الحاسبة

$mode \to 3 \to 1$

$6 = 6 = 6 = 6 = 6 = 6 = 6 =$ AC

$shift \to 1 \to 4 \to 4(S) = Zero$

وهذه هي الحالة الوحيدة التي يمكن للانحراف المعياري أن يساوي صفر عندما تكون جميع القيم متساوية

3 *Range* (المدى)

- ☑ و هو الفرق بين أكبر قيمة وأصغر قيمة في البيانات
- ☑ أسرع مقياس يعطي فكرة عن مدى تفاوت البيانات
- ☑ لا يفضل استخدامه في حالة وجود قيم شاذة
- ☑ قيمته موجبة دائما حتى لو كانت كل البيانات سالبة
- ☑ له نفس وحدة البيانات محل الدراسة

$$R = max - min$$

Example find the Range for

X: 4, 8, 7, 3, 5, 10, 24, 5

Y: 10, 7, 9, 11, 11, 8, 9, 7

Solution

For X: Range = max –min = 24 – 3 = 21

For Y: Range = 11- 7 = 4

المدى لبيانات Y أصغر من المدى لبيانات X وبالتالي فإنه يمكننا القول أن البيانات في Y أكثر تجانسا أو البيانات في X أكثر تشتتا

For frequency distribution

$R = X_K - X_1$ where

X_K is the midpoint for last class

X_1 is the midpoint for first class

Interquartile Range (المدى الربيعي)

Also called as mid-spread IQR = $Q_3 - Q_1$

4 Coefficient variation (معامل الإختلاف)

$C.V = \frac{S}{\bar{X}} \times 100$

Has no unite

دائما يستخدم لمقارنة تشتت مجموعتين من البيانات ليس لها نفس الوحدة

Z-score (الدرجة المعيارية)

$Z = \frac{X-\bar{X}}{S}$ where mean = 0 and standard deviation = 1

The CHEBYSHEV'S Rule قاعدة تشابيشيف

At least $1 - \frac{1}{K^2}$ of the data lies within K standard deviation of the mean

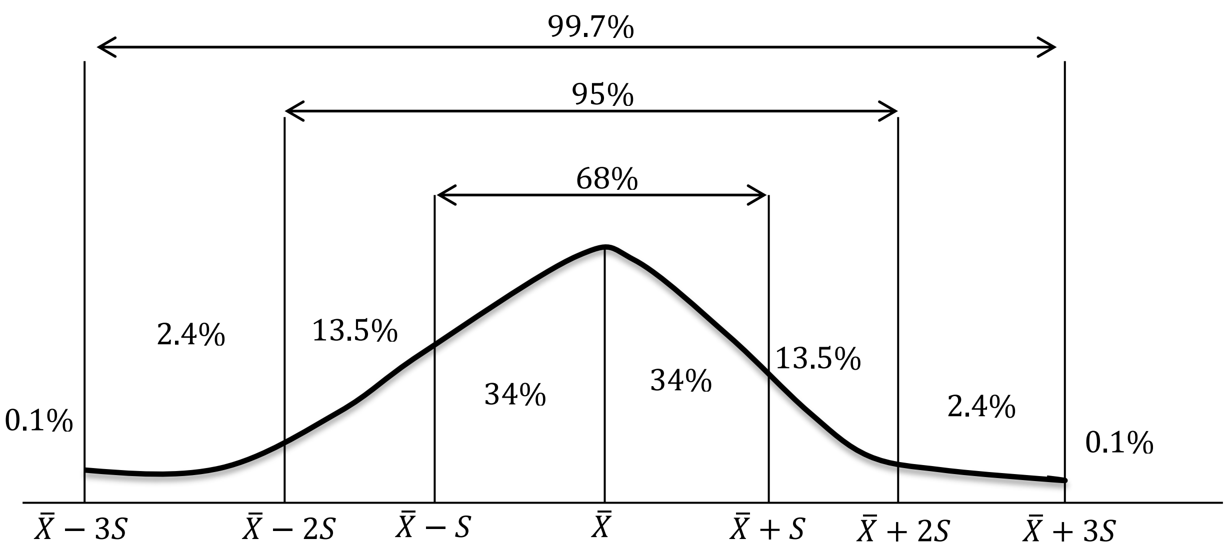
i.e in the interval $\bar{X} \pm KS$, where K whole number > 1

- At least 75% of data lie within two standard deviation, in the interval with end points $\bar{X} \pm 2S$

- At least 88.89% of the data lie within three standard deviation, in the interval with end points $\bar{X} \pm 3S$

The empirical rule (القاعدة التجريبية)

1- 68% of data lies within one standard deviation, in $\bar{X} \pm S$ interval
2- 95% of data lies within one standard deviation, in $\bar{X} \pm 2S$ interval
3- 99.7% of data lies within one standard deviation, in $\bar{X} \pm 3S$ interval



Example score of some tests have a bell-shaped distribution with mean $\bar{X} = 100$ and standard deviation S=10 discuss what the empirical rule concerning individuals with scores of 110, 120 and 130

Solution

❶ $68\% \rightarrow \bar{X} \pm S \Longrightarrow$ scores lie between (100-10 , 100+10)

$\Longrightarrow (90{,}110)$

❷ $95\% \rightarrow \bar{X} \pm 2S \Longrightarrow$ scores lie between (100-20 , 100+20)

$\Longrightarrow (80,120)$

❸ $99.7\% \rightarrow \bar{X} \pm 3S \Longrightarrow$ scores lie between (100-30 , 100+30)

$\Longrightarrow (70,130)$

Example the mean price of apartments in certain city 50000 with standard deviation 10000 , we will determine the price range for which at least 75% of the houses will sell

Solution

By Chebychev's Rule $\Longrightarrow$ at least 75% of the data fall within 2 standard deviation hence $(\bar{X} - 2S , \bar{X} + 2S) = (50000 - 2 \times 10000 , 50000 + 2 \times 10000)$

$= (30000 , 70000)$

i.e at least 75% of all apartments sold will have a price range from 30000 to 70000

EXERCISES

QESTION1: Classify each variable as quantitative or qualitative

1. Blood group
2. Time to finish exam
3. Height of student
4. Colors of flowers

- Qualitative (1)
- Quantitative (2)
- Quantitative (3)
- Qualitative (4)

QESTION2: Classify each variable as discrete or continuous

1. Weight of children
2. Number of students
3. Age of cats
4. Number of cars

- Continuous (1)
- Discrete (2)
- Continuous (3)
- Discrete (4)

QESTION3: Answer with true or false to the following sentence

- *Mode is defined for qualitative data*
- *The mean is sensitive to extreme value*
- *For a skewed distribution we have mode= mean = median*
- *Histogram with two peaks is multimodal*

- *True*
- *True*
- *False*
- *False*

QESTION4: Classify each variable as qualitative or quantitative

- The variable that record ID of students in an exam
- The variable that record weight of children in a school
- The variable that record phone number
- The variable that record sugar level

- Qualitative
- Quantitative
- Qualitative
- Qualitative

The variable that record blood pressure	Quantitative
The variable that record colors of cars	Qualitative
The variable that record number of flights	Quantitative
The variable that record time to finish exam	Quantitative

QESTION5: Classify each variable as Continuous or Discrete

The variable of heights of people	Continuous
The age of baby	Continuous
The number of children in school	Discrete
The weight of books	Continuous
The distance between cities	Continuous
Number of accidents	Discrete
The number of goals	Discrete
Temperature in a city	Continuous
Sex ratio in a country	Continuous

QESTION6: Put right word or symbol in its proper position:

statistic , variable , bar chart , Descriptive statistic.

[ ] is those statistical methods or techniques which are used for priesting and summarizing

[ ] is a function of a sample

[ ] is characteristic, feature or factor that varies from one individual to another in a population.

[ ] the frequency of each class in represented by a bar

The height of the bar doesn't matter

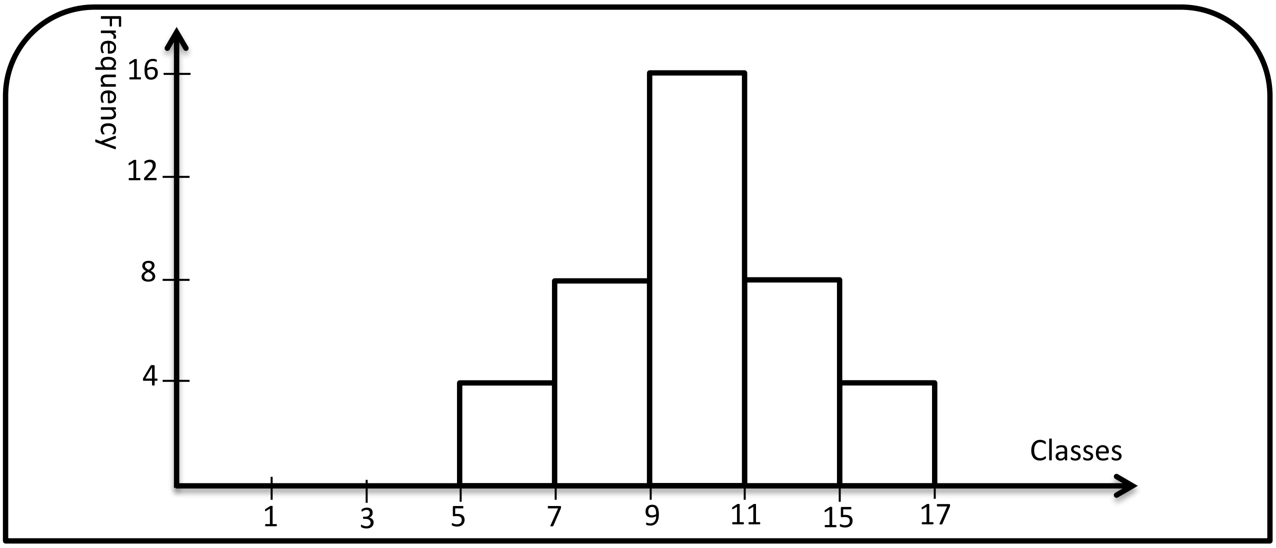
QESTION7: Consider the following data.

5.5	7.5	6	6.5	7	6.75	7	7.25	7.5	7.75	8
8.25	9	9.75	10	10.25	10.5	10.75	9	9.25	9.5	9.75
10.25	10.5	11	12	12.5	11.5	11.25	11.75	12.75	12.99	13.01
13.25	13.5	14.89	9.01	9.55	10.10	10.89				

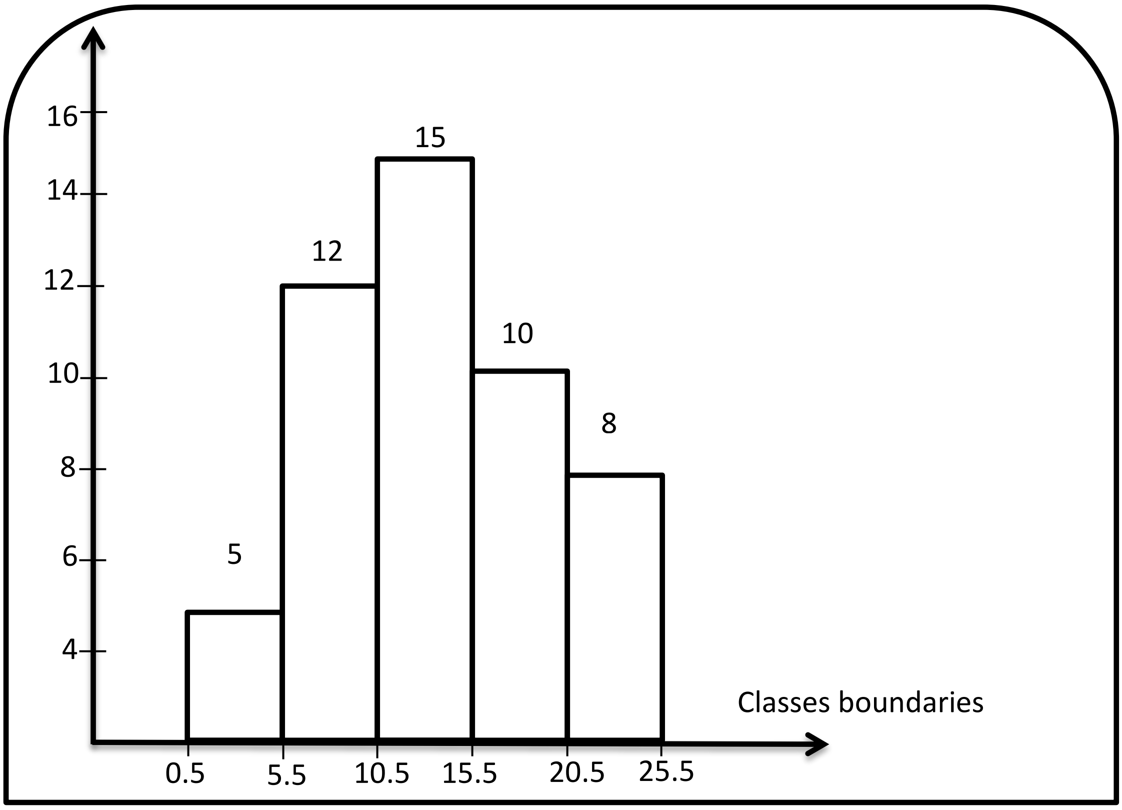
Construct the frequency distribution table

Class boundaries	Midpoint	Frequency	R.F	P.F	A.C.F
5-7		4			
7-9		8			
9-11		16			
11-13		8			
13-15		4			
Total		40			

Draw the histogram fir the data above



QESTION8: If we have data with the following histogram.



