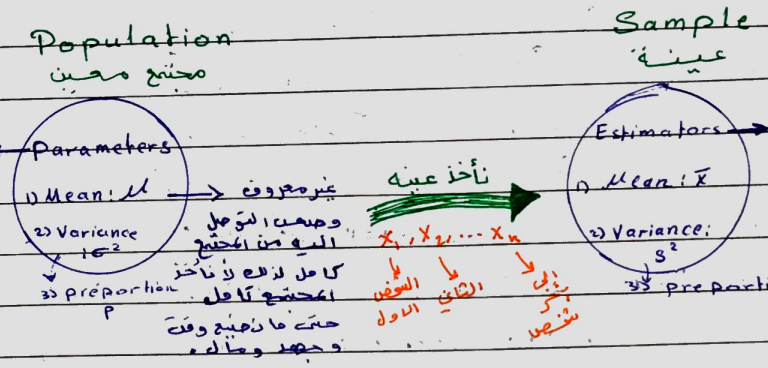
Principles of Statistics

lecture "1"

فرضًا بدي أعمل دراسة على مجتمع معين (مثلًا الأردن) مثلًا كم متوسط دخل الفرد في الأردن.



تُحتم تعريف لـ Parameters في الكتاب للتوضيح [Statistic] Examples هنا

* طريقة أخذ العينة :-

طبيعي اوخدها بالتركيز فقط من منطقة وخلها عالي او التوجه لمنطقة وخلها فيلي جداً (ثانية) وإذا الطريقة دايما (course) موجودة . ع نوكز على موضوع العينة العشوائية [Random of sample] يعني عينة من كل جزء بحسب نسبتها .

ضابط Mean نبع العينة $\bar{x}$ لتقدير Mean للمجتمع وهاي متساويان ، وإذا Population Mean و Estimator هي sample mean فـ . (error) بيكون في خطأ

• أحد فروع نوعين من [Branches of statistics] وهي [descriptive statistics] (جمع البيانات ← تنظيمها ← تلخيصها ← عرضها) ويتكون عنا باب first ثم [descriptive statistic] هو الاستنتاج اللي بنطلع منه بالاعتماد على [Inferential statistic] ع نتكلم عنها عادة . final . وبنا Propelity الاحتمالية ع يكون جزء منا بـ first و Second

• بدرجع البيانات بيتم تصنيفها ، فأنواع البيانات

لـ Estimators قبل حساب

*Types of data :

[1] Qualitative ⇒ يعتمد على النوع Quality — مدخلات نوعي غير عددية

[2] Quantitative ⇒ كمي — مدخلات عددية

Quantitative و Qualitative يعني سؤال صنف الذي

كم سيارة تدخل على البلد ؟ كمي // شو هم انواع السيارات اللي بتدخل : نوعي

نوعية ← له وصف

عددية ← كمية

عادة أعداد صحيحة — عادة بكون كسور

(Quantitative) → Discrete [countable] — Count — قابل للعد

→ Continuous [measurable] — measurements — قابل للقياس

لكن باختلاف الأجهزة المستخدمة للقياس وهو بكون فيه error حسب دقة الجهاز المستخدم

فمثلاً طول الطول 170 cm ممكن يكون 170.02 أو 170.005 أو 169.98 وهكذا

من هنا مهمة جديدة في Statistic اسمها:

الأطوال الآتية للشعبة مثلاً لو أخذت عدد معين 170, 171, 172, ...

تعتبر (Discrete data) . هل الأطوال قابلة للقياس؟ نعم / لكنها أعداد صحيحة. بعض الأسئلة ممكن الانتبه إلى ...

• في شيء مهم مشروع الـ بالصفحة التفاعلية وهو (level of measurement)

* Measurement level

1. Nominal [اسمي] — labelling فقط أسماء زي ذكر male / أنثى female — Qualitative only

2. Ordinal [رتبي] — Order — يكون فيه ترتيب — Qualitative أو Quantitative

مثل ترتيب نتائج الطلاب بالمعدل من الأول 1st والثاني 2nd والثالث 3rd

مثل ترتيب T-shirt small < med. < large

* متر القماش كذا 36 / 37 / 38 ليس له فيه عدد فقط لغير المتر (القماش)

* 3. Interval — في ترتيب لكن يوجد فيه عدد والفروقات لها معنى

درجة الحرارة أمس 2 واليوم 6 فيه ترتيب من الصغير للكبير

لكن قياس الحديث 32 / 33 / 34 ليس العرف ما له معنى لكن متلاً ضيف جديد صغير

8 / 8 / 10 — اليوف 4 درجات ناقص الفرق 4 درجات له معنى

4. Ratio — أيضاً يوجد فيه ترتيب والفروقات لها معنى لكن النقطة هو أن الصفر بحالة Interval متلاً درجة الحرارة صفر هم اصطلح اعتبره صفر واقل بالك

أما Ratio زي الأطوال لما الطول صفر يعني انعدام الطول يعني صفر متعلق

COLLEGE

small letter capital..

ك ، ك : سيجا حرف لاتيني .

• يكمل الفرق بين Interval و Ratio .

Ratio نسبة فاذا كان معك 10 و 20 و 90 وبعض هذه الاصفار ليس اذا اردت تحويله الى من 1 الى 2 لا تقول باضافة اصفار اقول اضافة الصفر .

• تأتي بالأشكال التالية: The data come in one of the following Forms

بيانات الخام

1) Raw data → 2, 7, 9, 3, 2, ...

2) Frequancy distribution → تكرار عناصر البيانات (5) (البسط)

x	1	2	3	4	5	sum
f	20	10	15	30	15	

summation $\sum f = n$

3) Group of frequancy distribution:-

Interval لعمل continuous data يفضل بحالة فترات

e.g.

I	0-4	5-9	10-14	15-19	sum
f	10	12	8	7	37

فمثلاً علامات الطلاب التي بجيب من 95 لـ 100 يوخذ A والتي من 80 لـ 94 يوخذ B^+ وهكذا او احياناً لما نستخدم ادوات قياس غير دقيقة 172.2 - 180.3 نضيف في error نضيف القيمة من الطول كذا اي كذا

الجذع ⑤ الورقة

4) stem and leaf diagram:

11, 12, 13, 21, 21, 22, 31, 33, 35

S	L
1	1 2 3
2	1 1 2
3	1 3 5

Key 1|2 represent 12

لاستخراج (كيف نقرأ الـ stem and leaf)

key: 1 | 2 represent 12

S to L

مفتاح قيمة

بينما لو كان

key: 1 | 2 represent 1.2

يعني 1.2 .../2.1/2.1/1.3/1.2/1.2

• بعد التصنيف نصير أحسب Estimators
مقاييس النزعة المركزية
• Measures of central tendancy وين تتجمع البيانات

1] Mean $\bar{x}$ (المتوسط الحسابي) [مثلاً جمعة العلامات وقسمت على عددهم]

