Statistics:

The Study of The Collection, analysis, it's a mathematical discipline To Collect, Summarize Data

Samples and Population:

it's a Collection of all Possible individuals → Population

A Sample is Portion of The Population of interest → Sample

- Ana bastar Random Sample w bsmchom Case el nas elly bstarha aw el 7alat elly bstrha. → Case

. El 7aga elly 3rftha 3n el Case bta3ty elly estrtha random deh esmha variable → variable

Data
↓ Case ↓ variable

. Cases + Variables = Data Matrix

Ana 3nd y no3en Mn el Data

* Numerical
. Discrete . Continuous
Deh arqam Mofsla (30, 35, 60, 42, ...)
Deh range (30 _ 60)

* Categorical or Qualitative
. Nominal it Can't be ordered and Measured → name
. ordinal it Can be ordered and Measured → Rank

Graphical Presentation of Data: There are Different Types of graphical Presentation:

[1] Dot Plot (Dot Chart) → graphical is display Data using dots x-axis

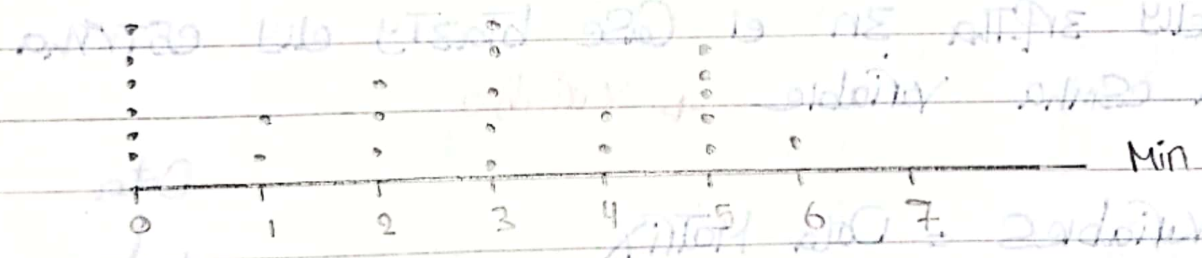
Ex: How long does it Takes To eat breakfast?

Min	0	1	2	3	4	5	6
People	6	2	3	5	2	5	1

1) What's The Sample Size?

Sample Size = 6+2+3+5+2+5+1 = 24

2) Construct a Dot Plot graph?



[2] Bar Graph (Bar Chart) → graphical display of Data using bars rectangle

How To Construct Bar graph?

x-axis → What is being Measured (case)

y-axis → Has a number for amount (Variable)

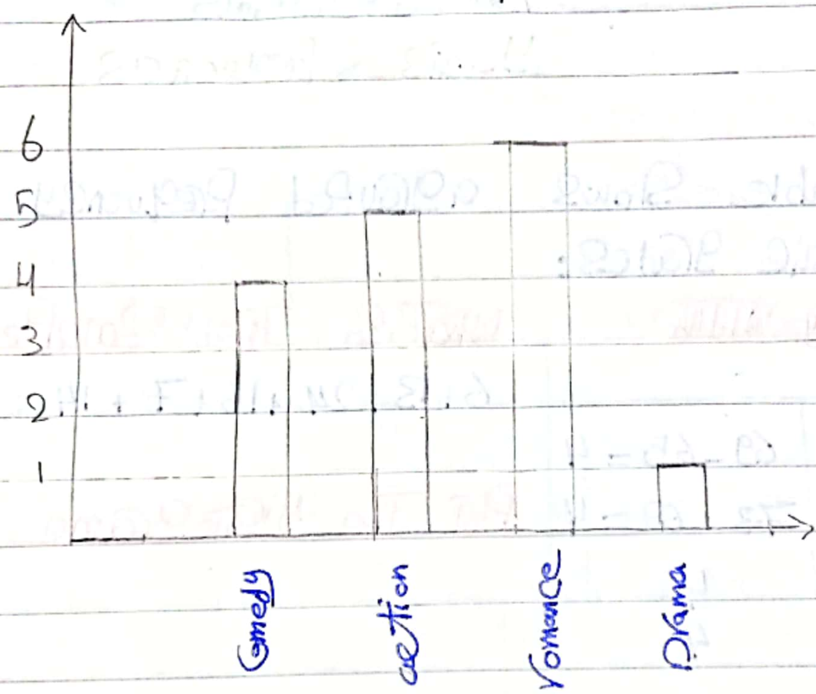
Note → There exist gaps between The bars

Ex:

Movie	Comedy	action	romance	Drame
People	4	5	6	1

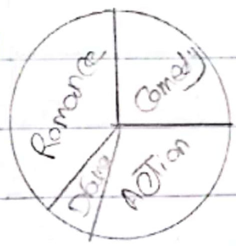
1) Sample Size = 4 + 5 + 6 + 1 = 16

2) Bar graph



3 Pie Chart (circle Chart)

movie	Comedy	Action	Romance	Drama	Total
People	4	5	6	1	16
Relative Size	(4/16)*100 25%	(5/16)*100 31.25%	(6/16)*100 37.5%	(1/16)*100 6.25%	