A) Complete the following frequency distribution

Solution

Class limit	Class boundaries	Midpoint	Frequency	R.F	P.F	A.C.F
			5			
			12			
			15			
			10			
			8			
Total			50			

B) Calculate the median for the given data

$\frac{n}{2} = \frac{50}{2} = 25 \qquad \bar{X} = L + \frac{\frac{\sum f}{2} - (F-f)}{f} \times c$

Median class $10.5 - 15.5 \Longrightarrow L = 10.5$

$F = 32 \qquad f = 15 \qquad c = 5$

$\bar{X} = 10.5 + \frac{25-(32-15)}{15} \times 5 \approx 13.17$

C) calculate the range for the given data

Solution

$R = X_K - X_1$, where X_K is midpoint for last , X_1 midpoint for first

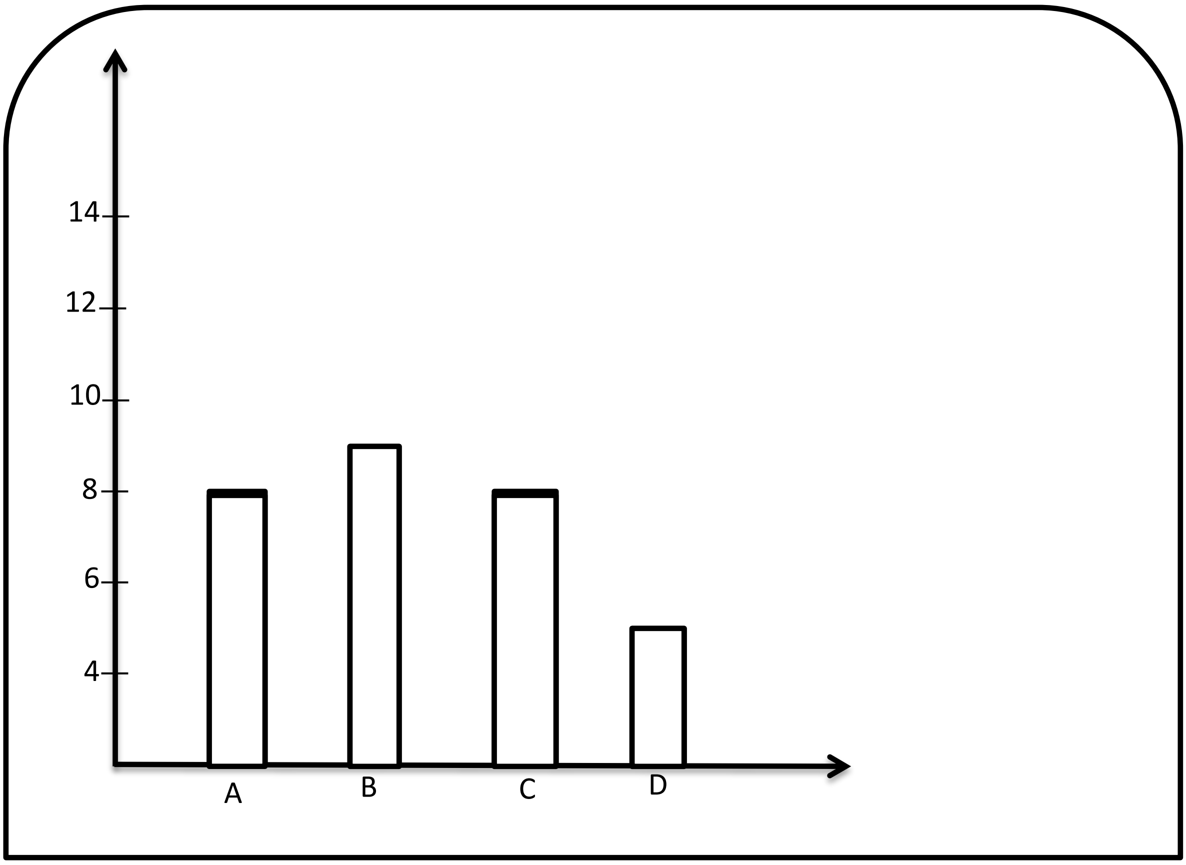
$R = 23 - 3 = 20$

QUESTION9: consider the following data

B	A	A	B	C	A	B	A	C	A	B
C	D	C	B	C	D	A	A	D	C	B
A	C	B	D	C	D	B	B			

Draw the bar graph for the given data

Solution



QUESTION: Using the data shown in the following table

No. of subjects	Frequency
0	8
1	18
2	12
3	2
Total	40

A. Represent them graphically using pie chart and bar chart

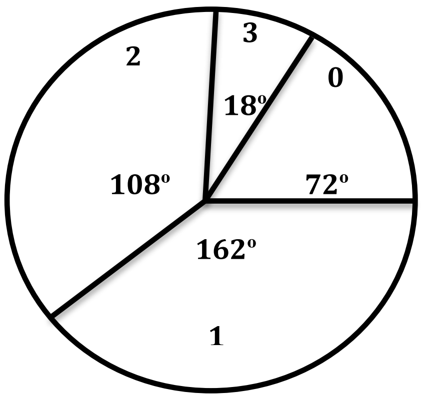
B. Compute the range of this data

C. Compute the mean of this data

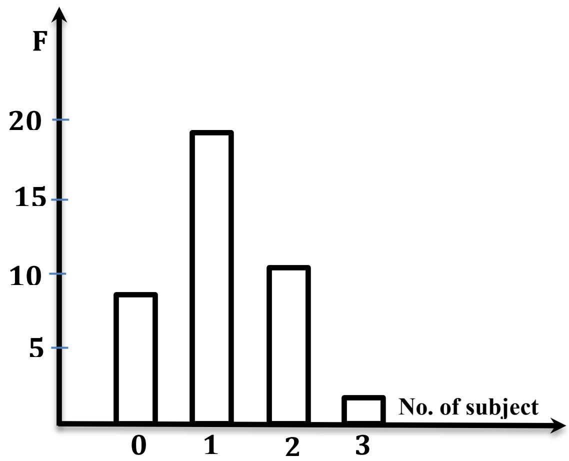
Solution

A.

No. of subjects X	Frequency F	R.F	FX	Degree
0	8	0.2	0	0.2x360=72
1	18	0.45	18	0.45x360=162
2	12	0.3	24	0.3x360=108
3	2	0.05	6	0.05x360=18
Total	40		48	



Pie-chart



B. The Range $= \max - \min = 3 - 0 = 3$

C. The $mean = \bar{X} = \frac{\sum FX}{\sum F} = \frac{48}{40} = 1.2$

QUESTION: the following data give the result of a sample survey A, B and C represent the three categories

C	A	C	B	A	C	B	C	C	C	B	C	B	C	B
C	B	C	C	B	C	C	C	A	B	C	C	B	C	A

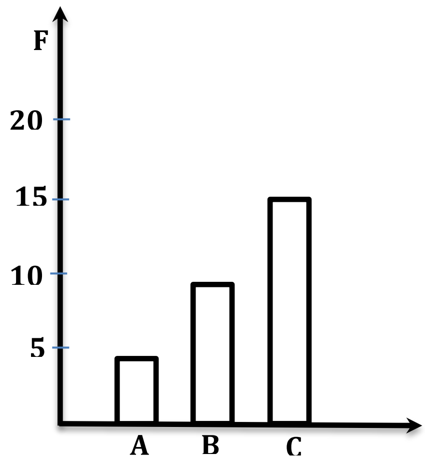
A. Prepare the frequency table
B. Calculate the relative frequency and percentage for all symbols
C. What percentage of elements belongs to category B?
D. Draw a bar chart and pie chart for frequency table

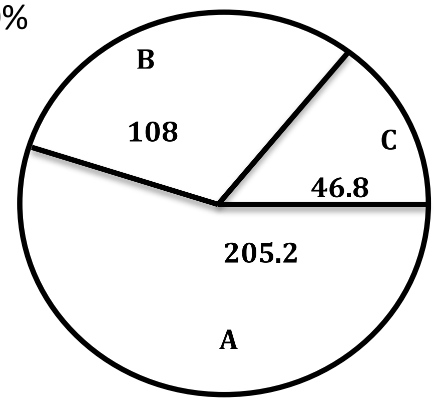
Solution

Letters	Frequency	Relative Frequency	Percentage Frequency
A	4	0.13	13%
B	9	0.3	30%
C	17	0.57	57%
Total	30	1	100%

C. The percentage of *B* Category is 30%

D.





A=0.13x360=46.8

B=0.3x360=108

C=0.57x360=205.2

QUESTION: A Company manufactures car batteries of a particular type the lives in years of 40 such batteries were recorded as follows

2.6	3	3.7	3.2	2.2	4.1	3.5	4.5	4.6	3.8	3.5	2.3	3.2
3.4	3.8	3.2	4.6	3.7	2.9	3.6	2.5	4.4	3.4	3.3	2.9	3
4.3	2.8	3.5	4.2	3.5	3.2	3.9	3.2	3.2	3.1	3.7	3.4	3.2
2.6												

Construct Frequency distribution using class of size 0.5, starting 2-2.5

Solution

Class interval	Frequency	Relative frequency	Percentage	ACF
$2 \to 2.5$	2	0.05	5%	2
$2.5 \to 3.0$	6	0.15	15%	8
$3.0 \to 3.5$	14	0.35	35%	22
$3.5 \to 4.0$	11	0.275	27.5%	33
$4.0 \to 4.5$	4	0.10	10%	37
$4.5 \to 5$	3	0.075	7.5%	40
Total	**40**	**1**	**100%**	

QUESTION: The distance (in km) of 40 engineers from their residence to place of work as follows

5	3	10	20	25	11	13	7	12	31
19	10	12	17	18	11	32	17	16	2
7	9	7	8	3	5	12	15	18	3
12	14	0.5	9	6	15	15	7	6	12

A. Construct a frequency distribution table with class size 5

B. Draw the histogram for the data

C. Draw the polygon for the data

D. Draw the ogive

E. Haw many engineers have residence at distance more than or equal 15km

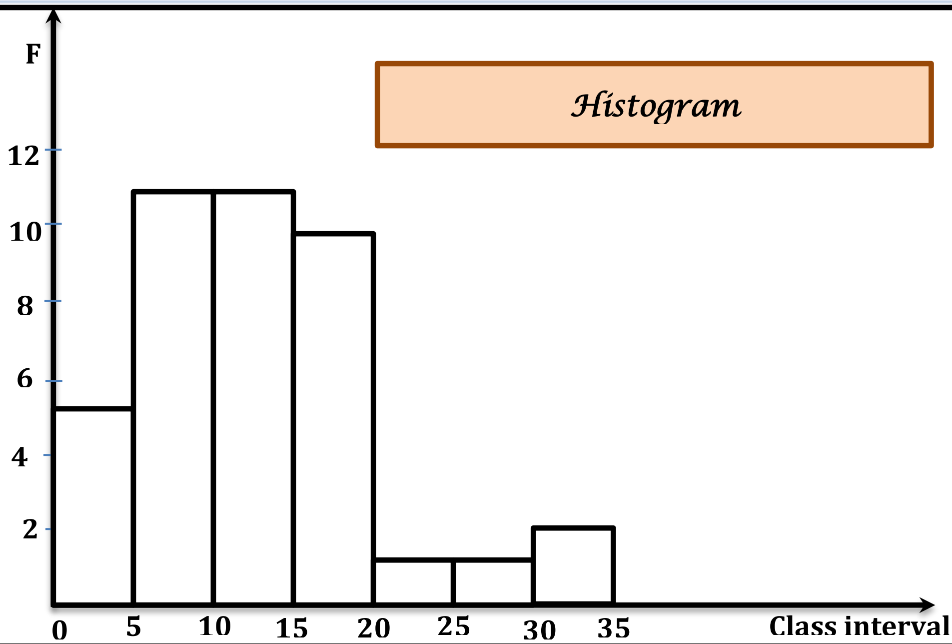
F. Haw many engineers have residence at distance less than 20km

Solution

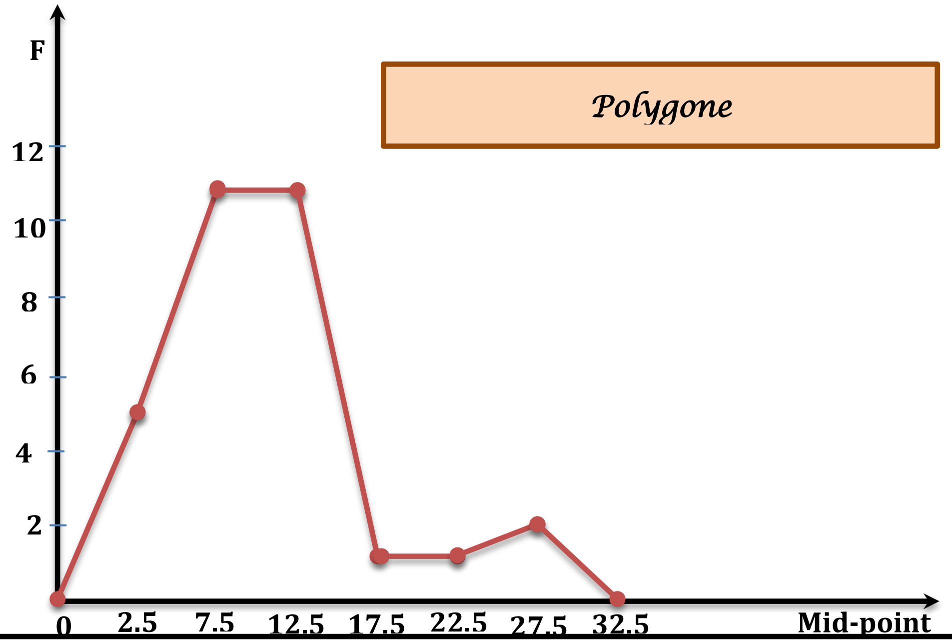
A.

Class interval	Frequency	Relative Frequency	Percentage	ASF
$0 \to 5$	5	0.125	12.5%	5
$5 \to 10$	11	0.275	27.5%	16
$10 \to 15$	11	0.275	27.5%	27
$15 \to 20$	9	0.225	22.5%	36
$20 \to 25$	1	0.025	2.5%	37
$25 \to 30$	1	0.025	2.5%	38
$30 \to 35$	2	0.05	5%	40
	40	1	100%	

B.

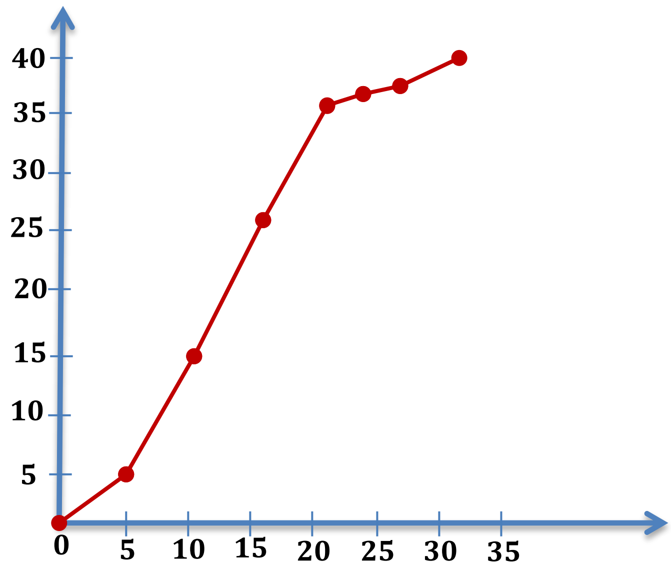


C.



D.

Less than	Frequency
0	0
5	5
10	16
15	27
20	36
25	37
30	38
35	40
"Less than ogive"	



E. Engineers have residence at distance more than or equal 15 km from their workplace are 13

F. Engineers have residence at distance less than 20km are 27

QUESTION: a sample of 100 children was asked haw many times they play computers games for a period of one week

No. of times	No. of children
$0 \to 3$	23
$4 \to 7$	40
$8 \to 11$	28
$12 \to 15$	6
$16 \to 19$	3
Total	100

A. Find the class midpoint

B. Do all classes have the same width? If so, what is this width?

C. Prepare the relative frequency and percentage distribution columns?

D. What percentage of these children play 8 or more times?

Solution

A.

