Date
Page

Probablity & Statistics - II   assignment - 1

Q1 $PQ = \dfrac{\hat{\theta} - \theta}{\sqrt{\dfrac{\hat{\theta}}{n}}}$, where $\hat{\theta} = \dfrac{\Sigma x_i}{n}$

$\Rightarrow \hat{\theta} = \dfrac{200}{1} = 200$

$\Rightarrow PQ = \dfrac{200 - \theta}{\sqrt{\dfrac{200}{1}}}$

$\Rightarrow P\left(-1.96 < \dfrac{200 - \theta}{\sqrt{200}} < 1.96\right) = 0.95$

$\Rightarrow P(-1.96 \times \sqrt{200} < 200 - \theta < 1.96 \times \sqrt{200}) = 0.95$

$\Rightarrow P(-200 - 1.96 \times \sqrt{200} < -\theta < 1.96 \times \sqrt{200} - 200) = 0.95$

$\Rightarrow P(200 + (1.96 \times \sqrt{200}) > \theta > 200 - (1.96 \times \sqrt{200}) = $

$\Rightarrow P(200 - (1.96 \times \sqrt{200}) < \theta < 200 + (1.96 \times \sqrt{200}))$

So 95% CI for θ = $[172.28, 227.71]$

Q2 Sample mean $= \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

Sample Variance $= S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$

i $E(\overline{X}) = E\left[\frac{\sum X_i}{n}\right] = \frac{1}{n} E(\sum X_i)$

$= \frac{1}{n} E\{X_1 + X_2 + X_3 + \cdots + X_n\}$

$= \frac{1}{n} \{E(X_1) + E(X_2) + E(X_3) + \cdots + E(X_n)\}$

$= \frac{1}{n} \{\mu + \mu + \mu + \cdots + \mu\}$

$= \frac{\cancel{n}\mu}{\cancel{n}} = \mu$

$V(\overline{X}) = V\left[\frac{\sum X_i}{n}\right] = \frac{1}{n^2} V[\sum X_i]$

$= \frac{1}{n^2} V\{X_1 + X_2 + X_3 + \cdots + X_n\}$

$= \frac{1}{n^2} \cancel{V}\{\sigma^2 + \sigma^2 + \sigma^2 + \cdots + \sigma^2\}$

$= \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$

ii $E(s^2) = \sigma^2$

$\Rightarrow E(s^2) = E\left[\frac{(n-1)s^2}{\sigma^2}\right] \times \frac{\sigma^2}{n-1}$

$= E\left[\chi^2_{n-1}\right] \times \frac{\sigma^2}{n-1}$

$= \cancel{n-1} \times \frac{\sigma^2}{\cancel{n-1}}$

$= \sigma^2$

iii $V(s^2) = V\left[\frac{(n-1)s^2}{\sigma^2}\right] \times \frac{\sigma^4}{(n-1)^2}$

$= V\left[\chi^2_{n-1}\right] \times \frac{\sigma^4}{(n-1)^2}$

$= 2\cancel{(n-1)} \times \frac{\sigma^4}{(n-1)^{\cancel{2}}}$

$= \frac{2\sigma^4}{(n-1)}$

IV $E(s^2) = \sigma^2 = 100$

$\Rightarrow$ 50 % of $E(s^2) = 50$

So, $P(-50 < s^2 < 50) = P\left[\frac{-5\cancel{0} \times 9}{10\cancel{0}} < s^2 < \frac{50 \times 9}{100}\right]$

$= P\left[-4.5 < \chi^2_9 < 4.5\right]$ $\left\{\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}\right\}$

$\Rightarrow P[\chi_9^2 < 4.5] - P[\chi_9^2 \leq -4.5]$

$\downarrow$

0 (because χ^2 cannot be -ve)

$\Rightarrow P[\chi_9^2 < 4.5] = 0.1245$

Q3

No. of claims	0	1	2	3	4
No. of policies	86	75	16	2	1

$X \sim Bin(4, p)$

$f(x) = [P(X=0)]^{86} \times [P(X=1)]^{75} \times [P(X=2)]^{16} \times [P(X=3)]^{2} \times [P(X=4)]^{1}$

$= \{p^0(1-p)^4\}^{86} \times \{p^1(1-p)^3\}^{75} \times \{p^2(1-p)^2\}^{16} \times \{p^3(1-p)^1\}^{2} \times \{p^4(1-p)^0\}^{1}$