2] The Mode (m) المنوال ما إله رمز بس بستعمل [أكثر علامة تكررت] highest frequency
A, B, C, D Qualitative data عادة أفضل جاية
ما منحتهم بتشوف إذا كزتكرارًا

3] The Median (Q_2) الوسط، القيمة المتوسطة بالنص [عرفت العلامة اللي بالوسط] middle value
second Quarter
↳ better if we have an outlier

* ليش ما منعتمد فقط على Mean ؟!
لو اعطيتكم علامات طلبة 20, 20, 20, 20, 0
وحسبنا $\bar{x} = \frac{80}{5}$ ← $\bar{x} = 16$ ← بتحس انه علامات الطلاب حول (16).
بصفي تمويه لأن غالبية الطلاب 20 إلا طالب واحد شاذ (0) بسبب ((outlier)). لو أسئلة بصيرا كلهم 20 فبصير الـ average = (20) وليس (16)
فالأفضل عند وجود الـ outlier استخدام الـ Median

20, 20, 20, 20, 0
Median ↑ ↳ outlier

Median = 20 $Q_2 = 20$
لبعض الطلاب اكثر من 20 ولعظم اقل من 20.
في median ما بتأثر بوجود outlier بخلاف ~ ص ~ عن الـ $\bar{x}$
أما الـ mode أفضل مقياس حالات (Qualitative data) منذكر انواع A, B, C, D، بنأخذ منه أفضل بكونه الأكثر تكرارًا وبعدين بنكتبه بصيغة mode

• فرضًا احيائي صف فيه 20 طالب (20 / $\bar{x}$ = 15) وصف تاني (20 / $\bar{x}$ = 15) هل علاماتهم ضروري نفس الشي لأكيد لأ والكلام عنا

14, 15, 16	10, 15, 20
$\bar{x} = 15$	$\bar{x} = 15$
قريبين للوسط	(أكثر تشتت)

COLLEGE

مقاييس التشتت

Measure of variation (4) عددهم

[1] The range مقياس بعاد او فروق البعض

[2] The inter-quartile range (IQR)

[3] Inter-percentile Range (IPR)

[4] The standard deviation سيجما الانحراف المعياري

[5] The variance $(S)^2$

• The range 4-1=3 بعد 1, 2, 3, 4 يوضح اكبر عدد واصغر عدد وفرقهم عن بعض كل ما كان (Range) اكبر كل ما كانوا فروق البعض بين الكل واذا كان في outlier مثل 1, 2, 3, 4, 100 بين الـ Range انهم بعاد بينهم في قيم فروق البعض الا (100) شاذة. لهيك على (IQR) وبتأثر بـ outlier بين بوعز IPR و IQR • data من 50٪ اقل من outlier يوجد (Variance) يستخدموا بالرياضيات العالية. اذا كان بقاربين مجتمعين تشتت اكبر.

Graphical representation • ممكن نمثل البيانات بالرسم

[1] Box and whiskers plot → Whiskers

[2] Histogram and Polygon المدرج والمضلع التكراري الاثنين نفس طريقة التعامل

[3] Cumulative frequency curve of curve (ogive)

[4] Pie chart [6] Time-series graph البيانات الماضي

[5] Bar graph غير متصلين [7] dot plot

Pareto

COLLEGE

lecture "2"

إذا عنا مجتمع بدي أعمل دراسة عليه ما بوخد كل المجتمع بوخد عينة بعدها لحتى measure of central tendancy و Measures of Variation احسب بعدها بدي أعمل Quantitative و Qualitative إلى data لكن قبل بمشان ال

Discrete Continuouse

① raw data ② Frequancy distribution البيانات بأحد الأشكال:
③ Group of frequancy ④ stem and leaf diagram
أو من خلال الرسم:
1 Box and whiskers 2 Histogram and Polygon
3 cumulative frequancy curve of curve

* Measures of central tendancy

1 The mean ($\bar{x}$)

1) for raw data and stem and leaf diagram:-

raw ⇒ 7, 3, 5, 7

stem	leaf
1	1 2 3
2	1 2 3

بدنا نتعرف على الرموز الرياضية.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

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

7 , 3 , 5 , 7
X_1 المشاهدة الأولى — X_2 — X_3 — X_4 المشاهدة الرابعة

الوسط الحسابي $\frac{X_1+X_2+X_3+X_4}{4}$ ← عددهم

$$\frac{\sum_{i=1}^{4} X_i}{4}$$

Σ: summation المجموع

COLLEGE

ex. Find the mean of:

2, 7, 5, 11, 5

sol) $\bar{x} = \frac{2+7+5+11+5}{5}$ over

solution

$= \frac{30}{5} = 6$

---

eg. If the mean of

a, a, 7, 11, 2

is 6 find a

sol) $\frac{a+a+7+11+2}{5} = 6$

$\frac{2a+20}{5} = \frac{6}{1}$

$30 = 2a + 20$

$2a = 10$

$\boxed{a = 5}$

---

eg) If the mean of:

x, y, 12 is 10, then find the mean of x and y

is $\frac{x+y}{2}$

Now $\frac{x+y+12}{3} = 10$

$x + y + 12 = 30$

$-12 \quad -12$

$x + y = 18$

so $\frac{x+y}{2} = \frac{18}{2} = 9$

---

e.g/

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

key 1/2 represents 12

$\bar{x} = \frac{11+11+12+21+22+23+11+11+12}{9}$

$\bar{x} = \ldots$

---

eg. $\boxed{\Sigma x = n \cdot \bar{x}}$

عدد طلاب الصف 20 طالب كان الاقريج 15 ضفنا طالبة جديدة للصف علامتها $\frac{20}{20}$ كم يصير الاقريج الجديد؟ برتمح لا بقل بي.

e.g. If the mean mark of 9 students is 15. Ahmed with mark 20 joind the class. find the new mean.

sol) $n = 9$

$\bar{x} = 15$

$15 = \frac{\Sigma x}{9}$

$\Sigma x = 135$

لو طاح طالب من الصف يصير

New Σx

135 - 20

New $n = 8$ و

New $\Sigma x = \frac{\Sigma x}{\text{من}} + 20$

$(\Sigma x = 155)$

New $n = 10$

New $\bar{x} = \frac{155}{10} = 15.5$

eg. If the mean mark of [10] boys is 12
and " " " " " [12] girls is 10
find " " " " the students all together

sol)