→ bgeb el relative Size (Sample / Total) *100

4 Histogram → graphical display Data by using bars rectangle and in Histograme There no gaps between The bars.

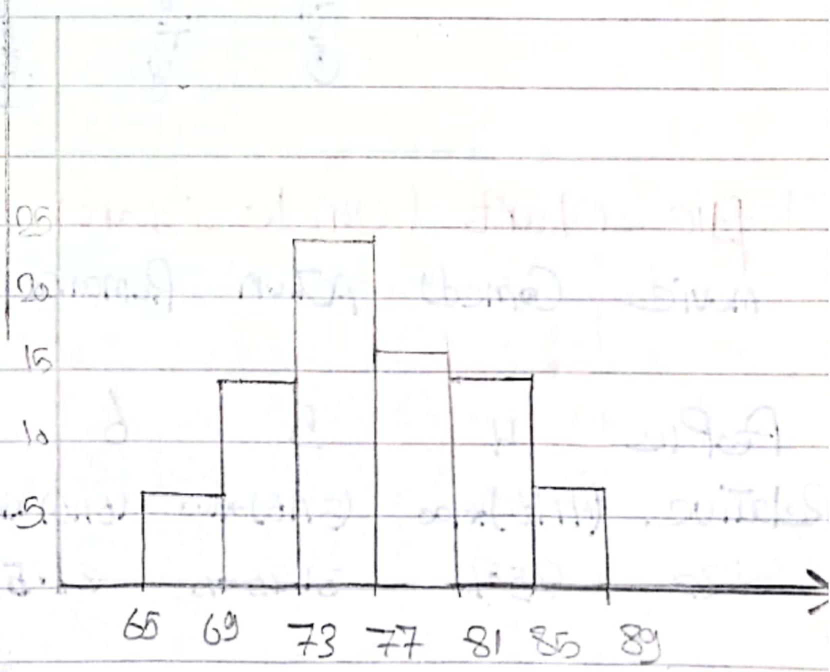
in Histograme we Have Two Cases
1) Histogram for grouped Data:
.intervals with equals ~~width~~ width
X-axis → intervals
Y-axis → frequences

Ex: The following Table Shows a grouped frequency distribution For The Statistic grades:

interval	frequency	width
65 - 69	6	69 - 65 = 4
69 - 73	13	73 - 69 = 4
73 - 77	24	4
77 - 81	16	4
85 - 89	7	4
81 - 85	14	4

What is The Sample Size?
$6+13+24+16+7+14 = 80$

Plot The Histograme:



• intervals with unequal width

X-axis → intervals

Y-axis → hight

Area = hight * width → hight = $\frac{\text{frequency}}{\text{width}}$

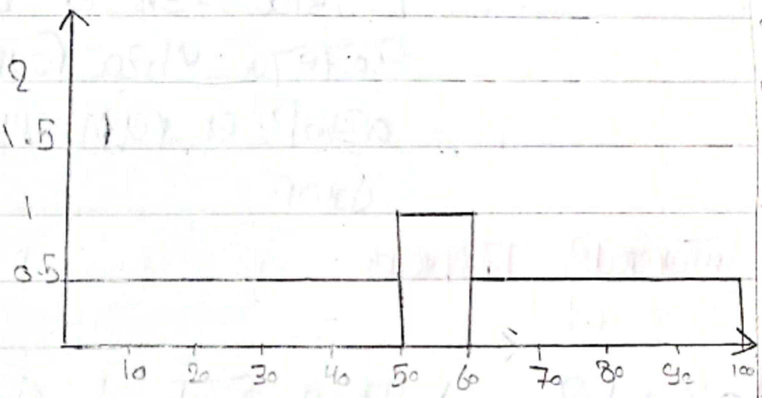
Ex: The following Table Shows a grouped frequency distributio for The Statistics grades:

intervals	frequency	width	Hight → $\frac{F}{w}$
0 - 50	25	50	0.5
50 - 60	10	10	1
60 - 100	20	40	0.5

what is The Sample Size?

25 + 10 + 20 = 55

Plot The Histogram:



2) Histogram for ungrouped Data:

We Should following 4 Steps 3shan a7wl el ungrouped Data, grouped Data w ageb el intervals 3ady.

1- Find The Smallest and largest Data Point.

2- Define Sample Size → n, $K = \sqrt{n}$ (num of intervals)

3- Define width of The intervals = (largest - smallest) / K

4- creat bin boundaries (n3ml el intervals ely hnshtghl behal)

Start with Smallest number or "less" and adding width

and The last intervals Should Contain The largest (5)

Ex: The Following Data represent The length of life in years Measured To The nearest Tenth of 30 Similar fuel Pumps:

~~2.6~~	~~3.0~~	~~0.3~~	~~3.3~~	~~1.3~~	~~0.4~~
~~0.2~~	~~6.0~~	~~5.5~~	~~6.5~~	~~0.2~~	~~2.3~~
~~1.5~~	~~4.0~~	5.9	~~1.8~~	~~4.7~~	~~0.7~~
~~4.5~~	~~0.3~~	~~1.5~~	~~0.5~~	~~2.5~~	~~5.0~~
~~1.0~~	~~6.0~~	~~5.6~~	6.6	~~1.2~~	~~0.2~~

