1. 95% Confidence interval is

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{3n}\sum_{i=1}^{3n} y_i - \Phi^{-1}(0.975)\,\sigma\sqrt{\frac{1}{n}+\frac{1}{3n}}\,,\right.$$

$$\Phi^{-1}(0.975) = \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \quad (\text{here } \alpha = 0.05)$$

$$\left.\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{3n}\sum_{i=1}^{3n} y_i + \Phi^{-1}(0.975)\,\sigma\sqrt{\frac{1}{n}+\frac{1}{3n}}\right).$$

Here $\sigma = 5$ & $\Phi^{-1}(0.975) = 1.96$

Hence, the width of the interval is

$$2\,\Phi^{-1}(0.975)\,5\sqrt{\frac{4}{3n}}$$

Solve $\cancel{2}\,\Phi^{-1}(0.975)\,5\sqrt{\frac{4}{3n}} = \cancel{4}\,2$

$$\Leftrightarrow \Phi^{-1}(0.975)\,5\sqrt{\frac{4}{3n}} = 2$$

$$\Leftrightarrow 25\,(1.96)^2\,\frac{\cancel{4}}{3n} = \cancel{4}$$

$$\Leftrightarrow n = \frac{25 \times (1.96)^2}{3}$$

$$\Leftrightarrow n = 32.01333\ldots$$

However, n cannot be fraction. So, we should consider $n = [32.01333] + 1 = 33$.

Given information:

2. $X_1, \dots, X_n \sim N(\mu, 1.2^2)$, here $n = 20$ (2)

X_i: weight of cement in the i-th bag. $i = 1, \dots, 20$.

a) $H_0: \mu = 50$ against $H_1: \mu = 49$.

b) Note that H_0 will be rejected when $\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{} < 49.5$. (say $\overline{X_n}$)

So, $P[\text{Type-I error}] = P[\overline{X_n} < 49.5 \mid \mu = 50]$

$$= P\left[\frac{\sqrt{n}(\overline{X_n} - 50)}{1.2} < \frac{\sqrt{n}(49.5 - 50)}{1.2}\right]$$

$$= \Phi\left(\frac{\sqrt{20}\,(49.5 - 50)}{1.2}\right)$$

$$= 0.0312\,0371.$$

c) $P[\text{Type-II error}] = P[\overline{X_n} > 49.5 \mid \mu = 49]$

$$= P\left[\frac{\sqrt{n}(\overline{X_n} - 49)}{1.2} > \frac{\sqrt{n}(49.5 - 49)}{1.2}\right]$$

$$= 1 - \Phi\left(\frac{\sqrt{20}\,(49.5 - 49)}{1.2}\right)$$

$$= 0.0312\,0371$$

d) Test-statistic : $Z = \dfrac{\overline{X}_n - \mu_0 \text{ (i.e., } \mu \text{ under } H_0)}{\underset{\substack{\| \\ 1.2}}{\sigma} / \sqrt{\underset{\substack{\| \\ 20}}{n}}}$ (3)

$$= \frac{\overline{X}_n - 50}{1.2/\sqrt{20}} = \frac{\overline{X}_n - 50}{0.2683},$$

Reject H_0 if $\dfrac{\overline{X}_n - 50}{0.2683} < \Phi^{-1}(0.02) = -2.0538$

e) If $\overline{X}_n = 49.27$, then $Z = -2.72 < -2.0538$

Hence, H_0 will be rejected at 2% level of significance.

3. Let μ be the average score in the exam.

a) $H_0 : \mu = 83$ against $H_1 : \mu \neq 83$.

Test-statistic $T_n = \dfrac{\overline{X}_n - \mu_0}{s/\sqrt{n}}$, where

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\mu_0 = 83$, $s = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2$.

Here $n = 8$, ~~and~~ Reject H_0 at $\underset{\substack{\| \\ \alpha = 0.1}}{10\%}$ level of significance if $|T_n| > t_{\cancel{0.05}\,\underset{\substack{\| \\ 0.05}}{\frac{\alpha}{2}},\ \underset{\substack{\| \\ 7}}{n-1}} = 1.895$.