$n = 10$, $\bar{x} = 12$, $\Sigma x = 120$

$n = 12$, $\bar{x} = 10$, $\Sigma x = 120$

$n = 22$, $\Sigma x = 240$, $\bar{x} = \frac{240}{22} = 10.9$

ii) For frequancy distribution $\bar{x} = \frac{\Sigma fx}{\Sigma f}$

Quize عمل علامات الطلاب لـ

eg)

x	1	2	3	4	5	sum → نجد المجموع
f	3	8	7	2	4	= 24
fx ← نضيف خانة	3	16	21	8	20	= 68

$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{68}{24}$

eg) if $\bar{x} = 3$ find k

x	1	2	3	4	5	sum
f	3	8	7	k	5	23+k
fx	3	16	21	4k	25	65+4k

$3 = \frac{65 + 4k}{23 + k}$

$69 + 3k = 65 + 4k$ ⇒ $k = 4$

الكتاب
section 2.3
* TRY IT YOURSELF *

x	w	x.w
2	3	6
2	4	8
1	1	1
4	3	12
3	2	6
4	3	9
sum	16	42

$\bar{x} = \frac{42}{16} = 2.62$ about 2.6

COLLEGE

iii) Group of frequancy distribution

eg. تقدير ← X $\frac{4+0}{2}=2$ | $\frac{9+5}{2}=7$ | 12 | 17

I	(0-4)	5-9	10-14	15-19	sum
f	3	8	7	2	20
xf	6	56	84	34	180

$\bar{x}=\frac{180}{20}$

$\bar{x}=9$

$\bar{x}=\frac{\Sigma fx}{\Sigma f}$, Where x is the mid interval value

midpoint $=\frac{(\text{lower class limit})+(\text{Upper class limit})}{2}$

$x=(a+b)/2$

eg. If $\bar{x}=9$ then find k

I	0-4	5-9	10-14	15-19	sum
f	3	8	k	2	13+k
x	2	7	12	17	96+12k
fx	6	56	12k	34	

$9=\frac{96+12k}{13+k} \Rightarrow 96+12k+117+9k$

$3k=21 \Rightarrow k=7$

* Measures of central tendancy:- The mean / The mode / The median

Done Before

The mode :-

The value that occurs mostly (or with highest frequancy)

eg. Find the mode for:-

i) 2, 7, 5, 11, 5

mode = 5 unimodal

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

No mode

e) 2, 7, 5, 11, 5, 2

mode = 2, 5 (Both) Bimodal

iv)

S	L
1	1 12
2	1 2 2
3	1 2 3

Key : 1|2 represent 12

11 / 22 ⇐ Bimodal

v)

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

2 repeated 8 times (The most.)

Unimodal: mode = 2

vi)

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

✓ الأصح modal class: 5-9

$$\text{mode} = \frac{5+9}{2} = 7 \leftarrow \text{midpoint}$$

لأنه class تكرارات

lecture "3"

[1] The mode (Done) [2] The mean (Done)

[3] The median (Q_2):

For raw data and stem and leaf diagram:-

The middle value after arranging the data in ascending order. يجب إترتيب تصاعدي او تنازلي or descending

نفضل تصاعدي

eg: Find the median for:-

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

sol) 2, 3, 5, (6), 7, 10, 11

$Q_2 = 6$ ⇒ لأن لما يجي 7 قيم كان Q_2 من احد العشاصات

ii) 2, 3, 5, (6, 8), 10, 10, 12 ⇐ هنا جاهزين مرتبين

two values ⇒ وحيدين, even ⇐ n لأن عدد الـ

هذا الـ median 7 وهو ليس من احد العشاصات

$$Q_2 = \frac{6+8}{2} = \frac{14}{2} = 7 \Rightarrow \text{mid point}$$

not practical لما يكونوا اعداد كبيرة او جدول هذه الطريقة موعلية

$$\frac{n}{2} = \frac{8}{2} = \frac{4^{th} + 5^{th}}{2} = \frac{6+8}{2} = \frac{14}{2} = 7$$

الطريقة التالية.

$\frac{n}{2}$

عدد كسري ← fraction ↓ next integer

عدد صحيح ← Whole no.(k) ↓ $\frac{k^{th} + (k+1)^{th}}{2}$

location of median

1st 2nd 3rd 4th 5th 6th 7th 8th

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

$\frac{n}{2} = \frac{9}{2} = 4.5 \rightarrow$ 5th is 12 $\rightarrow$ so $Q_2 = 12$

Fraction ⇓ ⇒ next Whole number

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

$\frac{n}{2} = \frac{8}{2} = \frac{4^{th} + 5^{th}}{2} = \frac{9+10}{2} = \frac{19}{2} = 9.5$

$Q_2 = 9.5$

v)

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

$\frac{n}{2} = \frac{9}{2} = 4.5 \rightarrow$

5th value is 22 $\quad Q_2 = 22$

key : 2|3 represent 23 بنأخذ الرقمين stem and leaf ← "Ecole

ii) For frequancy distribution:

Same as the previous case but after finding the cumulative frequancy (c.f), where $n = \Sigma f$.

التكرار التراكمي

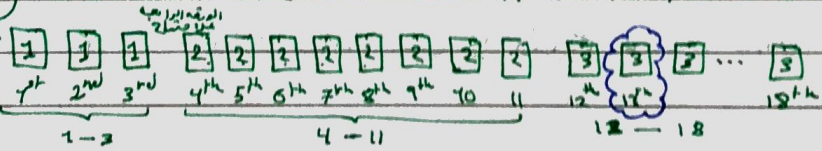
eg) Find the median for the following distribution:-

i)

x	1	2	3	4	5	sum
f	3	8	7	2	5	25
c.f	3	8+3 = 11	18	20	25	
	1st-3rd	4th-11th	12th-18th	19th-20th	21st-25th	

How many student achive 3 or less?? (7+8+3)

How many students achive 5 or less ?? all of the students = 25



$\frac{n}{2} = \frac{\Sigma f}{2} = \frac{25}{2} = 12.5 \rightarrow 13^{th}$ value $\Rightarrow Q_2 = 3$

(ii)

x	10	20	30	40	50	sum
f	3	8	7	2	2	22
c.f	3	11	18	20	22	
	1-3	4-11	12-18	19-20	21-22	

$\frac{n}{2}$ $\frac{22}{2}$ = $\frac{11^{th}\text{ value} + 12^{th}\text{ value}}{2}$ =

$Q_2 = \frac{20+30}{2} = \boxed{25}$

eg. For the following ordered set of data

a, 2, 3, 5, b, 9, 10, c

the mean is 7

the mode is 2 → الاكثر تكرارًا ⇒ $\boxed{a=2}$

the median is 6 Find a, b, and c.

SOL $\boxed{a=2}$ لان 2 هي الاكثر تكرارًا وهم مرتبين

median = $\frac{5+b}{2} = 6$ ⇒ $12 = 5+b$ → $\boxed{b=7}$

فعلًا بين 5 و 9

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

$38 + c = 56$ ⇒ $\boxed{c=18}$

فعلًا اكبر من 10 ✓

* ممكن يطلب بس c فنلزم نحسب a, b حتى نطلع c

* TRY IT YOURSELF 3 ← section 2.3 الكتاب

Find the median of the following data : 20, 48, 32, 24, 21, 29, 17, 27, 31, 31, 21, 34, 43, 28, 24, 34

نرتب الـ data 17, 20, 21, 21, 24, 24, 27, 28, 29, 31, 31, 32, 34, 34, 43, 48

ثم $\frac{n}{2}$ ← $\frac{16}{2} = 8$ ← $\frac{8^{th} + 9^{th}}{2}$ ← $\frac{28+29}{2}$ = 28.5

* Measures of Central tendancy.