① largest value = 6.5 Smallest value = 0.2

② n = Sample Size = 5x6 = 30 $K = \sqrt{n} = \sqrt{30} = 5.4 = 6$

note: K is The number of interval That will be used in Table

③ L = width of each interval = $\frac{6.5 - 0.2}{6} = 1.05 = 1$

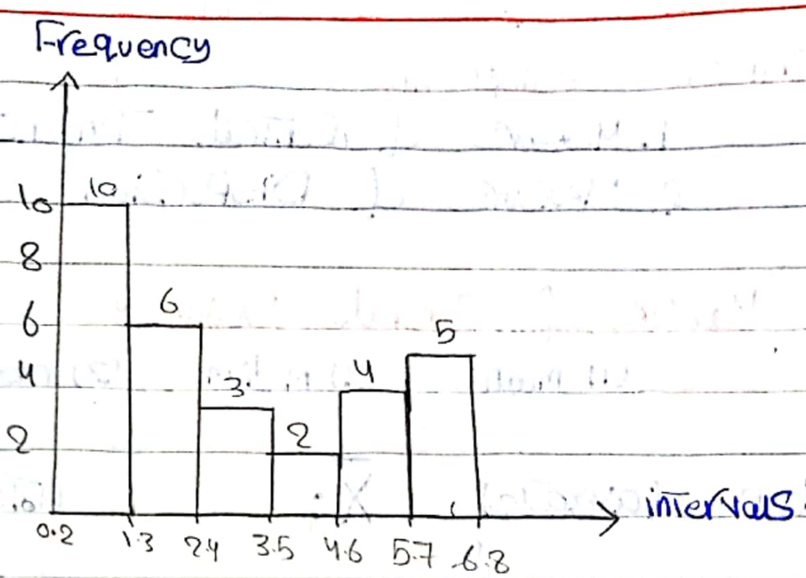
note: K dayman b3miha round up
L 3la 7sb el Data elly 3ndy lw kdkfa Sa7e7a yb2a round up Lw Ksor yb2a ashof el ra2m elly Tis hyt3mlo round up wala down

intervals Smalles + L	Frequency
0.2 - 1.2	
1.2 - 2.2	
2.2 - 3.2	
3.2 - 4.2	
4.2 - 5.2	
5.2 - 6.2 → X	

1. 3shan a7ot el 6 intervals elly 3ndy Hbtdy Mn el smallest value w a7ud 3leha el L(width)
0.2 - 1.2 → awl interval
akml l7d ma5ls kol 6 interval
1.2 - 2.2 → Tany interval l7d m awsl ll largest value b2a wa5ls el intervals el 6

2. la2m kull el intervals elly 3ndy ykon el Smalles w el largest value fehom TYP lw El 6 interval ET7to w lsa mwsltsh ll large aro7 Shayla el Tekrar → 0.2 - 1.2
1.3 - 2.3
2.4 - 3.4

intervals	Frequency	Relative
0.2 _ 1.2	10	0.33
1.3 _ 2.3	6	0.2
2.4 _ 3.4	3	0.1
3.5 _ 4.5	2	0.07
4.6 _ 5.6	4	0.13
5.7 _ 6.7	5	0.17



Relative Frequency = $\frac{F}{n}$ → n = Sample size , F = Frequency

lw 2aly ena 3ayz arsm el relative frequency Hmsh nfs el steps bs bdl M a7ot el frequency 3la y axis H7ot el relative frequency.

Lw 2aly 3ayz el nearest Tenth b2a :

① hgeb el MidPoint = $\frac{\text{lower limit} + \text{upper limit}}{2}$

Mid Point
$(0.2+1.2) \div 2 = 0.7$
$(1.3+2.3) \div 2 = 1.8$
$(2.4+3.4) \div 2 = 2.9$
$(3.5+4.5) \div 2 = 4$
$(4.6+5.6) \div 2 = 5.1$
$(5.7+6.7) \div 2 = 6.2$

② Hmsh. 3la el midPoint btaa3 kull interval w awsilhom bb3d

Descriptive Statistics:

1. Measure of Central Tendency
2. Measure of Dispersion

a) Measure of Central Tendency

(1) mean (2) median (3) mode

[1] Mean (average) $\bar{X}$:

If we have ungrouped Data $\rightarrow \bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

If we Have grouped Data $\bar{X} = \frac{\sum x \cdot f}{n}$

x: (Start + end) / 2

f: frequency

note: el mean byt2sr b kull el 9ym wel extreme values w msh shrt y kon f nos el Data btaSty.

[2] Median: It is The value Splits The Dataset in Half

1) order Your Data from Smallest To largest
2) A7DD el DataSet btaSty even wala odd

a) odd

$$X_{\left(\frac{n+1}{2}\right)}$$

b) even

$$\frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}$$

Ex: 10, 20, 30 $\quad X_{\left(\frac{3+1}{2}\right)} = X_2$

Ex: 10, 20, 30, 40

$$\frac{X_2 + X_3}{2} = \frac{20+30}{2} = 25$$

note: el Median msh byt2sr b kull el Data elly 3ndy wala el extreme values Hwa Mo3tmd 3la el Data elly f el nos

[3] Mode: The mode is The value That occurs The most frequently There Can be one mode, Multiple Mode, no Mode.

Ex: 10, 20, 20, 30, 50 $\quad$ mode = 20 $\quad$ Ex: 10, 20, 30, 40 $\quad$ no Mode

Note : el Mode 3la ek graphe Hwa el A3la frequency
el Mean Hwa elly Sensetive ll extreme value
el Median elly fe el nos been el values elly 3ndy

Mode < Median < Mean → Positively Skewed , right Skewed
Mode > Median > Mean → negatively " , left "
Mode = Median = Mean → Symmetric

When To use the Mean?
Symmatric Distribution, because it include all of The Data in Calculations

When You Have Skwed Distribution → The Median is a better Measure of The Central tendency Than The Mean

When You Have Categories Data → You must use The Mode

Ex: The Length of Time, in minutes That 10 Patients waited in a doctor's offic before to receiving Treatment records as following : 5, 11, 9, 5, 10, 15, 6, 10, 5, 10

5, 5, 5, 6, 9, 10, 10, 10, 11, 15

1) Mean → $\frac{\sum_{i=1}^{n} x_i}{n} = 8.6$

2) Median → even = $\frac{X_{(\frac{n}{2})} + (X_{(\frac{n}{2})+1})}{2} = \frac{X_5 + X_6}{2} = 9.5$

3) Mode → 5, 10 → Most frequency

feh Examples Tanya fe el See.