For the given data $T_n = \dfrac{\bar{X}_n - \mu_0}{s/\sqrt{n}}$

$$= -1.12$$

Hence $|T_n| = 1.12 < t_{0.05,7} = 1.895.$

<u>Conclusion</u>:- Do NOT reject H_0 at 10% level of significance.

b) p-value $\overset{*}{=} P[T_n > |t_0|]$

observed value of $T_n = -1.12$

$$* = P\left[\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{s} > |-1.12|\right].$$

$$= P[Y > 1.12]$$

Y follows t_7 distribution.

$$= 0.299 > \alpha = 0.1.$$

Hence, H_0 will not be rejected.

4. $H_0: \mu = 70$ USD against $H_1: \mu < 70$ USD. (5)

μ ↓ Mean parameter

Test-statistic (say T_n) $= \dfrac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \underset{\|}{\overset{70}{\mu_0}}\right)}{\underset{17.32}{\overset{\sigma}{\|}}}$. Here $n = 45$.

<u>Decision rule</u>:- Reject H_0 if $T_n < \Phi^{-1}(0.01)$
$= -2.326$.

For a given data, $\frac{1}{n}\sum_{i=1}^{n} x_i = 67.57$.

Hence, $T_n = -0.94$, which is greater than -2.326.

So, H_0 will NOT be rejected at 1% level of significance. In other words, there is ~~not~~ NOT enough evidence that the average amount spent by students is less than 70 \$.

5. $$\Lambda = \frac{L(\theta_0 | x_1, \ldots, x_n)}{L(\theta_1 | x_1, \ldots, x_n)}$$

likelihood ratio

$$= \frac{\dfrac{e^{-n\theta_0}\, \theta_0^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}}{\dfrac{e^{-n\theta_1}\, \theta_1^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}}$$

~~Reject~~ Reject H_0 if $\Lambda < c$, c is any constant. (can be determined from the level of significance).

Note that

$$\Lambda = e^{-n(\theta_0 - \theta_1)} \left(\frac{\theta_0}{\theta_1}\right)^{\sum_{i=1}^n x_i}$$

So, $\Lambda < c$

$$\Rightarrow e^{-n(\theta_0 - \theta_1)} \left(\frac{\theta_0}{\theta_1}\right)^{\sum_{i=1}^n x_i} < c$$

$$\Rightarrow -n(\theta_0 - \theta_1) + \sum_{i=1}^n x_i \left[\log \theta_0 - \log \theta_1\right] < \log c = c' \text{ (say)}.$$

$$\Rightarrow \sum_{i=1}^n x_i \left[\log \theta_0 - \log \theta_1\right] < c' + n(\theta_0 - \theta_1) = c'' \text{ (say)}.$$

$$\Rightarrow \sum_{i=1}^n x_i > c'' \quad (\text{since } \log \theta_0 - \log \theta_1 < 0).$$

Hence, Reject H_0 at 5% level of significance if $\sum_{i=1}^n x_i < c^0$ $\left(\text{or } \frac{1}{n}\sum_{i=1}^n x_i < K\right)$, where c is such that $P_{H_0}\left[\sum_{i=1}^n x_i < c\right] < \alpha$.

6\. a) $\Lambda = \dfrac{L(\theta_0)}{L(\theta_1)}$, where θ_0 is the value of θ under H_0, i.e., $\theta_0 = \frac{1}{3}$. And, θ_1 is the value of θ under H_1, i.e., $\theta_1 = \frac{2}{3}$.

If $X = 0$, $\Lambda = \dfrac{\frac{1}{4}\theta_0}{\frac{1}{4}\theta_1} = \dfrac{\theta_0}{\theta_1} = \dfrac{2}{3}$.

