Statistics, lecture 3:

*Measures of central tendency:

1. The Mean "done before"

2. The mode:

The value that occurs mostly or with highest frequency

e.g: Find the mode for:

i) 2, 7, 5, 11, 5

mode = 5 "unimodal"

ii) 2, 7, 5, 11, 5, 2

Bimodal: 2 and 5

ملاحظة: قيمة واحدة الأكثر تكراراً: Unimodal

قيمتين اكثر تكراراً: Bimodal

ثلاث قيم اكثر تكراراً: Trimodal

العديد من القيم اكثر تكراراً: Multimodal

iii) 2, 7, 5, 11, 3, 8

No mode

iv)

S	L		
1	1	1	2
2	1	2	2
3	1	2	3

Key: 1|2 represents 12

Bimodal: 11 and 22

v)

X	1	2	3	4	5
f	3	8	7	2	5

Mode: 2 "الرقم الذي له اعلى تكرار"

vi)

I	0-4	5-9	10-14	15-19
f	3	8	7	2

Modal class: 5-9

mode: $\frac{5+9}{2} = 7$ "القيمة في المنتصف"

3. The Median:

i) For raw data and steam and leaf-diagram:

The middle value arranging data in ascending order.

ترتيب تصاعدي

eg) Find the median for:

i) 2, 7, 5, 11, 3, 6, 10

2, 3, 5, 6, 7, 10, 11

Q2 = 6

لاحظوا: عدد القيم فردي، فالجواب احد المشاهدات

آلية الحل: حذفنا كل قيمتين معاً بعد الترتيب والقيمة المتبقية هي الوسيط

ii) 2, 3, 5, 6, 8, 10, 10, 12

2, 3, 5, 6, 8, 10, 10, 12

$Q2 = \frac{6+8}{2} = 7$

لاحظوا: عدد القيم زوجي، تبقى قيمتين في المنتصف فتكون قيمة الوسيط هي حاصل جمعهم مقسوم على 2 والجواب ليس من ضمن المشاهدات

مشكلة الـ Mode انه ليس unique اذ انه قد يكون موجود وقد يكون له اكثر من قيمة يكون الـ Mean والـ Median

الـ Mode دوماً يكون احد المشاهدات اما الـ Mean والـ Median فليس من الضروري ان ينطبق عليها ذلك.

الممسوحة ضوئيا بـ CamScanner

ملاحظة: الطريقة السابقة في الحل جيدة لكنها غير عملية عند وجود عدد كبير من القيم، لذلك نتابع طريقة أخرى.

Note: $\frac{n}{2} \rightarrow$ n: عدد القيم

Fraction → next integer
"اذا كان ناتج القسمة كسر، فدائماً 3.1 فالوسيط هو القيمة الرابعة"

Whole no. (k) → $\frac{k^{th} + (k+1)^{th}}{2}$

ii) For frequency distribution:-
Same as the previous case but after finding the comutative frequency (c.f)
where $n = \Sigma f$
e.g) Find the median for:

X	1	2	3	4	5	Sum
f	3	8	7	2	5	25
c.f.	3 (1-3)	11 (4-11)	18 (5-18)	20 (19-20)	25 (21-25)	

$\frac{n}{2} = \frac{\Sigma f}{2} = \frac{25}{2} = 12.5 \rightarrow 13^{th}$ value

$Q_2 = 3$

iii) 5, 5, 8, 10, 12, 12, 13, 13, 15

$\frac{n}{2} = \frac{2}{2} = 4.5 \rightarrow 5^{th}$ value

$Q_2 = 12$

ii)

X	10	20	30	40	50
f	3	8	7	2	2
c.f.	3 (1-3)	11 (4-11)	18 (5-18)	20 (19-20)	22 (21-22)

$\frac{n}{2} = \frac{22}{2} = \frac{11^{th} + 12^{th}}{2} = \frac{20+30}{2}$

$Q_2 = 25$

iv) 5, 7, 7, 9, 10, 10, 12, 14

$\frac{n}{2} = \frac{8}{2} = 4$

$Q_2 = \frac{4^{th} + 5^{th}}{2} = \frac{9+10}{2} = 9.5$

e.g) For the following orderd set of data
a, 2, 3, 5, b, 9, 10, c
the Mean is: 7
the Mode is: 2
the Median is 6, Find a, b, c

a = 2 اسهل شيء هو الـ Mode اذ ان قيمته 2 وبالتالي وذلك بناءً على ان الارقام مرتبة

