Step-3: Use tally

Step-4: Count the tally

$$\text{Relative frequency} : \frac{\text{this frequency}}{\text{total frequency}}$$

Percentage ⇒ Relative frequency × 100%.

⊛ Construct the frequency table with relative frequency.

⊛ For exclusive method, if a value comes with the boundary value, then put it in the 2nd class where the starting boundary is the same.

L-19/26.09.2023/

Different Type of Graph

(i) Bar Diagram

(ii) Pie

(iii) Histogram

Measures of Central Tendency:

(i) Mean

(ii) Median

(iii) Mode

Median, $M_e = L_0 + \frac{h}{f_0}\left(\frac{n}{2} - F\right)$

L-20/01.10.2023/

Mode

⇒ The number that occurs maximum time.

⇒ Can be multiple, if frequency same.

⇒ if all unique, then there is no mode.

Mode = 1 ⇒ Unimodal

Mode = 2 ⇒ Bimodal

mode = >2 ⇒ Multimodal

Mode from group Data

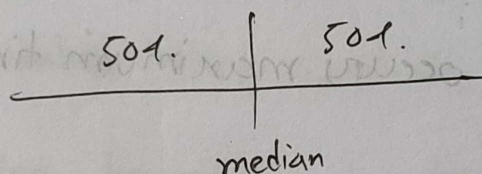
$$M_o = L_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} h$$

$$M_o = 5 + \frac{12-6}{(12-6)+(12-7)} 5$$

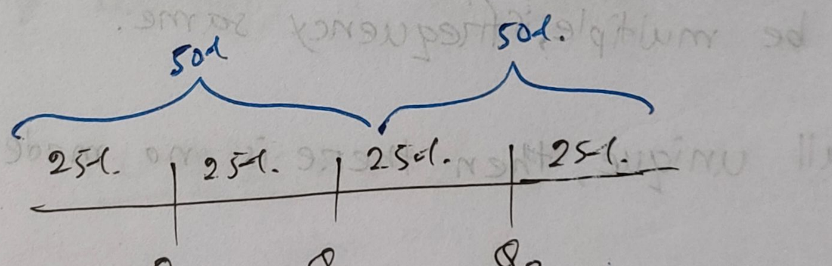
$$= 5 + \frac{6}{6+5} \cdot 5$$

$$= 7.73$$

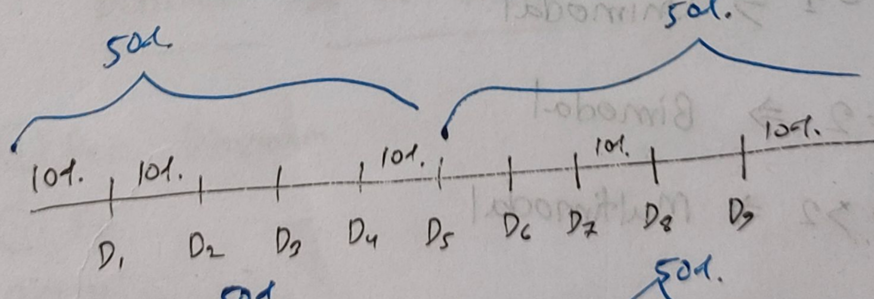
Median



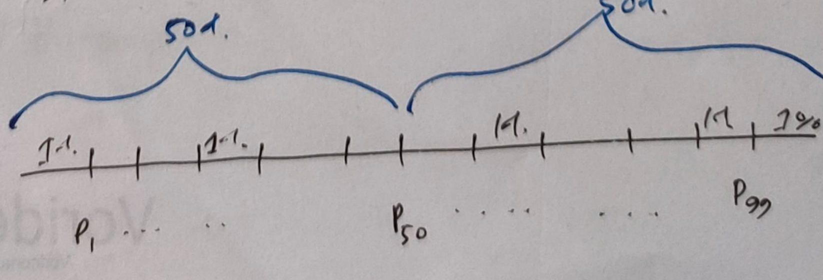
Quantile:



Decile:



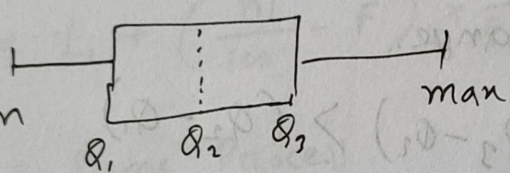
Percentile:



