

* Statistics:-

- What is statistics:-

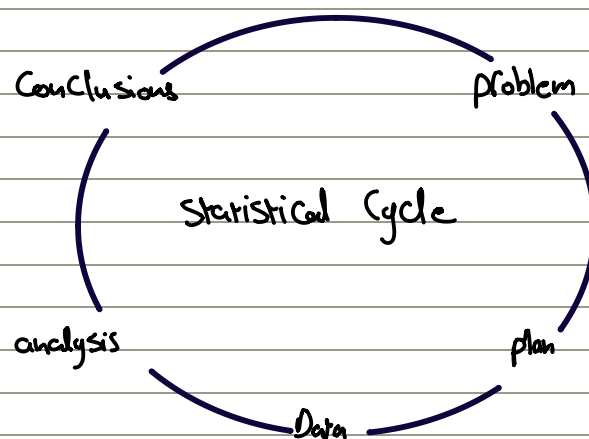
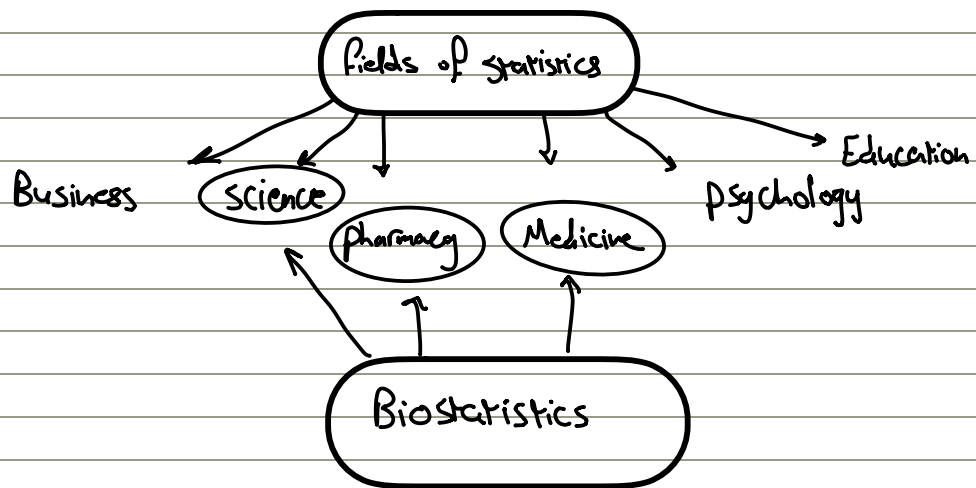
→ Science → ^{جمع} Collection, ^{تنظيم} organisation, ^{تلخيص} summarisation, ^{تحليل} analysis
→ Drawing of inferences

① جمع البيانات → ② تنظيها (بطبقها جدولاً) → ③ تلخيصها (الى بياني) → ④ تحليلها
⑤ اخريتي Conclusion (الاستنتاج)

note :-

Statistics → producing analges + predictions → by offering critical guidance

المعصاء → بقديم تحليل و تنبؤ لبيانات → بالاعتماد على وجهات فائدة لا تعد على رأي الشخص



Branches of statistical

- Descriptive :-
sample data or ^{نمونه} ^{بيانات} *
Collection, organisation,
summarisation, analysis.

inferences :-
يعتبر الاستنتاج على كل
population
- draw conclusions

* Vocabulary of statistics :-

→ population :- is entire group, Collection all items

هي مجموعة العناصر بكل صف الدراسة، لا يمكن إفحصها كلها

→ Sample :- is ^{نموة} subset of the population

هي مجموعة جزئية من ال population لها نفس الخلفيات

→ Sample should be selected represent whole population and unbiased

اختيار sample لازم يمثل كل population في اختيار لغاية معينة

→ Data :- the raw material of statistice ^{هي المادى الاصلى} steps

observation or measurement ^{ملاحظة أو قياس}

→ There are 2 Type of numbers :-
→ From measurement (معدنى) (E.x) :- Temperatur
→ From process Counting (عدى) (E.x) :- patient in hospital

→ Data obtained by :-

① Surveys ② Experiment ③ Routinely ④ External resources

→ Variables :- ^{هي المتغير} اللى بغير كذا الدراسة

E.x :- ^{العمر} age

→ 2 Type :- ① Qualitative :- Not measurable

E.x :- (Female, male)

Blood groupe (A, B, O...)