No. of times	Midpoint	F	R.F	P.F	A.C.F
$0 \to 3$	1.5	23	0.23	23%	23
$4 \to 7$	5.5	40	0.40	40%	63
$8 \to 11$	9.5	28	0.28	28%	91
$12 \to 15$	13.5	6	0.06	6%	97
$16 \to 19$	17.5	3	0.03	3%	100
Total		100		100%	

B. All classes have the same width and it is equal 4

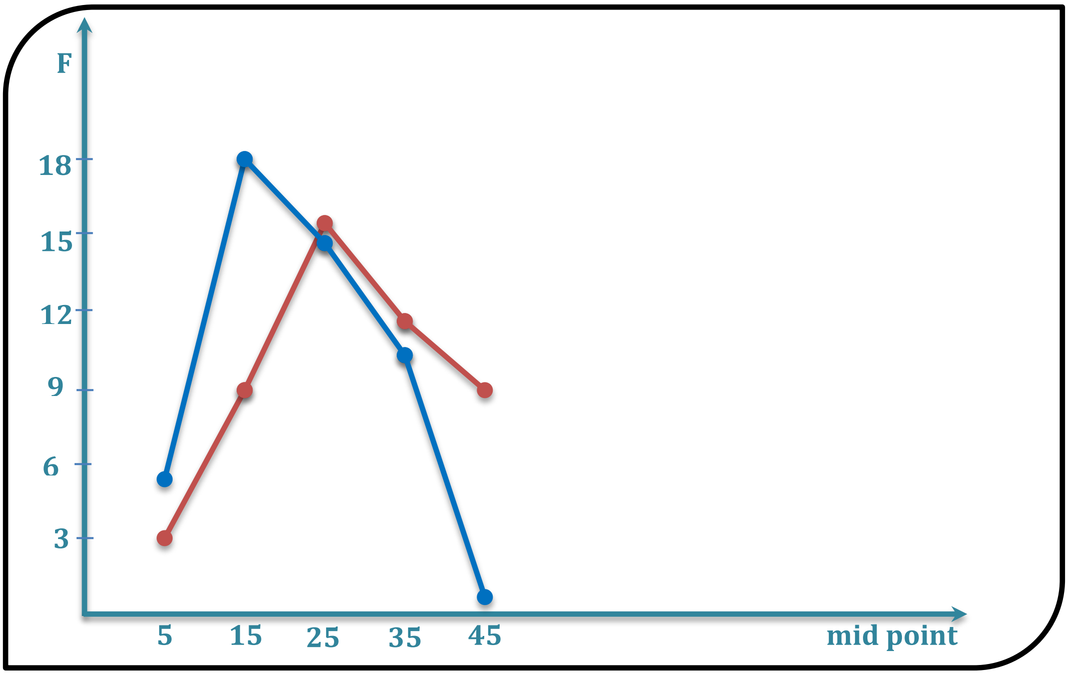
D. the percentage of children play 8 or more $\frac{37}{100} \times 100 = 37\%$

QUESTION: The table below gives the distribution of students of two sections according to the marks obtained

Marks	Section A	Section B
$0 \to 10$	3	5
$10 \to 20$	9	19
$20 \to 30$	17	15
$30 \to 40$	12	10
$40 \to 50$	9	1

Represent the marks on the same graph by two frequency polygons

Solution



QUESTION: Consider the following frequency distribution, representing the degree of an examination of 50 student of a class.

Class limit	Class Boundary	Class midpoint	Frequency	Relative Frequency	ACF
2-6			6		
7-11				0.24	
12-16					36
17-21				0.12	
22-26			8		
Total			50		

A. Complete the frequency distribution table

B. Draw the histogram, polygon and ogive

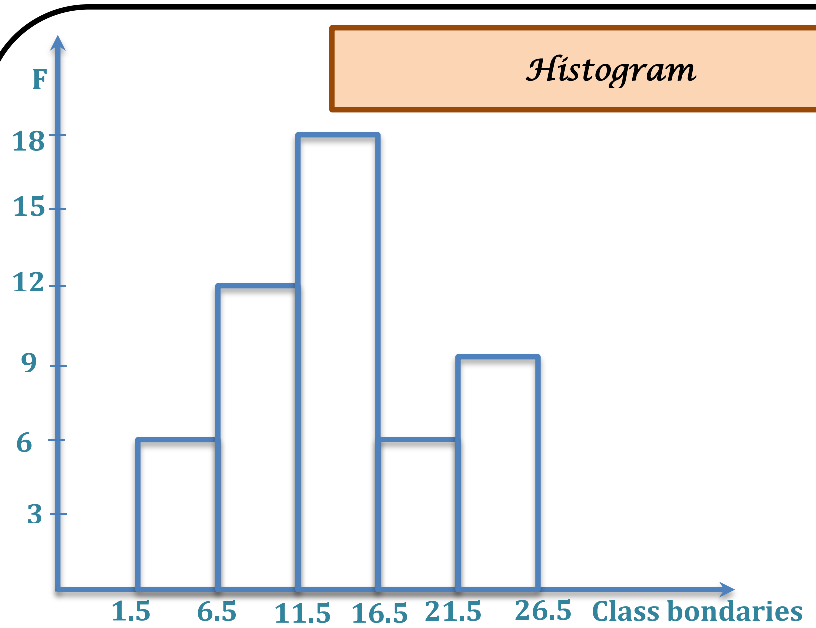
C. Calculate the mean, the median and mod

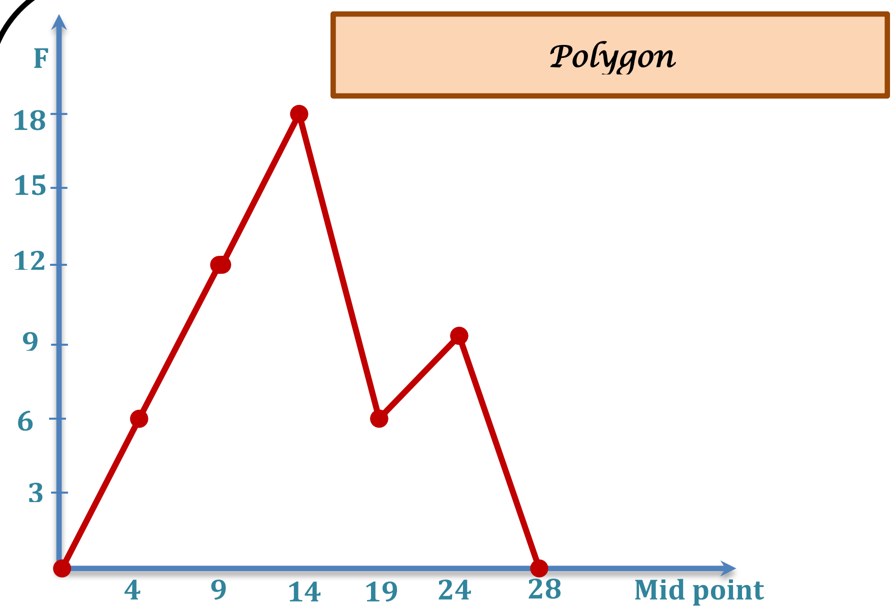
D. Calculate the standard deviation

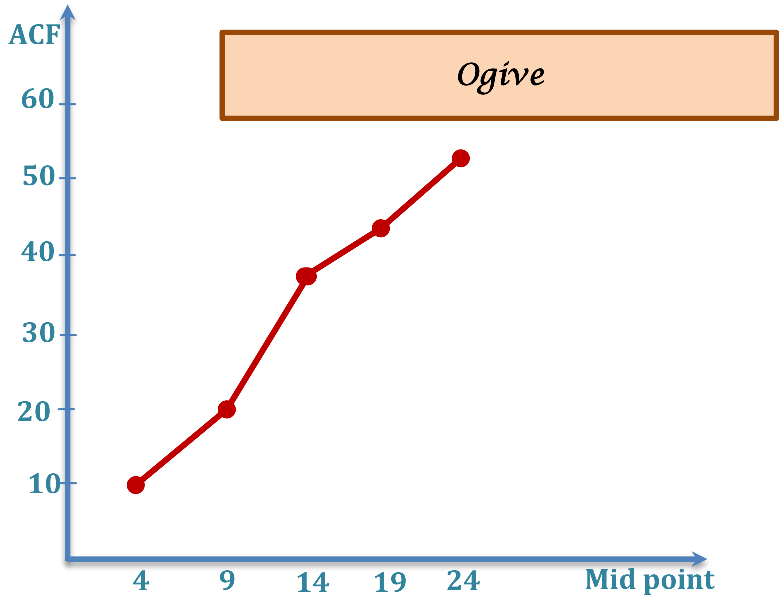
Solution

Class limit	Class boundaries	Class midpoint	Frequency	Relative Frequency	ACF
2-6	1.5-6.5	4	6	0.12	6
7-11	6.5-11.5	9	12	0.24	18
12-6	11.5-16.5	14	18	0.36	36
17-21	16.5-21.5	19	6	0.12	42
22-26	21.5-26.5	24	8	0.16	50
Total			50		

B.







C.

M	4	9	14	19	24	Total
F	6	12	18	6	8	50
FM	24	108	252	114	192	690

$$\bar{X} = \frac{\sum FM}{\sum F}$$

$$\bar{X} = \frac{690}{50} = 13.8$$

Class boundaries	Frequemcy	ACF
1.5-6.5	6	6
6.5-11.5	12	18
11.5-16.5	18	36
16.5-21.5	6	42
21.5-26.5	8	50
Total	50	

$$\frac{\sum F}{2} = \frac{50}{2} = 25$$

Class med in 11.5-16.5

L=11.5 f=18 F=36 C=5

$$\bar{X} = L + \frac{\frac{\sum f}{2} - (F - f)}{f} \times C$$

$$\bar{X} = 11.5 + \frac{25 - (36 - 18)}{18} \times 5 = 13.4$$

Class boundaries	Frequemcy
1.5-6.5	6
6.5-11.5	$12 \rightarrow f_p$
11.5-16.5	$18 \rightarrow f_m$
16.5-21.5	$6 \rightarrow f_n$
21.5-26.5	8
Total	50

$$\hat{X} = L + \frac{d_1}{d_1 + d_2} + C$$

$$d_1 = f_m - f_b = 18 - 12 = 6$$

$$d_2 = f_m - f_n = 18 - 6 = 12$$

The class mode 11.5-16.5

L=11.5 C=5

$$\bar{X} = 11.5 + \frac{6}{6+12} \times 5 = 13.2$$

D. يمكن إيجاد الانحراف المعياري بالحاسبة عن طريق إدخال قيم

الـ $X \leftarrow midpoint$ والـ $F \leftarrow frequency$

QUESTION: The point scored by a team are as follows

17, 2, 7, 27, 15, 5, 14, 8, 10, 24, 48, 10, 8, 7, 18, 28

A. Calculate the mean and standard deviation
B. Calculate the standard score of the value 7
C. Calculate the coefficient of variation
D. Calculate Q_1, Q_2, Q_3
E.

Solution

A. $\bar{X} = \frac{\sum X}{n} = \frac{17+2+7+27+15+5+14+8+10+24+48+10+8+7+18+28}{16} = \frac{248}{16} = 15.5$

- $\text{mod} \to 3 \to 1(\text{البيانات}) \to \text{AC}$
- $Shift \to 1 \to 4 \to 2(\bar{X})$

. . . الحاسبة

لإيجاد الانحراف المعياري ننشئ الجدول التالي:

X	17	2	7	27	15	5	14	8	10	24	48	10	8	7	18	28	$\sum X = 248$
X^2	289	4	49	729	225	25	196	64	100	576	2304	100	64	49	324	784	$\sum X^2 = 5882$

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{5882 - \frac{248^2}{16}}{15}} = 11.656$$

- $\text{mod} \to 3 \to 1(\text{data}) \to \text{AC}$
- $\text{Shift} \to 1 \to 4 \to 2(\bar{X})$ → $2(\bar{X})$ الوسط ؛ $4(S)$ الانحراف

. . . الحاسبة

B. standard score $Z = \frac{X-\bar{X}}{S}$

$for \quad X = 7 \qquad \bar{X} = 15.5 \qquad S = 11.656$

$$Z = \frac{7-15.5}{11.656} = -0.729$$

C. $C.V = \frac{S}{\bar{X}} \times 100 = \frac{11.656}{15.5} \times 100 = 75.2\%$

- The coefficient variation has no unit
- The variance has unit square

. . . خليك فاكر

D. $q_r = \frac{r(n+1)}{4}$, $where \quad r = 1,2,3$

$Q_r = X_K + S(X_{K+1} - X_K)$

يجب ترتيب البيانات في البداية

2, 5, 7, 7, 8, 8, 10, 10, 14, 15, 17, 18, 24, 27, 28, 48

$$q_1 = \frac{1 \times (16+1)}{4} = 4.25$$

أي أن الربيع الأول يقع بين القيمة الرابعة والخامسة وقيمته هي:

$$Q_1 = 7 + 0.25 \times (8 - 7) = 7 + 0.25 = 7.25$$

$$q_2 = \frac{2 \times (16+1)}{4} = 8.5 \Rightarrow K = 8 \quad S = 0.5$$

$$Q_2 = 10 + 0.5 \times (14 - 10) = 10 + 2 = 12$$

$$q_3 = \frac{3 \times (16+1)}{4} = 12.75 \Rightarrow K = 12 \quad S = 0.75$$

$$Q_3 = 18 + 0.75 \times (24 - 18) = 18 + 4.5 = 22.5$$

QUESTION: The daily sale of sugar (kg) in a certain grocery shop

Mon	Tus	Wed	Thur	Fri	Sat
15	120	12	50	70.5	140.5

A. Calculate the average

B. Calculate the variance and standard deviation

C. Determine the coefficient of variation

Solution

A. The average (mean): $\bar{X} = \frac{\sum X}{n}$

$$\bar{X} = \frac{75 + 120 + 12 + 50 + 70.5 + 140.5}{6} = 78kg$$

B.

X	75	120	12	50	70.5	140.5	$\sum X = 468$
X^2	5625	14400	144	2500	4970.25	19740.25	$\sum X^2 = 4737905$

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{47379.5 - \frac{468^2}{6}}{5}} = 46.637kg$$

$$S^2 = 2175.1\ kg^2$$

C. Coefficient variation $C.V = \frac{S}{\bar{X}} \times 100 = \frac{46.637}{78} \times 100 = 59.9\%$

QUESTION: Let the following data be marks of 10 students

45, 45, 63, 76, 67, 84, 75, 84, 75, 48, 62, 65

- **A.** Calculate Q1, Q2 and Q3
- **B.** Calculate IQR
- **C.** Have the given data extreme value?
- **D.** Construct the box plot

Solution

أولا: ترتيب البيانات

45, 45, 48, 62, 63, 65, 67, 75, 76, 84

$$q_r = \frac{r(n+1)}{4} \qquad where\ \ r = 1, 2, 3$$

$$Q_r = X_K + S(X_{K+1} - X_K)$$

$$q_1 = \frac{1 \times (10+1)}{4} = 2.75 \qquad \Rightarrow K = 2 \quad S = 0.75$$

$$Q_1 = 45 + 0.75 \times (48 - 45) = 47.25$$

$$q_2 = \frac{2 \times (10+1)}{4} = 5.5 \qquad \Rightarrow \quad K = 5 \qquad S = 0.5$$

$$Q_2 = 63 + 0.5 \times (65 - 63) = 64$$

$$q_3 = \frac{3 \times (10+1)}{4} = 8.25 \qquad \Rightarrow \quad K = 8 \qquad S = 0.25$$

$$Q_3 = 75 + 0.25 \times (76 - 75) = 75.25$$

B. IQR = Q3 – Q1 =75.25 – 47.25 = 28

C. Extreme value are

$X < Q_1 - 1.5 \times (Q_3 - Q_1)$

$X < 47.325 - 1.5 \times (28)$

$X < 5.25$

$X > Q_3 \pm 1.5 \times (Q_3 - Q_1)$

$X > 75.25 \pm 1.5 \times (28)$

$X > 117.25$

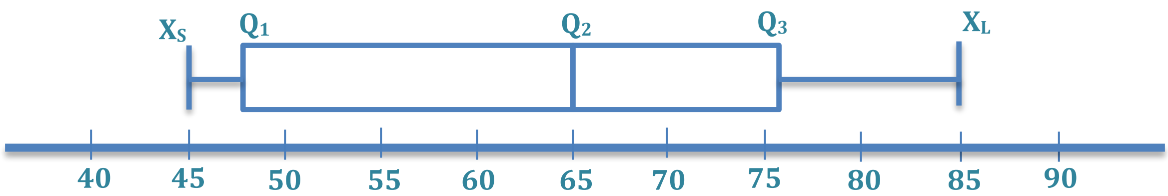
أي قيمة أقل من 5.25 أو أكبر من 117.25 هي قيمة شاذة وبالتالي فإن هذه البيانات لا تحتوي على قيم شاذة

D. To construct the box plot

Five numbers are

Smallest value, Q_1, Q_2, Q_3 largest value

45, 47.25, 64, 75.25, 84



QUESTION: Consider the following data

-15, 20, 40, 50, 65, 65, 70, 73, 75, 137

A. Have the given data extreme value

B. Use the suitable measure to calculate the average and dispersion for the give data

Solution

لمعرفة ما إذا كانت توجد قيم شاذة أم لا نوجد الربيعيات حيث:

$q_r = \frac{r(n+1)}{4}$ & $Q_r = X_K + S(X_{K+1} - X_K)$

$q_1 = \frac{1 \times (10 + 1)}{4} = 2.75 \quad \Longrightarrow K = 2 \quad S = 0.75$

$Q_1 = 20 + 0.75 \times (40 - 20) = 35$

$q_2 = \frac{2\times(10+1)}{4} = 5.5 \quad \Longrightarrow \quad K = 5 \quad S = 0.5$

$Q_2 = 65 + 0.5 \times (65 - 65) = 65$

$q_3 = \frac{3 \times (10+1)}{4} = 8.25 \quad \Longrightarrow \quad K = 8 \quad S = 0.25$

$Q_3 = 73 + 0.25 \times (75 - 73) = 73.5$

A) $X < Q_1 - 1.5 \times (IQR) = 35 - 1.5 \times (38.5) = -22.5$

$X > Q_3 + 1.5 \times (IQR) = 73.5 + 1.5 \times (38.5) = 131.25$

أي قيمة أقل من 22.75- أو أكبر من 131.25 هي قيمة شاذة

ولذلك 137 هي extreme value

لوجود قيمة شاذة في البيانات لا يفضل استخدام الوسط الحسابي لأنه يتأثر بوجود القيم الشاذة وكذلك لحساب التشتت لا يفضل استخدام الانحراف أو التباين لتأثرها بالقيم الشاذة أيضا.