انتشار البيانات

b) Measure of Dispersion: also called Variability, Scatter, Spread.

(1) Range, (2) interquartile Range, (3) variance

(4) Standard Devision

1) Range (R): The Differance between The largest and The Smallest observation in The Data

note: it very Sensitive To outliers (extreme) and Doesn't use all The observations in a Data Set

Ex: Stat grade (A): 20, 30, 35, 45 $R_A = 45 - 20 = 25$

(B): 1, 30, 35, 45 $R_B = 45 - 1 = 44$

(C): 20, 30, 35, 100 $R_C = 100 - 20 = 80$

Who Had The greater Spread of Marks?

$R_C > R_B > R_A$

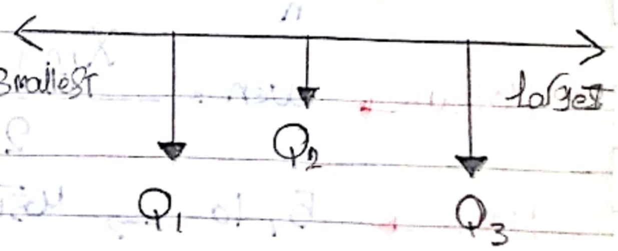
---

2) Interquartile Range (IQR): The IQR Discribes The Middle 50% of observation

note: it not effected by extreme value

1. Put The Data in order from Smallest To largest
2. Find Median of all Data: Q_2
3. Find Median of The Data before Q_2: Q_1
4. " " " " After Q_2: Q_3

$IQR = Q_3 - Q_1$



[3] Variance (S^2) and [4] Standard Deviation (S) :
it measure how far a set of number is spread out from their average value

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \qquad S = \sqrt{S^2}$$

Ex: 9, 2, 5, 4, 12, 7 $\qquad n=6$

Mean $\bar{x} = \frac{\sum_{i=1}^{6} x_i}{n} = \frac{39}{6} = 6.5$

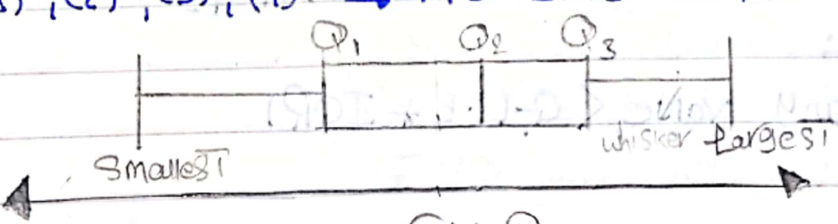
Variance $S^2 = \frac{\sum_{i=1}^{6} (x_i - \bar{x})^2}{n-1} =$

$\sum_{i=1}^{6} (x_i - \bar{x})^2 = (9-6.5)^2 + (2-6.5)^2 + \dots + (7-6.5)^2 = 65.5$

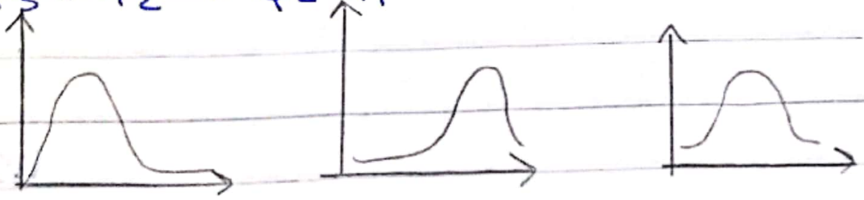
$S^2 = \frac{65.5}{5} = 13.1 \qquad$ Standard devesion $= \sqrt{13.1} = 3.6$

Box Blot: a way to show the spread and skewness of Data Set.

Steps: (1), (2), (3), (4) → nps steps IQR



$Q_3 - Q_2 > Q_2 - Q_1$ +ve, right skew
$Q_3 - Q_2 < Q_2 - Q_1$ -ve, left skew
$Q_3 - Q_2 = Q_2 - Q_1$ Symmetric



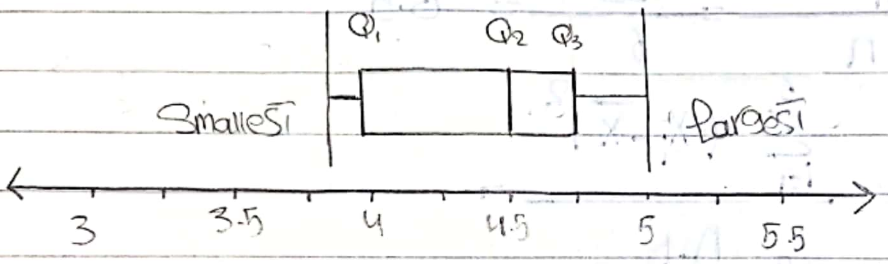
outliers: any value $> Q_3 + 1.5\, IQR$
any value $< Q_1 - 1.5\, IQR$

