Probability & Statistics

Assignment - 1

1) P, θ $\rightarrow$ $\frac{\hat{\theta} - \theta}{\sqrt{\theta / n}} \sim N(0,1)$

$$\hat{\theta} = \frac{\Sigma x_i}{n} = \frac{200}{1} = 200$$

C.I = 95%

$$P\left(-1.96 < \frac{\hat{\theta} - \theta}{\sqrt{\bar{X}/n}} < 1.96\right) = 0.9$$

$$P\left(-1.96\sqrt{\frac{\theta}{n}} < \theta - \hat{\theta} < 1.96\sqrt{\frac{\bar{X}}{n}}\right) = 0.9$$

95% CI for θ = (172.28, 227.72)

2(i) $\bar{x} = \frac{\Sigma x_i}{n}$ $\quad$ $E(x_i) = \mu$, $V(x_i) = \sigma^2$

$$E(\bar{x}) = E\left(\frac{\Sigma x_i}{n}\right) = \frac{1}{n} \cdot E(x_1 + x_2 + \dots x_n)$$

$$= \frac{1}{n}\left(E(x_1) + E(x_2) + \dots + E(x_n)\right)$$

$$= \frac{1}{n}\left[\mu + \mu + \mu \dots + \mu\right)$$

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

$$v(\bar{x}) = v\left(\frac{\Sigma x_i}{n}\right) = \frac{1}{n^2} v(x_i + x_2 + \dots x_n)$$

$$= \frac{1}{n^2}\left[v(x_1) + v(x_2) + \dots + v(x_n)\right]$$

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \dots + \sigma^2\right]$$

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

$$v(\bar{x}) = \frac{\sigma^2}{n}.$$

ii) $$S^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2$$

$$E(S^2) = \frac{1}{n-1} E\left(\Sigma x_i^2 + \bar{x}^2 - 2\Sigma x_i \bar{x}\right)$$

$$= \frac{1}{n-1} E\left(\Sigma x_i^2 + \bar{x}^2 - 2n\bar{x}^2\right)$$

$$= \frac{1}{n-1}\left[n E(x_i^2) + n E(\bar{x})^2 - 2E(\bar{x})^2\right]$$

$$= \frac{1}{n-1}\left[n E(x_i)^2 - E(\bar{x})^2\right]$$

classmate

Date

Page

3

$$E(x_i^2) = [V(x_i) + (E(x_i))^2]$$

$$= \sigma^2 + \mu^2$$

$$E(\bar{x}^2) = V(\bar{x}) + (E(\bar{x}))^2$$

$$= \frac{\sigma^2}{n} + \mu^2$$

$$= \frac{1}{n-1}\left[n\sigma^2 + n\mu^2 - n\mu^2 - \sigma^2\right]$$

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

$$= \sigma^2$$

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

iii) $x_i \sim N(\mu, \sigma^2)$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

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

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

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

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

classmate
Date
Page

iv) $P[-0.5\,E(S^2) < S^2 < 0.5\,E(S^2)]$

$E(S^2) = \sigma^2 = 100$, $n = 100$

$P[-50 < S^2 < 50]$

$$P\left(\frac{9 \times -50}{100} < \frac{(n-1)S^2}{\sigma^2} < \frac{9 \times 50}{100}\right)$$

$P(-45 < \chi^2_9 < 4.5)$

$= P(\chi^2_9 < 4.5) - P(\chi^2_9 \leq -4.5)$

$= 0.1245$

3) $X \sim Bin(4, P)$

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

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

$l = 603 \log(p) + 117 \log(p)$

$$\frac{dl}{dp} = \frac{603}{1-p} + \frac{117}{p}(-1) = 0$$

$603p - 117 + 117p = 0$

$720p - 117 = 0$

$\hat{p} = 0.1625$

ii) H_0 : Data for no. of claims confirms to be binomial distribution.