لأن البيانات السابقة لها أكثر من منوال يفضل استخدام الوسيط

~~-15,~~ ~~20,~~ ~~40,~~ ~~50,~~ 65, 65, ~~70,~~ ~~73,~~ ~~75,~~ ~~137,~~

$median \frac{65+65}{2} = 65$

أفضل مقاييس لتشتت في هذه الحالة هو المدى الربيعي

B) $IQR = Q_3 - Q_1 = 38.5$

QUESTION: The following data give the number of computer keyboards for a sample of 25 days

45	52	48	41	56
46	44	42	48	53
51	53	51	48	46
43	52	50	54	47
44	47	50	49	52

Prepare a box-plot and then comment the skewness of data

Solution

يجب ترتيب البيانات في البداية وحساب Q_1, Q_2, Q_3 :

Arrange data

41, 42, 43, 44, 44, 45, 46, 46, 47, 47, 48, 48, 48, 49, 50, 50, 51, 51, 52, 52, 52, 53, 53, 54, 56

$q_r = \frac{r(n+1)}{4}$ & $Q_r = X_K + S(X_{K+1} - X_K)$

$$q_1 = \frac{1 \times (25+1)}{4} = 6.5 \quad \Rightarrow K = 6 \quad S = 0.5$$

$$Q_1 = 45 + 0.5 \times (45 - 46) = 45.5$$

$$q_2 = \frac{2 \times (25+1)}{4} = 13 \quad \Rightarrow \quad K = 13 \quad S = 0$$

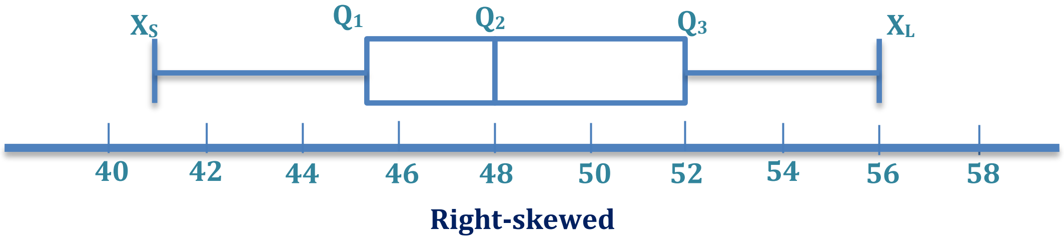
$$Q_2 = 48 + 0 \times (49 - 48) = 48$$

$$q_3 = \frac{3 \times (25+1)}{4} = 19.5 \quad \Rightarrow \quad K = 19 \quad S = 0.5$$

$$Q_3 = 52 + 0.5 \times (52 - 52) = 52$$

Five numbers:

Small value Q_1, Q_2, Q_3 large value **41, 45.5, 48, 52, 56**



QUESTION: The mean age of six persons is 49 years, the age of five of these six persons are 55, 39, 44, 51 and 45 fin the age of the sixth

Solution

$$\bar{X} = \frac{\sum X}{n}$$

$$49 = \frac{55 + 39 + 44 + 51 + 45 + X}{6}$$

$$234 + X = 294$$

$$X = 294 - 234 = 60$$

QUESTION: The following observation have arranged in ascending order

29, 32, 48, x, x+2, 72, 78, 84, 95

If the median of the data is 63, then

- *A·* Calculate the value of X
- *B·* Calculate mean and standard deviation
- *C·* Find the five number and construct the box plot

Solution

~~29,~~ ~~32,~~ ~~48,~~ ~~X,~~ **X+2,** ~~72,~~ ~~78,~~ ~~84,~~ ~~95,~~

A· the median $X + 2 = 63 \implies X = 61$

So the data will be

29, 32, 48, 61, 63, 72, 78, 84, 95

B· The mean $\bar{X} = \frac{\sum X}{n}$

$$\bar{X} = \frac{29+32+48+61+63+72+78+84+95}{9} = \frac{562}{9} = 62.4$$

To calculate standard deviation, we construct the table

X	29	32	48	61	63	72	84	95	$\sum X = 562$
X^2	841	1024	2304	3721	3969	5184	7056	8836	$\sum X^2 = 39019$

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} = \sqrt{\frac{39019 - \frac{562^2}{9}}{85}} = 22.15$$

PRINCIPLES OF

PROBABILITY & STATISTICS

CHAPTER 2

[redacted]

The fundamental principle of country

مبدأ العد التنازلي

Multiplicative rule (قاعدة الضرب)

إذا كان لدينا عدد K من ظاهرة $O_1 , O_2 , \dots\dots , O_K$ و هذه الظاهرة تحدث $n_1 , n_2 ,\dots n_K$ من الطرق بالترتيب وبالتالي فإن عدد الطرق لحدوث كل هذه الظواهر في نفس الوقت هو $n_1 x n_2 x \dots x n_K$ من الطرق.

Addition rule (قاعدة الجمع)

إذا كان لدينا عدد K من ظاهرة $O_1 , O_2 , \dots\dots , O_K$ و هذه الظاهرة تحدث بـ n_1 أو n_2 ...أو n_K من الطرق بالترتيب وبالتالي فإن عدد الطرق لحدوث كل هذه الظواهر في نفس الوقت هو $n_1+n_2+\dots+n_K$ من الطرق.

Example Haw many 6-digits Zip Codes are possible if

a) Digits can be repeated?
b) Digits can be repeated?

Solution

A- If digits can be repeated, then number of 6-digites equal to

$10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6$ codes

B- If digits can not be repeated, then the number of 6-digits equal to

$10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151200$ codes

Example Haw many book can we use Arabic or English characters?

Solution

by addition rule we can index 28+26= 54 book

Factorial notation (مضروب العدد)

$$n! = n \times (n-1) \times (n-2) \dots \dots 3 \times 2 \times 1$$

Note $0! = 1 \quad \& \quad 1! = 1$

Example $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

Permutations (التباديل)

$$nP_r = \frac{n!}{(n-r)!} \quad \text{where } 0 \le r \le n$$

خليك فاكر

Key هناك كلمات مفتاحية نعرف من خلالها أنه يجب استخدام التباديل حيث أن الترتيب فيها مهم

words ⟹ order, arrangement , array

Example Haw many ways one can arrange in order any three of the first 8 letters of L, m, o, p, q, r, s, t

$$8P_3 = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8 \times 7 \times 6 \times 5!}{5!} = 8 \times 7 \times 6 = 336$$

Combinations (التوافيق)

$$nC_r = \frac{n!}{r!\,(n-r)!} \quad \text{where } 0 \le r \le n$$

خليك فاكر

- في التوافيق الترتيب غير مهم
- يمكن كتابة التوافيق أيضا على الصورة $\left[{n \atop r}\right]$

Example Haw many different unorder groups of any three of six letter L, m, n, o, p and q?

Solution

$$6C_3 = \frac{6 \times 5 \times 4}{3 \times 2 \times 1} = 20$$

Example Haw many ways can you arrange 4 out of 7 books on a shelf.

Solution

لاحظ وجود كلمة arrangeوهي من الكلمات الدالة على وجوب الترتيب لذلك نستخدم التبادل

$$7P_4 = 7 \times 6 \times 5 \times 4 = 840$$

7	Shift	X	4	=	840	بالحاسبة

Example Haw many possible different hands of 5 cards each can be dealt from a standard deck of 52 cards.

Solution

This is a combination because order is not important

$$n = 52 \quad , \quad r = 5$$

$$52C_5 = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2598960$$

52 | Shift | / | 5 | = | 2598960 ← بالحاسبة

Example If a man owns 5 pairs of parts, 7 shirts and four pairs of shoes, haw many outfits can be assembled?

Solution

By multiplicative rule $5 \times 7 \times 4 = 140$ outfit

Example In a group of 10 people, a 20 $, 10$ and 5$ prize will be given, Haw many ways can the prize be distributed

Solution

لا حظ أن الترتيب هنا مهم

لأنه من العشرة أفراد سوف يحصل شخص عن جائزة الـ 20$ ومن التسعة أفراد الباقية سوف الباقية سوف يحصل شخص على جائزة الـ 10$ وبالتالي فإن فرد من الثمانية الباقين سوف يحصل على 5$

$$10P_3 = 10 \times 9 \times 8 = 720$$

Example In a group of 10 people, three 5$ prizes will be given, Haw many ways can the prizes be distributed

Solution

Order does not matter $10C_3 = \frac{10\times9\times8}{3\times2\times1} = 120$

- ❖ and → multiplicative rule
- ❖ or → additional rule
- ❖ ordered matter → permutation
- ❖ ordered does not matter → combination

خليك فاكر

Type of experiment

Regular	Random
Which we knaw the results in advance	which we don't knaw its exact out come, but we can determine the set of all out come
Example $H_2 + O \rightarrow H_2O$	**Example** **tossing coin**

خليك فاكر

في التجربة العشوائية يقال لحدثان لهما نفس الفرصة في الظهور equally likely

Probability science: is a branch of mathematics that deals with theoretical models of random experiments

Note probability space of any random experiment is a tripple have from

$[\Omega , \mathcal{A} , p]$

- Ω is the set of all outcome
- $\mathcal{A}$ is algebra of events
- p is called probability measure

أولا: Ω هي مجموعة كل النواتج الممكنة لتجربة عشوائية

Example In the experiment of flip a coin and roll adie , The set of all outcome

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

Event is a subset of sample space

Example Tossing a coin three times and determine the Sample space, then determine the following events

E_1: The event of two heads

E_2: The event that at least two heads

E_3: The event that at most two heads

E_4: The event that a heads is the first toss

Solution

$\Omega = \{HHH,\ HHT,\ HTH,\ THH,\ TTT,\ TTH,\ THT,\ HTT\}$

$E_1 = \{HHT,\ HTH,\ THH\}$

$E_2 = \{HHT,\ HTH,\ THH,\ HHH\}$

$E_3 = \{HHT,\ HTH,\ THH,\ TTT,\ TTH,\ THT,\ HTT\}$

$E_4 = \{HHH,\ HHT,\ HTH,\ HTT\}$

ثانيا: Algebra of events:

A collection $\mathcal{A}$ of subset of Ω is said to be an algebra on Ω if

- $\Omega \in \mathcal{A}$
- If A and B $\in \mathcal{A}$, then $A \cup B \in \mathcal{A}$
- If $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$

✓ هناك نوعان من الحوادث
✓ حوادث بسيطة تحتوى على عنصر واحد فقط
✓ حوادث مركبة تحتوى أكثر من عنصر

Example In tossing die, the sample space is $\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$

The events $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\} \implies$ Simple

Even events $\{2,\ 4,\ 6\} \implies$ Compound

Odd events $\{1,\ 3,\ 5\} \implies$ Compound

Note if Ω is a set, then the set of all subset in Ω is 2^{Ω}

Example $\Omega = \{a,\ b,\ c\} \quad \implies |\Omega| = 3$

Then the set of all elements is $2^{|\Omega|} = 2^3 = 8$

$\phi, \{a\}, \{b\}, \{c\}, \{a,\ b\}, \{a, c\}, \{b, c\}, \{a,\ b,\ c\}$

بعض العمليات على الأحداث

Union (الاتحاد)

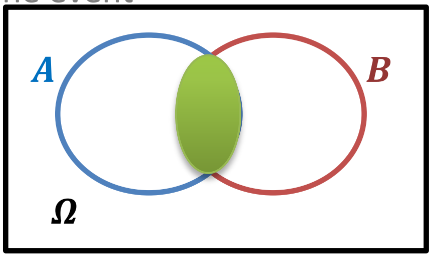
Is denoted by $A \cup B$, is an event containing all element in A or B or both

i.e $A \cup B$ is the occurrence of at least one event

Example $\Omega = \{1, 2, 3, 4, 5, 6\}$

$A = \{1, 2\}$, $B = \{2, 5, 6\}$

Then $A \cup B = \{1, 2, 5, 6\}$



$A \cup B$

Intersection (التقاطع)

$A \cap B$ or A and B

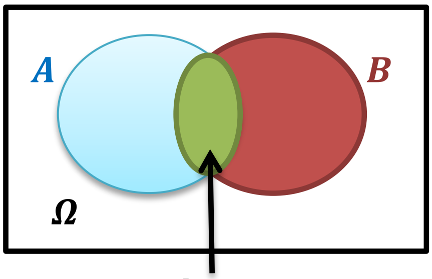
$A \cap B$ is an event containing all elements in A and B in the same time

Example tossing coin two time

$\Omega = \{HH, HT, TH, TT\}$

$A = \{TH, TT\}$, $B = \{HT, TT\}$

$A \cap B = \{TT\}$



$A \cap B$

Complement ($\bar{A}$) (المتممة)

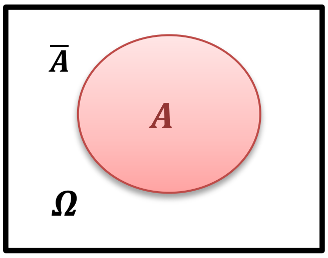
The complement of an event that occurs if A does not

Example Tossing adie , determine the following event

a. E_1: the event of even or prime

b. E_2: the event of prime

c. E_3: the event of prime does not occur



Solution

$\Omega = \{1, 2, 3, 4, 5, 6\}$

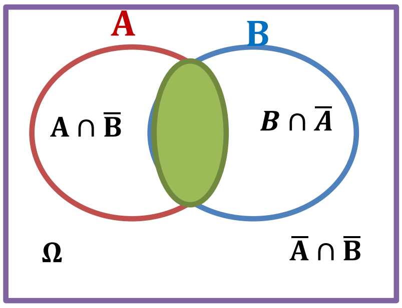
$E_1 = \{2, 4, 6\} \cup \{2, 3, 5\} = \{2, 3, 4, 5, 6\}$

$E_2 = \{2, 3, 5\}$

E_3 is the complement of $E_2 \implies E_3 = \{1, 4, 6\}$

Difference between two events (الفرق)

يرمز لها بأحد الرموز $A\backslash B$ أو $A \cap \bar{B}$ وهي تعني العناصر الموجودة في A وغير موجود في B

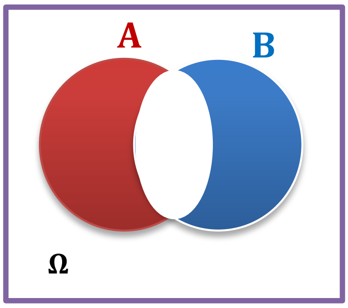


Example let $\Omega = \{1, 2, 3, 4, 5, 6\}$

A={1, 2} B={2, 3, 5}

A\B = {1}

Exactly one event



If A and B are two event from Ω, then:

$(A\backslash B) \cup (B\backslash A)$or $(A \cap \bar{B}) \cup (B \cap \bar{A})$

Example let A={1, 2, 3, 4} , B={3, 4, 5, 6}, C={2, 4, 6, 7}

Find $A\Delta B\Delta C$

Solution

First we find the symmetric of first two set A and B

$A\Delta B = \{1,2,5,6\}$ then

$A\Delta B\Delta C = \{1, 4, 5, 7\}$

Impossible event (الحدث المستحيل)

هو الحدث الذي لا يمكن وقوعه مثال ظهور العدد 7 عند القاء زهر أو ظهور عدد سالب لنفس التجربة كما أن تقاطع أي حدث مع متممته مثال على الحوادث المستحيلة

Example let $\Omega = \{1, 2, 3, 4, 5, 6\}$

A the event of even number $A = \{2, 4, 6\}$ So that the event of odd number $\bar{A} = \{1, 3, 5\}$ then $A \cap \bar{A} = \emptyset \Longrightarrow$ impossible event

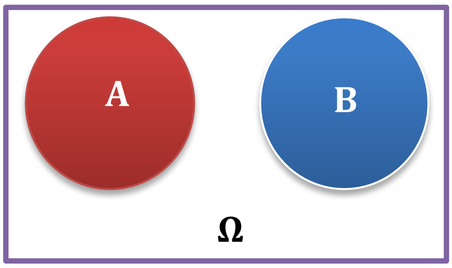
Certain event (الحدث المؤكد)

هو ذلك الحدث الذي لا بد من حدوثه بشكل مؤكد ومثال ذلك احتمال وقوع الحدث اتحاد متممة ذلك الحدث هو sure event

$A \cup \bar{A} = \Omega$ for any event A and $\bar{A}$

Mutually exclusive events (الأحداث المتنافية)

يقال لحدثان A و B أنهما متنافيان أو disjoint إذا كانت $A \cap B = \emptyset$



Example let $\Omega = \{1, 2, 3, 4, 5, 6\}$

A={1, 2} B={3, 4, 6} then $A \cap B = \emptyset$

So we said A and B are disjoint or mutually

EXERCISES

QESTION1: If an automobile license plate must consist of three letters followed by three single digit numbers.

haw many different license plates are possible?

Solution

According to the fundamental principle of counting the possible number of license plate $= 26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17576000$

QESTION2: The science club has challenged the math club each club's team should be comprised of 2 boys and 3 girls, there are 20 boys and 15 girls in the science club and 25 boys and 30 girls in the math club

Solution

For science

$20C_2 \times 15C_3 = 86450$

For math

$25C_2 \times 30C_3 = 1218000$

QESTION3: A manager must choose five secretaries from a many 12 applicants and assign them to different stations haw many different arrangement are possible?

Solution

لاحظ أن الترتيب هنا مهم في هذا المثال ولذلك نستخدم التباديل

$n = 12 \quad r = 5 \quad \Rightarrow nP_r = 12P_5 = 95040$

CONCEPT OF PROBABILITY

Probability (الاحتمال)

Is the numerical measure of likelihood that a specific event will occur

الاحتمال هو مقياس عددي لوقوع حدث ما

التكرار النسبي هو تكرار لفئة مقسومة على المجموع الكلي للتكرار و هو نفس التعريف الكلاسيكي للاحتمال

حيث : $F_A = \frac{n(A)}{N} \Rightarrow P(A) = \frac{n(A)}{N}$

خليك فاكر

Laplace principle of probability

For simple event $E \Rightarrow P(E) = \frac{1}{|\Omega|}$

For compound event $A \Rightarrow P(A) = \frac{|A|}{|\Omega|}$

In the experiment of rolling a fair die find the

A: probability of event number

B: probability of odd number

C: Probability of number less than two

D:probability of number more than six

Solution

$\Omega = \{1, 2, 3, 4, 5, 6\} \Rightarrow |\Omega| = 6$

$A = \{2, 4, 6\} \Rightarrow |A| = 3$ & $B = \{1, 3, 5\} \Rightarrow |B| = 3$

$C = \{1\} \Rightarrow |C| = 1$ & $D = \phi \Rightarrow |D| = 0$

$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$ حدث مركب

$P(B) = \frac{|B|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$ حدث مركب

$P(C) = \frac{|C|}{|\Omega|} = \frac{1}{6}$ حدث بسيط

$P(D) = \frac{|D|}{|\Omega|} = \frac{0}{6} = 0$ حدث مستحيل

خليك فاكر

1. $0 \leq P(A) \leq 1$
2. $\sum P(A) = 1$
3. $P(\emptyset) = 0$
4. If $A \cap B = \emptyset \implies P(A \cap B) = 0$ *disjoint*
5. $P(\Omega) = 1$
6. $P(A) + P(\bar{A}) = 1$
7. If $A \subset B, \implies P(A) \leq P(B)$
8. $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$
9. $P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$

(8 and 9: De_Mogan's laws)

10. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ additive rule

If A and B are disjoint $\implies P(A \cap B) = 0$ *then,*

$$P(A \cup B) = P(A) + P(B)$$

EXAMPLE:

Let $P(A) = \frac{3}{8}$ $P(B) = \frac{1}{2}$ $P(A \cap B) = \frac{1}{4}$

Find: *A)* $P(A \cup B)$ *B)* $P(\bar{A}),\ P(\bar{B})$ *C)* $P(\bar{A} \cap \bar{B})$

D) $P(\bar{A} \cup \bar{B})$ *E)* $P(A \cap \bar{B})$ *F)* $P(\bar{A} \cap B)$

Solution

A) $P(A \cup B) = P(A) + P(B) - P(A \cap) = \frac{3}{8} + \frac{1}{2} - \frac{1}{4} = \frac{5}{8}$

B) $P(\bar{A}) = 1 - P(A) = 1 - \frac{3}{8} = \frac{5}{8}$

$P(\bar{B}) = 1 - P(B) = 1 - \frac{1}{2} = \frac{1}{2}$

C) $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - \frac{5}{8} = \frac{3}{8}$

D) $P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B}) = 1 - P(A \cap B) = 1 - \frac{1}{4} = \frac{3}{4}$

E) $P(A \cap \bar{B}) = P(A) - P(A \cap B) = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$

F) $P(\bar{A} \cap B) = P(B) - P(A \cap B) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

EXAMPLE: Three students A, B, and C are in a swimming race. A and B have the same probability of swimming and each is twice as likely to win as C.

Find the probability that B or C wins.

Solution

Let $P(C) = x$, so $P(A) = P(B) = 2x$

since $\sum P(x) = 1 \implies P(A) + P(B) + P(C) = 1$

$2x + 2x + x = 1 \implies 5x = 1$

So $P(A) = P(B) = \frac{2}{5}$ and $P(C) = \frac{1}{5}$

$$P(B \cup C) = P(B) + P(C) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5}$$

Example: Let adie is weighted, so the even number have the same chance of appearing, the odd number have the same chance, and each even number is twice odd.

Find the probability that:

A) An even number appears
B) Odd number appears
C) A prime number appears
D) An odd number but not prime

Solution

Let $\Omega = \{1, 2, 3, 4, 5, 6\}$

Let $\underset{even}{A = \{2, 4, 6\}}$ & $\underset{odd}{B = \{1, 3, 5\}}$ & $\underset{prime}{C = \{2, 3, 5\}}$

$P(2) + P(4) + P(6) + 2P(1) + 2P(3) + 2P(5) = 1$

$$x + x + x + 2x + 2x + 2x = 1$$

$9x = 1 \implies x = \frac{1}{9}$, then $P(1) = P(3) = P(5) = \frac{1}{9}$

$$P(2) = P(4) = P(6) = \frac{2}{9}$$

A) $P(A) = P(2) + P(4) + P(6) = \frac{6}{9}$

B) $P(B) = P(1) + P(3) + P(5) = \frac{3}{9}$

C) $P(C) = P(2) + P(3) + P(5) = \frac{4}{9}$

D) $P(B \cap \bar{C}) = P(B) - P(B \cap C) = \frac{3}{9} - \frac{2}{9} = \frac{1}{9}$

Example: Let $P(A) = 0.8 \quad P(B) = 0.55 \quad P(A \cup B) = 0.9$

Find :

A) Occurrence of A and B
B) Occurrence of only A and not B
C) Non occurrence of A and B
D) Occurrence of only A or only B

Solution

A) $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.8 + 0.55 - 0.9 = 0.45$

B) $P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.8 - 0.45 = 0.35$

C) $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.9 = 0.1$

D) $P(A \cap \bar{B}) \cup P(\bar{A} \cap B) = P(A) - P(A \cap B) + P(B) - P(A \cap B) = 0.35 + 0.10 = 0.45$

Example: $P(A) = \frac{1}{2} \quad P(B) = \frac{1}{8} \quad P(C) = \frac{1}{4}$

Where A, B and are mutually exclusive, Find :

A) $P(A \cup B \cup C)$
B) $P(\bar{A} \cap \bar{B} \cap \bar{C})$

Solution

A) $P(A \cup B \cup C) = P(A) + P(B) + P(C) = \frac{1}{2} + \frac{1}{8} + \frac{1}{4} = \frac{7}{8}$

B) $P(\bar{A} \cap \bar{B} \cap \bar{C}) = P(\overline{A \cup B \cup C}) = 1 - P(A \cup B \cup C) = 1 - \frac{7}{8} = \frac{1}{8}$

Example: A certain family owns two television sets one color and one black and white, let A be the event of color and B is the event of black and white

if $P(A) = 0.4 \quad P(B) = 0.3 \quad P(A \cup B) = 0.5$ Find :

A) Both sets are on
B) The color set on and other is off
C) Exactly one set is on

D) neither set is on

Solution

A) $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.3 - 0.5 = 0.2$

B) $P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.4 - 0.2 = 0.2$

C) $P(A \cap \bar{B}) \cup P(\bar{A} \cap B) = 0.2 + P(B) - P(A \cap B) = 0.2 + 0.3 - 0.2 = 0.3$

D) $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.5 = 0.5$

EXAMPLE: Let $\boldsymbol{P(A) = 0.4}$ $\boldsymbol{P(A \cup B) = 0.6}$ find the value of $\boldsymbol{P(B)}$ which make A and B mutually excusive

Solution

Since A and B are disjoint $P(A \cup B) = P(A) + P(B)$

$$0.6 = 0.4 + P(B) \Longrightarrow P(B) = 0.2$$

Example: There are two traffic lights on the route used by pickup Andropov to go from home to work, let E denote the event that pickup must stop at the first light and F in a similar manner for second light, suppose $\boldsymbol{P(E) = 0.4}$, $\boldsymbol{P(F) = 0.3}$ and $\boldsymbol{P(E \cap F) = 0.15}$ what is the probability that he:

A) Must stop for at least one light?

B) Doesn't stop at either light?

C) Must stop at exactly one light

D) Must stop just at the first light

Solution

A) $P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.4 + 0.3 - 0.15 = 0.55$

B) $P(\bar{E} \cap \bar{F}) = P(\overline{E \cup F}) = 1 - P(E \cup F) = 1 - 0.55 = 0.45$

C) $P(E \cap \bar{F}) \cup P(F \cap \bar{E}) = P(E) - P(E \cap F) + P(F) - P(E \cap F) = 0.4 - 0.15 + 0.3 - 0.15 = 0.4$

D) $P(E \cap \bar{F}) = P(E) - P(E \cap F) = 0.4 - 0.15 = 0.25$

Example: In a hospital, there are 12 nurses and 4 doctors

A) if 4 nurses and 2 doctors to be chosen, haw many possibilities?

B) If a committee contains 6 person, what is the probability that two doctors in this committee?

Solution

الترتيب غير مهم بالسؤال ولذلك نستخدم التوافيق

A) $nc_r = \frac{n!}{r!(n-r)!}$

$$12C_4 = \frac{12!}{4!\,(12-4)!} = \frac{12 \times 11 \times 10 \times 9}{4 \times 3 \times 2 \times 1} = 495$$

$$4C_2 = \frac{4!}{2!\,(4-2)!} = \frac{4 \times 3}{2 \times 1} = 6$$

All possibilities are 495.6=2970

B) $P(A) = \frac{n(A)}{N} = \frac{12C_4 \times 4C_2}{16C_6} = \frac{2970}{8008}$

Example: Let A, B, and C $\in 2^{\Omega}$, where

$\mathbf{P(A-B)=0.15}$, $\mathbf{P(B-A)=0.30}$, $\mathbf{P(C-A)=0.35}$
$\mathbf{P(A\cap B)=0.10}$, $\mathbf{P(A\cap C)=0.15}$, $\mathbf{P(B\cap C)=0.20}$
and $\mathbf{P(A\cap B\cap C)=0.05}$ then find

A)

$\mathbf{P(A)}$	$\mathbf{P(A-C)}$	$\mathbf{P(B)}$	$\mathbf{P(C-B)}$
$\mathbf{P(C)}$	$\mathbf{P(B-C)}$	$\mathbf{P(A\backslash B)}$	$\mathbf{P(a\backslash B\cap C)}$

B) If you know that

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

$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Find $P(A \cup B \cup C)$, $P(\bar{A} \cap \bar{B} \cap \bar{C})$

C) Are the events A, B and C independent

Solution

A)

$P(A-B) = P(A) - P(A \cap B) \Rightarrow 0.15 = P(A) - 0.10 \Rightarrow P(A) = 0.25$

$P(A-C) = P(A) - P(A \cap C) = 0.25 - 0.15 \Rightarrow P(A-C) = 0.10$

$P(B-A) = P(B) - P(B \cap A) \Rightarrow 0.30 = P(B) - 0.10 \Rightarrow P(B) = 0.40$

$P(C-A)=P(C)-P(C\cap A)\Longrightarrow 0.35=P(C)-0.15\ \Longrightarrow P(C)=0.50$

$P(B-C)=P(B)-P(B\cap C)\Longrightarrow P(B-C)=0.4-0.2=0.2$

$P(C-B)=P(C)-P(C\cap B)=0.5-0.2\ \Longrightarrow P(C-B)=0.3$

$$P(A\backslash B)=\frac{P(A\cap B)}{P(B)}=\frac{0.10}{0.40}=0.25$$

$$P(A\backslash B\cap C)=\frac{P(A\cap B\cap C)}{P(B\cap C)}=\frac{0.05}{0.20}=0.25$$

B)

$P(A\cup B\cup C)=0.25+0.40+0.5-0.10-0.15-0.20+0.05=0.75$

$P(\bar{A}\cap\bar{B}\cap\bar{C})=P(\overline{A\cup B\cup C})=1-P(A\cup B\cup C)=1-0.75=0.25$

C)

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

$0.05=(0.25)(0.40)(0.5)$

$0.05=0.05\Longrightarrow$ The event are independent

❶ $P(A\cap B)=P(A)\bullet P(D)$ independent

$P(A\cap B)=(0.03)(0.08)=0.0024$

❷ at least one machines $\Longrightarrow P(A\cup B)$

$P(A\cup B)\quad =P(A)+P(B)-P(A\cap B)=0.03+0.08+0.002$
$=0.1076$

❸ only first $\Longrightarrow P(A\cap\bar{B})$

$P(A\cap\bar{B})=P(A)-P(A\cap B)=0.03-0.0024$

❹ $P(A\backslash B)=\frac{P(A\cap B)}{P(A)}=\frac{P(A)\bullet P(B)}{P(A)}=P(B)=0.0800$