Ex: Draw the box plot for the following set of data:
4.7, 3.8, 3.9, 3.9, 4.6, 4.5, 5

1) 3.8, 3.9, 3.9, 4.5, 4.6, 4.7, 5

2) $Q_2 = 4.5$, $Q_1 = 3.9$, $Q_3 = 4.7$, The largest = 5

The Smallest = 3.8



$Q_3 - Q_2 = 4.7 - 4.5 = 0.2$ ∴ $Q_3 - Q_2 < Q_2 - Q_1$ left skew

$Q_2 - Q_1 = 4.5 - 3.9 = 0.6$ −ve

outliers: $IQR = Q_3 - Q_1 = 4.7 - 3.9 = 0.8$

$1.5 * IQR = 1.5 \times 0.8 = 1.2$

any value $> (1.5 * IQR) + Q_3$

∴ any value > 5.9

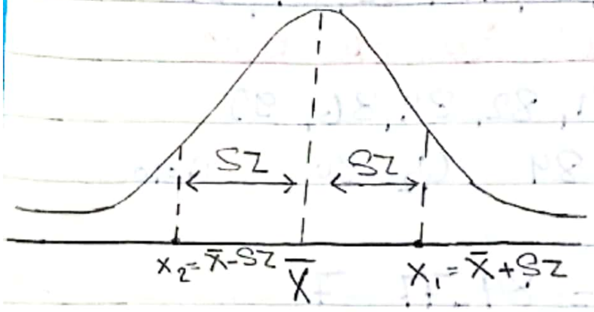
any value $< Q_1 - (1.5 * IQR)$

∴ any value < 2.7

∴ No outliers

Z-SCORE: How Many Standard devision above or below The Mean it is.

$\bar{X} \rightarrow$ Mean
$S \rightarrow$ Standard Devision



let: $x = \bar{x} + SZ$

$x - \bar{x} = SZ$

$\therefore Z = \dfrac{x - \bar{x}}{S}$

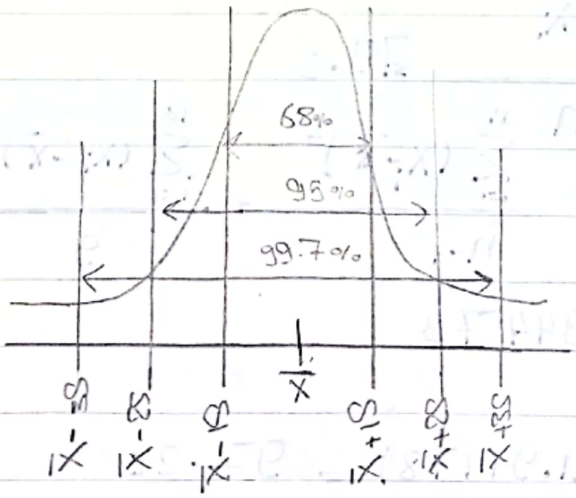
Z Score if $= 0 \rightarrow x = \bar{x}$

Z-Score if Positive $\rightarrow x > \bar{x}$

" " negative $\rightarrow x < \bar{x}$

Empirical Rule: almost all observed Data will fall within Three Standard deviation of The mean

- 68% of The observations fall with in +/- 1 Standard Devision from The mean
- 95% $\rightarrow$ Fall with in +/- 2 Standard Devision from $\bar{x}$
- 99.7% $\rightarrow$ Fall with in +/- 3 Standard Devision from $\bar{x}$



Ex: Student grades on a Chemistry exam were:

77, 78, 76, 81, 86, 51, 79, 82, 84, 99

1) Are there Any outliers? if So which Score are They?

51, 76, 77, 78, 79, 81, 82, 84, 86, 99

~~$Q_3=(76+77)/2$~~ $Q_1 = 77$ $\quad Q_3 = 84$ $\quad Q_2 = 80$ $\quad n = 10$

$IQR = Q_3 - Q_1 = 84 - 77 = 7$

$IQR * 1.5 = 10.5$

$\therefore$ Any value $> Q_3 + 1.5IQR$

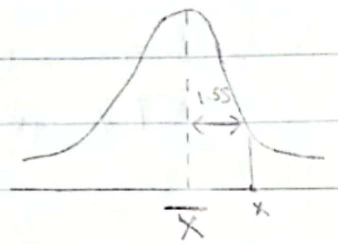
Any value $> 94.5 \rightarrow$ outliers 99

$\therefore$ Any value $< Q_2 - 1.5IQR$

Any value $< 66.5 \rightarrow$ outliers 51

2) What would be The Score That exceeds The mean by 1.5 Standard Deviations?

$X = \bar{X} + ZS$



$X = \bar{X} + 1.5S$

$\bar{X}$ = average = $\dfrac{\sum_{i=1}^{n} x_i}{n} = 79.3$

$S = \sqrt{S^2}$, $\quad S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \dfrac{\sum_{i=1}^{10}(X_i - \bar{X})^2}{9} = 142.68$

$S = \sqrt{142.68} = 11.94478$

$X = 79.3 + (1.5)(11.94478) = 97.22$

CORRELATION: Show The Strength and The Direction of a relationship between Two Variables and is expressed numerically by The Correlation Coefficient "r"

[I] Algebraic: . Correlation be assumed To be linear (Following line)

$-1 \leq r \leq 1$

$r = 0$ is no Correlation

r = Sign Value

Sign → Direction: +ve (علاقة طردية), -ve (علاقة عكسية)