If $X = 1$, $\Lambda = \dfrac{\frac{1}{2}(1-\theta_0)}{\frac{1}{2}(1-\theta_1)} = \dfrac{1-\theta_0}{1-\theta_1} = \dfrac{1-\frac{1}{3}}{1-\frac{2}{3}} = 2.$

If $X = 2$, $\Lambda = \dfrac{\frac{1}{4}(2+\theta_0)}{\frac{1}{4}(2+\theta_1)} = \dfrac{2+\theta_0}{2+\theta_1} = \dfrac{2+\frac{1}{3}}{2+\frac{2}{3}} = \dfrac{7}{8}.$

b) Note that under H_0,
follows from (a)

$$P\left[\Lambda = \tfrac{1}{2}\right] = P[X = 0] = \frac{\theta_0}{4} = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}.$$

$$P[\Lambda = 2] = P[X = 1] = \frac{1-\theta_0}{2} = \frac{1-\frac{1}{3}}{2} = \frac{1}{3}.$$

$$P\left[\Lambda = \tfrac{7}{8}\right] = P[X = 2] = \frac{1}{4}(2+\theta_0) = \frac{1}{4}\left(2+\frac{1}{3}\right) = \frac{7}{12}.$$

Hence, the distribution of Λ under H_0 is

$\Lambda = \frac{1}{2}$ with probability $\frac{1}{12}$.

$\Lambda = \frac{1}{3}$ with probability $\frac{1}{3}$.

$\Lambda = \frac{7}{8}$ with probability $\frac{7}{12}$.

(i) <u>p-value</u>

$$p\text{-value} = P_{H_0}[\Lambda \leq \Lambda_{\text{observed}}]$$

Since $X = 2$ observed $\Rightarrow \Lambda = \frac{7}{8}$ observed.

Hence, $p\text{-value} = P_{H_0}\left[\Lambda \leq \frac{7}{8}\right]$

$$= P_{H_0}\left[\Lambda = \frac{1}{2}\right] + P_{H_0}\left[\Lambda = \frac{7}{8}\right]$$

$$= \frac{1}{12} + \frac{7}{12} = \frac{8}{12} = \frac{2}{3}.$$

7. For the model $Y_i = \beta_0 + \beta_1 x_i + e_i$, where $e_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $i = 1, \ldots, n$

least square estimators are

$$\hat{\beta}_{0,n} = \bar{y}_n - \hat{\beta}_{1,n} \bar{x}_n \quad \&$$

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

Note that if $\frac{1}{n}\sum_{i=1}^{n} x_i \equiv 0$ (i.e., $\bar{x}_n = 0$), we have

$$\hat{\beta}_{0,n} = \bar{y}_n \quad \& \quad \hat{\beta}_{1,n} = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y}_n)}{\sum_{j=1}^{n} x_j^2}.$$

Now, $$\text{Cov}\left(\hat{\beta}_{0,n}, \hat{\beta}_{1,n}\right) = \text{Cov}\left(\bar{y}_n, \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y}_n)}{\sum_{j=1}^{n} x_j^2}\right)$$

$$= \text{Cov}\left(\bar{y}_n, \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{j=1}^{n} x_j^2}\right) - \text{Cov}\left(\bar{y}_n, \frac{\sum_{i=1}^{n} \bar{y}_n x_i}{\sum_{j=1}^{n} x_j^2}\right)$$

$$= \cancel{\sum_{i=1}^{n}} = \frac{\sum_{i=1}^{n} x_i \, Cov(\overline{y_n}, y_i)}{\sum_{j=1}^{n} x_j^2} - \frac{\sum_{i=1}^{n} x_i \, Cov(\overline{y_n}, \overline{y_n})}{\sum_{j=1}^{n} x_j^2}$$

$$= \frac{\sum_{i=1}^{n} x_i \left[Cov\left(\frac{1}{n}\sum_{k=1}^{n} y_k, y_i\right)\right]}{\sum_{j=1}^{n} x_j^2} - \frac{\sum_{i=1}^{n} x_i \, Var(\overline{y_n})}{\sum_{j=1}^{n} x_j^2}$$

$$= \frac{\sum_{i=1}^{n} x_i \left(\frac{1}{n} Var(y_i)\right)}{\sum_{j=1}^{n} x_j^2} - \frac{\sum_{i=1}^{n} x_i \frac{\sigma^2}{n}}{\sum_{j=1}^{n} x_j^2}$$

since $y_i \sim N(\beta_0 + \beta_1 x, \sigma^2)$ $\Rightarrow \overline{y_n} \sim N(\beta_0 + \beta_1 x, \frac{\sigma^2}{n})$