$\frac{5+b}{2} = 6 \Rightarrow b = 7$

$\bar{x} = \frac{2+2+3+5+7+9+10+c}{8} = 7 \Rightarrow c = 18$

v)

S	L
1	1 1 2
2	1 2 2
3	1 2 2

Key: 2|3 represents 23

$\frac{n}{2} = \frac{9}{2} = 4.5 \rightarrow 5^{th}$ value

$Q_2 = 22$

"طريقة تكوين Steam and leaf تكون البيانات مرتبة جاهزة"

الممسوحة ضوئيا بـ CamScanner

iii) For grouped frequency distribution

Q_2: $(\frac{n}{2})^{th}$ value after finding the comutative frequency (c.f.) and the upper real boundaries (U.R.B) then we use the linear interpolation method wher $n = \Sigma f$

e.g)

real boundaries: -0.5, 4.5, 9.5, 14.5, 19.5

I	0-4	5-9	10-14	15-19
f	3	8	7	2
c.f.	3	11	18	20
URB	4.5	9.5	14.5	19.5

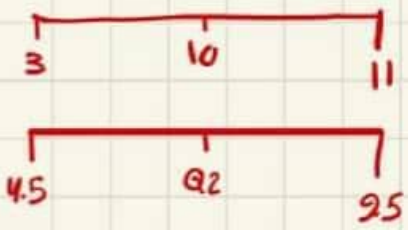
$\frac{n}{2} = \frac{20}{2} = 10^{th}$ value

↓

11 و 3 بين.

بما أن Median إذا

4.5 - 9.5 بين



$\frac{10-3}{11-3} = \frac{Q_2-4.5}{9.5-4.5}$

$Q = 8.875$

e.g)

I	$0 \le x \le 6$	$6 \le x \le 12$	$12 \le x \le 18$	$18 \le x \le 24$
f	2	8	7	3
c.f.	2	10	17	20
U.R.B	6	12	18	24

$\frac{n}{2} = \frac{20}{2} = 10^{th}$ value

$Q_2 = 12$

H.W:

x	1	2	3	4	5
f	8	6	k	3	5

If the mode is 1 and the median is 3, Find k.

Sol) $k < 8 \Rightarrow$ Mode is 1

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

8 6 k 3 5

14 + 6 + = 14

$k = 7$

الممسوحة ضوئيا بـ CamScanner

Data come in the following types:-

i) Raw data :

2, 5, 7, 11, 9, 10, 15

ii) Frequency distribution:

X	1	2	3	4	5
f	3	8	7	2	5

افضل من ال raw عندما يكون عدد البيانات كبير .

iii) Grouped frequency distribution

I	10-20	21-30	31-40	41-50
f	3	4	5	7

ملاحظة :- هذه الطريقة مفيدة جداً في حالة تواجد Continuous data .... فمثلاً لو قاس احد الاشخاص طوله ووجد انه 170 cm ، غالباً سيكون طوله الفعلي يتراوح بين 170.5 - 169.5 " حسب دقة الخطأ في القياس " بالتالي اصبح لدينا فترة