$\therefore Me = Q_2 = D_5 = P_{50}$

Box Plot → Must in Final

(*) Given Data Set
⇒ construct Box Plot → Need to Organize, maximum value, minimum value, Q_1, Q_2, Q_3 → Median
⇒ Hence comment on it



⇒ $P_{80} = ?$

$L_p = \frac{P}{100}(n+1)$

$L_{80} = \frac{P}{100}(12+1)$

$= 10.4$

$P_{80} = 6050 + (6130 - 6050) \times 0.4$

$= 6082$

(*) min = 5710

max = 6325

$L_{25} = \frac{25}{100}(12+1) = 3.25$

$L_{50} = \frac{50}{100}(12+1) = 6.5$

$L_{75} = \frac{75}{100}(12+1) = 9.75$

$Q_1 = 5850 + (5880 - 5850) \times 0.25$

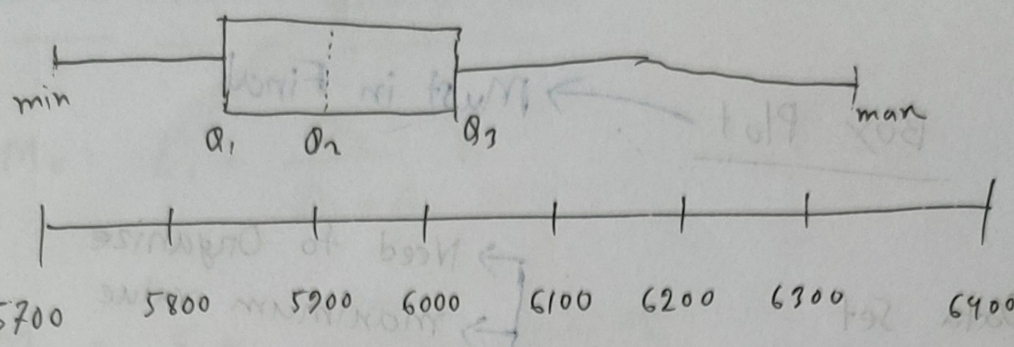
$= 5857.5$

$Q_2 = 5890 + (5920 - 5890) \times 0.5$

$= 5905$

$Q_3 = 5950 + (6050 - 5950) \times 0.75$

$= 6025$



<u>Comment:</u>

min =

man =

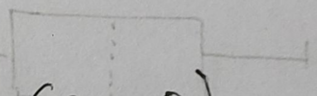
Q_1 =

Q_2 =

Q_3 =

Inter Quartile Range $= Q_3 - Q_1$

Middle 50%. observation lies in this range.



$(Q_3 - Q_2) > (Q_2 - Q_1)$

Positively skeweed distribution

if inverse then,

negatively skeweed.

$$P_i = L_1 + \left(\frac{ni}{100} - F\right) \times \frac{h}{f}$$

$M_e = P_{50}$

$$\frac{ni}{100} = \frac{5763 \times 50}{100} = 2881.5$$

$L_1 = 24.5$

$F = 2585$

$h = 29.5 - 24.5 = 5$

$f = 1737$

$\therefore P_{50} = L_1 + \left(\frac{ni}{100} - F\right) \times \frac{h}{f}$

$= 24.5 + (2881.5 - 2585) \times \frac{5}{1737}$

$= 25.35$

$\therefore Q_1 = ? = P_{25}$

$\frac{ni}{50} = \frac{5763 \times 25}{100} = 1440.75$

then, $P_{25} = L_1 + \left(\frac{ni}{100} - F\right) \times \frac{h}{f}$

= Same Process

L-21 / 03.10.2023 /

Correlation Analysis

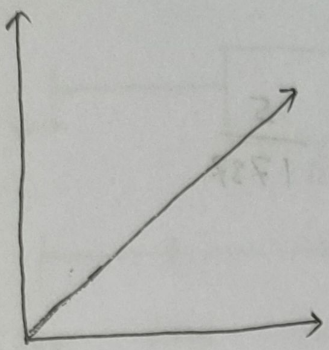
X	Y
$(x_1,$	$y_1)$
$(x_2,$	$y_2)$
$\vdots$	
$(x_n,$	$y_n)$

Pearson's correlation coefficient

$r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$

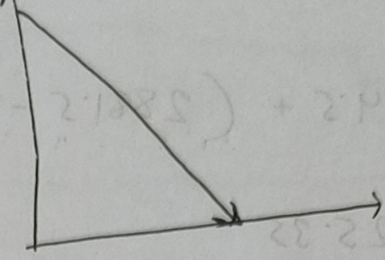
$= \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2 \; \Sigma (y-\bar{y})^2}}$

$= -1 \leq r_{xy} \leq 1$



$r_{xy} = 1$

Perfect positive



$r_{xy} = -1$

perfect negative

⊛ Graphical Representation of Relationship.