② Quantitative :- measurable, Can be ordered, ranked ^{نمى}

E.x :- age

→ Parameter :- describing some characteristic of Population

→ Statistics :- basis information of Sample

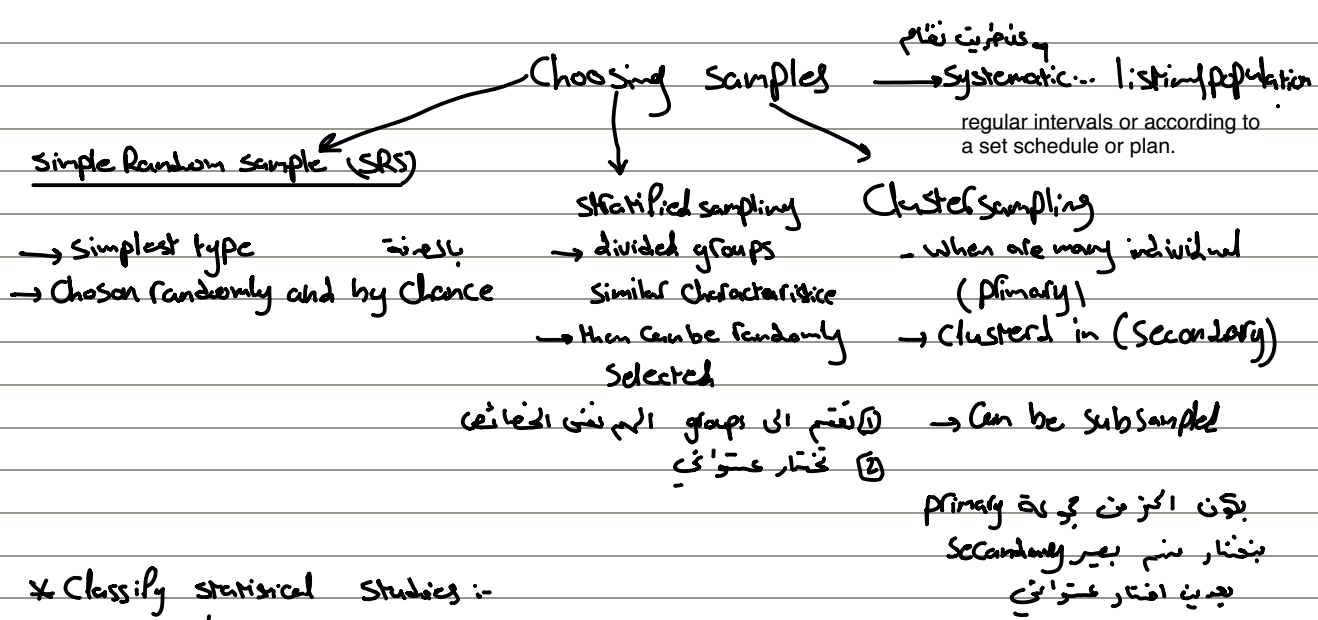
* why we study sample rather than population :- ^{دالة في بدل}

