ASSIGNMENT-1

Q.1) $\hat{\lambda} = \frac{\Sigma x_i}{n} = 200$, θ

$P(z_{2.5\%} < z < z_{97.5\%}) = 0.95$

$P(-1.96 < z < 1.96) = 0.95$

$P\left(-1.96 < \frac{\hat{\lambda}-\lambda}{\sqrt{\frac{\hat{\lambda}}{n}}} < 1.96\right) = 0.95$

$P\left(\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}}{n}} < \lambda < \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}}{n}}\right) = 0.95$

95% confidence interval for $\lambda = \left(200 - 1.96\sqrt{200},\ 200 + 1.96\sqrt{200}\right)$

$= (172.2814,\ 227.7185)$

Q.2) $\bar{x} = \frac{\Sigma x_i}{n}$, $s^2 = \frac{1}{n-1}\left(\Sigma x_i^2 - n\bar{x}^2\right)$

(i) $E(\bar{x}) = E\left(\frac{\Sigma x_i}{n}\right) = \frac{1}{n}\Sigma E(x_i) = \frac{1}{n} \times n\mu = \mu$

$V(\bar{x}) = V\left(\frac{\Sigma x_i}{n}\right) = \frac{\Sigma V(x_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

$\therefore E(\bar{x}) = \mu$ and $V(\bar{x}) = \frac{\sigma^2}{n}$

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

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

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

$= \frac{1}{n-1}\left(\Sigma\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right)$

$$E(s^2) = \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)$$

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

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

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

$$var\left(\frac{(n-1)s^2}{\sigma^2}\right) = var(\chi^2_{n-1})$$

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

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

(iv) $P(0.5\sigma^2 < s^2 < 1.5\sigma^2)$

$= P(s^2 < 1.5\sigma^2) - P(s^2 < 0.5\sigma^2)$

$= \cancel{P(s^2<150) - P(s^2<50)}$ ... $(\cancel{\sigma^2=100, n=10})$

$= \cancel{P}$

$= P\left(\frac{9s^2}{\sigma^2} < 1.5\times 9\right) - P\left(\frac{9s^2}{\sigma^2} < 0.5\times 9\right)$

$= P(\chi^2_9 < 13.5) - P(\chi^2_9 < 4.5)$

$= 0.8587 - 0.1245$

$= 0.7342.$

$P(0.5\sigma^2 < s^2 < 1.5\sigma^2) = 0.7342.$

Q.3) (i) $L = \prod_{i=1}^{180} f(x_i)$

$= P(X=0)^{86} * P(X=1)^{75} * P(X=2)^{16} * P(X=3)^{2} * P(X=4)^{1}$

$= \left({}^4C_0\, p^0 (1-p)^4\right)^{86} * \left({}^4C_1\, p^1 (1-p)^3\right)^{75} * \left({}^4C_2\, p^2 (1-p)^2\right)^{16}$

$* \left({}^4C_3\, p^3 (1-p)\right)^2 * \left({}^4C_4\, p^4 (1-p)^0\right)^1$

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

$\log L = 117 \log p + 603 \log(1-p) +$ constant

$$\frac{d \log L}{dp} = \frac{117}{p} - \frac{603}{1-p}$$

equating it with 0.

$$117(1-\hat{p}) - 603\hat{p} = 0$$
$$117 - 117\hat{p} - 603\hat{p} = 0$$
$$\hat{p} = 0.1625$$

again differentiating to check whether it is maxima ~~or minima~~ or not,

$$\frac{d^2 \log L}{dp^2} = \frac{-117}{p^2} + \frac{603}{(1-p)^2}$$

$$\left(\frac{d^2 \log L}{dp^2}\right)_{\text{at } p=0.1625} = -3571.067738 < 0$$

so it is maxima at $p = 0.1625$

(ii) H_0: no. of claims conforms to the binomial model.
H_1: no. of claims does not conforms to the binomial model.