Quiz - 2

10.10.2023

Syllabus = L17 - L20

L-22/08.10.2023/

* Correlation

* Regression Analysis (Simple Linear)

Model,

$y = \alpha + \beta x + \epsilon$

y = Dependent Variable
α = Constant
β = Regression coefficient or slope
x = Independent Variable
ϵ = Random error term

$$\hat{\beta} = \frac{\sum_{n=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{n=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \beta \bar{x}$$

So, $\hat{y} = \hat{\alpha} + \hat{\beta} x$

$\boxed{\text{Do practice from Anderson's Book 14th Edition}}$

L-23 / 10.10.2023 /

Quiz-2

L-24 / 15.10.2023 /

Confidence interval

* Estimation:

1. Point estimate: single best value.
2. interval estimate: confidence interval

⇒ You need to add the confidence level in percentage (Probability)

$\boxed{\text{Problem from Slide}}$

$$\bar{x} \pm Z_{\alpha/2} \frac{\hat{\sigma}_x}{\sqrt{n}}$$

$n = 35$

$\bar{x} = 78$

$SD = \hat{\sigma}_x = 9$

$\alpha = 0.10$

$\frac{\alpha}{2} = 0.05$

$Z_{\alpha/2} = 1.65$

$$= 78 \pm 1.65 \frac{9}{\sqrt{35}}$$

$$= 78 \pm 2.5101$$

$$= (75.4899, 80.5101)$$

From Side - Page 43

a)

$n = 49$

$SD = \hat{\sigma}_x = 2.50$

margin of error (ME) $= Z_{\alpha/2} \dfrac{\hat{\sigma}_x}{\sqrt{n}}$

$= 1.96 \dfrac{2.5}{\sqrt{49}} = 0.7$

95%

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$Z_{0.025} = 1.96$ (from table)

2)

$n = 40$

$SD = \hat{\sigma}_x = 0.3$

$\bar{x} = 14.1$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$Z_{0.025} = 1.96$

$14.1 \pm 1.96 \dfrac{0.3}{\sqrt{40}}$

$= 14.1 \pm 0.093$

$= (14.007 - 14.193)$

Do more from slide

⊛ Sample size less than 30 then small sample otherwise large sample.

For small sample,

$$\bar{x} \pm t_{\alpha/2, (n-1)} \frac{\hat{\sigma}_x}{\sqrt{n}}$$

$n = 20$

$\bar{x} = 67.3$

$\hat{\sigma}_x = 3.6$

95%

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$t_{0.025, 19} = 2.093$

$\Rightarrow 67.3 \pm 2.093 \frac{3.6}{\sqrt{20}}$

$= 67.3 \pm 1.64$

$= (65.61 - 68.98)$

Assignment ⇒ Page - 44 ⇒ 3, 4

Page - 47 ⇒ 1, 3

More problem will be added later

L-25 / 17.10.2023 /

Hypothesis Testing

Solved Some Problem from Lecture Slide

(*)

$H_0 : \mu = 89$

$H_1 : \mu < 89$ (one trial test)

level of significance $\alpha = 5\%$
$= 0.05$

Test statistics:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Here

$\bar{x} = 92.2$, $\mu_0 = 89$

$\sigma = \sqrt{144}$, $n = 12$

$Z_{cal} = 0.09237$

$|Z_{cal}| = 0.09237$

Critical Value

$|Z_{tab}| = |Z_{\alpha}| = |Z_{0.05}| = 1.645$

Decision Rule:

Since $|Z_{cal}| < |Z_{tab}|$

$\Rightarrow$ H_0 is accepted at 5% level of significance

i.e. the average time of production is 89 mins.

(*) When,

σ = known $\Rightarrow$ Z_{cal} Test

σ = unknown & $n > 30$ $\Rightarrow$ Z_{cal} Test

σ = unknown & $n \leq 30$ $\Rightarrow$ t_{cal} Test

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim t_{(n-1)}$$