① ↓ less Cost ^{تكلفة أقل} ② preservation from loss

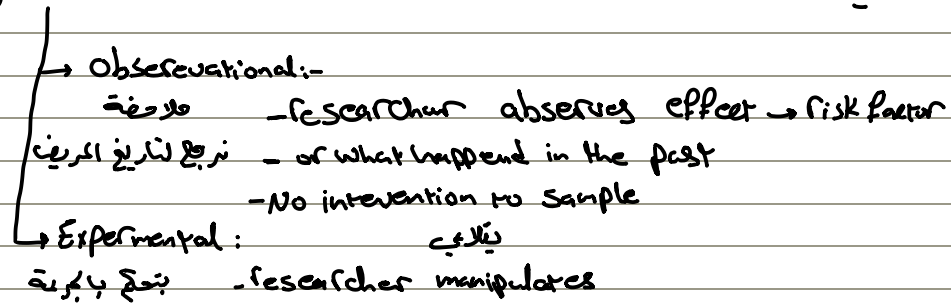
③ save time

④ More accuracy

⑤ population is so huge



* Classify statistical studies :-



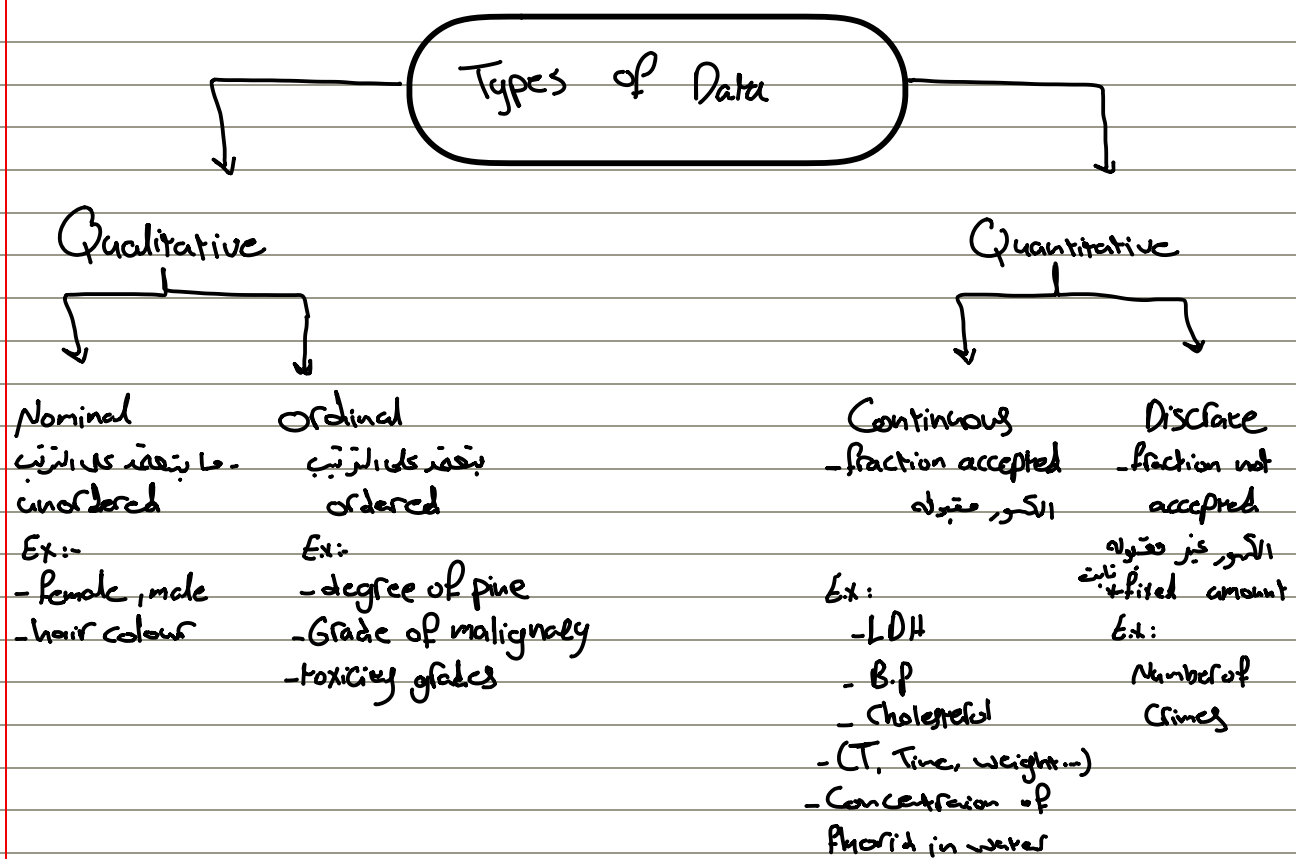
* Dependency :-

- ① independent → Called → experimental
 - Can be controllable
- ② dependant → Called → outcome variable
 - effect by independent

* Computer and Calculators :-

The most commonly used statistical :-

- * GraphPad prism
- * SPSS
- * STATA
- * SAS
- * MINITAB



* $\text{Relative frequencies} = \frac{\text{observation}(n)}{\text{Total}}$

Total Relative frequencies = 1 ∴ note

* $\text{Cumulative frequencies} = \text{The relative frequencies before} + N (\text{observation})$