$= (1-p)^{344} \times [p^{75}(1-p)^{225}] \times [p^{32}(1-p)^{32}] \times [p^6(1-p)^2] \times [p^4]$

$= p^{117}(1-p)^{603}$

$\Rightarrow L = \log(p^{117}(1-p)^{603})$

$\Rightarrow \quad l = 117 \log p + 603 \log (1-p)$

$\Rightarrow \quad \frac{dl}{dp} = \frac{117}{p} \cancel{+} \frac{603}{1-p} = 0$

$\Rightarrow \quad 117 - 117p \cancel{+} 603p = 0$

$\Rightarrow \quad \cancel{117 + 486} \; 117 = 720p$

$\Rightarrow \quad \hat{p} = 0.1625$

ii

H_0 : Number of claims confirm for binomial model

H_1 : Number of claims does not confirm for binomial model

$P(X=0) = {}^4C_0 \, p^0 (1-p)^4 = {}^4C_0 \, 1 \cdot (0.4919) = \cancel{0.4919}$ 0.4919

$P(X=1) = {}^4C_1 \, p^1 (1-p)^3 = {}^4C_1 \times 0.1625 \times 0.5874$

$= 0.38181$

$P(X=2) = {}^4C_2 \, p^2 (1-p)^2 = {}^4C_2 \times (0.1625)^2 \times 0.7014$

$= 0.1111$

$P(X=3) = {}^4C_3 \, p^3 (1-p)^1 = 4 \times (0.1625)^3 \times 0.8375$

$= 0.0144$

$P(X=4) = {}^4C_4 \, p^4 (1-p)^0 = 0.0007$

No. of C	0	1	2	3	4	
O_i	86	75	16	2	1	= 180
~~E_i~~	~~42.30~~	~~28.63~~	~~1.77~~	~~0.02~~		
E_i	88.54	68.72	19.99	2.59	0.12	
~~$(O_i-E_i)^2$~~						

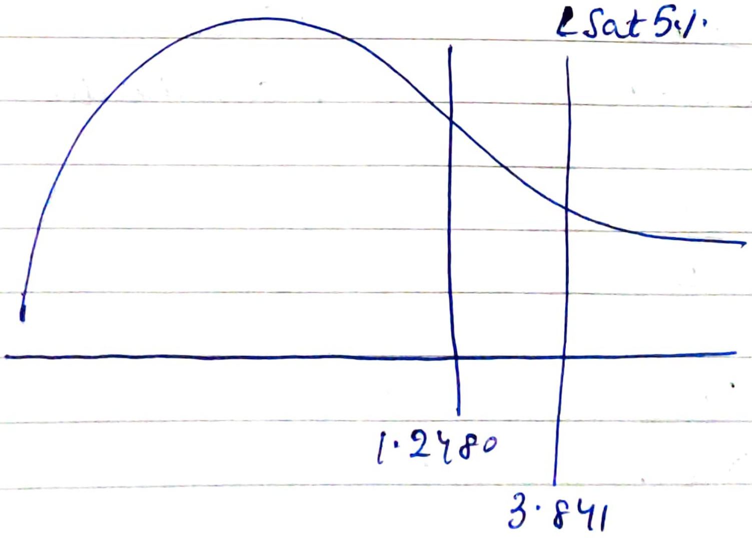
⇒	0	1	2+	
O_i	86	75	19	= 180
E_i	88.54	68.72	22.7	
$(O_i-E_i)^2$	6.45	39.43	13.69	

$$\Rightarrow TS = \frac{6.45}{88.54} + \frac{39.43}{68.72} + \frac{13.69}{22.7}$$

$$= 1.2480$$

$$\Rightarrow TS \sim \chi^2_{5-1-1-2}$$

$$\Rightarrow 1.2480 \sim \chi^2_1$$



Since TS lies in AR

We have insufficient evidence to reject Ho at 5% LS

Q4 PDF = $f(x) = \dfrac{C\beta^3}{(x+\beta)^4}$ $\quad x > 0 \quad \beta > 0$

i ⇒ a> $\displaystyle\int_0^\infty \frac{C\beta^3}{(x+\beta)^4}\,dx = 1$

$\Rightarrow \displaystyle C\beta^3 \int_0^\infty \frac{1}{(x+\beta)^4}\,dx = 1.$