When $k \neq i$, $Cov(y_k, y_i) = 0$ since y_i's are indep. with each other.

$$= \frac{\frac{\sigma^2}{n} \sum_{i=1}^{n} x_i}{\sum_{j=1}^{n} x_j^2} - \frac{\frac{\sigma^2}{n} \sum_{i=1}^{n} x_i}{\sum_{j=1}^{n} x_j^2} = 0.$$

8. For the model $Y_i = \beta x_i + e_i$, where $e_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$

Define

$$S(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2$$

$$\hat{\beta}_n = \underset{\beta}{\operatorname{argmin}}\ S(\beta)$$

Now, $\dfrac{dS(\beta)}{d\beta} = 0 \Rightarrow 2 \sum_{i=1}^{n} (y_i - \beta x_i)(-x_i) = 0$

$$\Rightarrow \hat{\beta}_n = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Note that $\left.\dfrac{d^2 S(\beta)}{d\beta^2}\right|_{\beta = \hat{\beta}_n} = 2 \sum_{i=1}^{n} x_i^2 > 0$

Hence, $S(\beta)$ will be minimized at $\beta = \hat{\beta}_n$.

Hence, the least-square estimator of β is $\hat{\beta}_n = \dfrac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$.

(b) $$E(\hat{\beta}_n) = E\left(\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}\right)$$

$$= \left(\frac{\sum_{i=1}^{n} x_i E(y_i)}{\sum_{i=1}^{n} x_i^2}\right) = \frac{\sum_{i=1}^{n} x_i \beta x_i}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\beta \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2} = \beta.$$

(c) $$Var(\hat{\beta}_n) = Var\left(\frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n} x_i^2\right)^2} Var\left(\sum_{i=1}^{n} y_i x_i\right)$$

$$= \frac{1}{\left(\sum_{i=1}^{n} x_i^2\right)^2} \sum_{i=1}^{n} x_i^2 \underbrace{Var(y_i)}_{\sigma^2}$$ since $y_i \perp\!\!\!\perp y_j \Rightarrow Cov(y_i, y_j) = 0 \ \forall i \neq j$.

$$= \sigma^2 \times \frac{1}{\cancel{\left(\sum_{i=1}^{n} x_i^2\right)^2}_{\left(\sum_{i=1}^{n} x_i^2\right)}} \times \cancel{\left(\sum_{i=1}^{n} x_i^2\right)} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

(d) Note that $\sum_{i=1}^{5} x_i y_i = 203.09$

& $\sum_{i=1}^{5} x_i^2 = 60.4$.

Hence, $$\hat{\beta}_5 = \frac{\sum_{i=1}^{5} x_i y_i}{\sum_{i=1}^{5} x_i^2} = \frac{203.09}{60.4} \approx 3.36.$$

Further, note that $$\sum_{i=1}^{5} \hat{e}_i^2 = \sum_{i=1}^{5} (y_i - \hat{\beta}_5 x_i)^2 = 34.96.$$

Now, $$\hat{\sigma}^2 = \frac{\sum_{i=1}^{5} \hat{e}_i^2}{\underset{5}{n} - 1} = \frac{34.96}{4} = 8.74.$$

(Not 2, since one unknown parameter β)

Hence, $$\sqrt{\widehat{Var}(\hat{\beta}_5)} = \sqrt{\hat{\sigma}^2 \frac{1}{\sum_{i=1}^{5} x_i^2}} = \sqrt{\frac{8.74}{60.4}} \approx 0.38.$$

Hence, 95% C.I. of β is

$$\left(\hat{\beta}_5 - t_{4,0.975}\sqrt{\widehat{Var}(\hat{\beta}_5)},\ \hat{\beta}_5 + t_{4,0.975}\sqrt{\widehat{Var}(\hat{\beta}_5)}\right)$$

$$= (3.36 - 2.776 \times 0.38,\ 3.36 + 2.776 \times 0.38)$$

$$= (2.31,\ 4.42).$$