Value → Strength: Perfect (1), Strong (0.7, 0.8, 0.9), Moderate (0.4, 0.5, 0.6), weak (0.1, 0.2, 0.3)

Lma Tgely el So2al la2m a2ol el Direction wel Strength F nfs el w2t: Direct Moderat, inverse weak, +ve Perfect

$$r = \frac{\sum Z_x Z_y}{n-1}, \quad Z_x = \frac{X - \bar{X}}{S_x}, \quad Z_y = \frac{Y - \bar{Y}}{S_y}$$

X	Z_x	y	Z_y	$Z_y Z_x$

→ la2m a3mel el Table dah f ay msa2la Correlation

Ex: Calculate The Correlation Coefficient For The Following Data:

$\bar{X} = \frac{\sum_{i=1}^{6} X_i}{6} = 5$ $\quad S_x = 4.3818$

$\bar{y} = \frac{\sum_{i=1}^{6} y_i}{6} = 4$ $\quad S_y = 2.2804$

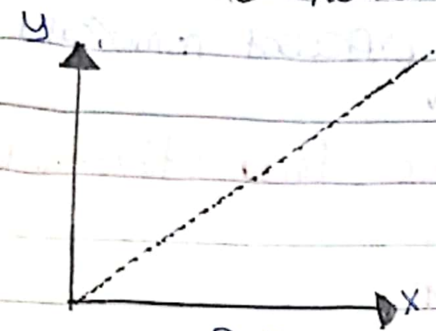
$Z_x = \frac{X - \bar{X}}{S_x}$ $\quad Z_y = \frac{y - \bar{y}}{S_y}$

$Z_x Z_y$	Z_y	Z_x	X	y
-2.10	-1.3	1.5	12	1
0.9008	1.3	0.68	8	7
0	0	0	5	4
-0.4003	0.87	0.45	3	6
0	0	0.68	2	4
1.0007	-0.87	1.14	0	2
-0.6005				

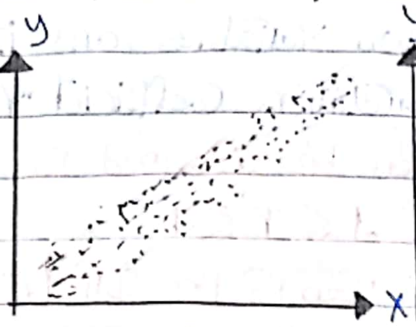
negative weak

$r = \frac{-0.6005}{5} = 0.1201 \simeq -0.1$

2 Graphical Method (Scatter Diagram)

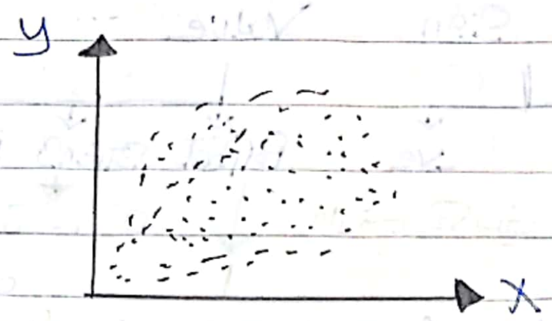


+ve Perfect

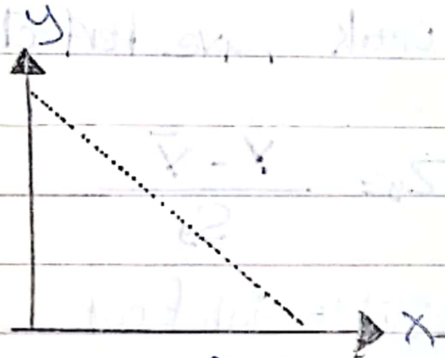


+ve Strong

+ve moderate



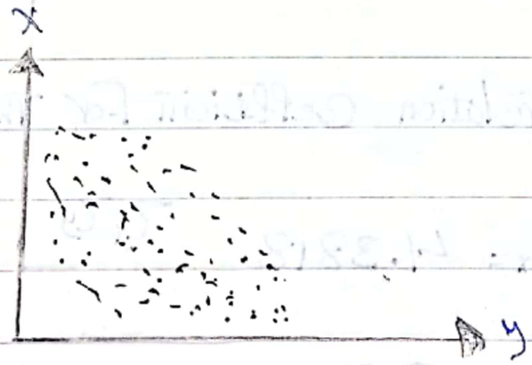
+ve weak



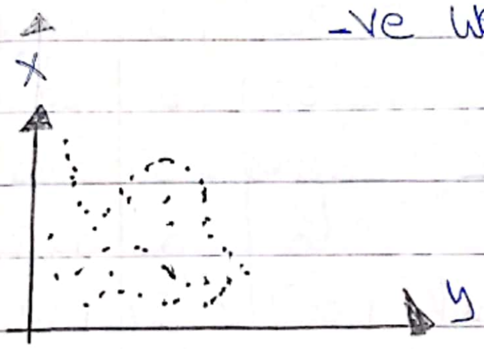
-ve Perfect

-ve Strong

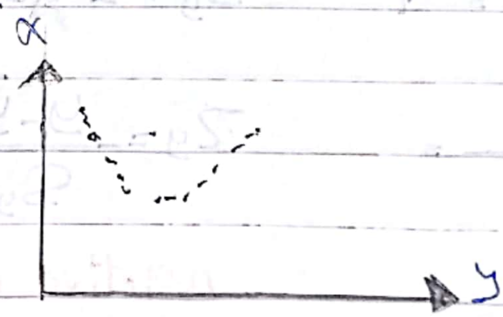
-ve moderate



-ve weak



no Correlation



non linear Correlation

Simple Linear Regression: It's used to describe or predict the relationship between two variables