	0	1	2	3	4
O_i	86	75	16	2	1
E_i	88.542	68.724	19.998	2.578	0.108

$\hat{p} = 0.1625$ $\quad P(X=0) = 0.4818 = {}^4C_0\, p(1-p)^4$

$P(X=1) = 0.3818$

$P(X=2) = 0.1111$

$P(X=3) = 0.0143$

$P(X=4) = 0.0006$

O_i	86	75	19
E_i	88.542	68.724	22.68

$$T.S. = \frac{\sum (O_i - E_i)^2}{E_i} \sim \chi^2_{5-1-1-2}$$

$$1.24322 \sim \chi^2_1$$

We have insufficient evident to reject H_0 at 5% level of significance. Therefore, no. of claims conform to the binomial model.

4) $f(x) = \dfrac{c\beta^3}{(x+\beta)^4}$; $x>0, \beta>0$

(i) $PDF = \dfrac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}$ comparing above $f(x)$ to P.D.F of Pareto distribution (2 parameter version)

Thus, β corresponds to λ.

$\Rightarrow$ Thus, $c=3$ (as c corresponds to α here)

$\Rightarrow E(X) = \dfrac{\lambda}{\alpha-1}$ $\quad E(X) = \dfrac{\beta}{2}$ $\quad V(X) = \dfrac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = \dfrac{3\beta^2}{4}$

(ii) $E(X) = \frac{\beta}{2}$ ~~$\bar{x}$~~ $\qquad \bar{X} = \frac{\hat{\beta}}{2}$

$$\therefore 2\bar{X} = \hat{\beta}$$

$$Bias = E(\hat{\beta}) - \beta = 2E\left(\frac{\Sigma X_i}{n}\right) - \beta = \frac{2}{n}\Sigma E(X_i) - \beta = \frac{2}{n} \times \frac{n\beta}{2} - \beta = 0$$

hence it is unbiased.

$$MSE = \cancel{V E \hat{\beta}}\ Var(\hat{\beta}) + Bias^2$$
$$= Var(\hat{\beta}) = Var(2\bar{X}) = 2\,Var\left(\frac{\Sigma X_i}{n}\right) = \frac{2}{n^2} Var(\Sigma X_i) = \frac{3\beta^2}{n}$$

as $n \to \infty$, $MSE \to 0$, hence $\hat{\beta}$ is consistent.

(iii) $\hat{\beta}_2 = b\bar{X} \qquad E(\hat{\beta}_2) = E(b\bar{X}) = \frac{b}{n}E(\Sigma X_i) = \frac{b}{n} \times \frac{n\beta}{2} = \frac{b\beta}{2}$

$$Var(\hat{\beta}_2) = \frac{b^2}{n^2} Var(\Sigma X_i) = \frac{b^2}{n^2}\, n * Var(X_i) = \frac{b^2}{n} \times \frac{3}{4}\beta^2 = \frac{3b^2\beta^2}{4n}$$

$$MSE = Var(\hat{\beta}_2) + \cancel{E(\hat{\beta}_2)}\ bias^2$$
$$= \frac{3b^2\beta^2}{4n} + \left(\frac{b\beta}{2} - \beta\right)^2$$

$$MSE = \frac{\beta^2}{4n}\left(3b^2 + nb^2 + 4n - 4nb\right)$$

$$\frac{d\,MSE}{db} = \frac{\beta^2}{4n}\left(6b + 2nb - 4n\right)$$

~~db~~ equating it with 0;

$$\frac{\beta^2}{4n}\left(6\hat{b} + 2n\hat{b} - 4n\right) = 0$$

$$\hat{b} = \frac{2n}{3+n}$$

again differentiating to check whether it is minima or not,

$$\frac{d^2 MSE}{db^2} = \frac{\beta^2}{4n}(6+2n) > 0 \text{ hence it is minimum.}$$

thus $b = \frac{2n}{3+n}$, minimises the MSE.