iv) Steam and leaf diagram

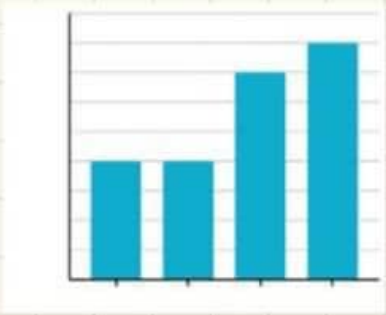
S	L
1	1 1 2
2	1 2 2
3	1 2 3

Key : 1|2 represents 12

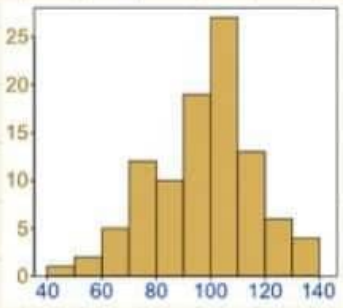
* Graphical representation

i) Pie chart

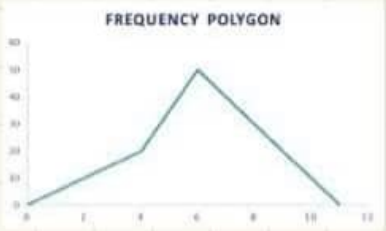
2- Bar graph



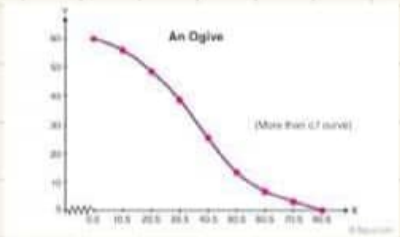
iii) Histogram



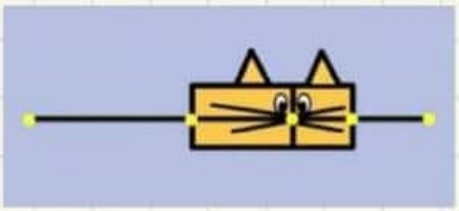
iv) Polygon



v) Ogive (c.f curve)

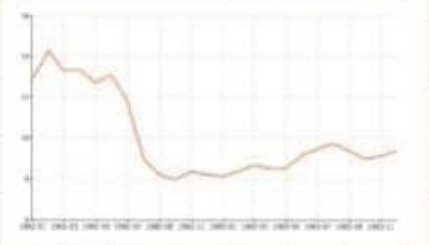


vi) Box and wiskers plot

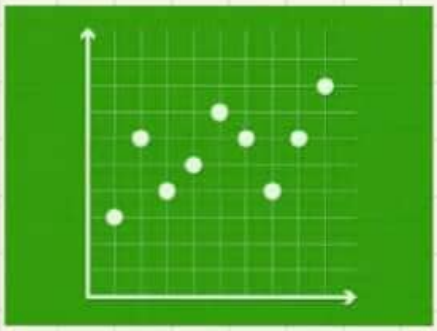


الممسوحة ضوئيا بـ CamScanner

vii) Time series graph



viii) Dot plot



* نعود الى موضوعنا "مقاييس النزعة المركزية" ، تحدثنا سوياً عن الوسط الحسابي ، اذاً ما اهمية الوسيط والمنوال ؟ انظروا للمثال:

e.g) 20, 20, 20, 20, 0
↓
outlier

$\bar{X} = 16$

هذه القيمة تعني ان مجمّع العلامات حول الـ 16 على الرغم من ان جميع الطلبة علاماتهم كانت واحدة وهي صاحب العلامة صفر، اذاً نلاحظ ان وجود outlier يسبب Miss-leading وان الوسط الحسابي يتأثر به

$Q_2 = 20$ لاحظوا ان الوسيط لا يتأثر بال outlier بالتالي نستنتج ان عند وجود outlier يكون الوسيط افضل من الوسط الحسابي .

Qualitative data أمّا المنوال فهو افضل عند وجود

A	B	C
3	10	2

9

* Measures of variation:-

هي مقاييس التشتت ، تقيس مدى تقارب او تباعد البيانات عن بعضها البعض

14, 15, 16 ; $\bar{X} = 15$

10, 15, 20 ; $\bar{X} = 15$

على الرغم من ان مجموعتي البيانات لهما نفس الوسط الحسابي الا ان بيانات المجموعة الاولى اكثر تقارباً بالتالي ستختلف طريقة المعالجة لكل مجموعة

i) The Range, R = Max - Min

سهل و يُعطي فكرة سريعة لكن مشكلته ان يتأثر بال outliers

ii) The inter-quartive range "IQR"

IQR: better when we have an outlier

Q: Which of the following is the best measure of variation

2, 3, 5, 1000

Sol) IQR

iii) The IPR

iv) The standard deviation

v) Variance

10

الممسوحة ضوئيا بـ CamScanner