$$y = B_0 + B_1 x + e \rightarrow B_1 = r\left(\frac{S_y}{S_x}\right)$$

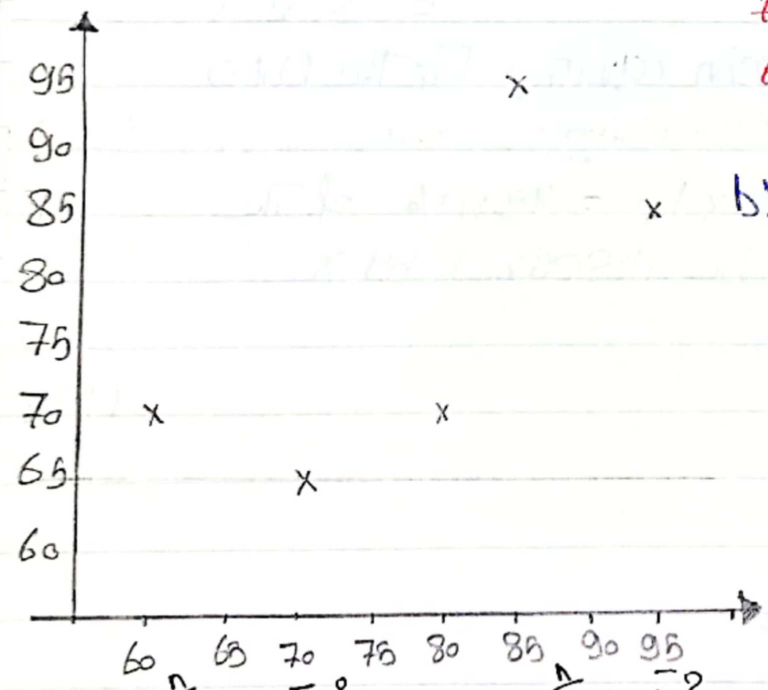
y = The dependent variable
x = The independent variable
B_0 = Is the y-intercept of the regression line
B_1 = Is the slope

Ex:

X_i	Y_i	Z_x	Z_y	$Z_x Z_y$
95	85	1.258	0.6376	0.8022
85	95	0.5182	1.4343	0.7433
80	70	0.1480	-0.5578	-0.0826
70	65	-0.592	-0.9562	0.5663
60	70	-1.352	-0.5578	0.7432

Total = 2.7724

a) Draw the scatter plot representing the data:



b) What is the linear regression equation best predicts statistics performance based on math aptitude scores?

$$\hat{y} = B_0 + B_1 x + e$$

$$B_1 = r\left(\frac{S_y}{S_x}\right), \quad B_0 = \bar{y} - B_1 \bar{x}$$

$$S_y = \frac{\sum_{i=1}^{n}(y - \bar{y})^2}{n-1} \qquad S_x = \frac{\sum_{i=1}^{n}(x - \bar{x})^2}{n-1} \qquad r = \frac{\sum Z_x Z_y}{n-1}$$

$n = 5 \quad \bar{y} = 77$
$S_y = 12.5499$

$n = 5 \quad \bar{x} = 78$
$S_x = 13.5093$

$r = \frac{2.7724}{4}$
$r = 0.6931$

$B_1 = 0.6931 * \frac{12.5499}{13.5093}$
$B_1 = 0.644$
$B_0 = 77 - (0.644 * 78)$
$B_0 = 26.768$

$$\hat{y} = 26.768 + 0.644x$$

c) if a student made an 80 on the aptitude test, what grade would we except her to make in statistics:

at $X = 80$

$$\hat{y} = \beta_0 + \beta_1 X$$

$$\hat{y} = 26.768 + 0.644 * 80 =$$

$$\hat{y}_{80} = 78.288$$

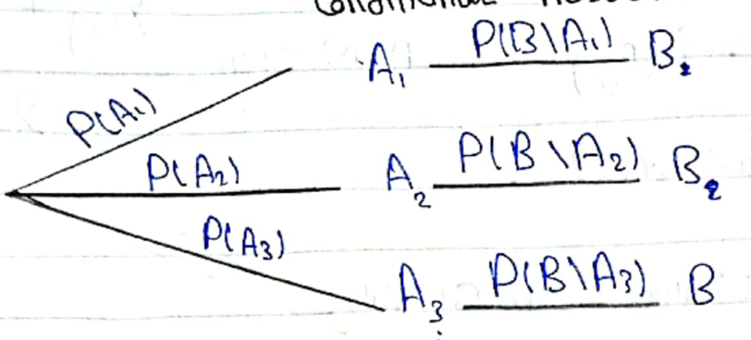
$$\text{Error} = |\hat{y} - y| = |78.288 - 70|$$

$$= 8.288$$

d) How well does the regression equation fit the data?

$r^2 = (0.6931)^2 = 0.4804 \times 100 = 48.04\%$ of the variation in y can be described by x

Bayes' Theorem: is a mathematical formula for determining Conditional Probability



$P(A_1 \cap B) = P(A_1) P(B \setminus A_1)$

$P(A_2 \cap B) = P(A_2) P(B \setminus A_2)$

$P(A_3 \cap B) = P(A_3) P(B \setminus A_3)$

$P(B) =$ Sum

$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$

$P(B) = P(B \setminus A_1) P(A_1) + P(B \setminus A_2) P(A_2) + P(B \setminus A_3) P(A_3)$

Ex: All tractors made by a Company are Produced on one of Three assembly lines, named Red, Blue, White. The Chanes That a Tractor will not Start when it rolls off a line are 6%, 11%, 8% for lines Red, Blue, White, Respectively. 48% of The Company's Tractors are made on The Red line and 31% are made on The Blue line.