(iv) $E(\hat{\beta}_2) = \frac{b\beta}{2} = \frac{n\beta}{(n+3)}$

$$\text{Bias} = E(\hat{\beta}_2) - \beta = \frac{n\beta - n\beta - 3\beta}{n+3} = \frac{-3\beta}{n+3}$$

~~Thus, we ha~~ Thus it is underestimately bias (negatively bias)

$$MSE = \frac{\beta^2}{4n}(3b^2 + nb^2 + 4n - 4nb)$$

$$= \frac{12n\beta^2}{n+3}$$

as $n \to \infty$, MSE don't tend to 0, Hence it is inconsistent.

Q.5) (i) $X \sim U(a,b)$

$E(X) = \frac{a+b}{2}$ $\qquad$ $V(X) = \frac{(b-a)^2}{12}$

$\bar{x} = \frac{a+b}{2}$ $\qquad$ $12S^2 = (b-a)^2 = 4(\bar{x}-a)^2$

$a = 2\bar{x} - b$ ~~vnlis~~ $\qquad$ $\sqrt{3}S = \bar{x} - a$

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

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

(ii) $b = 2\bar{x} - a$

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

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

(iii) $\bar{x} = 42.2$ $\quad$ $\hat{a} = -67.906$

$s = 63.57$

$$\hat{b} = \bar{x} + \sqrt{3}s = 152.3064$$

$\hat{b}$ is very close to sample value but $\hat{a}$ is very far from the true sample value, it is the potential weakness of mom.

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

$$L = \prod f(x_i) = \left(\frac{1}{b}\right)^n$$

$$\log L = -n \log b$$

$$\frac{d \log L}{db} = \frac{-n}{b}$$

equating it with zero,

$$\frac{-n}{\hat{b}} = 0$$, gives $b = \infty$ which is not possible

Also,

$$\frac{d^2 \log L}{db^2} = \frac{n}{b^2} > 0$$

thus we say, $\hat{b} = \max(x_i)$

6) $H_0: p = 0.1$
$H_1: p > 0.1$

(i) P(Type 1 error) = P(rejecting H_0 | H_0 is true)

X be no. of hits in first 12 minutes
Y be no. of hits in step 12 min.

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

$$= (1-0.8891) + 0.2824(1-0.8857) + (0.6580-0.2824) * (1-0.9744) + (0.8891-0.6590)(1-0.8881)$$
$= 0.147$
$= 14.7\%$

(ii) Probability of regretting null hypothesis when $p = 0.3$

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

$$= (1-0.2528) + 0.0130(1-0.7237) + (0.0850-0.0198)(1-0.4825) + (0.2528-0.0850)(1-0.2528)$$

$= 0.9125$

$= 91.25\%$

(iii) P(type 2 error) $= 1 - 0.91253$
$= 0.08747$
$\approx 9\%$ (when $p = 0.3$)

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

$n = 120, x = 40$

$\hat{p} = \dfrac{40}{120} = \dfrac{1}{3} = 0.333$

$z_{10\%} = 1.2816$

Lower bound of 90% right tailed confidence interval for p is,

$\hat{p} - 1.2816\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 0.3333 - 1.2816\sqrt{\dfrac{1/3\,(1-1/3)}{120}} = 0.2782.$

(v) $p > 0.278$ at 10% LOS

$H_0: p = 0.278.$

$H_1: p > 0.278$

One sample binomial model.

$\dfrac{x - np_0}{\sqrt{np_0q_0}} \sim N(0,1)$

$\dfrac{39.5 - 120(0.278)}{\sqrt{120(0.278)\times 0.7222}} = 1.2511$

We donot have sufficient evidence to reject H_0 at 10% level of significa

(vi) Lower bound of confidence interval implies that their is only a 10% ~~change~~ chance of $p \leq 0.2782$ whereas from the hypothesis test, 'p' could be less than or equal to 0.2780 with probability of more than 10%.