The probability of A given B defined as

$$P(A\backslash B)=\frac{P(A\cap B)}{P(B)} \qquad ,P(B)>0$$

Note:

1. If A_1 , and A_2 are mutually exclusive, then

$$P(A_1\cup A_2\backslash B)=P(A_1\backslash B)+P(A_2\backslash B)$$

2. If $P(B)>0)$, then $P(\Omega\backslash B)=\frac{P(\Omega\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$

3. If $B\subseteq A$, then $A\cap B=B$,

Therefore

$$P(A\backslash B)=\frac{P(A\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$$

$$P(B\backslash A)=\frac{P(A\cap B)}{P(A)}=\frac{P(B)}{P(A)}$$

So, $P(B)=P(A)\bullet P(B\backslash A)$

Example: Let $\Omega=\{1,2,3,4,5,6\}$

$$A_1=\{1,5\}\Rightarrow P(A_1)=\frac{1}{3}\ \&\ A_2=\{2,4,6\}\Rightarrow P(A_2)=\frac{1}{2}$$

$$B=\{1,2,3\}\Rightarrow P(B)=\frac{1}{2}$$

$$A_1\cap B=\{1\}\quad\Rightarrow\quad P(A_1\cap B)=\frac{1}{6}$$

$$A_2\cap B=\{2\}\quad\Rightarrow\quad P(A_2\cap B)=\frac{1}{6}$$

So, $P(A_1\cup A_2\backslash B)=P(A_1\backslash B)+P(A_2\backslash B)$

$$=\frac{P(A_1\cap B)}{P(B)}+\frac{P(A_2\cap B)}{P(B)}=\frac{1/6}{3/6}+\frac{1/6}{3/6}=\frac{2}{3}$$

Example: A student is randomly selected from a class where 35% is left hand, and 50% are sophomores, we know that 5% of the class consists of left handed sophomore, Given that a randomly selected student is a sophomore, what is the probability that he is left handed

Solution

Let $P(A) = 0.35$, $P(B) = 0.5$, $P(A \cap B) = 0.05$

Then $P(A \backslash B) = \frac{P(A \cap B)}{P(B)} = \frac{0.05}{0.05} = 0.1$

يمكن استنتاج قواعد أخرى من الاحتمال الشرطي كما يلي

If $B(> 0)$

$$P(A \backslash B) = \frac{P(A \cap B)}{P(B)} \Longrightarrow P(A \cap B) = P(B) \bullet P(A \backslash B)$$

If $A(> 0)$

بالمثل

$$P(B \backslash A) = \frac{P(A \cap B)}{P(A)} \Longrightarrow P(A \cap B) = P(A) \bullet P(B \backslash A)$$

Independent event (الأحداث المستقلة)

Example: In tossing coin two times, let A means get a head on the first flip, B means get a head on the second flip, what the probability that you get a head on the second flip given that you had a head on the first flip?

Solution

$\Omega = \{HH, HT, TH, TT\}$

$$A = \{HH, HT\} \Longrightarrow P(A) = \frac{2}{4} = \frac{1}{2}$$

$$B = \{HH, TH\} \Longrightarrow P(B) = \frac{2}{4} = \frac{1}{2}$$

$$P(A \cap B) = \{HH\} \Longrightarrow P(A \cap B) = \frac{1}{4}$$

$$P(A \backslash B) = \frac{P(B \cap A)}{P(A)} = P(A) \bullet P(B) \Longrightarrow \frac{1}{4} = \frac{1}{2} \bullet \frac{1}{2} = \frac{1}{4}$$

So A and B are independent

Example: In tossing coin three times, let A get a head on first toss and B get a head on the second toss, C is get head in third toss, what is the probability of getting a head of third toss given that the previous two flips were heads?

Solution

$\Omega = \{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT\}$

$A = \{HHH, HHT, HTH, HTT\} \Rightarrow P(A) = \frac{4}{8}$

$B = \{HHH, HHT, THH, THT\} \Rightarrow P(B) = \frac{4}{8}$

$A \cap B = \{HHH, HHT\} \Rightarrow P(A \cap B) = \frac{2}{8}$

$C = \{HHH, HTH, THH, TTH\} \Rightarrow P(C) = \frac{4}{8}$

$P(C \backslash A \cap B) = \frac{P(C \cap B \cap A)}{P(A \cap B)} = \frac{1/8}{2/8} = \frac{1}{2}$

- **Mutually exclusive** $\Rightarrow P(A \cap B) = 0$
- **Independent** $\Rightarrow P(A \cap B) = P(A) \bullet P(B)$

خليك فاكر

Example: Give that a student studied, the probability of passing quiz is o.gg. Given the student did not study, the probability of passing quiz iso.05, Assume that the probability of studding is 0.7. A student flunks the quiz, what is the probability that he is did not study

Solution

$P(\text{pass} \backslash \text{studied}) = 0.99 \quad \Rightarrow \quad P(A \backslash B) = 0.99$

$P(\text{pass} \backslash \text{not study}) = 0.05 \quad \Rightarrow \quad P(A \backslash \bar{B}) = 0.05$

$P(\text{study}) = 0.7 \quad \Rightarrow \quad P(B) = 0.7$

$P(\text{no study} \backslash \text{no pass}) \quad \Rightarrow \quad P(\bar{B} \backslash \bar{A})$

$P(\bar{B} \backslash \bar{A}) = \frac{P(\bar{B} \cap \bar{A})}{P(\bar{B})}$

$P(A \cap B) = P(B) \bullet P(A \backslash B) = 0.7 \times 0.99 = 0.693$

$P(A \backslash \bar{B}) = \frac{P(A \cap \bar{B})}{P(\bar{B})} = \frac{P(A) - P(A \cap B)}{P(\bar{B})} \Rightarrow 0.05 = \frac{P(A) - 0.693}{0.3}$

P(A)=0.708

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.708+0.3-0.693$$

Then $P(\bar{B}\backslash\bar{A})=\frac{P(\bar{B}\cap\bar{A})}{P(\bar{A})}=\frac{P(\overline{B\cup A})}{P(\bar{A})}=\frac{1-P(A\cup B)}{P(\bar{A})}=\frac{1-0.715}{1-0.708}=0.976$

Example: In a certain college 0.25 of students failed math, 0.15 of them sailed in chemistry and 0.10 failed in both

A) If he failed chemistry, what the probability that he failed in math?

B) If he failed in math, what the probability that he failed in chemistry?

C) What is the probability that he failed in math or chemistry?

Solution

$P(M)=0.25$ $\quad P(C)=0.15$ $\quad P(M\cap C)=0.10$

A) $P(M\backslash C)=\frac{P(M\cap C)}{P(C)}=\frac{0.10}{0.15}=\frac{2}{3}$

B) $P(C\backslash M)=\frac{P(C\cap M)}{P(M)}=\frac{0.10}{0.25}=\frac{2}{5}$

C) $P(M\cup C)=P(M)+P(C)-P(M\cap C)=0.25+0.15-0.10=0.30$

BAY'S THEOREM

$$P(A\backslash B)=\frac{P(A)\bullet P(B\backslash A)}{P(A)\bullet P(B\backslash A)+P(\bar{A})\bullet P(B\backslash\bar{A})}$$

$$P(A\backslash B)=\frac{P(A)\bullet P(B\backslash A)}{P(A)\bullet P(B\backslash A)+P(\bar{A})\bullet P(B\backslash A)}$$

Example: In a statistics class it was found that 60% of student attend class on Thursday, from past data it was found 98% of those who went to class on Thursday pass the course, while 20% of those who did not go to class on Thursday passed the course.

A) What percentage of students is expected to pass the course?

B) Given a student passes the course, what is the probability that he attend classed on Thursday

Solution

A_1: student attend class

B_1: student pass the course

A_2: student did not attend

B_2: do not pass the course

A)

$P(A_1)=0.6$ $\quad P(A_2)=0.4$ $\quad P(B_1\backslash A_1)=0.98$ $\quad P(B_2\backslash A_2)=0.2$

$P(B_1)=P(B_1\cap A_1)+P(B_1\cap A_2)$

Since $P(B_1\backslash A_1)=\frac{P(B_1\cap A_1)}{P(A_1)}\Longrightarrow P(B_1\cap A_1)=P(B_1\backslash A_1)\bullet P(A_1)$

$P(B_1\backslash A_2)=\frac{P(B_1\cap A_2)}{P(A_2)}\Longrightarrow P(B_1\cap A_2)=P(B_1\backslash A_2)\bullet P(A_2)$

So $P(B_1)=P(B_1\backslash A_1)\bullet P(A_1)+P(B_1\backslash A_2)\bullet P(A_2)$

$=0.98\times 0.6+0.2\times 0.4=0.668$

B) $P(A_1\backslash B_1)=\frac{P(B_1\backslash A_1)\bullet P(A_1)}{P(B_1\backslash A_1)\bullet P(A_1)+P(B_1\backslash A_2)\bullet P(A_2)}=\frac{0.6\times 0.98}{0.6\times 0.98+0.4\times 0.2}=0.854$

CHAPTER 3

[redacted] **cs 1**

- ❖ Basic
- ❖ Organization

CONCEPT OF RANDOM VARIABLES AND THEIR DISTRIBUTION

Example: In Case of tossing a coin twice, find the random variable x which denotes the number of heads

Solution

$S = \{HH, HT, TH, TT\}$

$X = \{2, 1, 0\}$, Then the distribution

X	0	1	2	$\sum$
P(X)	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{4}{4}$

Example: In Case of tossing a coin twice, find the distribution function (DF) of x

Solution

$$f_x X = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & 0 \le x < 1 \\ \frac{3}{4} & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

ملاحظات

- $0 \le F_x(x) \le 1$
- $\lim_{x \to -\infty} F_x(x) = 0$ & $\lim_{x \to \infty} F_x(x) = 1$

DISCRETE RANDOM VARIABLE AND THEIR DISTRIBUTION

Example: A coin tossed three times, X is discrete random variable denotes the number of head,

Find the probability distribution and mass function

Solution

$\Omega = \{HHH, HHT, HTH, THH, TTT, TTH, THT, HTT\}$

$X = \{0, 1, 2, 3\}$

X	0	1	2	3	$\sum$
P(X)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

$$M.f \begin{cases} 0 & x < 0 \\ \frac{1}{8} & 0 \le x < 1 \\ \frac{1}{8} + \frac{3}{8} = \frac{4}{8} & 1 \le x < 2 \\ \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8} & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

Probability mass function دالة الكتلة الاحتمالية

Let X be a discrete random variable on the probability space $[\Omega, \mathcal{A}, P]$, Then

$P_x : \mathbb{R} \rightarrow [0,1]$ such that

❶ $P_x = P(X = x) \ge 0$ i.e $0 \le P(x) \le 1$

❷ $\sum P_x = 1$

Is said to be probability mass function p.m.f

Example: Let x be a discrete random variable representing the sum of two numbers on throwing two identical balanced dice for one time, The

a) Find the possible values of random variable X
b) Determine the probability mass function
c) Determine the distribution function

Solution

X	2	3	4	5	6	7	8	9	10	11	12
$P_x = P(X + x)$	$\frac{1}{21}$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{3}{21}$	$\frac{3}{21}$	$\frac{2}{21}$	$\frac{2}{21}$	$\frac{1}{21}$	$\frac{1}{21}$
$F_x(x)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{6}{21}$	$\frac{9}{21}$	$\frac{12}{21}$	$\frac{15}{21}$	$\frac{17}{21}$	$\frac{19}{21}$	$\frac{20}{21}$	$\frac{21}{21}$

d) Determine the mean and the variance

Solution

$$mean = \mu = \sum xp(x) = \left(2.\frac{1}{21}\right) + \left(3.\frac{1}{21}\right) + \left(4.\frac{2}{21}\right) + \left(5.\frac{2}{21}\right) + \left(6.\frac{3}{21}\right) + \left(7.\frac{3}{21}\right) + \left(8.\frac{3}{21}\right) + \left(9.\frac{2}{21}\right) + \left(10.\frac{2}{21}\right) + \left(11.\frac{1}{21}\right) + \left(12.\frac{1}{21}\right) = 7$$

$$variance = \sigma^2 = \sum x^2 p(x) - \mu^2$$

$$\sum x^2 P(x) = \left(4.\frac{1}{21}\right) + \left(9.\frac{1}{21}\right) + \left(16.\frac{2}{21}\right) + \left(25.\frac{2}{21}\right) + \left(36.\frac{3}{21}\right) + \left(49.\frac{3}{21}\right) + \left(64.\frac{3}{21}\right) + \left(81.\frac{2}{21}\right) + \left(100.\frac{2}{21}\right) + \left(121.\frac{1}{21}\right) + \left(144.\frac{1}{21}\right)$$

$$= 55.66$$

$$\sigma^2 = \sum X^2 p(x) - \mu^2 = 55.66 - 7^2 = 6.66$$

Example: Consider rolling a balanced die twice, and let X be the maximum of two numbers

a) Determine the probability mass function and distribution function

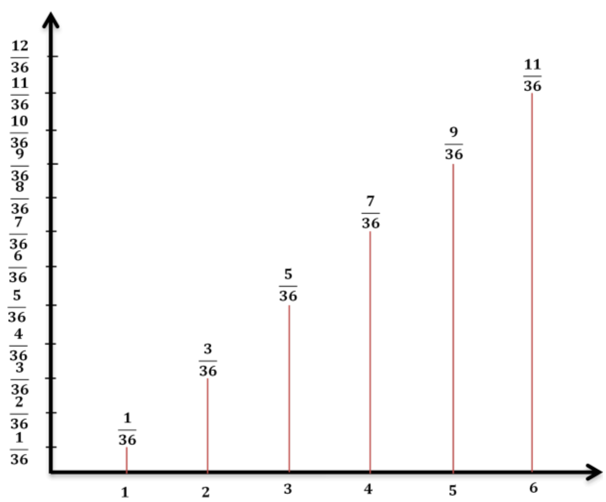
Solution

X	1	2	3	4	5	6
$P_x = P(X = x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$
$F_x(x)$	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	1

b) Sketch the function in part a)

Solution

$$F_x(x) \begin{cases} 0 \rightarrow X < 1 \\ \frac{1}{36} \rightarrow 1 \le X < 2 \\ \frac{4}{36} \rightarrow 2 \le X < 3 \\ \frac{9}{36} \rightarrow 3 \le X < 4 \\ \frac{16}{36} \rightarrow 4 \le X < 5 \\ \frac{25}{36} \rightarrow 5 \le X < 6 \\ 1 \rightarrow X \ge 6 \end{cases}$$



Example: Consider a discrete random variable x with the following probability mass function

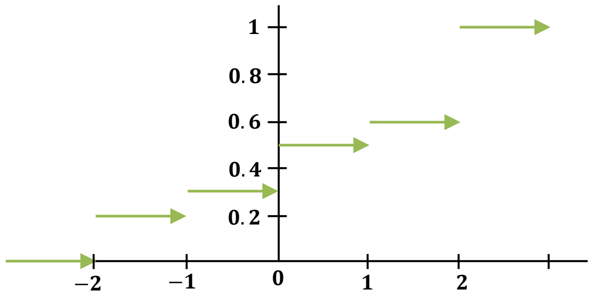
X	-2	-1	0	1	2
P_x=P(X=x)	0.20	0.15	0.15	0.1	0.4

a) Determine the distribution function F_x and draw the p.m.f. and D.F. for this variable

Solution

X	-2	-1	0	1	2
$P_x = P(X = x)$	0.20	0.15	0.15	0.1	0.4
$F_x(x)$	0.20	0.35	0.5	0.6	1

$$F_x(x) = \begin{cases} 0 \rightarrow X < -2 \\ 0.20 \rightarrow -2 \leq X < -1 \\ 0.35 \rightarrow -1 \leq X < 0 \\ 0.50 \rightarrow 0 \leq X < 1 \\ 0.6 \rightarrow 1 \leq X < 2 \\ 1 \rightarrow X \geq 2 \end{cases}$$



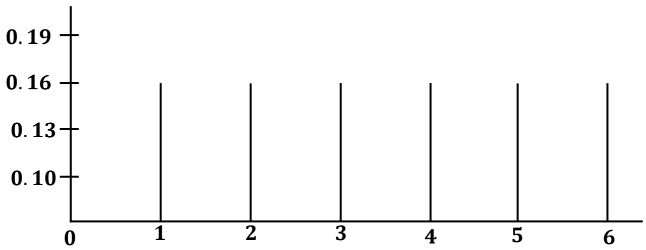
Example: Consider a random variable x, assume the values x1,x2,…x6 with equal probabilities.