<u>Toxicity</u>	<u>N</u>	<u>Relative Freq</u>	<u>Cumulative Freq</u>
None	68	.34	.34
Mild	92	.46	.80
Moderate	24	.12	.92
Severe	12	.06	.98
Life-threatening	4	.02	1.00

* note:- Ordinal data → Can also use percentiles → divide distribution equal of unequal
 Qualitative J median ← لا يسقط الحساب ← median ← 50%, 50%
 ارقام

$$* \text{mean} = \frac{\sum \text{observation}}{N} \rightarrow \frac{\text{جمع البيانات}}{\text{عدد هم}}$$

* median :-

1] order observations (Smallest $\rightarrow$ largest)
رتب البيانات

2] Calculate median

- odd $\rightarrow j = \frac{(N+1)}{2} \rightarrow$ الرقم الناتج هو ترتيب ال median
اذا كانت $N =$ عدد فردي
- even $\rightarrow j = \frac{N}{2} \rightarrow j$ and $j+1$
اذا كانت $N =$ عدد زوجي
 $\frac{(j) + (j+1)}{2} = \text{median}$
هو ترتيب (موقع)

Ex: 5 10 15 20 30
median $\leftarrow$ (15)

$N=5$ odd $\leftarrow j = \frac{5+1}{2} = 3 \rightarrow$ هو الرقم ترتيب 3 = 15

Ex: 5 10 15 20 30 40

$N=6$ even $\rightarrow j = \frac{6}{2} = 3$ and $j+1 = 4$
رقم 15 رقم 20
 $\frac{15 + 20}{2} = 17.5$ median

مقاييس التشتت

Measures of dispersion

- Range = largest observation - smallest observation
- Interquartile Range = $Q_3 - Q_1 = P_{75} - P_{25}$
- Standard Deviation = $\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$
 $\rightarrow$ SD presented with mean value

Ex: $\frac{145}{\text{mean}} \pm \frac{15}{\text{SD}}$

* raw data $\rightarrow$ not been organized or otherwise manipulated
 $\rightarrow$ non ordered

* Ordered array $\rightarrow$ is listing of collection (smaller to largest)
 $\rightarrow$ ordered

* Class Intervals $\rightarrow$ non-overlapping غير متداخلة

$$\rightarrow \text{Range} = (\text{largest} - \text{smallest}) + 1$$

$$\rightarrow \text{Interval width} = (\text{upper class limit} - \text{lower class limit}) + 1$$

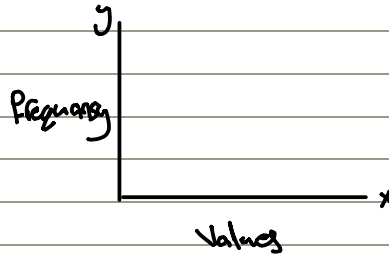
$\rightarrow$ Relative Frequency: نسبة التكرار في الفئة إلى عدد

$$= \frac{\text{number of values}}{\text{Total number of values}}$$

Histogram:

(x) axis $\rightarrow$ values

(y) axis $\rightarrow$ frequency



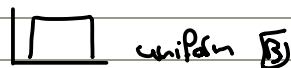
* $k = 1 + 3.322 (\log_{10} n)$
 $\downarrow$ $\downarrow$
 number of interval number of observations

interval width = Range / k

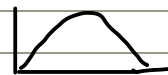
* Data $\rightarrow$ مجموعة البيانات

middle of data set is located تقع في منتصف البيانات (1) Centre

Variation $\rightarrow$ التباين (2)



uniform (3)



bell-shaped (4) Distribution



skewed (5)

median $\rightarrow$ الوسيط (6)

mean $\rightarrow$ المتوسط (7) + outliers

mode $\rightarrow$ لايئاز (8)

most commonly measures

Mean Median mode

properties

uniqueness

simplicity

Robust measure

little used

مقياس قوي

مقياس قوي

* normal distribution $\rightarrow$ mean = median = mode

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

sample

population

number of values

mean

* $\rightarrow$ مقياس في التوزيع

$$x_i = \frac{\text{lower class limit} + \text{upper class limit}}{2} = \text{midpoint}$$

$$\text{mean} = \frac{\sum (f \cdot x_i)}{n}$$

(3)

(f x i) = frequency x midpoint (2)