H_1 = " "

$p = 0.1625$, $P(X = x) = {}^nC_x p^x (1-p)^{n-x}$

$P(X=0) = {}^nC_0 (0.1625)^0 . (0.8375)^4 = 0.4920$

$P(X=1) = {}^4C_1 (0.1625)^1 . (0.8375)^3 = 0.3818$

$P(X=2) = {}^4C_2 (0.1625)^2 . (0.8375)^2 = 0.1118$

$P(X=3) = {}^4C_3 (0.1625)^3 . (0.8375)^1 = 0.0144$

$P(X=4) = {}^4C_4 (0.1625)^4 . (0.8375)^0 = 0.0007$

No. of Claims	→	0	1	2	3	4
O_i	→	86	75	16	2	1

= 180

E_i → 180 × 0.492

180 × 0.3813   180 × 0.1111   180 × 0.0144   180 × 0.0007

~~$603p - 117 + 117p = 0$~~

~~$720p - 117 = 0$~~

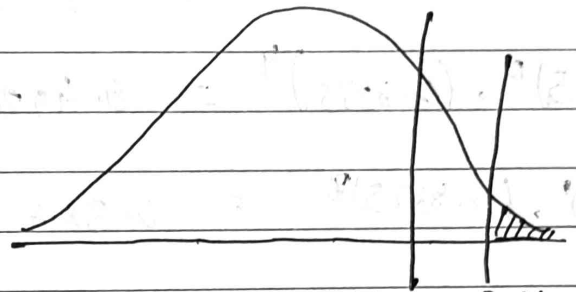
~~$\hat{p} = 0.1625$~~

$E_i \rightarrow$ 88.56 68.72 19.99 2.59 0.13

	0	1	2
O_i	86	75	19
E_i	88.56	68.72	22.71

$$\sum_{1}^{n} \left[\frac{(O_i - E_i)}{E_i} \right] = 1.2540 \sim \chi^2_{5-1-1,2}$$

$$= 1.2540 \sim \chi^2_1$$



Since, T.S lies in A.R, we have insufficient evidence to reject H_0 at 5% L.S.

$\therefore$ It concludes that data confirms to follow Binomial distribution.

iv) $f(x) = \frac{C\beta^2}{(x+\beta)^4}$; $x>0$, $\beta>0$

$$\int_0^\infty f(x)\,dx = 1$$

$$CB^3 \int_0^\infty \frac{1}{(x+B)^4} = 1$$

classmate Date ___ Page ___

$$CB^3 \left[\frac{(n+B)^{-3}}{-3} \right]_0^\infty = 1$$

$$\frac{CB^3}{-3} \left[0 - B^{-3} \right] = 1$$

$$C = 3$$

$$E(x) = \int_0^\infty x f(x)\, dx$$

$$= 3B^3 \int_0^\infty \frac{x}{(x+B)^{+4}}\, dx$$

$$= 3B^3 \int_0^\infty \frac{x+\beta}{(x+\beta)^{+4}} - \frac{B}{(x+\beta)^{+4}}\, dx$$

$$= 3\beta^3 \left[\frac{(x+\beta)^{-2}}{-2} - \frac{\beta(x+B)^{-3}}{-3} \right]_0^\infty$$

$$= 3\beta^3 \left[0 - 0 - \left[\left(\frac{\beta^{-2}}{-2} \right) - \left(\frac{\beta^{-2}}{-3} \right) \right] \right]$$

$$= 3\beta^3 \left[\frac{\beta^{-2}}{2} - \frac{\beta^{-2}}{3} \right]$$

$$= \frac{\beta^3}{2} \left[\beta^{-2} \right]$$

$$= \frac{\beta}{2}$$

$$E(x^2) = \int_0^\infty x^2 \; f(x)\, dx$$

$$= 3\beta^3 \int_0^\infty \frac{x^2}{(x+\beta)^4}\, dx$$

Using Pareto Dist^{n} $\rightarrow$ $\frac{3\beta^2}{5}$ $\Rightarrow$ $V(x)$

where $\alpha = 3$
$\lambda = \beta$

ii) $\beta = x_1, x_2, x_3, \dots x_n$

$E(x) = \bar{x}$

$\frac{\beta}{2} = \bar{x}$

$\hat{\beta} = 2\bar{x}x$

$E(\hat{\beta}),\ E(2\bar{x}_x) = 2E\left(\frac{\Sigma x_i}{n}\right)$

$\frac{2}{\cancel{x}n} \Sigma \cancel{E}(x_i) = \frac{\cancel{2}}{\cancel{n}} \times \cancel{n} \times \frac{\beta}{\cancel{2}} = \beta$