$\Rightarrow \dfrac{C\beta^3}{-3}\Big[(x+\beta)^{-3}\Big]_0^\infty = 1$

$\Rightarrow \dfrac{C\beta^3}{-3}\left\{(\infty+\beta)^{-3} - (0+\beta)^{-3}\right\} = 1$

$\Rightarrow \dfrac{C\cancel{\beta^3}}{3}\,\cancel{(\beta)^{-3}} = 1 \quad \Rightarrow \quad C = 3$

b> $f(x) = \dfrac{C\beta^3}{(x+\beta)^4}$. $\quad$ [This follow Pareto distribution]

$\Rightarrow E(X) = \mu = \dfrac{\beta}{2}$

$\Rightarrow V(X) = \dfrac{3\beta^2}{(2)^2(1)^2} = \dfrac{3\beta^2}{4}$

ii a) By MOM

$E(X) = \overline{X}$

$\Rightarrow \dfrac{\beta}{2} = \overline{X} \Rightarrow \hat{\beta} = 2\overline{X}$

b) To show, $E(\hat{\beta}) = \beta$

$\Rightarrow E(\hat{\beta}) = E(2\overline{X}) \Rightarrow 2E(\overline{X})$

$\Rightarrow 2 \times E\left[\dfrac{\beta}{2}\right] \Rightarrow \dfrac{\cancel{2} \times \beta}{\cancel{2}} = \beta$

Hence the above is unbiased

c) $MSE = V(\hat{\beta})$

$= V(2\overline{X}) \Rightarrow 4V(\overline{X})$

$\Rightarrow 4 \times V\left[\dfrac{\Sigma X_i}{n}\right] \Rightarrow \dfrac{4}{n^2} V[\Sigma X_i]$

$\Rightarrow \dfrac{\cancel{4}}{n^2} \times \dfrac{3\beta^2}{\cancel{4}} = \dfrac{3\beta^2}{n^2}$

So as $n \rightarrow$ increases, $MSE \rightarrow$ decreases.

iii $MSE[b\overline{X}_n] = Var[b\overline{X}_n] + (Bias[b\overline{X}_n])^2$

$= b^2 \dfrac{3}{4} \dfrac{\beta^2}{n} + (E[b\overline{X}_n] - \beta)^2$

$= b^2 \dfrac{3}{4} \dfrac{\beta^2}{n} + \left(\beta\left(\dfrac{b}{2} - 1\right)\right)^2$

$\Rightarrow \dfrac{d(MSE)}{db} = \dfrac{\beta^2}{n}\left(\dfrac{3}{2}b + n\left(\dfrac{b}{2} - 1\right)\right)$

$$\Rightarrow \frac{\beta^2}{m}\left(\frac{3}{2}b + m\left(\frac{b}{2}-1\right)\right) = 0 \Rightarrow \frac{3}{2}b = -m\left(\frac{b}{2}-1\right)$$

$$\Rightarrow \frac{3b}{2} + \frac{mb}{2} = 1 \Rightarrow b(3+m) = 2 \Rightarrow b = \frac{2}{(3+m)}$$

IV It is seen that $\hat{\beta} = 2\bar{X}_n$ is an unbiased estimator

As $n \to \infty$, b tends to 2. So the estimator is also consistent in view of ii

Q5 i a) $E(X) = \bar{x}$

$$\Rightarrow \frac{a+b}{2} = \bar{x} \Rightarrow \hat{a} = 2\bar{x} - b$$

b $V(X) = s^2$

$$\frac{(b-a)^2}{12} = s^2 \Rightarrow (b-a)^2 = 12s$$

$$\Rightarrow b - a = 2\sqrt{3} \cdot s$$

$$\Rightarrow b - 2\sqrt{3}s = a$$

So $2\bar{x} - b = b - 2\sqrt{3}s$

$$\Rightarrow 2b = 2\bar{x} + 2\sqrt{3}s$$

$$\Rightarrow b = \bar{x} + \sqrt{3}s$$

So $a = 2\bar{x} - \bar{x} - \sqrt{3}s$

$$\Rightarrow \hat{a} = \bar{x} - \sqrt{3}s$$

ii $2\bar{x} - b = b - 2\sqrt{3}s$

$\Rightarrow \cancel{2}b = \cancel{2}(\bar{x} + \sqrt{3}s)$

$\Rightarrow \hat{b} = \bar{x} + \sqrt{3}s$

iii

IV $f(x) = \frac{1}{b-a} \Rightarrow f(x) = \frac{1}{b}$ $\{$ because $a=0$ $\}$

$\Rightarrow L = \frac{1}{b^n} \Rightarrow \log L = \log\left[\frac{1}{b^n}\right]$

