29.8.2023

The main purposes of statistics are classified under three headings : description, analysis and prediction. Individual data are not important but they are considered as a means to measure a certain physical property of interest, the test of hypothesis on the prediction of future occurences under given conditions.

Whatever the final objectives of the experiment; statistical methods are techniques of inductive inference in which a particular set (or sets) of data - " the so-called realization of the sample" is used to draw inferences of general nature on a "population" under study.

Problems for which erroneous results occurs → So we need to be cautions while collecting data

Inductive inference drawn from incomplete information may be wrong even if the original information is not. In the field of statistics this possibility is often related to the process of data collection on the one hand and on the other hand, to the fact that we can only make probabilistic statements or predictions. It is evident that insufficient or biased or / and the failure to consider an important influencing factor in the experiment may lead incorrect conclusion.

Steps to follow

- State the problem
- analyze the data (after collecting data)
- interprete " " and make decision
- implementing & varifying decisions
- Plan for next decision/step

Essential steps to solve problems in industry and business :

1. State the problem or question
2. collect & analyze data.
3. Inteprete the data and make decision
4. Implement & varify decisions.
5. plan next action.

There are two types of data :

→ quantitative (numerical)

→ qualitative (categorical)

Data in the form of numerical measurements or counts are referred to an quantitative data.

Data in the form of classifications into different groups or categories are referred to as qualitative data.

sample, $\bar{x}_1 \xrightarrow{\text{size}} n_1$

$\bar{x}_2 \xrightarrow{"} n_2$

Pg 44 → Measures of center

* Mean → $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ / $\bar{x}_w = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$ ← weighted avg
* Median → data set has to be ordered
* Standard deviation

5.9.2023

Q

				67	73	79	80	83	81
60	63	64	71				80	77	78
86	90	96	98	98	99	89			
						90	78	79	79
71	79	74	84	85	82				

Summarize the data into a freq. distribution table and calculate mean, median & mode.

60 63 64 67 71 71 73 74 77 78 78 79

79 79 79 80 80 81 82 83 84 85 86 89

90 90 96 98 98 99

listing

Taking 2nd row,

X	frequency
77	1
78	1
80	1
86	1
89	1
90	1
98	1
98	1
99	1

Mean $(\bar{X}) = \frac{\sum_{i=1}^{10} X_i}{n}$ $[n=10]$

Median $= \frac{89+90}{2} = 89.5$

Mode $= 98$

যে event maximum times occur করে।

- variance $(\sigma^2 \text{ or } s^2) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
- Standard deviation $= \sigma = s = \sqrt{\sigma^2} = \sqrt{s^2}$

class → estimated mean
range এ কাজ করে " median
" mode

Grouping

class	f frequency	x_i Midpoint	fx_i	Cumulative freq.
60 – 64	3	62	186	3
65 – 69	1	67	67	4
70 – 74	4	72	288	8
75 – 79	~~8~~ 7	77	~~616~~ 539	15
80 – 84	6	82	492	21
85 – 89	~~2~~ 3	87	~~174~~ 261	24
90 – 94	2	92	184	26
95 – 99	4	97	388	30
	$n = 30$		$\Sigma = 2405$	

$$\text{Mean } (\bar{X}) = \frac{\sum_{i=1}^{n} fx_i}{n} = \frac{2405}{30} = 80.17$$

$\bar{X}$ = Mean = 80.17

Modal Class - 75 – 79 (As we find max freq. in this class)

Mode - 79 (frequency = 4)

$$\text{Estimated median} = L + \frac{\left(\frac{n}{2} - cf\right)}{f} \cdot w$$

L = lower class boundary of the group contain

$$\text{Estimated mode} = L + \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times w$$

L = lower class boundary of the modal class

f_{m-1} = freq. of the group before modal class

f_m = " " " modal class

f_{m+1} = " " " group after modal class

$L = 75$

$f_m = 7$

$f_{m-1} = 4$

$f_{m+1} = 6$

$w = 5$ (group width)

$$\text{Estimated mode} = 75 + \frac{7-4}{(7-4)+(7-6)} \times 5$$

$$= 78 \cdot 75$$

$n = 30$ = total number of values.

Median class 75 - 79

cf = cumulative f. just before the median class

f = f. of median class

$$\text{Estimated mean} = 75 + \frac{(15-8)}{7} \times 5$$

$$= 80$$

mean sqr. deviation $\longrightarrow$ variance 12.9.2023

class	midpoint (x_i)	freq	$x_i - \bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
60 - 64	62	3	-18.17	330.15	990.45
65 - 69	67	1	-13.17	173.45	173.45
70 - 74	72	4	-8.15	66.42	256.68
75 - 79	77	7	-3.15	9.92	69.44
80 - 84	82	6	1.83	3.35	20.1
85 - 89	87	3	6.83	46.65	139.95
90 - 94	92	2	11.83	139.95	279.9
95 - 99	97	4	16.83	283.25	11.33
					3062.97

Mean, $\bar{x} = 80.17$

Mode $= 79$

Median $= 80$

variance, $s^2 / \sigma^2 = \dfrac{\sum_{i=1}^{n} f(x_i - \bar{x})^2}{n-1}$

Standard variance, $\sqrt{s^2} = \sigma$

mean square deviation $\therefore$ variance, $s^2 = \sigma^2 = \dfrac{3062.97}{30-1} = 105.62$

standard deviation, $\sigma = \sqrt{s} = 10.28$

* Range, R = largest measurement − smallest measurement

Quartiles (we can use cumulative freq to determine this)