Find the probability mass function and distribution function and their graphs.

Solution

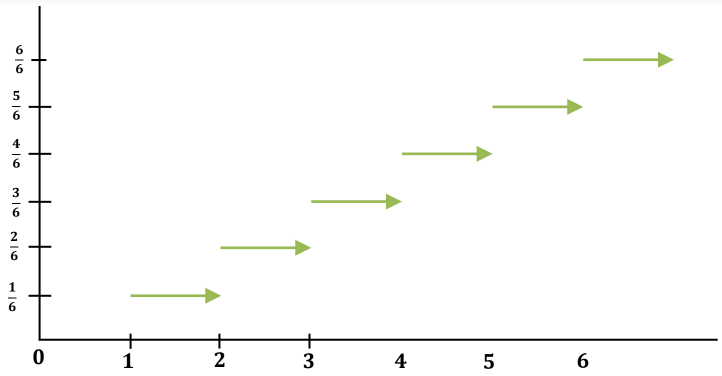
Probability mass function:

X	1	2	3	4	5	6
$P_x = P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



The distribution function $F_X(X)$

$$F_x(x) \begin{cases} 0 \rightarrow X < 1 \\ \frac{1}{6} \rightarrow 1 \leq X < 2 \\ \frac{2}{6} \rightarrow 2 \leq X < 3 \\ \frac{3}{6} \rightarrow 3 \leq X < 4 \\ \frac{4}{6} \rightarrow 4 \leq X < 5 \\ \frac{5}{6} \rightarrow 5 \leq X < 6 \\ \frac{6}{6} \rightarrow X \geq 6 \end{cases}$$



خليك فاكر

- ✓ $mean = \mu = E(X) = \sum XP(X)$
- ✓ $variance = \sigma^2 = E(X) = \sum X^2P(X) - \mu^2$
- ✓ $E(ax + b) = aE(x) + b$
- ✓ $v(a) = 0$
- ✓ $E(a) = a$
- ✓ $v(ax + b) = a^2v(x)$

BIONNOMIAL DISTRIBUTION

$$P(K) = \binom{n}{k} P^K (1 - P)^{n-K}$$

حيث:

n: sample size

p: نسبة النجاح

q: نسبة الفشل

K: المطلوب

mean: $\mu = np$ & $variance = np(1 - p)$

Example: if the mean and variance of binomial are 16 and 18, then

a) Determine the mass function

$$P(K) = \binom{n}{k} P^K (1 - P)^{n-K}$$

$$\mu = np = 16 \Rightarrow n = \frac{16}{p}$$

$$\sigma^2 = np(1 - p) = 8 \Rightarrow \frac{16}{9p} p(1 - p) = 8$$

$$16(1 - p) = 8 \Rightarrow p = 0.5 \qquad \Rightarrow \qquad n = \frac{16}{0.5} = 32$$

$P(K) = \binom{32}{K}(0.5)^K(0.5)^{32-K}$

b) Calculate $\mu(x=0)$ $\quad P(0) = \binom{32}{0}(0.5)^0(0.5)^{32-0} = (0.5)^{32}$

c) Calculate $P(x \geq 2)$

$$P(x \geq 2) = 1 - p(x < 2) = 1 - [p(x=0) + p(x=1)]$$

$$= 1 - \left[(0.5)^{32} + \binom{32}{1}(0.5)^1(0.5)^{31}\right] = 0.99$$

POISSON DISTRIBUTION

$P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0,1,2,\ldots$

$\mu = \lambda \qquad \sigma^2 = \lambda$

Continous Random variables and their distribution

Continuous random variable: is random variable whose set of it possible values are uncountable set.

Example: Life length, temperature, blood pressure, …

P.d.f $\Longrightarrow p(a \leq x \leq b) = \int_a^b f(x)dx$

D.f $\Longrightarrow F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f_x dx$

$f_x(x) = \dfrac{dF_x}{dx}$

أي أن دالة كثافة الاحتمال *(p.d.F)* هي مشتقة دالة التوزيع *D.F*

Example: **Let x be a continuous random variable with distribution function F_x, then find the pdf of x**

a) $F_x(x) = \dfrac{x^4}{16} \qquad 0 < x < 2 \qquad \text{pdf} = \dfrac{d}{dx}F_x = \dfrac{d}{dx}\left(\dfrac{x^4}{16}\right) = \dfrac{4x^3}{16} = \dfrac{x^3}{4}$

b) $F_x(x) = 1 - e^{-5x} \qquad x \geq 0$

$$\text{pdf} = \frac{d}{dx}(F_x) = \frac{d}{dx}(1 - e^{-5x}) = 0 - (-5)e^{-5x} = 5e^{-5x}$$

Expected value $E(x) = \mu = \int_{-\infty}^{\infty} Xf_x\,dx$

Variance $= \sigma^2 = E(x^2) - (E(x))^2$

Example:

$$F_x(x)\begin{cases}\dfrac{1}{b-a} & a \le X \le b\\ 0 & \text{otherwase}\end{cases}$$

Calculate the mean and variance

$$\mu = E(x) = \int_{-\infty}^{\infty} X f_x\, dx = \int_a^b x\left(\frac{1}{b-a}\right) = \frac{1}{b-a}\int_a^b \frac{x^2}{2} = \frac{1}{b-a}\left[\frac{b^2}{2} - \frac{a^2}{2}\right]$$

$$\frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}$$

2) $\sigma^2 = E(x^2) - (E(x))^2 = E(x^2) - (\frac{a+b}{2})^2$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f_x dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a}\int_a^b x^2 dx$$

$$\frac{1}{b-a}\left[\frac{x^3}{3}\right]_b^a = \frac{1}{b-a}(b^3-a^3) = \frac{1}{b-a}\left(\frac{(b-a)(b^2+ba+a^2)}{3}\right) = \frac{b^2+ba+a^2}{3}$$

$$v(x) = \frac{b^2+ba+a^2}{3} - (\frac{b+a}{2})^2$$

The exponential distribution

$$f_x(x)\begin{cases}\lambda e^{-\lambda x} & X \ge 0\\ 0 & x < 0\end{cases}$$

$F_x(x) = 1 - e^{-\lambda x}$

$\mu = \frac{1}{\lambda}$ $\sigma^2 = \frac{1}{\lambda^2}$

Normal distribution

$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{X-\mu}{\sigma})^2}$ $X \sim N(\mu\,,\sigma^2)$

Standard normal distribution

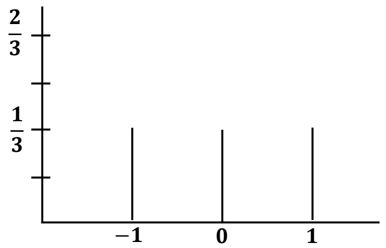
$Z = \frac{X-\mu}{\sigma}$ $Z \sim N(0\,,1)$

Example: Let $[\Omega , \mathcal{A} , P]$ be the probability space of rolling die one time , and x is r.v defined

$$X: \underset{w_1, w_2, w_3, w_4, w_5, w_6}{\Omega = \{1, 2, 3, 4, 5, 6\}}$$

$$w \to X(w) \begin{cases} -1 & \text{for } w_1, w_2 \\ 0 & \text{for } w_3 , w_4 \\ 1 & \text{for } w_5 , w_6 \end{cases}$$

1) What type is this random variable? discrete
2) What the name of this random variable? Uniform
3) Draw the graphical representation of this random variable



4) Determine the distribution function of x

$$F_x(x) \begin{cases} 0 & X < -1 \\ \frac{1}{3} & -1 \le X < 0 < 2 \\ \frac{2}{3} & 0 \le X < 1 \\ 1 & X \ge 1 \end{cases}$$

5) Calculate the mean of X

$$\text{Mean} = E(x) = \mu = \sum xp(x) = -1\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right) + 1\left(\frac{1}{3}\right) = 0$$

DEFINITIONS AND CONCEPTS

Estimator (المقدر):

Is a statistic whose value depends on sample.

Example:

- $\overline{X}$ is an estimator for population mean
- S^2 is an estimator for population variance
- $\widehat{P}$ is an estimator for population proportion

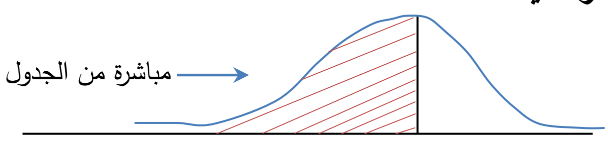
خليك فاكر

المقدر هو إحصاءة محسوبة من العينة وتستخدم للتنبأ بمقدار المعلمة في المجتمع

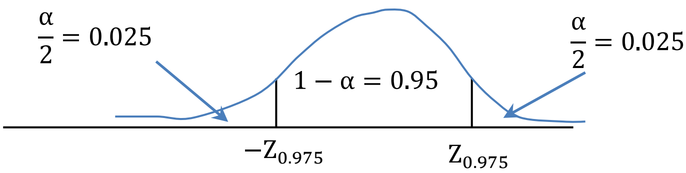
(Z-value)

Example: if $Z_\alpha = 0.57 \Longrightarrow$ find $p(Z < Z_x)$

لاحظ أن وجود Z دائما يعطي المساحة ليسار القيمة



Example: if $\alpha = 0.05 \Longrightarrow 1 - \alpha = 1 - 0.05 = 0.95$



جدول للقيم المشهودة

$100(1-\alpha)\%$	$Z_{1-\frac{\alpha}{2}}$
90%	1.65
95%	1.96
98%	2.33
99%	2.58

Central limit theorem (نظرية النهاية المركزية):

Let Ω be the sample space of a population which described by a r.v X with mean μ and standard deviation σ and $x = (x_1, x_2, \ldots, x_n)$ is a random sample of Ω, then for sufficiently large sample size of x. the sampling distribution of $\bar{X}$ follows a normal distribution with mean $\mu_{\bar{X}} = \mu$a and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ (standard error)

Example: let X is r.v with normal distribution has mean μ=3 and standard deviation $\sigma = 2.25$ is a sample from this population, we find:

- The mean of the sample size $\mu_{\bar{Z}} = \mu = 3$
- The standard deviation of the sample size $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{225}} = 0.17$
- The probability of the sample average exceeds

$$p(\bar{X} \geq 3) = p(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}})$$

$$p(Z \geq \frac{3-3}{0.17}) = p(Z \geq 0) = 0.5$$

Sampling distribution of the sample proportion

P: النسبة في المجتمع　　　　　$\hat{P}$: النسبة في العينة

$\mu_{\bar{X}}$　　　　　$\sigma_X = \sqrt{\frac{P(1-P)}{n}}$

$n \geq 30$　　*or*　　$np \geq 5$　　$n(1-P) \geq 5$

Example: let $\hat{P} = 0.20$　　　　$n = 225$

According to the central limit theorem, we note the sampling distribution of P approximate the normal distribution, so

$\mu_{\bar{X}} = \hat{p} = 0.20$ **and**

$$\sigma_{\bar{X}} = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = \sqrt{\frac{0.20(0.80)}{225}} = 0.0267$$

For $P(P < 0.25) = P(\frac{P-\mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{0.25-0.20}{0.0267})$

$P(Z < 1.87) = 0.9693$

ESTIMATION OF THE POPULATION MEAN

Point estimate (نقطة التقدير):

Is the a summarization of the sample by a single number is an estimate of the population parameters.

خليك فاكر

نقطة التقدير هي إما متوسط القيمة $\bar{X}$ أو النسبة في العينة $\hat{P}$

Interval estimation (فترة التقدير):

Is the interval that predicted to contain the parameter

Confidence interval (فترة الثقة):

هي تلك الفترة التي تؤمن بأن معلمة المجتمع تقع فيها

$$C.I = \bar{X} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

خليك فاكر

هناك حالات يكون الانحراف المعياري للمجتمع غير معلوم والجدول التالي هو تلخيص لكل الحالات

Population	Sample size	Standard deviation	Confidence interval
Normal	Any size	Known	$\bar{X} \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$
Any pop	$n \geq 30$	Known	$\bar{X} \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$
Any pop	$n \geq 30$	Unknown	$\bar{X} \pm Z_{1-\alpha/2}\frac{S}{\sqrt{n}}$

Interval estimation (فترة التقدير):

Is the maximum error (σ_μ)

$$\sigma_\mu = \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Example: the following are measurements for the drying time

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	28
4.4	40	5.2	3.0	4.8

Suppose that population standard deviation is known as $\sigma = 0.96$

A) Find 99% confidence interval

B) The marrying of error

Solution

A) $\alpha = 1 - 0.99 = 0.01 \Rightarrow 1 - \frac{\alpha}{2} = 1 - \frac{0.01}{2} = 0.995$

نبحث عن المساحة الأخيرة تحديد في جدول Z

$$Z_{0.995} = \frac{2.57+2.58}{2} = 2.575$$

$C.I = \bar{X} \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ $\quad\quad \sigma \Rightarrow known \quad\quad n \Rightarrow any\ sample\ size$

يمكن حساب متوسط العينة من خلال البيانات المعطاة

$$\bar{X} = \frac{\sum X}{n} = \frac{57}{15} = 3.8$$

$$C.I = 3.8 \pm 2.575 \times \frac{0.96}{\sqrt{15}} = 3.8 \pm 0.638 = (3.162\ , 4.438)$$

B) Merging of error $\sigma_\mu = \pm Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = \pm 0.638$

Example: a random sample of 120 students from a large university yields mean GPA 2.71, with sample standard deviation 0.51 construct a 90% confidence interval

Solution

C) σ unknown $\quad n = 120 \geq 30$ large

$$C.I = \bar{X} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$90\% \Rightarrow \alpha = 0.10 \Rightarrow Z_{1-0.10/2} = Z_{0.95}$$

$$Z_{0.95} = \frac{1.64+1.65}{2} = 1.645$$

$$C.I = 2.71 \pm 1.645 \times \frac{0.51}{\sqrt{120}} = 2.71 \pm 0.0766 = (2.63\,, 2.75)$$

Example: thirty-six cars of the same model are driven the same distance, the gas mileage for each is recorded, the result give $\bar{X} = 18$ with standard deviation $S = 3$, give a 95% confidence interval

Solution

$\bar{X} = 18$	$S = 3$	$n = 36$
σ unknown	$n = large \geq 30\,, 50$	

$$C.I = \bar{X} \pm Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$95\% \Rightarrow \alpha = 0.05 \Rightarrow Z_{1-\frac{0.05}{2}} = Z_{0.975}$$

$$Z_{0.975} = 1.96$$

So $C.I = 18 \pm 1.96 \times \frac{3}{\sqrt{36}} = 18 \pm 0.98 = (17.02\,, 18.98)$

نحن نثق بنسبة 95% أن متوسط استهلاك البيانات يقع في تلك الفترة

Determine of the sample size $n = [\frac{\sigma Z_{1-\alpha/2}}{\sigma_{\mu}}]^2$

Estimation of the population proportion:

Confidence interval for population proportion

$$C.I = \hat{P} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = (\hat{P} \pm \sigma_P)$$

Where $\sigma_P = Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$ is merging of error

Example: assume the medical researcher collected a random sample size 1000 from Saudi adults and found 320 of them are diabetics

- Find point estimate
- Find 90% confidence interval
- Find the merging of error
- Comment on the result

Solution

$$\hat{P} = \frac{k}{n} = \frac{320}{1000} = 0.32 \Longrightarrow 1 - \hat{P} = 0.68$$

point estimate is $\hat{P} = 0.32$

$90\% \longrightarrow Z_{1-\alpha/2} = 1.645$

$$C.I = \hat{P} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.32 \pm 1.645\sqrt{\frac{0.32\times0.68}{1000}}$$

$$= 0.32 \pm 0.02427 = (0.344\,, 0.296)$$

The merging of error $\sigma_P = Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 1.645\sqrt{\frac{0.32\times0.68}{1000}} = 0.024$

So we can believe that between 30% and 34% of Saudi adults are diabetic, we trust the 90% if we consider the point estimate 32% , we can sure with 90% that this estimate will be in error by 2.4%

Example: a random sample of 125 individuals working in a large city indicated that 42 dissatisfied with their working conditions construct a 90% lower confidence interval.