$\therefore$ Unbiased

$Var(\hat{\beta}) = Var(2\bar{x}n) = 4Var(\bar{x}n) = 4\,Var(x)/n$

$= 4 \times \frac{3\beta^2/n}{4} = \frac{3\beta^2}{n}$

Hence $MSE(\hat{\beta}) = \frac{3\beta^2}{n} + 0 = \frac{3\beta^2}{n}$

the estimatior is consistent since MSE tends to zero as $n \rightarrow \infty$

classmate

Date
Page

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

$$= b^2 \times \frac{3}{4} \times \frac{\beta^2}{n} + [E(b\bar{x}_n) - \beta]^2$$

$$= b^2 \times \frac{3}{4} \times \frac{\beta^2}{n} + \left[\beta\left(\frac{b}{2} - 1\right)\right]^2$$

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

Diff the MSE respect to b:

$$\frac{d(MSE)}{dt} = \left[\frac{\beta^2}{n}\left[\frac{3}{2}b + n\left(\frac{b}{2} - 1\right)\right]\right]$$

Setting this equal to 0 gives $b = \frac{2n}{n+3} = \frac{2}{1+3/n}$

Diff MSE for 2^{nd} times wrt b:

$$\frac{d^2}{dt^2}(MSE) = \frac{\beta^2}{n}\left(\frac{\alpha}{2} + \frac{n}{2}\right) > 0$$

$\therefore$ Min value for MSE at $b = \frac{2}{1+3/n}$

iv) It is seen in (ii) that $\beta = 2\bar{x}_n$ is $2n$ unbiased estimator. The estimator $b\bar{x}_n$ in iii) where $b = \frac{3}{1+3/n}$ is $-$vely biased.

estimator for $b < 2$, $x \geq 1$
As $n \to \infty$, b tends to 2, so estimator in iii is also Consistent as in ii)

5 i) $E(x) = \bar{x}$

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

$a = 2\bar{x} - b$

$V(x) = s^2$

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

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

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

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

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

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

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

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

iii) Sample $\rightarrow$ 1, 2, 3, 4, 50, $\bar{x} = 12$, $s = 21.23$

$\hat{a} = 12 - \sqrt{3}(21.27) = -24.84$

$\hat{b} = 12 + \sqrt{3}(21.27) = 48.84$

f $U(a, b)$, the prob of a sample pt being less than 'a' or greater than 'b' is zero and we have a sample value so that is greater than our estimate 'b'

It highlights 2 potential w.r.t to MOM

iv) $X \sim U(0, b)$

$f(x) = \frac{1}{b-a} = \frac{1}{b}$

$L = f(x) \cdot f(x_2) \ldots f(x_n)$

$= \frac{1}{b} \cdot \frac{1}{b} \ldots \frac{1}{b} = \frac{1}{b^n}$

$l = -n \log b$

11

$$\frac{dl}{db} = -\frac{n}{b} = 0 , b \to \infty$$

$$\frac{d^2 l}{dt^2} = \frac{n}{b^2} > 0 , \text{minimum at } b \to \infty$$

So using common sense, $L(\theta)$ is maximum when θ is minimum $b > \max x_i$

$$\hat{b} = \max x_i$$

6(i) $P(\text{type 1 error}) = P(\text{Reject } H_0 / H_0 \text{ is T})$

X be no. of hits in 1st step 12 missiles

Y be no. of hits in 2nd step 12 missiles

$$P(\text{Type 1 error}) = P(X \geq 3 / p = 0.1) + P(X = 0/p = 0.1) \cdot P(Y \geq 5/p = 0.1) + P(X = 1/p = 0.1) \cdot P(Y \geq 4/p = 0.1) + P(X = 2/p = 0.1) \cdot P(Y \geq 3/p = 0.1)$$

$$= (1 - 0.8891) + (0.2824)(1.09957) + (0.6590 - 0.2824)(1 - 0.9944) + (0.8891 - 0.6590)(1 - 0.8891)$$

$$= 0.14727 \approx 15\%$$

ii) $P(X \geq 3 / p = 0.3) + P(X = 0/ p = 0.3) \cdot P(Y = 5), P(0.3) + P(X = 1/p = 0.3) \cdot P(Y \geq 4/p = 0.3) + P(X = 2/ p = 0.3) \cdot P(Y = 3/p = 0.3)$

$$= 0.91253 \approx 91\%$$

iii) P(type II error) is the prob. of accepting H_0 where H_0 is F.

