Quantitative Variables :-

- Center (measures of center)
- spread (measures of spread)
- shape
- outliers

spread → انتشار بعد النقطة عن النقطة الأخرى

Measures of spread :- How far are points from one another.

- Range
- Interquartile Range (IQR)
- Standard deviation
- Variance

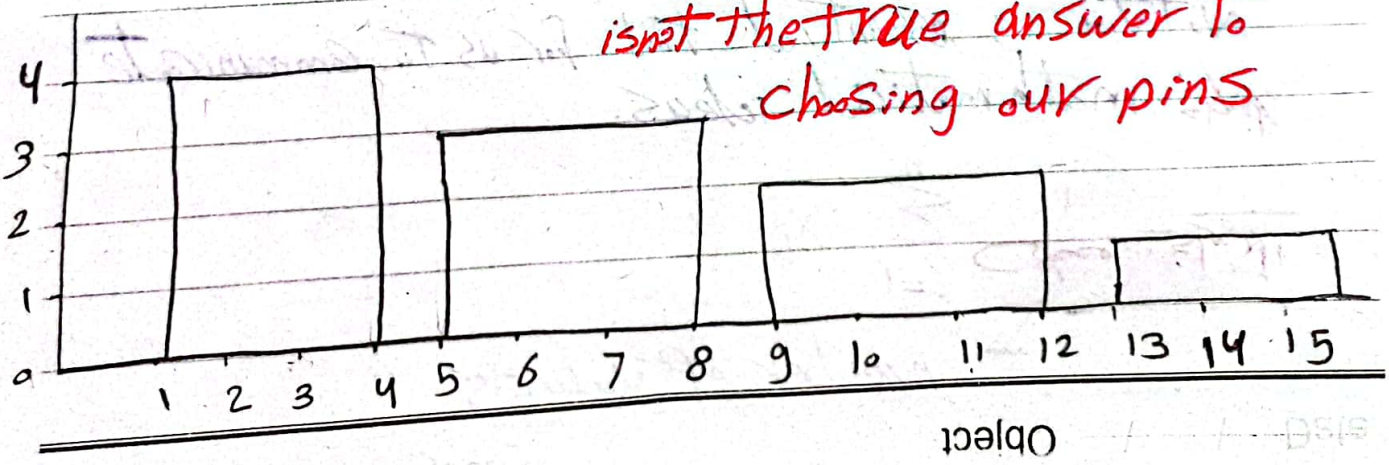
Histogram :- The most common visual for quantitative Data

1, 2, 2, 4 — 4 (1-4)

5, 7, 8 — 3 (5-8)

9, 12 — 2 (9, 12)

15 — 1 (13-16)



is not the true answer to choosing our pins

IRQ

* one of the most common ways to measure the spread is **five number summary**

five number summary:-
- first quartile
- second " (median)
- third "
- minimum
- maximum

5, 8, 3, 2, 1, 3, 10

$Q_1 = 25\%$ → $Q_3 = 75\%$

1, 2, 3, 3, 5, 8, 10

min → 1

Q_2 (median) → 3

max → 10

Range = max − min = 10 − 1 = 9

$IQR = Q_3 - Q_1 = 8 - 2 = 6$

Q_2 (median) → 50% of our Data fall below the middle value

Q_1, Q_3 → The median of the Data on either side of Q_2

25% 75%

Object

Date / /

15-

5, 8, 3, 2, 1, 3, 16, 105

1, 2, 3, 3, 5, 8, 10, 105

min

$Q_1 = \frac{2+3}{2} = 2.5$ $Q_2 = \frac{3+5}{2} = 4$ $Q_3 = \frac{10+8}{2} = 9$

max

$= 2.5$

$IQR = Q_3 - Q_1 = 9 - 2.5 = 6.5$

Box plot:- useful for quickly comparing the spread of two Data set

*The most common way that professionals measures the spread of A data with a single value is:

Standard deviation and variance

Standard deviation:- on average, how much each point varies from the mean of the point

mean

mean $\bar{X}$

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

Object

Date / /

15.7 Data Set

10 6 14 6 10 6

$\bar{X} = 10$ ↳ mean

$X_i - \bar{X} =$

$10 - 10 = 0$

$14 - 10 = 4$

$10 - 10 = 0$

$6 - 10 = -4$

$0 + 4 + 0 - 4 = 0$ X

There is no spread here we should make this values positive

Variance $= \frac{1}{n}\sum_{i=1} (X - \bar{X})^2 = \frac{1}{4}(16 + 16 + 0 + 0) = 8$

$S.D = \sqrt{8} = 2.8$

Variance: average Squared Difference of each observation From the mean

$s.d = \sqrt{\text{Variance}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X - \bar{X})^2}$

S.D used To Compare the spread of Different groups

18.7 Shape How to use Histograms to Determine shape associated with Data

* symmetric distribution = normal distribution = Bell curve

mean = median = mode

* left skewed
mean < median

* Right skewed
mean > median

Quick plot : when you are working with data you can Build A quick plot to see the shape.

outliers :- Data points that fall very far from the rest of the values in our Data set

mean X median → a better measure of center

Standard Deviation X

*when Data includes outlier values S.D & mean became more greater.

Object

Date / /

١٨,٠٣

statistics < descriptive
inferential.

descriptive statistics describing collected Data

inferential statistics Drawing Conclusion ABout A population Based on Data Collected from a sample of individuals from that population

* النتيجة الإحصائية عن Population نتيجة إحصائية التي تعتمد على Sample.