[1] The mean ($\bar{x}$) [2] The mode [3] The median (Q_2)

a) For raw data and stem and leaf diagram. ✓

b) For frequancy distribution ✓

c) For grouped frequancy distribution :-

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

lower R.B — upper R.B — upper Boundery (9.5) — mid point $\frac{14+15}{2}$ (14.5) → Real Bonderies 19.5 → 0.5

e.g) — إذا كانوا العينة نفس الرقم عادي ولا منقص 0.5

الحدود الفعلية بزيد 0.5

مش ما نطرح 0.5 time, length, weight إذا كان مستحيل فيه سالبة ويقف عند (0).

$\frac{n}{2} = \frac{\sum f}{2} = \frac{20}{2} = 10^{th}$ value — لفرقاعنا السابق اذا طلع Frachion ما بوخذ العدد الصحيح اللي بعده

وإذا طلع عدد صحيح نركب 4 تنشه ما بعمل $\frac{k^{th} + (k+1)^{th}}{2}$

c.f بـ 11 الـ 3 بس موجودة بس 10th value بسأل

وبكون Q_2 بين 4.5 و 9.5

إذا كان بسم نسبة

3 — 10 — 11

4.5 — Q_2 — 9.5

نفس النسبة بين جدول

بحسب الـ deference بين 10 و 3

Total defer... على الـ $\frac{10-3}{11-3} = \frac{7}{8}$

$$\frac{7}{8} = \frac{Q_2 - 4.5}{9.5 - 4.5}$$

$$\frac{7}{8} = \frac{Q_2 - 4.5}{4.5}$$

$$8Q_2 - 36 = 35 \rightarrow \frac{8Q_2}{8} = \frac{71}{8} \rightarrow Q_2 = 8.875$$

$(\frac{n}{2})^{th}$ value after finding the cumulative frequancy and upper real boundies (U.R.B), then we use the interpolution method. — الطريقة

linear

COLLEGE

Find the median:-

← gap ie 6

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

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

$\boxed{Q_2 = 12}$

* Measures of Variation

lecture "4"

دائماً قيم موجبة.

I) The range : المدى

"أسهل شيء"

"أقل يعبر الكميات اكثر تقاربا لبعضهم"

range = max - min.

= max (U.R.B) - min (L.R.B)

[المسافة بين اكبر مشاهدة وأصغر مشاهدة]

e.g) Find the range for:-

متغيرات صفيت

$\bar{x}$ = 15 — Rang = 13 — صدور لاعبين تقاربا علاما لهم

$\bar{x}$ = 15 — Rang = 15 — أكثر تباعدا.

i) 2, 7, 5, 11, 3, 8. Sol) Range = 11 - 2 = 9

ii)

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

Sol) Range = 32 - 11 = 21

key: 1|2 represent 12

iii)

x	1	2	3	4	5
f	100	2000	13	7	25

Sol) Range = 5 - 1 = 4

متغير له علاقة بالتكرار.

iv)

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

-0.5 — 4.5 — 9.5 — 14.5 — 19.5

Sol) Range =

Range = max (U.R.B) - min (L.R.B)

= 19.5 - (-0.5)

= 20

اذا مقياس يكون Negative نقفز عن الصفر.

Note: Class length (c.L) or class width (c.w) is :

(c.L. = U.R.B - L.R.B)

e.g)

I	0-3	4-9	10-11	12-20
f	3	7	8	9
C.L	4	6	2	9

-0.5 — 3.5 — 9.5 — 11.5 — 20.5

الكتاب
*section 2.4
TRY IT YOURSELF
R = 58 - 23 = 35
R(A) اكبر من R(B)

20, 20, 20, 20, 0
Rang = 20 !!!
لو حذفنا 0 بعد
Range = 0 wow

لان Range تتأثر بـ outlier حتى أفضل آخر Measure (IQR).
هذا لما كنا نحكي عن مقاييس النزعة المركزية كان الأفضل median بوجود outliers

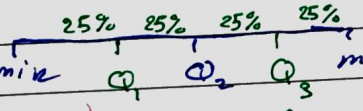
[2] The Inter-quartile range (IQR):

* Section 2.5

$IQR = Q_3 - Q_1$, where

Q_1: Lower quartile (25% of the data lie below Q_1) الربيع الأقل (الأول)

Q_3: upper quartile (75% of the data lie below Q_3) الربيع الأعلى (الثالث)



لو رتبنا الـ data

Q_1, Q_2 and Q_3 divide an ordered data set into four equal parts

IQR

⇒ The middle 50% of the data lie between Q_1 and Q_3

$(Q_3 - Q_1)$

كيف منحسبهم نفس طرق الـ median

One can find Q_1 and Q_3 in the same ways as finding Q_2 but we use:

Q_1: $\left(\frac{n}{4}\right)^{th}$ value

Q_3: $\left(\frac{3n}{4}\right)^{th}$ value.

eg) 2, 7, 5, 11, 3, 8, 10

sol) (2, 3, 5), 7, (8, 10, 11)
Q_1 Q_2 Q_3

$Q_1 = \left(\frac{n}{4}\right)^{th} = \frac{7}{4} = 1.75 \rightarrow 2^{nd}$ value $= 3$

$Q_3 = \left(\frac{3n}{4}\right)^{th} = \left(\frac{21}{4}\right) = 5.25 \rightarrow 6^{th}$ value $= 10$

$IQR = Q_3 - Q_1 = 10 - 3 = 7$

COLLEGE

eg) 2, 3, 5, 7, 7, 9, 11, 12 ترتيب ::

sol) $Q_1 = \left(\frac{n}{4}\right)^{th} = \left(\frac{8}{4}\right)^{th} = \frac{(2)^{nd} + 3^{rd}}{2} \Rightarrow Q_1 = \frac{3+5}{2} = \frac{8}{2} = 4$

$Q_3 = \left(\frac{3n}{4}\right)^{th} = \frac{(6)^{th} + (7)^{th}}{2} = \frac{9+11}{2} = \frac{20}{2} = 10$

First / lower quartile is 4 // third / upper quartile is 10

$\therefore IQR = Q_3 - Q_1 = 10 - 4 = 8$

eg)

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

$\frac{12+13}{2} = \frac{25}{2} = 12.5$ Q_1

Q_2

$(32+32) \div 2 = 32$ Q_3

Key: 1|2 represents 12

sol) $Q_1 = \left(\frac{n}{4}\right)^{th} = \left(\frac{9}{4}\right)^{th} = 2.25 \rightarrow (3)^{rd}$ value