$$= 1 - 0.91253 = 0.08747 = 9\% \quad (p = 0.3)$$

iv) 10 space agencies fire 120 missiles and record 40 strike.

$$X \sim Bin(120, 1/3) \quad \hat{p} = \frac{40}{120} = \frac{1}{3}$$

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

$$P\left[-1.2816 < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < 1.2816\right]$$

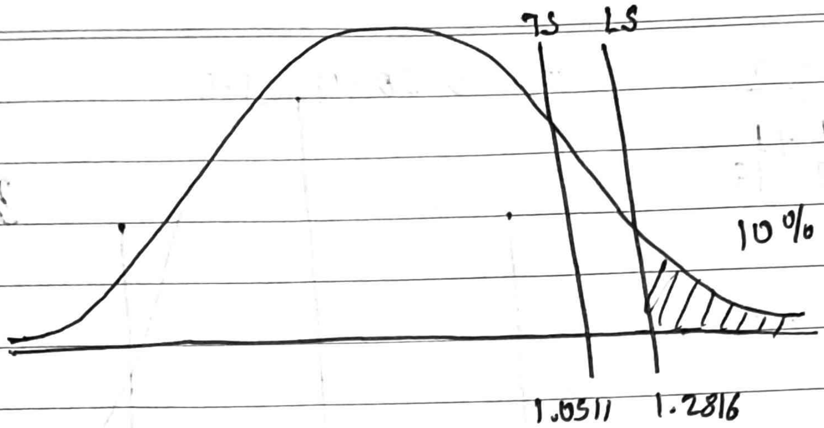
$[C]$ 90% $p \in (2.333, 0.2782)$

v) $p > 0.278$ at 10% LS

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

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

$$\frac{39.5 - 120 \times 0.278}{\sqrt{120 \times 0.278 \times 0.722}} = 1.2511$$



As T.S < 1.2816

We have insufficient evidence to reject H_0 at L.S 10%

vi) When Bound of CI implies that there is only a 10% change 'p' ≤ 0.2782, whereas from the hypothesis test 'p' could be less 0.2780 with prob. more than 10% (approx 10.5% corresponding to 1.2511)
The minor disconnect b/w CI, H.T at some level is due to.

1) Applying continuity correction in the T.
2) Use of sample proportion $\hat{p}$ to estimate popn variance in calculating CI

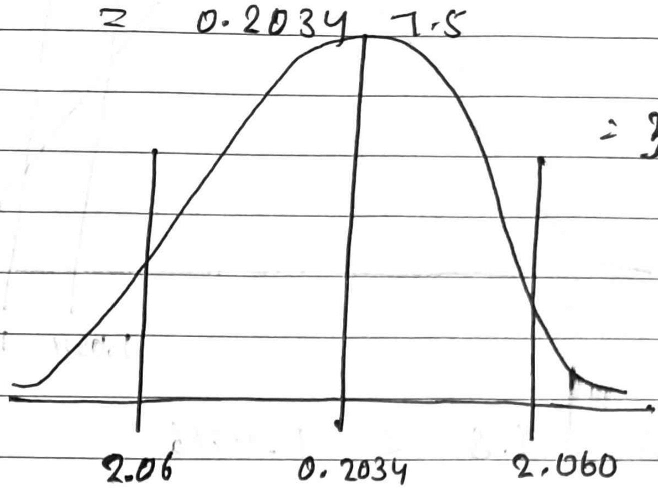
vii) $H_0 : \mu = \mu_0$ $H_1 : \mu_1 \neq \mu_2$

$$P : Q \Rightarrow \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

$$S_p = \frac{1}{25}\left(11 \times 2916 + 14 \times 4900\right) = 4027.04$$

$$\frac{65-70-6}{63.454\sqrt{\frac{1}{12}+\frac{1}{15}}} = 0.2034 \quad T.S$$

$= 5\ 2.5\%$



Since T.S lies in AR, we have inefficient evidence to reject H_0 at 5% L.S.