* حساب median في Table بدون فترات :-

Median :- $cf = \frac{n}{2}$ or $cf > \frac{n}{2}$ [3]

[1] حساب Cumulative Frequency (CF)

[2] Total $\leftarrow \frac{\sum of frequency}{2}$ $\frac{n}{2}$ حساب

* حساب median في الفترات :-

first class have :- $cf = \frac{n}{2}$ or $cf > \frac{n}{2}$ [3]

[1] حساب Cumulative Frequency (CF)

[2] Total $\leftarrow \frac{\sum of frequency}{2}$ $\frac{n}{2}$ حساب

$$Median = L_m + \left(\frac{\frac{n}{2} - F_m}{f_m} \right) i$$

$\nwarrow$ limit $\quad \downarrow$ f of class $\quad \nearrow$ cf before class

Mode :-

- all values are different $\rightarrow$ no mode Ex: 1, 5, 7, 10, 12

- there are more than one mode $\rightarrow$ Ex: 1, 2, 2, 5, 4, 4

mode = 2, 4

unimodal one mode $\quad$
 bimodal 2 mode $\quad$
 multimodal more than 2 mode

- Properties :-

[1] is not unique

يكون فريدي اراكثر

[2] robust measur

مقياس قوي

[3] describing $\rightarrow$ qualitative
 $\rightarrow$ quantitative

but (mode + median) $\rightarrow$ Calculated quantitative

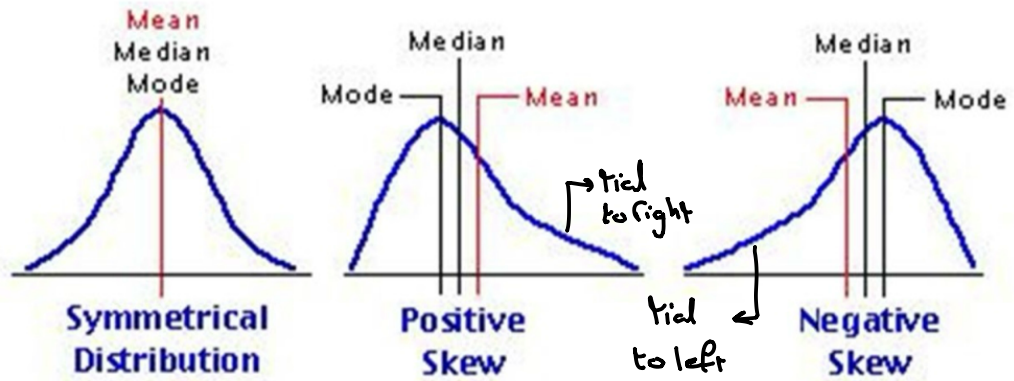
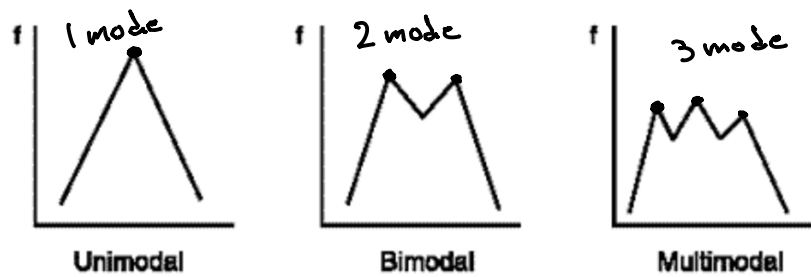
اكثر تكرار $\leftarrow$