$Q_1 = 13$

$Q_3 = \left(\frac{3n}{4}\right)^{th} = \left(\frac{3(9)}{4}\right)^{th} = 6.75 \rightarrow (7)^{th}$

$Q_3 = 32$

$IQR = Q_3 - Q_1 = 32 - 13 = 19$

بعض الكتب تقول $Q_3 - Q_1 = 32 - 12.5 = 19.5$

eg)

x	1	2	3	4	5	sum
f	3	8	7	2	5	25
c.f	3	11	18	20	25	
	1-3	4-11	12-18	19-20	21-25	

sol) $Q_1 = \left(\frac{n}{4}\right)^{th} = \left(\frac{25}{4}\right)^{th} = 6.25 \rightarrow 7^{th}$ value $\rightarrow Q_1 = 2$

$Q_3 = \left(\frac{3n}{4}\right)^{th} = \left(\frac{3(25)}{4}\right)^{th} = 18.75 \rightarrow 19^{th}$ value $\rightarrow Q_3 = 4$

$IQR = 4 - 2 = 2$

eg)

	-0.5	4.5	9.5	14.5	19.5	
I	0-4	5-9	10-14	15-19		sum
f	5	7	3	5		20 = n
c.f	5	12	15	20		
U.R.B	4.5	9.5	14.5	19.5		

$Q_1 = \left(\frac{n}{4}\right)^{th} = \left(\frac{20}{4}\right)^{th} = (5)^{th}$ value $\Rightarrow \boxed{Q_1 = 4.5}$

$Q_3 = \left(\frac{3n}{4}\right)^{th} = \left(\frac{3(20)}{4}\right)^{th} = (15)^{th}$ value $\Rightarrow \boxed{Q_3 = 14.5}$

$IQR = Q_3 - Q_1 = 14.5 - 4.5 = \boxed{10}$

Steps for IQR for grouped distribution

* The outliers

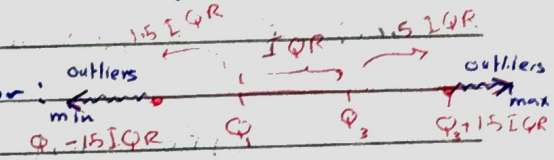
Lecture (5)

2, 3, 7, 5, 1000

كيف أحدد الـ outlier (outlier) هو شرط الأعلى أو الأكبر.

Any value less than $Q_1 - 1.5 \cdot IQR$

Or greater than $Q_3 + 1.5 \cdot IQR$



eg) Find the outliers (if any) for:

i) 2, 7, 8, 10, 12, 20

$Q_1 = \left(\frac{n}{4}\right)^{th} = \left(\frac{6}{4}\right)^{th} = 1.5 \rightarrow 2^{nd}$ value $\Rightarrow \boxed{Q_1 = 7}$

$Q_3 = \left(\frac{3n}{4}\right)^{th} = \left(\frac{3(6)}{4}\right)^{th} = 4.5 \rightarrow 5^{th}$ value $\Rightarrow \boxed{Q_3 = 12}$

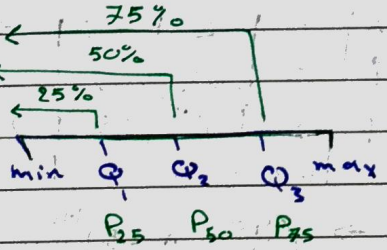
$IQR = 12 - 7 = 5$

- $Q_1 - 1.5\,IQR = 7 - 1.5*5 = -0.5$ ← أول قيمة في الداتا تقبل (غير سالب)
- $Q_3 + 1.5\,IQR = 12 + 1.5*5 = 19.5$ ← (20 is greater than 19.5)

since $20 > 19.5$

$\therefore$ 20 is an outlier ← Outliers ممكن يكون في مشاهدات هامشية

* Percentile:



So P_k (k^{th}-percentile): The value that k% of the data lie below it.

$P_k = \left(\frac{k}{100} n\right)^{th}$ value

$Q_3 = P_{75} = \left(\frac{75}{100} n\right)^{th}$ ← أي

$Q_3 = P_{75} = \left(\frac{3}{4} n\right)^{th}$

Note: the m% to n% inter-percentile range is:

$IPR = P_n - P_m$

eg) Find the 65th percentile for:

2, 3, 5, 7, 9, 10, 11, 12

Sol) $P_{65} = \left(\frac{65}{100} n\right)^{th} = \left(\frac{65}{100}(8)\right)^{th} = 5.2 \rightarrow 6^{th}$ value

$\therefore \boxed{P_{65} = 10}$

eg) Find the 35% ~~per~~ to 70% inter-percentile range for:

3, 5, 5, 6, 8, 8, 9, 10, 11, 12

$IPR = P_{70} - P_{35} = \left(\frac{70}{100} * 10\right)^{th} - \left(\frac{35}{100} * 10\right)^{th}$

$(7)^{th} - (3.5)^{th}$

$\left(\frac{7^{th} + 8^{th}}{2}\right) - (4^{th}) = \frac{9+10}{2} - 6 = \frac{19}{2} - 6 = 9.5 - 6$

The 35% to 70% IPR is $P_{70} - P_{35} = \boxed{3.5}$

* Find the 40th percentile

eg)

x	1	2	3	4	5	sum
f	3	8	7	2	5	25
c.f	3	11	18	20	25	
	1-3	4-11	12-18	19-20	21-25	

c.f: مرتبا أعطانا (f) وبدأنا بجمعها على بعض

Sol) $P_{40} \rightarrow \left(\frac{40}{100} * 25\right)^{th} = \frac{(10)^{th} + (11)^{th}}{2} = \frac{2+2}{2} = 2$

eg) Find the 60th percentile for:

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

(-0.5, 4.5, 9.5, 14.5, 19.5)

$P_{60} = \left(\frac{60}{100} * 20\right) = 12^{th}$ value

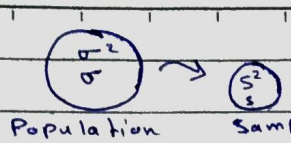
$\frac{12-11}{18-11} = \frac{P_{60} - 9.5}{14.5 - 9.5} \rightarrow \frac{1}{7} = \frac{P_{60} - 9.5}{5}$

(11) (12) (18) $P_{60} = 10.2$

9.5 P_{60} 14.5

COLLEGE

The standard deviation (s) and variance (s^2)



* a) For raw data and stem and leaf diagram:

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$ (variance) ← sample variance

or $$s^2 = \frac{\sum x^2}{n-1} - \frac{(\sum x)^2}{n(n-1)}$$ sample variance

standard deviation = $s = \sqrt{\text{variance}}$

eg) Find the variance and standard deviation for:

1) 2, 7, 5, 11, 5

sol) method I : $\bar{x} = \frac{2+7+5+11+5}{5} = \frac{30}{5} = 6$

						sum
$x-\bar{x}$	-4	1	-1	5	-1	0
$(x-\bar{x})^2$	16	1	1	25	1	44

Note: $\sum (x-\bar{x}) = 0$

variance = $s^2 = \frac{44}{5-1} = \frac{44}{4} = 11$

standard deviation = $s = \sqrt{11} \approx 3.32$

Method II : الطريقة الثانية

						sum
x	2	7	5	11	5	$30 = \sum x$
x^2	4	49	25	121	25	$224 = \sum x^2$

$$s^2 = \frac{\sum x^2}{n-1} - \frac{(\sum x)^2}{n(n-1)} = \frac{224}{4} - \frac{(30)^2}{5(4)} = 11$$

standard deviation = $s = \sqrt{11} \approx 3.32$

lecture (6)

eg) If the mean mark of 10 students is 15 and 4 students is 15 and the standard deviation is 3. Ahmad with mark 18 joined the class. Find the new mean and the new standard deviation.