$\Rightarrow l = -n \log b \Rightarrow \frac{dl}{db} = \frac{-n}{b} = 0$

Using common sense, we will find a value of b that maximises l, which is at minimum value of 0

$\Rightarrow \hat{b} = \max(x_i)$

Q7 i a) Public

Sample Mean = $\frac{\sum x_i}{n}$ = 30

Sample Variance = 64.3125

b) Private

Sample Mean = $\frac{\sum y_i}{n}$ = 35.05

Sample Variance = 67.402

ii

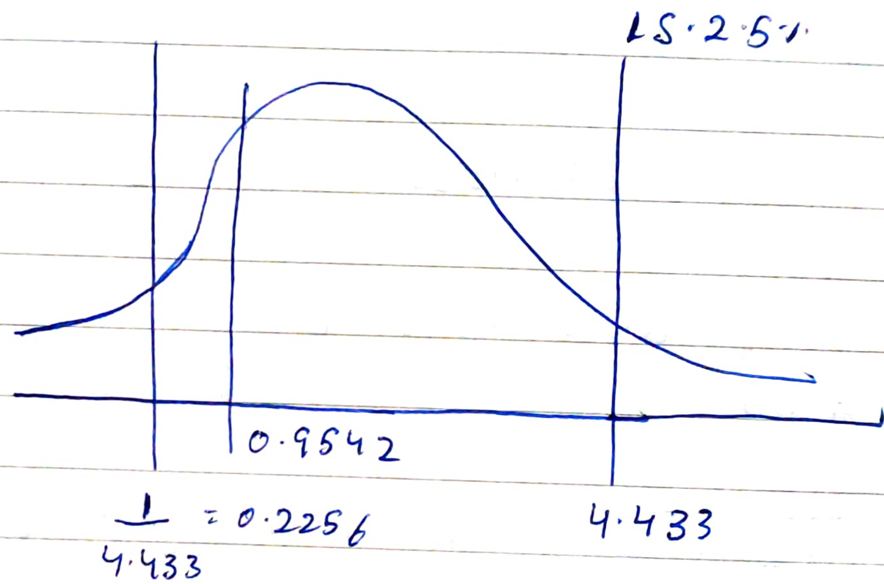
$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

$$T.S. = \frac{S_1^2}{S_2^2} \times \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,\, n_2-1}$$

$$\Rightarrow \quad = \frac{64.312}{67.402} \sim F_{8,8}$$

$$\Rightarrow \quad = 0.9542 \sim F_{8,8}$$



We have insufficient evidence to reject H_0 at LS ~~2.5%~~ 5%.

$H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$

$H_1: \mu_2 > \mu_1 \Rightarrow \cancel{\mu_1 - \mu_2 < 0}\ \mu_2 - \mu_1 > 0$

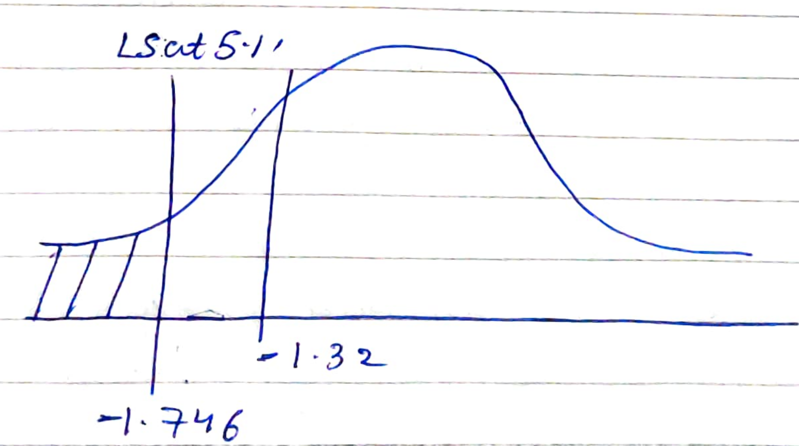
$$TS = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{S_p^2\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \sim t_{n_1+n_2-2}$$

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

$$= \frac{64.31 \times 8 + 67.42 \times 8}{16}$$

$$= 65.86$$

$$\Rightarrow TS = \frac{(35.05 - 30)}{\sqrt{65.86\left[\frac{1}{9} + \frac{1}{9}\right]}} = -1.32$$



We have insufficient evidence to reject H_0 at 5%.