(vii) $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

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

$$S_p^2 = \frac{1}{25}\left[11(2916)+14(4800)\right] = 4027.04.$$

$$T.S = \frac{65-70}{63.458\sqrt{\frac{1}{12}+\frac{1}{15}}} = -0.2034.$$

Their is ± 2.060 (at $t_{25, 2.5\%}$) so we have insufficient evidence to reject H_0 at 5% level of significance.

Q.7) (i) Public (ii) B

$\bar{x}_1 = 30$ $\bar{x}_2 = 35.055$

$s_1^2 = 64.3125$ $s_2^2 = 67.4027$

(ii) 2 sided t-test can be applied in case of samples come from populations with equal variance

We are testing $H_0: \sigma_1^2 = \sigma_2^2$

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

Test statistics is $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1+n_2-1}$

Value of test statistics $= \frac{64.31}{67.42} = 0.9542.$

$F_{8,8}$ values at 5% level are 0.2256 and 4.433 since the value of the T.S lie b/w the above value, we have insufficient evidence to reject our hypothesis and conclude that the population variance are equal.

(iii) H_0 $\mu_{pr} = \mu_{pub}$
H_1 $\mu_{pr} > \mu_{pub}$

$$T.S. = \frac{(\bar{x}_2 - \bar{x}_1) - (\mu_2 - \mu_1)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{35.055 - 30}{8.1152 \sqrt{\frac{2}{9}}}$$

$$= 1.32151$$

tabulated value $= t_{18, 5\%} = 1.746$.
Since, T.S cal < T tabulated value.
We have insufficient evidence to reject null.

Q.8) (i) Unbiased estimator
⇒ An unbiased estimator is an estimator whose expected value is equal to the parameter's value.

(ii) Biased estimator
⇒ A biased estimator is one which delivers an estimate which is consistently different from the parameters to be estimated.

(iii) The MSE of an estimator is $\hat{\theta}$ is defined by,

$$MSE(\hat{\theta}) = Var(\theta) + Bias^2$$

(iv) The estimator which has less MSE is more consitent

v) They both will be considered to be consistent when MSE minimise such that when $n \to \infty$, $MSE \to 0$

(vi) (i) method of moments (mom)
(ii) Maximum likelihood estimator (MLE)

Q.4) (i) Variable t_k is defined as $t_k = \frac{N(0,1)}{\sqrt{\chi^2_k / k}}$ where k denotes DOF and the 2 random var $N(0,1)$ and χ^2_k are also independent

(ii) For $k > 2$ $\quad E(t_k) = 0$

$$V(t_k) = \frac{k}{k-2}$$

(iii) a) $\frac{\bar{x}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$$P\left(\bar{x} - t_{9,0.5\%}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{9,0.5\%}\frac{S}{\sqrt{n}}\right) = 0.99$$

$t_{9,0.5\%} = 3.250$

99% CI for $\mu = \left(50 - 3.25\sqrt{\frac{48.667}{10}},\ 50 + 3.25\sqrt{\frac{48.667}{10}}\right)$

$= (42.83, 57.17)$

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

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

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

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

critical values LOC is 2.58

$CI = \left(50 - 2.58\sqrt{\frac{48.667 \times 9}{10 \times 7}},\ 50 + 2.58\sqrt{\frac{48.667 \times 9}{70}}\right)$

$CI = (43.54, 56.45)$

10) $n = 1000$ $\quad X \sim Poi(n\lambda)$

$\lambda = 3.$ $\quad X \sim Poi(3000)$

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

$= P(X > 3100 \mid X \sim Poi(3000))$

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

$= 1 - P(Z < 1.825)$

$= 3.39\%.$

(ii) Type 2 error = event of accepting the hypothesis when it is false.

(iii) Power of dist = $P(\text{rejecting } H_0 \mid H_0 \text{ is false})$

$= P(X > 3100 \mid X \sim Poi(n\lambda))$

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

The value of power of test will depend on the value of parameter values after nature hypothesis.

