Page No. :__

-0-		decement	
Date	:	/	

- STATISTICS Formula Sheet - ADITYA

Sample mean $\overline{x} = x_1 + x_2 + x_3 + \dots + x_n$

Population mean $\overline{u} = x_1 + x_2 + \dots + x_N$

Adding a constant > old mean + constant = new mean Multiplying a constant > old mean X constant = new mean

Mean !- Four grouped data [frequency is given.

 $\overline{\chi} = f_1 \chi_1 + f_2 \chi_2 + \dots + f_n \chi_n = \sum_{i=1}^n f_i \chi_i$

where n = fi+f2+f3+---+fn

Mean i- Foor grouped data [Class interval is given]

 $\overline{N} = fim_1 + f_{2m_2} + - - - - + f_{nmn} - \sum_{i=1}^{n} f_{imi}$

where, $n = f_1 + f_2 + - - - + f_n$ and m = mid point of interval e.g. m[60-70] = 65.

Mean is sensitive to outliers.

Mode: - Observation with heighest forequency
[Most forequent value of the data set]

Adding a constant \Rightarrow old mode + C = new mode. Multiplying a constant \Rightarrow all mode $\times C = new mode$.

Page No	.:		7
Date :	/	1	7

Median !	- Middle val	lue of -	the date	set	
				oordered	fagin

$$\left[\frac{n+n}{2} + \frac{n+1}{2} \right]$$
 observation

Metian is not sensitive to outliers

Range > Max value of data set - Min value of data set

Adding a constant > old range = new range

Multiplying a constant > old range x C = new range

Range is sensitive to outliers.

Variance i- Variability / Spread of data set.

Population Vay $(\sigma^2) = (\chi_1 - \mu)^2 + (\chi_2 - \mu)^2 + - - - + (\chi_W - \mu)^2$

Sample Var
$$(s^2) = (x_1 - u)^2 + (x_2 - u)^2 + - - - + (x_1 - u)^2$$

Adding a constant \Rightarrow odd variance = new variance. Multiplying a constant \Rightarrow odd variance $\times c^2 = new$ variance

1	Date: / /
	Standard deviation ! - Measure of spread of data in the
	same unit as osiginal data.
dian	1 - 1 - 1 - Arte Parlatile and Ox non accomplished to the
	SD = I Variance
Mar. and	Set The Burne 1999 - or all burned to the Bull of Burne Differ To the
	Adding a constant > old SD = new SD
	Multiplying a constant ⇒ old SDXC = new SD
	Pencentile: For computing pencentile, finst we have
	to arrange the data in increasing order.
	n = total no. of observations, p = percentile
a Miles	CRETE TO MAND NOT BUT BUT TO TO TO THE PARTY OF THE PROPERTY O
	Computing percentile = np
-	What properties win possensite outfolder, and the
(3.20)	If np is an integer then the average of (npth+(np+1)th) obser
	is the snequested percentile value.
	If no is not an integer then the smallest integer
(4.0.10)	greater than np. The data value in that position is
2.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4	the orequired percentile.
14	Ex!- 38, 35, 61, 68, 66, 70, 68, 47, 79, 58
A Dr.	Increasing order :- 35,38,47,58,61,66,68,68,70,79.
	3
	For 25th percentile, $n = 10$, $\rho = 0.25$
	=> np = 10x 5 = 2.5 [not an integor]
	So, the smallest integer greater than 2.5 is 3 then
	25th percentile = 3 rd observation = 47
	(B-iR) (E-ix) = (RIK) NO)-2000005
	Foor, 50th percentile, n=10, p=0.5, np=5[Integer]
	So, 50th percentile = 5th obser. + 6th obser 61+66 _ 63.5

	Date: / /
ent y	The Five Number Summary!
E	Minimumies as fixed across
	Q1 :- first Quartile ou lower quaritile - 25th percentile
	Q2: Second Quartile on Median - 50th percentile.
	Bz: Third Quartile ou upper quartile - 75th percentile
	Maximum = 0 1 1 1 4 Ayoung out to
	Majuan = 12018 blo & frames a graphy Mi
	The Interquartile Range !-
aven!	ou trait aliteraring principality percentillar fibrate and
tra L	IBR = Q3-Q1 = 75th perc 25th perc.
	elebration = 9 . Reducer Factor for the State of any
	Outliers '- Q3+1.5 IQR < Outliers, Q1-1.5 IQR > Outliers.
	SCI = Cottone made of the many of the
6.84	Row relative torequency ton Contingency table!- Divide each cell torequency in a now by its now
io -	Divide each cell forequency in a now by its now
	total. suite directing breature of is
F.3	Column relative frequency for Contingency table:
dil	Divide each cell forequency in a column by its column
	total.
	Covariance - quantifies the storength of the linear
10.	association between two numerical variables.
	N
	Population $Cou(x,y) = \sum_{i} (x_{i} - \overline{x})(y_{i} - \overline{y})$
Treas	the man and e.c. = 2.5 tool = an a

N

Sample - Cov $(x_iy) = \sum_{i=1}^{\infty} (x_i - \overline{x})(y_i - \overline{y})$

n-1

1 h	
	Conselation !- The correlation measure always lies
7	between -1 and +1.
v - 1911	FRANKET IN A MANAGER NAME OF MARKET OF THE MARKET AND SORIO
	$\sum_{i} (x_i - \overline{x}) (y_i \overline{y})$
	i=1 $i=1$ $Cov(x,y)$
	$\int_{i=1}^{\infty} (x_i - \overline{x})^2 \times \int_{i=1}^{\infty} (y_i - \overline{y})^2 $ $S \times S y$
1. popl	where, $Sx = standard deviation of and Sy = SD of y.$
	Association between categorical and numerical variables!
Annean en a de participa de la constante de la	Point Bi-Serial Correlation Coefficient:
4	$\mathcal{F}_{Pb} = \left(\frac{\overline{Y}_{o} - \overline{Y}_{1}}{S_{x}}\right) \sqrt{P_{o} P_{1}}$
	Where, $S_x = 8$ tondard deviation of numerical variable
The control of the co	Yo = mean value of group affata associated with. O
	Yz = mean value of genoup of data associated with I.
Calcar	$p_o = no. of observations associated with 0 total no. of observations$
16 IV	p = no. of observations associated with 1
	total no. of observations
10	water strain of the decomposition was the readour of with the
Por	When on is closed to 0 -> no association
	When sipp is closer to 0 -> no association when sipp is closer to -1 -> negatively associated (storongly When sipp is closer to 1 -> positively associated
	' V