Q6 i Let X be the no. of hits in first step 12 missiles & Y be the no. of hits in second step 12 missiles

=> P(Type 1 error) = P(Rejecting H_0 | H_0 is true)

$$= P(X \geq 3 | p=0.1) + P(X=0 | p=0.1)\, P(Y \geq 5 | p=0.1) + P(X=1 | p=0.1)\, P(X \geq 4 | p=0.1) + P(X=2 | p=0.1)\, P(Y \geq 3 | p=0.1)$$

$$= (1-0.8891) + (0.2824)(1-0.9957) + (0.6590-0.282)(1-0.9744) + (0.8891-0.6590)(1-0.8891)$$

$$= 0.14727 \approx 15\%$$

ii Probablity of rejecting the nule hypothesis when $p=0.3$

$$\Rightarrow P(X \geq 3 | p=0.3) + P(X=0 | p=0.3)\, P(Y \geq 5 | p=0.3) + P(X=1 | p=0.3)\, P(Y \geq 4 | p=0.3) + P(X=2 | p=0.3)\, P(Y \geq 3 | p=0.3)$$

Using table book

$$\Rightarrow (1-0.2528) + (0.0138)(1-0.7237) + (0.0850-0.0138)(1-0.4925) + (0.2528-0.0850)(1-0.2528)$$

$$= 0.91253 \approx 91\%$$

iii P(Type 2 error) = P(accepting H_0 | H_0 is false)

$$= 1 - 0.91253$$

$$= 0.08747 \text{ or } 9\%$$

IV

$$\frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

=> Lower bound of 90% right tailed CI for p is

$$\hat{p} - 1.2816\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.3333 - 1.2816\sqrt{\frac{0.3333(1-0.3333)}{120}}$$

$= 0.2782$

V

$H_0: p = 0.278$
$H_1: p > 0.278$

$$\frac{X - np_0}{\sqrt{np_0 q_0}} \sim N(0,1)$$

$$\Rightarrow \frac{39.5 - 120(0.278)}{\sqrt{120(0.278)(0.722)}} = 1.2511$$

We have insufficient evidence to reject H_0 at LS 10%.

VI

Lower bound of CI implies that there is only 10% chance of $p \leq 0.2782$ whereas from the hypothesis test $p \leq 0.2780$ with probablity more than 10%.

- Use of sample proportion $\hat{p}$ to estimate population variance in calculating CI
- Applying continuity correction in hypothesis testing

VII

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

$$PQ = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{1}{25}((11 \times 2916) + (14 \times 4900)) = 4027.04 \Rightarrow S_p$$

$$\Rightarrow \frac{65 - 70}{63.459\sqrt{\frac{1}{12} + \frac{1}{15}}} = 0.2034$$

We have insufficient evidence to reject H_0 at LS 5%.
Conclusion: There is no significant difference in mean scores for the population.

Q10i $P(\text{Type I error}) = P(\text{Rejecting } H_0 \mid H_0 \text{ is true})$

$H_0: X = 3000$

$H_1: X > 3100$

Rejecting $H_0 \rightarrow X > 3100$
Accepting $H_0 \rightarrow X < 3100$

$\Rightarrow P(X > 3100 \mid X \sim Poi(3000))$
By using Normal approximation
$X \sim N(3000, 3000)$

CI for μ is thus = $\left[50 - 3.25\sqrt{\frac{48.667}{10}}, 50+3.25\sqrt{\frac{48.667}{10}}\right]$

ie $(42.83, 57.17)$

b) $t_{n-1} \sim N\left[0, \sqrt{\frac{n-1}{n-3}}\right] \sim N\left[0, \sqrt{\frac{9}{7}}\right]$ ie

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim N\left[0, \sqrt{\frac{9}{7}}\right]$$

$$\Rightarrow \frac{\overline{X}-\mu}{S/\sqrt{\frac{7n}{9}}} \sim N(0,1)$$

Value of confidence is 2.58

CI for μ is thus $\left[50 - 2.58\sqrt{\frac{48.667}{10} \times \frac{9}{7}}, 50+2.58\sqrt{\frac{48.66}{10} \times \frac{9}{7}}\right]$

$= (43.54, 56.45)$.