(iv) $\hat{\mu}$ is estimator of μ, $\hat{\mu}$ follows $N\left(\mu, \frac{\hat{\mu}}{n}\right)$

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

Or $P\left(-2.5758 < \frac{2.9 - \mu}{\sqrt{2.9/1000}} < 2.5758\right) = \cancel{0.82029}\ 0.98$

Hence the confidence interval for μ is $(2.7613, 3.0389)$

Q.11) new sample mean:

$$\Sigma X_1 = \Sigma X - 11{,}000 - 48000 + 27500 + 31500 = 213\,200$$

$$\bar{X}_1 = \frac{213200}{12} = 17{,}767.$$

new standard deviation:

$$\Sigma X_1^2 = 4818\,860{,}000 - (11{,}000)^2 - (48000)^2 + (27500)^2 + (31500)^2$$
$$= 4\,243\,360{,}000$$

$$S_1^2 = \frac{1}{n-1}\left(\Sigma X_1^2 - n(\bar{X}_1)^2\right)$$

$$= 41396775.64$$

$$S_1 = 6434.032617$$

The sample mean remains the same ~~it~~ after the inclusion of ~~2 parameter~~ 2 employees as the 2 values removed where extreme values and the 2 values included were around the sample mean.

The sample ~~sd~~ S.D. reduces as the 2 values now included do not deviate very much from the sample mean causing variance to reduce and hence the reduction of the sample s.d.

Q.12) (i) ~~the~~ CRLB holds very general condition expect the range of the distribution involves the paramters, such as the uniform ~~distn~~ discontinuity, to the derivative in the formula make sense

(ii) $Y = \max X_i$

for $0 < y < \theta$

$$P(Y < y) = P[\max X_i < y]$$

$= \prod P(X_i < y) \cdots$ (X_i are independent)

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

$= \left(\frac{y}{\theta} \right)^n$

thus the probability density funcn of y will be,

$g_Y(y) = \frac{n}{\theta^n} y^{n-1}$ for $0 < y < \theta$.

now, for any non negative real no. k.

$E[Y^k] = \int_0^\theta y^k \frac{n}{\theta^n} y^{n-1} dy$

$= \frac{n}{n+k} \left(\frac{\theta^{n+k} - 0}{\theta^n} \right)$

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

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

$= E(cY) - \theta$

$= c\,E(Y) - \theta$

$= \frac{nc\theta}{n+1} - \theta$

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

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

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

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

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

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

(iv) For $\hat{\theta}(c)$ to be an unbiased estimator of θ, we need

$$\text{Bias}(\hat{\theta}(c_\mu)) = 0$$

$$\left(\frac{c_\mu n}{n+1} - 1\right)\theta = 0$$

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

(v) in order to minimise $MSE[\hat{\theta}(c)]$, we need:

$$0 = \frac{d}{dc} MSE(\hat{\theta}(c))$$

$$0 = \frac{d}{dc}\left(c^2\frac{n\theta^2}{n+2} - c\frac{2n\theta^2}{n+1} + \theta^2\right)$$

$$0 = \frac{2cn\theta^2}{n+2} - \frac{2n\theta^2}{n+1}$$

thus $c_m = \dfrac{n+2}{n+1}$

note that:

$$\frac{d^2}{dc^2} MSE(\hat{\theta}(c)) = \frac{2n\theta^2}{n+2} > 0$$ ~~as~~ so it is minima.

(vi) in order to minimise error in estimator, it is preferred to get for an estimator which has lower MSE among diff estimators. So here $\hat{\theta}(c_m)$ will be preferred over $\hat{\theta}(c_\mu)$

as n becomes large, $\hat{\theta}(CM) = \frac{n+1}{n} Y \to Y$

similarily $\hat{\theta}(UM) = \frac{n+2}{n+1} Y \to Y$

Thus the two estimators are ~~consistent~~ same ~~as~~ when $n \to \infty$, ~~MSE~~