Any numeric summary calculated from the sample is called statistic

Any numeric " of the population is known as a parameter

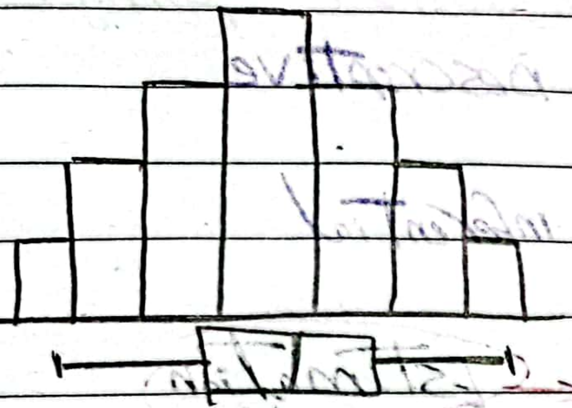
Drawing Conclusion regarding a parameter based on our statistics is known as inference.

Object

١٢٥

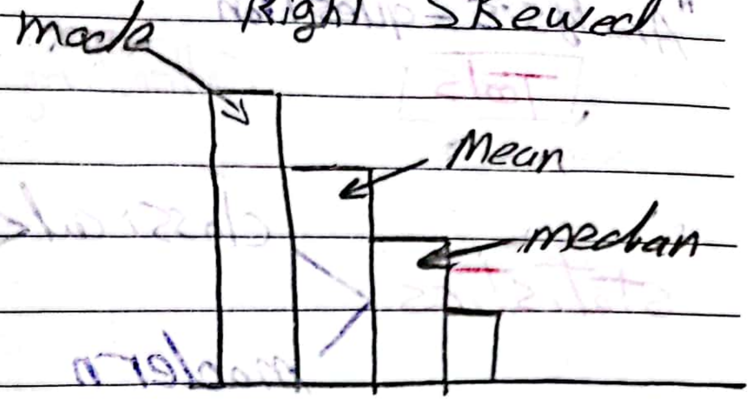
Shapes:-

Symmetrical



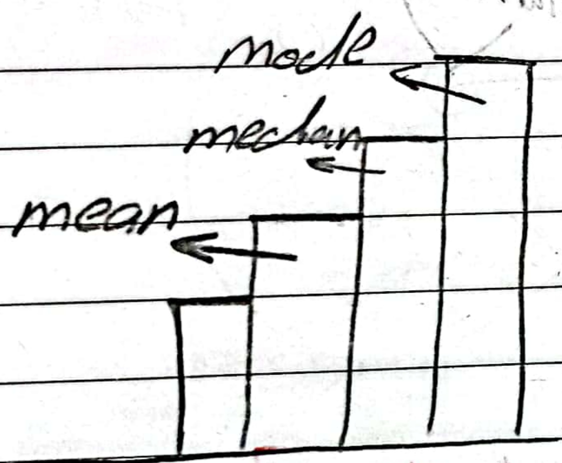
mean = median = mode

"Right Skewed"



Mean > median

left Skewed



mean < median

outlier ⟹ بعيد عن قيم الداتا اللي عندي
بتبقى غريبة عن القيم اللي عندي *

١٢٠٥م

statistics → تحمل الخطأ

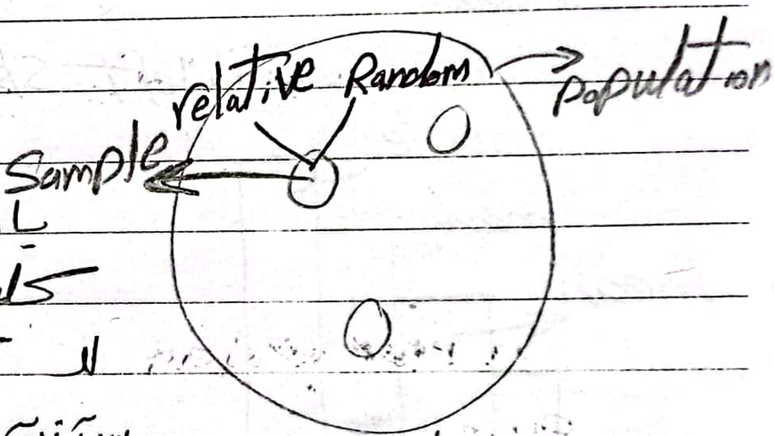
"Analysis = question"

[Tools] Collecting , Visualize , Analyze

statistics < classical < Descriptive / inferential
statistics < modern

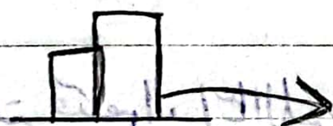
Estimation → انا مبعرفش اوصل للنتيجه بالتحديد بيحاول الاقرب للنتيجه الحقيقيه
بيعمل تقدير للداتا اللي انا بشتغل عليها

× مقدرش ادرس الـ population
لان العدد كبير جداً فانا
باخد عينه تمثل الفئة الكامل
كله لازم تكون عشوائيه وممثله
لـ data كامله .
بستنتج من Sample كل حاجه



[Descriptive statistics]

frequency Distripution عدد التكرارات .

× Histogram  → بحاول انت عدد التكرارات