Q8 i $\hat{\Theta}$ is said to be unbiased when $E(\hat{\Theta}) = \Theta$

ii Measure of the bias is given by $E(\hat{\Theta}) - \Theta$

iii Mean square error of this estimator $\hat{\Theta} = (E(\hat{\Theta}) - \Theta)^2$

IV $\tilde{\Theta}$ is efficient as an estimator with lower MSE is said to be more efficient than one with higher MSE.

V An estimator is termed as consistent if MSE converges to 0 as sample size tends to ∞

VI $\hat{\Theta}$ can be estimated using:

a) Method of moments
b) Maximum likelyhood method

Q9 i Variable t_k is defined as $\Rightarrow \dfrac{N(0,1)}{\sqrt{\dfrac{\chi_k^2}{k}}}$

where k denotes degree of freedom and $N(0,1)$ and χ_k^2 are two independent R.V.

ii Mean $= 0$

Variance $= \dfrac{k}{k-2}$

iii a) We know,

$\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ $\Rightarrow$ CI for $t_9 = 3.25$

$$\Rightarrow P(X > 3100) = P\left[Z > \frac{3100-3000}{\sqrt{3000}}\right]$$

$$= 1 - P\left[Z < 1.825\right] = 0.0339 / 3.4\%$$

ii Type II error = Event of accepting the hypothesis when it is false

iii Power of test = P(Rejecting Ho / Ho is false)

$$\Rightarrow P(X > 3100) = 1 - \left(Z < \frac{3100 - n\mu}{\sqrt{n\mu}}\right)$$

So the value of parameter decides the power of test

IV pq

$$\cancel{TS} = \frac{\hat{\mu} - \mu}{\sqrt{\frac{\hat{\mu}}{n}}} \sim N(0,1)$$

$$\Rightarrow P\left[-2.5758 < \frac{2.9 - \mu}{\sqrt{2.9/1000}} < 2.5758\right] = 0.99$$

$\Rightarrow$ CI for 99% is $(2.7613, 3.0387)$

Q11 Revised sample mean = $(213200 - 11000 - 48000 + 27500 + 31500)/12$

$$= \frac{213200}{12} = 17767$$

Revised $\Sigma x^2 = 4919860000 - 121000000 - 2304000000 + 756250000 + 992250000$

$$= 4243360000$$

$$\Rightarrow \text{Revised SD} = \left\{\frac{1}{11}\left(4243360000 - \frac{213200^2}{12}\right)\right\}^{1/2}$$

$$= (41409697)^{1/2} = 6435$$

Comment - There has been no change in sample mean. However there is deduction Sample SD.

The reduction in SD is because salary of Permanent employee who replaced temporary employee are closer to sample mean.

Q12) The cramer-rao lower bound result hold under very general conditions except where the range of the distribution involves the parameter, such as uniform dist. in this case

iii) Bias of $\hat{\theta}(c)$ is given as

$$\text{Bias}[\hat{\theta}(c)] = E[\hat{\theta}(c) - \theta]$$

$$= E[cY - \theta]$$

$$= c \cdot E(Y) - \theta \quad = \quad c \cdot \frac{n\theta}{n+1}$$

$$= \left(\frac{cn}{n+1} - 1\right) \times \theta$$

MSE of the estimator $\hat{\theta}(c)$ is given as

$$MSE[\hat{\theta}(c)] = E[\{\hat{\theta}(c) - \theta\}^2]$$

$$= E[(cY - \theta)^2]$$

$$= E[c^2Y^2 - 2c\theta Y + \theta^2]$$

$$= c^2 E(Y^2) - 2c\theta E(Y) + \theta^2$$

$$= c^2 \cdot \frac{n\theta^2}{n+2} - \frac{c \cdot 2n\theta^2}{n+1} + \theta^2$$

IV) For $\hat{\theta}(c)$ to be an unbiased estimator

To show: $\text{Bias}[\hat{\theta}(c_u)] = 0$

$$\Rightarrow \left(\frac{c_u n}{n+1} - 1\right) \times \theta = 0 \Rightarrow c_u = \frac{n+1}{n}$$

VI In order to minimize error in estimation, it is preffered to opt for an estimator which has lower MSE among different competing estimators

As n becomes large $\hat{\theta}_{C}(ln) = \frac{n+1}{n} Y \rightarrow Y$

similarly $\hat{\theta}_{C}(m) = \frac{n+2}{n+1} Y \rightarrow Y$

Thus two estimators ~~too~~ become same as $n \rightarrow \infty$