* حساب mode في الفترات :- $(f) mode class - (f) class before$

$$mode = L_m + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) i$$

$\nwarrow$ limit $\quad \downarrow$

$(f) mode class - (f) class after$

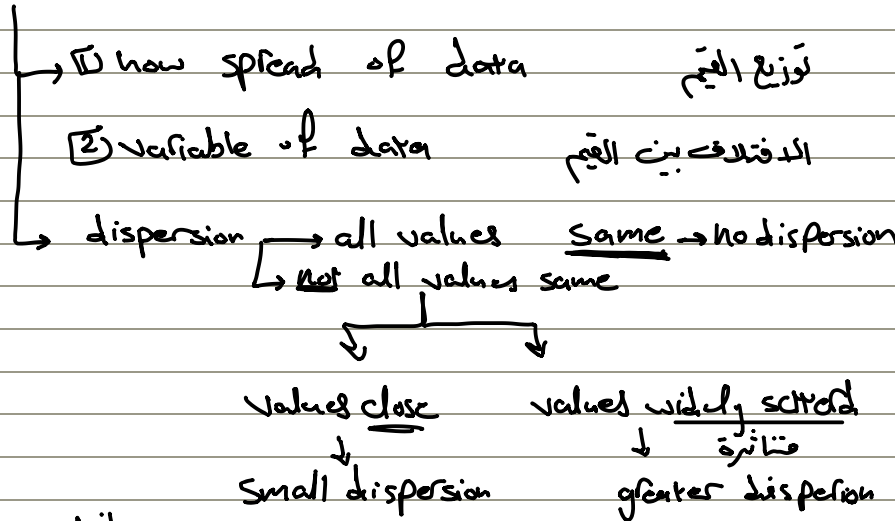


Mean = median = mode

Mean > median > mode

mean < median < mode

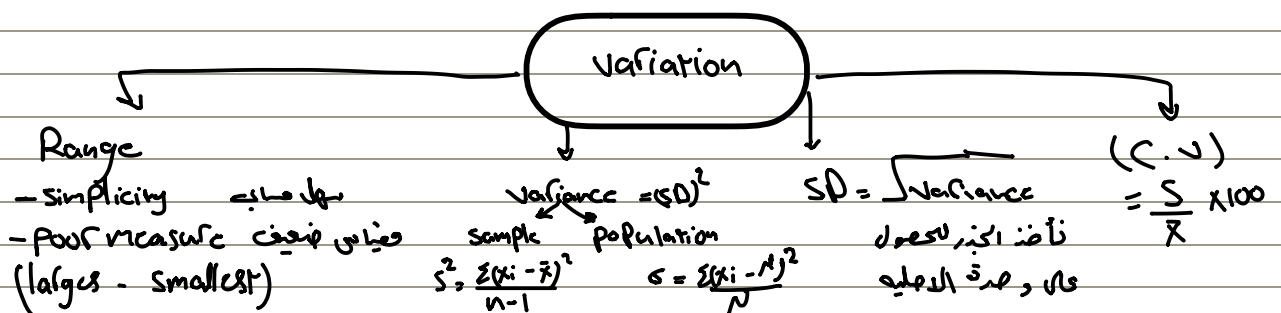
* Measures dispersion use:



سبب اختلاف في كل تجربة

Variation in experiments or in any assay is due to several reasons as

- [?] The instrument used for analysis. → أداة التحليل
- [?] The analyst performing the assay. → المحلل
- [?] The particular sample chosen. → عينة
- [?] Unidentified error commonly known as noise. → خطأ غير عرر



SD :- (A) له وحدة نفس وحدة ال data
(B) رمز (+) ← يستخدم لتجربته (more or less) ← SD

(C) ↑ variation , ↑ SD

لا C :- يستخدم للمقارنة التنت بين مجموعتين من البيانات
use to compare the dispersion in two sets of data

إنه مفيد في مقارنة النتائج التي حصل عليها الأشخاص المختلفون الذين يقومون بإجراء تحقيقات تتعلق بنفس

Unitless ← لأن ناتج قسمة $\frac{SD}{\bar{x}} \times 100$
Unit of $\bar{x}$ = Unit of SD
Ex: 5ml 3ml

$$\frac{3ml}{5ml} \times 100$$

حساب SD في الفترات :-

$$x_i = \frac{\text{lower class limit} + \text{upper class limit}}{2} = \text{midpoint}$$

$$(f \times x_i) = \text{Frequency} \times \text{midpoint}$$

(5) تعويض بالقانون :-

$$(\sum f \times x_i)^2$$

$$x_i^2$$

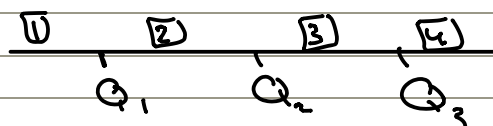
$$SD = \sqrt{\frac{\sum f x^2 - \frac{(\sum f x)^2}{n}}{n-1}}$$