Solution

$$\hat{P} = \frac{k}{n} = \frac{42}{125} = 0.336 \quad , \qquad 90 \Longrightarrow Z_{1-\alpha/2} = 1.645$$

Lower limit of $C.I = \hat{P} - Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.336 - 1.645\sqrt{\frac{0.336\times(1-0.336)}{125}}$

$$= 0.2665$$

Example: out of a random sample of 100 students, 82 stated that they were nonsmokers, construct 99%confidence interval

Solution

$90\% \rightarrow Z_{1-\alpha/2} = 2.575$

$$\hat{P} = \frac{k}{n} = \frac{82}{100} = 0.82 \qquad \Longrightarrow 1 - \hat{P} = 0.18$$

$$C.I = \hat{P} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.82 \pm 2.575\sqrt{\frac{0.82\times0.18}{100}}$$

$$= 0.82 \pm 0.0989 = (72.1\%\,,91.9\%)$$

Example: The New york times reported that a poll indicated that 46% of population was in favor of the way that president bush was handling the economy, with margin of error of ± 3 , what does this mean? Haw many people were questioned?

Solution

إذا لم يعطى مستوى الثقة في السؤال نستخدم درجة ثقة 95%

$$95\%C.I = \hat{P} \pm Z_{1-\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = 0.46 \pm 1.96\sqrt{\frac{0.46\times0.54}{n}}$$

$$\Longrightarrow 1.96\sqrt{\frac{0.46\times0.54}{n}} \Longrightarrow n = 1.96^2\,\frac{0.46\times0.54}{0.03^2} = 0.03$$

About 1060 were sampled and 46 percent were favor bush

Sample size determining when estimating P

$$n \geq [\frac{Z_{1-\alpha/2}}{\sigma_P}]^2\hat{P}(1-\hat{P})$$

When the value of $\hat{P}(1-\hat{P})$is mot known it is equal 0.25 and hence

$$n \geq 0.25[\frac{Z_{1-\alpha/2}}{\sigma_P}]^2$$

Example: Haw large a sample is needed to ensure that the maximum error of the 95% confidence interval estimate of σ_P less than 0.01?

Solution

لاحظ أنه لم يعطي قيمة $\hat{P}(1-\hat{P})$

ولذلك يتم التفويض عنها بـ $0.5\times0.5 = 0.25$

$$n \geq 0.25[\frac{Z_{1-\alpha/2}}{\sigma_P}]^2 = 0.25[\frac{1.96}{0.01}]^2 = 9604$$

ولذلك لابد من اختيار عينة على الأقل من 9604 شخص لتأكيد أن 95% مستوى ثقة على الأكثر لن يزيد عن 0.01 وكلما زاد حجم العينة يقل حجم الخطأ.

INTRODUCTION TO HYPOTHESES TESTING

Statistical Hypothesis (الغرض الإحصائي):

هو فرض محدد لحالة معينة قد يكون صحيح وقد يكون خاطئ

أنواع الفروض

فرض العدم: Null Hypothesis H_0

فرض البديل: Alternative Hypothesis H_1

Test statistic

σ known normal

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

σ known any pop

$n \geq 30$ $\quad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

σ unknown any pop

$n \geq 30$ $\quad Z = \frac{\bar{X} - \mu}{S/\sqrt{n}}$

Type of error

Type I error

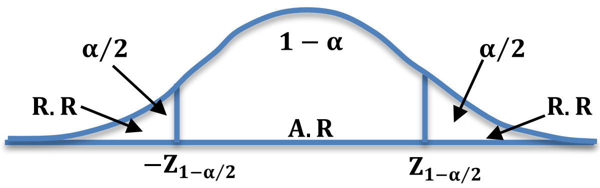
Reject Ho/While H_0 is true

Type II error

accept Ho/While H_0 is

خليك فاكر

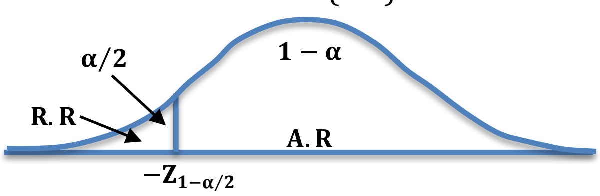
إذا الفرض البديل $(\mu \neq \mu_0)$ يكون الاختيار من طرفين Two-tailed



- إذا وقعت قيمة الـ T.S في المنتصف نقبل الفرض العدم H_0 وإذا وقعت في الأطراف نرفض فرض العدم.
- إذا كان الفرض البديل $\mu > \mu_0$ ⟸ طراف واحد يمين.
- إذا كان الفرض البديل $\mu < \mu_0$ ⟸ طراف واحد يسار.

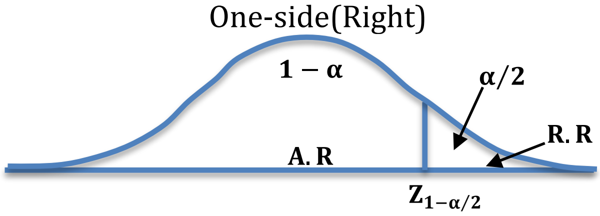
$H_1: \mu < \mu_0$

One-side(left)



$H_1: \mu > \mu_0$

One-side(Right)



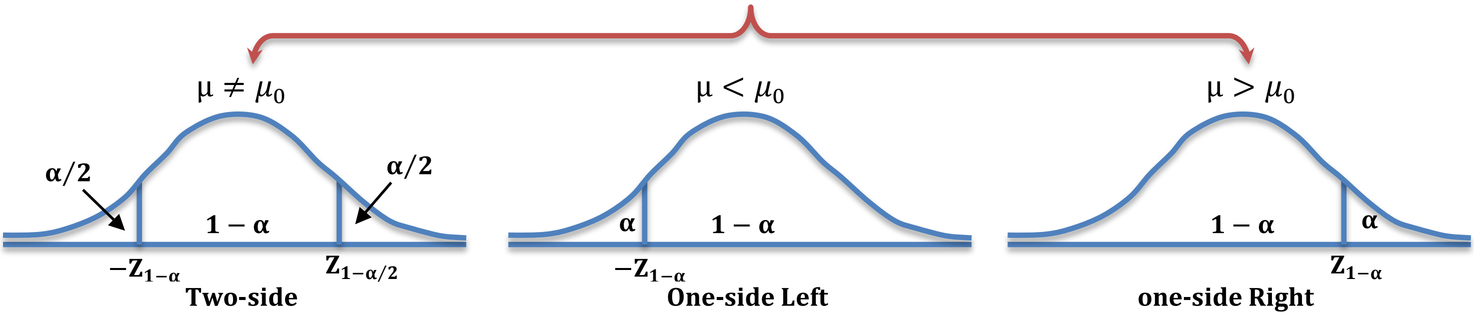
Hypothese testing for the population mean

خطوات تحديد الفروض

تحديد الفروض فرض العدم وفرض البديل

فرض العدم دائما يساوي متوسط القيمة المعطاة $H_0: \mu = \mu_0$

الفرض البديل يكون احتمال من ثلاث حسب معطيات السؤال



تحديد قيمة test statistic حسب معطيات كل سؤال

$$Z_0 = \begin{cases} \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} & \sigma \text{ known} \\[2ex] \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} & \sigma \text{ unknown} \end{cases}$$

تحديد المنطقة الحرجة critical region وهي منطقة الرفض كالتالي:

- إذا وقعت قيمة T.S في منطقة الرفض نرفض H_0 ونقبل H_1
- إذا وقعت قيمة T.S في منطقة القبول نقبل H_0 ونرفض H_1

P-value

يمكن حساب قيمة P-vale من قيمة Z المحسوبة كما يلي:

$H_1 = \mu \neq \mu_0$	$H_1 = \mu < \mu_0$	$H_1 = \mu > \mu_0$
$P(Z > \lvert Z_0 \rvert)$	$P(Z > -\lvert Z_0 \rvert)$	$P(Z > Z_0)$

- ❖ إذا كانت قيمة $\alpha \leq P-\text{value}$ نرفض H_0
- ❖ إذا كانت قيمة $\alpha \geq P-\text{value}$ لا نرفض H_0 وهذا لا يعني بالضرورة قبول H_0

Example: Suppose an editor claims that mean time to write a textbook is at most is month , a sample of 16 text book selected and it is found the mean time was 12.5 , let the standard deviation is known to be 3.6 , use 0.025 significance level.

Would you conclude the editor's claim is true?

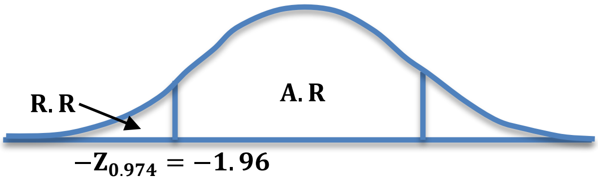
Solution

$\sigma = 3.6$ $\bar{X} = 13.5$ $n = 16$

▶Hypothesis $H_0: \mu = 15$ $H_1: \mu < 15$

▶T.S $Z_0 = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{13.5-15}{3.6/\sqrt{16}} = -1.67$

▶Critical region



الاختبار طرف واحد يسار تبعا لفرض البديل

▶Decision

قيمة T.S تقع في منطقة A.K ولذلك لا نرفض H_0 وهذا يعني أنه لا توجد معلومات كافية تقود إلى رفض H_0 وبالتالي يعتبر هذا دليل كافي لقبول فرض العدم

▶P-value $= P(Z < -|Z_0|) = P(Z < -1.67) = 0.047 > \alpha$

So, we accept H_0

Example: Suppose we would like to determine if the typical amount spent per customer for dinner at new restaurant is more than 200$. A sample of 49 customers was rwndomly selected and the average was 22.6 , assume that the standard deviation to be 2.5 , using 0.02 level of significance , would we conclude the typical amount spent more than 20$?

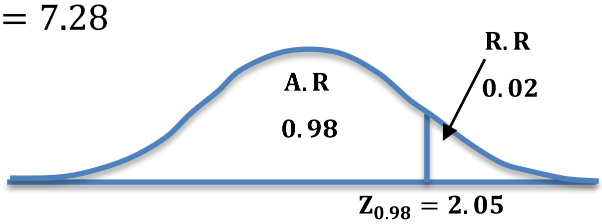
Solution

$\sigma = 2.5$ $\bar{X} = 22.6$ $n = 49$

▶Hypothesis $H_0: \mu = 20$ $H_1: \mu > 20$

▶T.S $Z_0 = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{22.6-20}{2.5/\sqrt{49}} = 7.28$

▶Critical region



اختبار طرف واحد يسار تبعا للفرض البديل $H_1: \mu > \mu_0$

▶Decision

حيث أن قيمة T.S تقع في منطقة الرفض فإن القرار هو رفض H_0 وقبول H_1

▶P-value $= P(Z > Z_0) = P(Z > 7.28) = 0 < \alpha = 0.02$

وبالتالي يوجد دليل كاف لرفض H_0 ويمكن استنتاج أن المبلغ الذي ينفقه العميل أكبر من 20$

Example: Test the claim that on average are three T.V sets on each U.S.home , assume $\sigma = 1$ a sample of 100 households are taken and found the average to be 3.2 can you conclude that this claim is false at $\alpha = 0.05$?

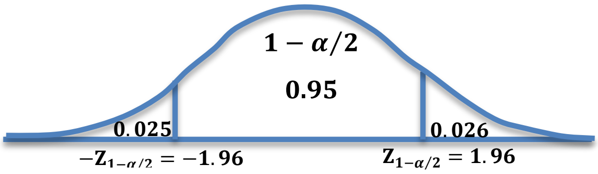
Solution

$\sigma = 1$ $\qquad$ $\bar{X} = 3.2$ $\qquad$ $n = 100$

►Hypothesis $H_0: \mu = 3$ $\qquad$ $H_1: \mu \neq 3$

►T.S $Z_0 = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{3.2-3}{1/\sqrt{100}} = 2$

►Critical region



اختبار من طرفين تبعا لفرض البديل $H_1: \mu \neq \mu_0$

►Decision

حيث أن قيمة Z تقع في المنطقة الحرجة فإنه يتم رفض H_0

►P-value $= 2P(Z > -|Z_0|) = 2P(Z > 20) = 0.045 < \alpha = 0.05$

So, we accept H_0

Hypothese testing for the population propotion

نفس خطوات الاختبار السابقة ما عدى حساب قيمة ~ المحسوبة وتكون كما يلي:

$$\mathbf{T.S:}\quad \mathbf{Z_0} = \frac{\mathbf{\hat{P} - P_0}}{\sqrt{\frac{\mathbf{P_0(1-P_0)}}{\mathbf{n}}}}$$

Example: A computer manufacture claims that at most 2 percent are defective , a sample of 400 of these chips taken if there are 13 defective , does this disprove at 5% level of significance the manufacturer's claim?

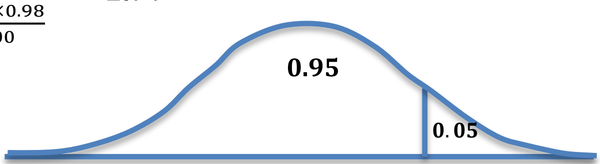
Solution

$P_0 = 0.02$ $\quad \hat{P} = \frac{13}{400} = 0.0325$ $\quad \alpha = 0.05 \quad \Longrightarrow Z_{1-\alpha/2} = 1.645$

▶Hypothesis $\mathrm{H_0: P = 0.02}$ $\quad \mathrm{H_1: P > 0.02}$

▶T.S $Z_0 = \frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.0325-0.02}{\sqrt{\frac{0.02\times0.98}{400}}} = 1.79$

▶Critical region



▶Decision

Since test statistic falls in critical region, so, we reject $\mathrm{H_0}$

▶P-value $= \mathrm{P(Z > 1.79) = 0.0367 = 0.05}$

Then, we reject $\mathrm{H_0}$

Example: Historical data indicate that 4 percen of components are defective if a random sample of 500 items indicated 16 defective (3.2 percent) is this evidence at 5% to conclude that a change has occured

Solution

$P_0 = 0.04$ $\quad n = 500$ $\quad \hat{P} = \frac{16}{500} = 0.032$ $\quad \alpha = 0.05$

$Z_{1-\alpha/2} = Z_{0.975} = 1.96$

▶Hypothesis $\mathrm{H_0: \mu = 0.04}$ $\quad \mathrm{H_1: \mu \neq 0.04}$

▶T.S $Z_0 = \frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.032-0.04}{\sqrt{\frac{0.04\times0.96}{500}}} = -0.913$

▶Critical region

Decision

Test statistic does not fall in critical region , so we cannot reject $\mathrm{H_0}$