لو كانت left منقل العكس (-ve)

Sol) $n = 10$, $\bar{x} = 15$ $\Rightarrow \Sigma x = n \cdot \bar{x} = 150$

$+18$

New $n = 11$ $\Rightarrow$ New $\Sigma x = 168$ $\Rightarrow$ New $\bar{x} = \frac{168}{11} = 15.3$

New $S^2 = \frac{\Sigma x^2}{n-1} - \frac{(\Sigma x)^2}{n(n-1)}$ $\Rightarrow 9 = \frac{\Sigma x^2}{9} - \frac{(150)^2}{10*9}$

$9 = \frac{\Sigma x^2}{9} - 250$ $\Rightarrow \Sigma x^2 = 259 * 9 = 2331$

$+250$ $+250$

New $\Sigma x^2 = 2331 + (18)^2 = 2655$

New $n = 11$

New $\Sigma x = 168$

New $S^2 = \frac{2655}{11-1} - \frac{(168)^2}{11*10} = 8.92$

New $S = \sqrt{8.92} \approx 2.99$

COLLEGE

eg) If the mean mark of 10 boys is 12 and the standard deviation is 2 while the mean mark of 12 girls is 10 and the standard deviation is 3, find the mean mark and the standard deviation for the students all togeter. sol)

Boys (1): $n_1 = 10$, $\bar{x}_1 = 12$, $S_1 = 2$, $\Sigma x_1 = 120$

Girls (2): $n_2 = 12$, $\bar{x}_2 = 10$, $S_2 = 3$, $\Sigma x_2 = 120$

all together: $n = 22$, $\Sigma x = 240$, $\bar{x} = \frac{240}{22} = 10.9$, $S = ??$, $\Sigma x^2 = ??$

Boys $\rightarrow S^2 = \frac{\Sigma x_1^2}{n-1} - \frac{(\Sigma x_1)^2}{n(n-1)}$

$2^2 = \frac{\Sigma x_1^2}{9} - \frac{(120)^2}{10*(9)}$

$4 = \frac{\Sigma x_1^2}{9} - 160$

$\therefore \boxed{\Sigma x_1^2 = 1476}$

Girls $\rightarrow 9 = \frac{\Sigma x_2^2}{11} - \frac{(120)^2}{12(11)}$

$9 = \frac{\Sigma x_2^2}{11} - 109.1$

$\Rightarrow \boxed{\Sigma x_2^2 = 1299}$

* All together

$\Sigma x^2 = \Sigma x_1^2 + \Sigma x_2^2 = 1476 + 1299 = 2775$

$S^2 = \frac{2775}{21} - \frac{(240)^2}{20(21)} = 7.47 \Rightarrow S = \sqrt{7.47} \approx 2.73$

S^2 دائماً موجب لأن S

b) For frequancy distribution and grouped ...

$\rightarrow n = \Sigma f$

Two Formulas

$$S^2 = \frac{\Sigma f (x - \bar{x})^2}{\Sigma f - 1} \quad \text{or} \quad S^2 = \frac{\Sigma f x^2}{\Sigma f - 1} - \frac{(\Sigma f x)^2}{\Sigma f (\Sigma f - 1)}$$

x	1	2	3	4	5	sum
f	3	8	7	4	5	27
fx	3	16	21	16	25	81
x^2	1	4	9	16	25	
$f.x^2$	3	32	63	64	125	287

$S^2 = \frac{287}{26} - \frac{(81)^2}{81*(80)}$

$S^2 = 1.69$

Thus, $S = \sqrt{1.69} \approx 1.30$

x = mid point

I	0-4	5-9	10-14	15-19	sum
f	3	8	7	2	20
x	4+0/2 = 2	7	12	17	
fx	6	56	84	34	180
x^2	4	49	144	289	
$f \cdot x^2$	12	392	1008	578	1990

$$s^2 = \frac{\Sigma f x^2}{\Sigma f - 1} - \frac{(\Sigma f x)^2}{\Sigma f(\Sigma f - 1)} = \frac{1990}{19} - \frac{(180)^2}{20(19)} \approx 19.5$$

$$s = \sqrt{19.5} \approx 4.41$$

هدول اهنطة Measures.. لطن بحالة outlier لاهنطن (IPR) ← دخل وفي هنا

Note: relative frequency (r.f)

$$r.f = \frac{f}{\Sigma f}$$

eg.

x	1	2	3	4	5	sum
f	3	8	7	2	5	25
r.f	3/25	8/25	7/25	2/25	5/25	
	0.12	0.32	0.28	0.08	0.20	1

$\Sigma r.f = 1$

$f = 25 * 0.12$

$$r.f = \frac{f}{\Sigma f} \Rightarrow f = r.f * (\Sigma f)$$

$\Sigma f = n$

رح نحكي عن تأثير تعديل البيانات على هذه الـ Measures
مثلاً لو زدت درجة كل طالب علامتين اوعشر درجات ...

$\sum \frac{f}{n} = 1$ دائماً

eg) I: 0-4 | 5-9 | 10-14 | 15-19

C.f: 3 | 11 | 18 | 20

Find the r.f for the class 10-14

$r.f = \frac{f}{\Sigma f} = \frac{7}{20} = 0.3$

f	3	8	7	2	sum 20

$\frac{18-11}{20} = \frac{7}{20}$

COLLEGE

lecture (7)

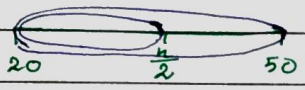
e.g] If $Q_2 = 40$, then find P.

No gap

I	$0 \le x < 10$	$10 \le x < 30$	$30 < x \le 60$	$60 < x \le 80$	$80 < x < 90$
f	5	15	30	P	2

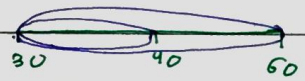
$n = \sum f = 52 + P$

c.f	5	20	$\frac{n}{2}$	50	50+P	52+P
U.R.B	10	30	*	60	80	90



$Q_2 = 40$

$$\frac{\frac{n}{2} - 20}{50 - 20} = \frac{40 - 30}{60 - 30} \rightarrow \frac{\frac{n}{2} - 20}{30} = \frac{10}{30}$$



$$\frac{n}{2} - 20 = 10 \quad (+20 \text{ on both sides}) \rightarrow \frac{n}{2} = 30 \rightarrow n = 60$$

$$60 = 52 + P \quad (-52 \text{ on both sides}) \rightarrow P = 8$$

e.g] Find the percentage of observation that are less than 6.