Measures of position:-

are used to describe the relative location of an observation

وصف علاقة الموقع بالبيانات

* Quantiles system → divided to quartiles + ordering



Ex: median, quartiles

Useful where there are quite
→ large number of observation

number of quartiles → number of ranges of data

$$N \longrightarrow N+1$$

$$4 \text{ quartiles} \longrightarrow 4+1 = 5 \text{ ranges}$$

Measures of Position

↓

Quartiles

↓

Percentiles

↓

Five Number Summary

↓

Boxplot

* Measures of position :-

→ percentile

$$L = \frac{k}{100} \cdot n$$

→ located

$$(L + .5)$$

تقریبی اقل

$P_{75} \rightarrow 75\%$ less
75% more

- divide data
4 equal part

25% 25% 25% 25%

Q_1 Q_2 Q_3

P_{25} P_{50} P_{75}

→ Q_1 :- (lower)
below the median
جزئیات سے کم

→ Q_2 :- (median)
نصف و نصف میں
مساوی

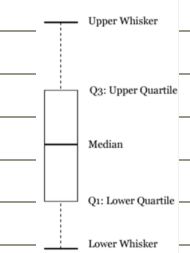
→ Q_3 (upper)
median سے زیادہ

- provide information about
data set Q_1, Q_2, Q_3



* helps :-

- 1) distribution shape
- 2) determine outlier



helps:-

- 1-Comparing between several groups
- 2-Identifying outliers
- 3-Identifying the approximate shape

$$\text{Lower Fence (LF)} = Q_1 - 1.5(IQR)$$

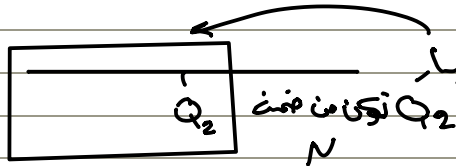
outlier ← (LF) اقل سے

$$\text{Upper Fence (UF)} = Q_3 + 1.5(IQR)$$

outlier ← (UF) زیادہ سے

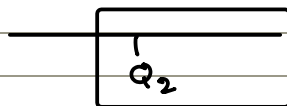
• or * ← outlier

* حساب Q_1, Q_2, Q_3 ← بدون Table



Median = Q_1 ← data set to left

Median = Q_2



Median = Q_3 ← data set to right

* حساب Q_1, Q_2, Q_3 ← بدون Table

$$\frac{3n}{4}, \frac{n}{2}, \frac{n}{4}$$

CP

$$CP > \frac{n}{4} \text{ or } CP = \frac{n}{4} \Rightarrow \text{اولیہ} = Q_1$$

$$CP > \frac{n}{2} \text{ or } CP = \frac{n}{2} \Rightarrow \text{دو لہ فیہ} = Q_2$$

$$CP > \frac{3n}{4} \text{ or } CP = \frac{3n}{4} \Rightarrow \text{تیسریہ} = Q_3$$

* حساب Q_1, Q_2, Q_3 + Table + فنرلے

مقیاسی قری لا فتلاف
Interquartile Range = $Q_3 - Q_1$

Semi-interquartile Range = $\frac{Q_3 - Q_1}{2}$

Midquartile Range = $\frac{Q_3 + Q_1}{2}$

$$Q_1 = L_1 + \left(\frac{\frac{n}{4} - c f_1}{p_1} \right) x_i$$

$$Q_2 = L_2 + \left(\frac{\frac{n}{2} - c f_2}{p_2} \right) x_i$$

$$Q_3 = L_3 + \left(\frac{\frac{3n}{4} - c f_3}{p_3} \right) x_i$$

Normal Distribution

(Quartile 3 - Quartile 2) = (Quartile 2 - Quartile 1)



Positive Skew

(Quartile 3 - Quartile 2) > (Quartile 2 - Quartile 1)



Negative Skew

(Quartile 3 - Quartile 2) < (Quartile 2 - Quartile 1)



Rana alkhawalda

[Handwritten signature]

$$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x}$$

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

$$\sum x_i - \cancel{n} \left(\frac{\sum x_i}{\cancel{n}} \right)$$

$$\sum x_i - \sum x_i = 0$$

$$\frac{(0.5 \times 11) + (1 \times 9) + (2 \times 5) + (3 \times 10) + (4.5 \times 15)}{50}$$

50

5 1 2 5 9 5 4 5 10 11 5 12 5