→ 1st quartile − 25% of values are at or below the lower quartile (Q_L)

→ 2nd " / Middle (50% or below)

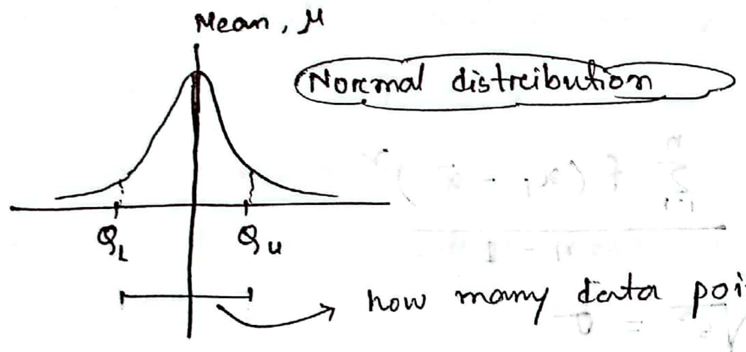
→ 3rd " / upper (75% or below) (Q_U)

Percentiles :

কত % value 'x' value এর নিচে ?

Interquartile range :

$$\boxed{IQR = Q_U - Q_L}$$

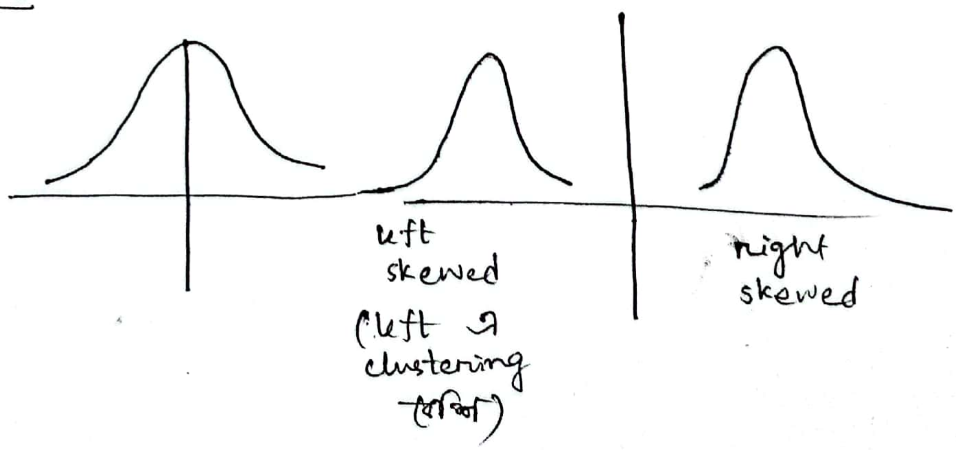


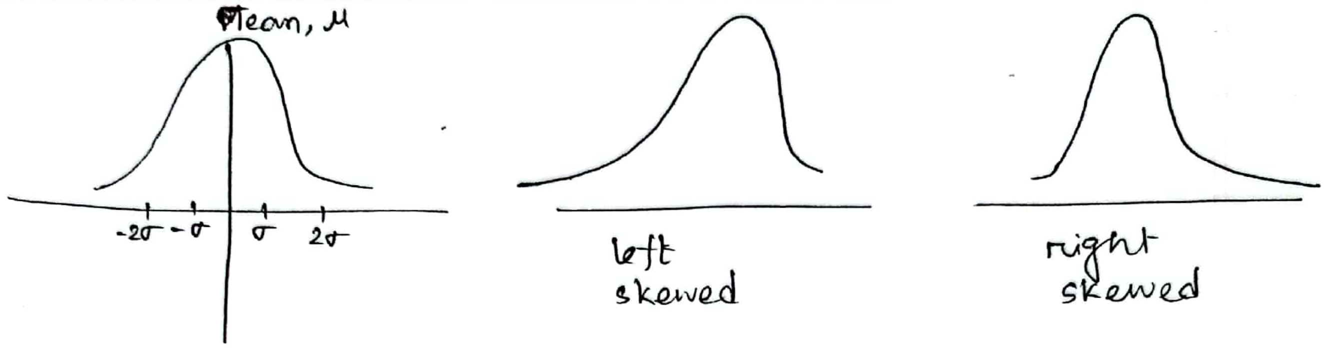
how many data points are within this range ?

Translating & Shifting :

↳ To shift unit from one to another

Note :





$x_i \to$ cm
$y_i \to$ inches
$y_i = 0.4 x_i$

If $x_1, x_2 \ldots x_n \longrightarrow$ mean $\bar{x}$
variance s_x^2

$y_1, y_2 \ldots y_n \longrightarrow$ mean $\bar{y}$
variance s_y^2

$\bar{y} = a\bar{x} + b$ and $s_y^2 = a^2 s_x^2$

If we shift measurements by adding or subtracting a constant, then the measure of center gets shifted by same amount, but the measure of variance is unaffected by any shift in measurement

When multiplying $\longrightarrow$ both center & variation are affected

Homework

Ex : 1.14 (Pg 54 of 839

a) $y_i = x_i + 50$; $\bar{y} = \bar{x} + 50 = 550\$$
standard deviation unchanged

100 এর 10% বাড়ে
5000 ~ 10% বেশি

b) $y_i = 1.10\, x_i$; $\bar{y} = 1.10\bar{x} = 1.10(500) = 550\$$
Standard deviation, $s_y = \sqrt{1.10^2 \cdot s_x^2}$
$= \sqrt{1.10^2 \times (125)^2} = 137.5 > 125$