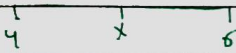
I	3-5	6-8	9-11	Sum
f	4	2	4	10

K% ← ; K = ??

$P_K = 6$

c.f	4	X	6	10	
U.R.B	5.5	6	8.5	11.5	

$$P_K = \left(\frac{K * n}{100}\right)^{th} = 4.3\overline{3} \rightarrow K = 43.\overline{3}$$



$$\frac{x - 4}{6 - 4} = \frac{6 - 5.5}{8.5 - 5.5}$$

$$\frac{x - 4}{2} = \frac{0.5}{3} \rightarrow 1 = 3x - 12 \quad (+12 \text{ on both sides}) \rightarrow \frac{13}{3} = \frac{3}{3}x$$

$x = \frac{13}{3} = 4.3\overline{3}$

$\frac{4.3\overline{3}}{10} * 100\% = 43.\overline{3}\%$ the percentage

COLLEGE

Coding

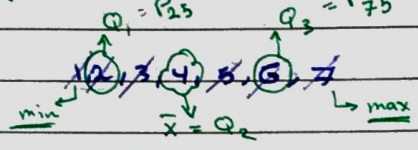
ضفّ لكل اختبار و حسبنا mean, mode ... وبعدها قررت اضيف لكل طالب علامتين هل يرجع احسن من جديد او لا يقدر احسبهم من العلامة اللي منحتها

* Coding (Liner Transformation) :

If $Y = aX + b$, then

(الجديد بعد التعديل) ← (بنضيف درجة) ← (العلامة قبل التعديل) ← (عشان مافارق)

عرضًا raw data لـ :



$\bar{x} = \frac{1+2+3+4+5+6+7}{7} = \frac{28}{7} = 4$

اذا جمعت لكل قيمة (2) ⟸ (2+x)

بدها تصير كالتالي : 3, 4, 5, 6, 7, 8, 9

$\bar{x} = 6$ / $Q_2 = 6$, $Q_1 = 4$ / $Q_3 = P_{75} = 8$

الـ mean ② median ③ min ④ max ⑤ quartiles ⑥ mode ⑦ Percentiles

تأثروا بالجمع .

طيب لو ضربت بشيء (2x) إذًا كلهم كذلك تأثروا بالضرب والقسمة والطرح والجمع

2, 4, 6, 8, 10, 12, 14

$\bar{x} = Q_2$

وإذا في حالة وحدة لو ضربنا بسالب (-1x) بصير اللي بالسالب ينعكس الترتيب

-1, -2, -3, -4, -5, -6, -7

فتصاعديًا -7, -6, -5, -4, -3, -2, -1

الـ quartiles تتبدلوا .

min, Q_1 (P_{25}), $\bar{x} = Q_2$, Q_3 (P_{75}), max

$Q_1 = 2$ و $Q_3 = 6$ كانت

$Q_1 = -6$ و $Q_3 = -2$ أصبحت

If $Y = ax + b$, then

1. $\bar{Y} = a\bar{x} + b$
2. Mode of $Y = a$ (mode of x) $+ b$
3. Q_2 of $Y = a$ (Q_2 of x) $+ b$
4. Q_1 of $Y = a(Q_1$ of $x) + b$, $a > 0$
   $= a$ (Q_3 of x) $+ b$, $a < 0$
5. Q_3 of $Y = a$ (Q_3 of x) $+ b$, $a > 0$
   a (Q_1 of x) $+ b$, $a < 0$
6. P_k of $Y = a(P_k$ of $x) + b$, $a > 0$
   $= a(P_{100-k}$ of $x) + b$, $a < 0$

في اللي بصير

$Q_1 \rightarrow Q_3$

$Q_3 \rightarrow Q_1$

$P_{20} \rightarrow P_{80}$

$P_{60} \rightarrow P_{40}$

COLLEGE

Mesures of variation

1, 2, 3, 4, 5, 6, 7

R = 6 = max − min = 7 − 1 = 6

IQR = $Q_3 - Q_1$ = 6 − 2 = 4

3, 4, 5, 6, 7, 8, 9 (x+2) 2 لو جمعنا

R = 9 − 3 = 6

IQR = 8 − 4 = 4

* R / IQR they are not affected by addition

2, 4, 6, 8, 10, 12, 14 (2x) 2 لو ضربنا

R = 14 − 2 = 12

IQR = 12 − 4 = 8

-14, -12, -10, -8, -6, -4, -2 (-2x) -2 لو ضربنا

R = −2 ⟷ 14 = 12

IQR = −4 ⟷ 12 = 8

[a] Range of y = |a| (Range of x) (absolute value)

[b] IQR of y = |a| · (IQR of x)

[c] S_y of y = |a| · S_x

[d] S^2_y of y = $a^2 \cdot S^2_x$

ملاحظة: لو أخذنا الجذر التربيع نتيجة عملية تصبح a^2

e.g) If y = 2 − 3·x and ($\bar{x}$ = 10 / Q_2 = 9 / Q_1 = 5 / Q_3 = 12 / P_{20} = 3 / P_{80} = 15 / S^2_x = 3) for x

Find the previous 9 for y

Sol) $\bar{y}$ = 2 − 3*10 = 2 − 30 = −28 / Q_2 of y = 2 − 3*9 = 2 − 27 = −25

Q_1 of y = 2 − 3 * Q_3 of x = 2 − 3*12 = 2 − 36 = −34

Q_3 of y = 2 − 3 * Q_1 of x = 2 − 3*5 = 2 − 15 = −13

P_{20} of y = 2 − 3 * P_{80} of x = 2 − 3*15 = 2 − 45 = −43

P_{80} of y = 2 − 3 * P_{20} of x = 2 − 3*3 = 2 − 9 = −7

S^2_y = $(3)^2 S^2_x$ = 9 * 8 = 27 (ما تأثر بالجمع والطرح)

eg) If the mean mark of a group of students is 15 and the variance is 2. The teacher decided to decrease each mark by 10% and then add 2 marks to each transcript. Find the new mean and new variance. $(y = 0.90x + 2)$