7) $\Sigma X_i = 270, \quad \bar{X}_{pu} = 30$

$\Sigma X_{p.ni} = 315.5, \quad \bar{x}_{pni} = 35.06$

$\Sigma X^2_{pu} = 2614.5, \quad S_1^2 = \frac{8614.5 - 9\times 30^2}{8}$

$= 64.37$

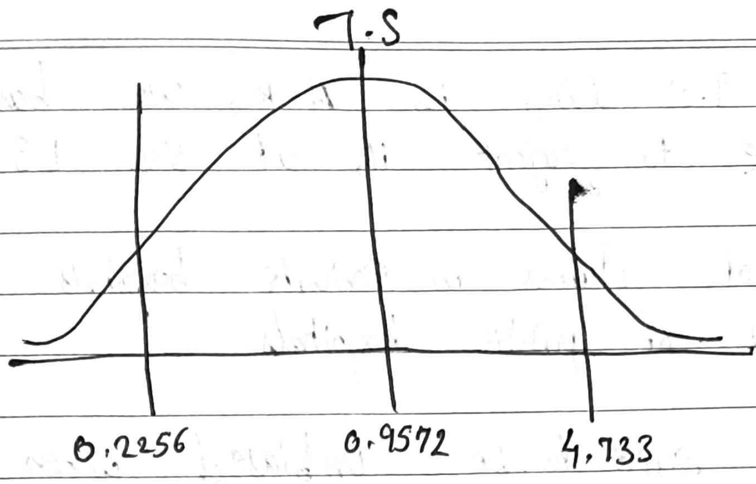
$S_2^2 = \frac{11599.25 - 9\times 35.06^2}{8} = 67.42$

ii) $H_0 : \sigma_1^2 = \sigma_2^2, \quad H_1 : \sigma_1^2 \neq \sigma_2^2$

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

$\frac{64.31}{67.42} = 0.9592 \quad \sim F_{8,8}$

$L. = 5.5\%$



Since T.S lies in A.R, we have insuff. evidence to reject H_0 at L.S 5%

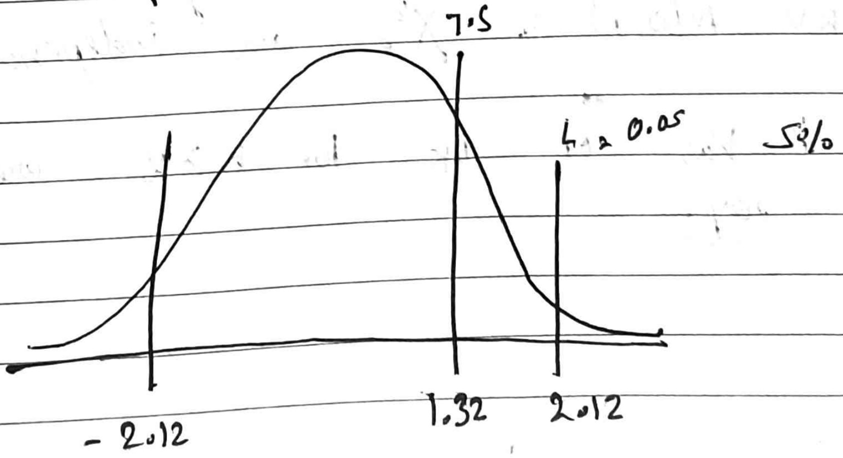
∴ Pop^n, Var are

iii) $H_0 : \mu = \mu_2$; $H_1 : \mu_1 < \mu_2$

$$T.S = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{64.31 \times 8 + 67.42 \times 8}{16} = 65.86$$

$$TS \rightarrow \frac{(35.06 - 30) - 0}{\sqrt{65.86}\sqrt{\frac{1}{9} + \frac{1}{9}}} = 1.32$$