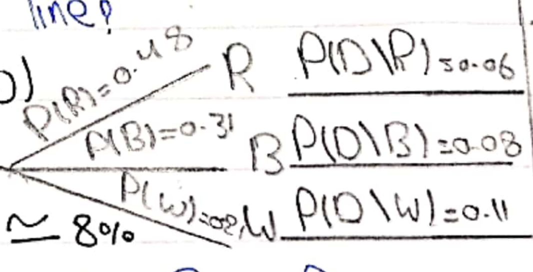
Red line: R, Blue line: B, White line: W, Not Start: Defective D

a) What Fraction of The Company's Tractors Don't Start when They roll off an assembly line?

$P(D) = P(R \cap D) + P(B \cap B) + P(W \cap D)$

$0.0288 + 0.0248 + 0.0231 =$

$100 * 0.0767 \simeq 7.67 \simeq 8\%$



$\cap$
0.0288
0.0248
0.0231
Total = 0.0767

b) What is The Probability That a Tractor Came From The Red Company Given That it was defective?

$$P(R \setminus D) = \frac{P(R \cap D)}{P(D)} = \frac{0.0288}{0.0767} = 0.3755$$

feh b2et msa2l f el Section

Sec 5,6,7,8

- Bionomial Distribution:

.To Study the number of Successes in a Sequence of n independent experiments.

$$f(x) = P(X=r) = \binom{n}{r} p^r q^{n-r}$$

$E(x) = np$

$var(x) = npq$

- normal Distribution:

- Data Tends To be around a Central value with no bias left or Right → It get Close To normal Distribution

$$Z = \frac{X-\mu}{S}$$ → 3mlna Convert "X" mn normal Distribution To Standard normal "Z"

note: if $P(Z<a)=1 \therefore a > 3.49$

if $P(Z<a)=0 \therefore a < -3.4$

- Central limit theorem:

X is normal Distribution ← . Sample Distribution of Sample Mean

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

.lma tgely Masala w msh 3arf emta astbdm anhy law bta3 el mean w normal Dis ashof el givens elly f el Soal w el 7aga elly talbha katb

$$Z = \frac{X-\mu}{S}$$

- Probability of random
- Probability of Mean of Random

- Sample Distribution of Difference Between Two Means

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Sample Distribution of The Sample Proportion ← X is Bionomial Distribution

$$Z = \frac{\hat{P} - P}{\sqrt{\frac{qp}{n}}}$$

$\mu_{\hat{p}} = P$

$\sigma_{\hat{p}} = \sqrt{\frac{Pq}{n}}$

Note: The Sample Distribution of Sample Proportion $\hat{P}$ is approximately normal if $nP \geq 15$ and $nq \geq 15$

- Sampling Distribution of Difference Between Two Pro..

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

- Confidence Intervals (CI) of The Population Mean "μ"

CI: $\bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{x}}$ there is A Confidence interval for mean

Step1: $n, \bar{X}, \sigma$

Step2: $1 - \alpha = \ldots\%$

Step3: if σ known → Z-Table $CI = \bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{x}}$

if σ unknown → if $n < 30$ → T-Table

if $n \geq 30$ → Z-Table

• Confidence intervals of The Difference between Two Population means

$\bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{X}}$

Step1: n_1, n_2 $\bar{X}_1, \bar{X}_2$ σ_1, σ_2 or S_1, S_2

Step2: $1 - \alpha = \_\%$

Step3: if σ_1, σ_2 known → use Z-table

$$\bar{X}_1 - \bar{X}_2 \mp Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

if σ_1, σ_2 unknow

- $\sigma_1 = \sigma_2$ → T-Table
- $\sigma_1 \neq \sigma_2$
  - $n_1, n_2 > 30$ → Z-Table
  - $n_2, n_1 < 30$ → T-Table

$$\bar{X}_1 - \bar{X}_2 \mp T_{\frac{\alpha}{2}, V} \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

note: $V = V_1 + V_2 = n_1 - 1 + n_2 - 1 = \boxed{n_1 + n_2 - 2}$

Pooled Variance S_p^2:

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

- Confident interval of The Population Proportion $\hat{P}$

$$CI = \hat{P} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{q}\hat{P}}{n}}$$

Step 1: $q, n, P, \hat{P}$ $\rightarrow \hat{P} = \frac{X}{n} \rightarrow X$ num of Defective

Step 2: $1-\alpha = -\%$

Step 3: Calculate CI: $\hat{P} \pm Z_{\frac{\alpha}{2}} \sigma_{\hat{P}}$

- Confidence interval of The difference between Two Population Proportions:

$$CI = (\hat{P}_1 - \hat{P}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}$$

Step 1: $P_1, P_2, q_1, q_2, n_1, n_2, X_1, X_2$

Step 2: $1-\alpha = -\%$

Step 3: Calculate CI: $(\hat{P}_1 - \hat{P}_2) \pm Z_{\frac{\alpha}{2}} (\sigma_{\hat{P}_1} - \sigma_{\hat{P}_2})$