$\bar{x} = 15$, $S_x^2 = 2$

فرضنا عندي مجموعة جعل علما خصم 10% بعني كانت 100% اصبح خصم 10% اصبح 90% من العدد الاصلي كذلك اي عدد بعد اضافة 10% بعني 90% وكذلك العلامات بعد اضافة 2 درجات بعني $(y = 0.90x + 2)$

$$\bar{y} = \frac{90}{100}(15) + 2 = 13.5 + 2 = 15.5 \quad [\text{Average}]$$

$$S_y^2 = (0.90)^2 S_x^2 = \frac{81}{100}(2) = 1.62$$

eg). If $y = ax + b$, $\bar{x} = 10$, $\bar{y} = 5$, $S_x = 3$, $S_y = 1$

find a , b

$$\bar{y} = a\bar{x} + b \Rightarrow \boxed{5 = a10 + b} \quad \cdots (1)$$

$$S_y = |a| S_x \rightarrow 1 = a3 \rightarrow \boxed{|a| = \frac{1}{3}} \rightarrow a = \pm\frac{1}{3}$$

نعوض $a = \frac{+1}{3}$

$$5 = \frac{10}{3} + b \rightarrow b = 1.66 \approx 1.7$$

$$\frac{-10}{3} \quad \frac{-10}{3}$$

$$\boxed{b = \frac{5.1}{3}} \Rightarrow y = \frac{+1}{3}x + \frac{5}{3} \checkmark$$

نعوض $a = \frac{-1}{3}$

$$5 = -\frac{10}{3} + b \rightarrow b = 8.33 \quad / \quad \boxed{b = \frac{25}{3}}$$

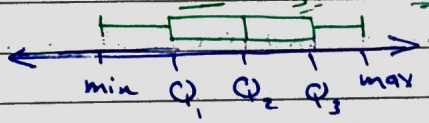
$$\Rightarrow y = \frac{-1}{3}x + \frac{25}{3} \checkmark$$

COLLEGE

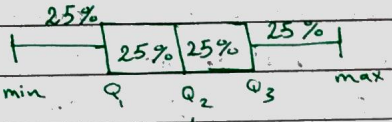
Graphical Representation

* Graphical Representation:.

(1) Box & whiskers plot: continuous و discrete تستعمل للـ



(min, Q_1, Q_2, Q_3, max) قانون الخمس أرقام Five numbers rule



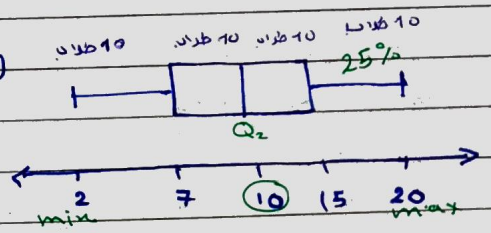
median = Q_2

$IQR = Q_3 - Q_1$

Range = max − min

مثلاً لو عدد الصف ٢٠ وسألت اصغر عدد الطلاب اللي حصلوا اكثر من Q_3 يعني (20)25% = $\frac{1}{4}$ ⇒ $\frac{1}{4}$ (العدد الكلي)

eg)



n=40 The above are the marks of 40 students, Find:

(i) The median (ii) the range

(iii) The IQR (iv) the no. of students that achieved more than 15

(v) an estimation for the mean

Sol)

(i) $Q_2 = \boxed{10}$

(ii) max − min = $\boxed{18}$ = R

(iii) $IQR = Q_3 - Q_2$

$= 15 - 7 = \boxed{8}$

(iv) 25%(20) = $\frac{1}{4}$(40)

= 10 students

~~$Q_3 = 15$~~ ~~n = 40~~ لأجل

$\boxed{n = 20}$

(v) "mean" لإيجاد

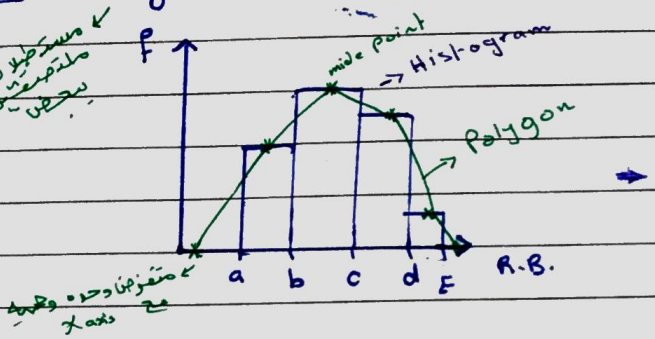
I	2−7	7−10	10−15	15−20	sum
f	10	10	10	10	40
x (mid point)	4.5	8.5	12.5	17.5	
fx	45	85	125	175	430

$\bar{x} = \frac{\sum fx}{\sum f} = \frac{430}{40}$

$= \boxed{10.75}$

[2] Histogram & Polygon :

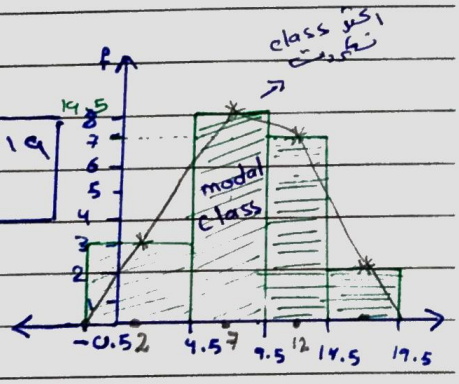
continuous متقطعة لـ classes بدون فجوات



eg: Draw a histogram & Polygon :

	-0.5	4.5	9.5	14.5	19.5
I	0-4	5-9	10-14	15-19	
f	3	8	7	2	

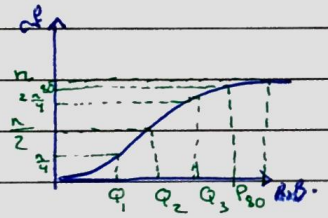
mide point of (0-4) $\frac{0+4}{2} = \frac{-0.5+4.5}{2}$



[3] The c.f. curve : (ogive)

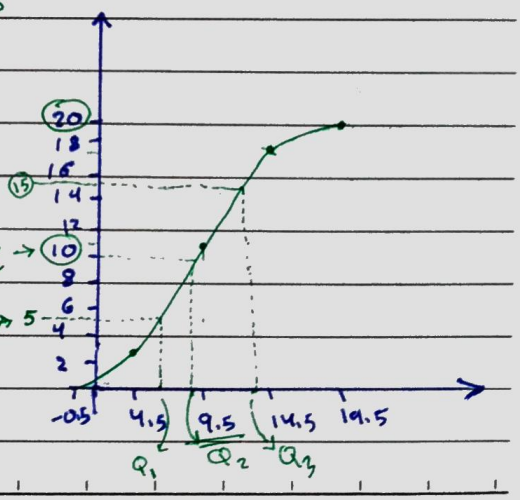
Percentile لـ quantile بشكل سريع

Q* Median of a given frequancy distribution is found with the help of a. A) Bar graph (B) Ogive C) Histogram d) None of them



eg Draw of c.f curve :

	-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	
U.R.B	4.5	9.5	14.5	19.5	



* What percentile is a score 9.5 ?

$\frac{10}{20} * 100\% = 50\%$

$P_{50} = 9.5$

COLLEGE