Since, T.S lies in A.R, we have insuff. evidence to reject H_0 at 5% LS

∴ Cost of claims in private hospitals is similar to that of public hospitals

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

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

iii) M.S.E of this estimator $\hat{\theta}$ - $(E(\hat{\theta}) - \theta)^2$

iv) $\hat{\theta}$ is efficient as an estimator with lowes MSE is said to be more efficient than one with higher MSE

v) θ can be estimated using:

a) Method of Moments
b) MLE

9) i) Variable 1 is defined as $t_k = \frac{N(0,1)}{\sqrt{\chi^2_k / k}}$

where k = degree of freedom

& 2 RV $N(0,1)$ & χ^2_k are independent.

ii) Mean & Var of t_k for $k > 2$ are 0 and $\frac{k}{k-2}$ resp.

classmate

Date

Page

iii) $\frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$, $t_9 = 3.25$

CI for μ is thus $\left(50 - 3.25\sqrt{\frac{48.5067}{10}}\right.$,

$50 + 3.2$ i.e. $\left.(42.83 , 57.17\right)$

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

$\frac{\bar{x} - \mu}{S/\sqrt{n}} \sim N(0, \sqrt{9}/\sqrt{7})$

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

CI for μ $\in \left[50 - 2.58\sqrt{\frac{48.667 \times 9}{10 \times 7}}\right.$,

$\left.50 + 2.58\sqrt{\frac{48.667}{10} \times \frac{9}{7}}\right]$

ie $(43.54, 56.45)$.

10) i) $P(\text{type I error}) = p(\text{Rej } H_0 / H_0 \text{ is T})$

$X \sim Poi(n\mu)$ ie $Poi(3000)$,

H_0, $X \sim Poi(3000)$

$P[\text{Reject } H_0 (H_0 \text{ is T})] = P(X > 3100$ where $X \sim Poi(3000))$

$X \sim N(3000, 3000)$

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

$1 - P(Z < 1.825) = 3.39\%$

ii) P(type II error) = P(Accepting H_0 / H_0 is F)

iii) Power of test - Prob of rejecting / H_0 is f

$P(\text{reject } H_0 / H_0 \text{ is } F) = P(X > 3100$ where $X \sim Poi(n\mu) \sim N(n\mu, n\mu))$

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

∴ Power of test depends on value of this parameter.

iv) $\hat{\mu} \sim N(\mu, \hat{\mu}/n)$

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

$$P\left(-2.5758 < \frac{2.9 - \mu}{\sqrt{2.9/1000}} < 2.5758\right) = 0.99$$

$\mu \in (2.7613, 3.0387)$

v) revised sample mean $= \dfrac{213200 - 11000 - 48000 + 27500 + 31500}{18}$

$= 11,767$

$\Sigma x^2 = 4243360000$

$\therefore$ Revised $S.D = \left[\frac{1}{14}\left(424336000 - \frac{213200^2}{12}\right)\right]$

$= 6435$

12) i) The CRLB result holds under very general condition except where range of $dist^n$ involves the parameter such as uniform $dist^n$ in this case.

11) $Y = \text{Max}\ X_i$

Since X_i lies b/w 0 & θ, the support for Y will be $0 \times \theta$, for $0 < y < \theta$

$P(Y < y) = P(\text{Max}\ X_i < y)$

$= P\left[\bigcap_{i=1}^{n} (X_i < y)\right]$

$= \prod_{i=1}^{n}\left[\int_0^{y} \frac{dx}{\theta}\right] = \left(\frac{y}{\theta}\right)^n$

$= \frac{n}{\theta^n} y^{n-1}$ for $0 < y < \theta$

now for $K \geq 0$

$E(Y^K) = \int_0^{\theta} y^K \frac{n}{\theta^n} y^{n-1} dy$

$= \frac{n\theta^k}{n+K}$

20

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

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

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

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

$$MSE(\theta(c)) = E((\theta(c) - \theta)^2)$$

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

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

vi) In order to minimize error in estimation, it is preferred to opt for an estimate which has lower MSE among diff. competing estimators so here $\hat{\theta}(c_m)$ will be preferred over $\hat{\theta}(c_u)$ As n because large $\hat{\theta}(c_u) = \frac{n+1}{n} y \to y$ ii, $\hat{\theta}(c_m) = \frac{n+2}{n+1} y \to y$,

the two estimator becomes one & same as $n \to \infty$