Probability Assignment-I
& Statistics

1) Ans :-

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

$$\hat{\lambda} = \frac{\Sigma x_i}{n}$$

$$\hat{\lambda} = 200$$

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

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

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

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

95% C.I. of λ when $\hat{\lambda} = 200$ and $n = 1$,

95% C.I. for $\lambda = \left(200 - 1.96\sqrt{200},\ 200 + 1.96\sqrt{200}\right)$

$= (172.28141,\ 227.7185)$

2) Ans :-

(i) Sample mean : $\bar{x} = \frac{\sum x_i}{n}$

$E[\sum x_i] = \sum E[x_i] = \sum \mu = n\mu$ $\quad \because$ (identically distributed)

$var[\sum x_i] = \sum Var[x_i] = n\sigma^2$ $\quad \because$ (iid)

$E[\bar{x}] = \mu$

$Var[\bar{x}] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$

(ii) Sample variance : $S^2 = \frac{1}{n-1}\left[\sum x_i^2 - nX^2\right]$

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

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

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

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

$$= \sigma^2$$

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

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

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

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

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

(iv) We need,

$P(0.5\sigma^2 < S^2 < 1.5\sigma^2)$

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

$n = 10$

$\sigma^2 = 100$

$P(S^2 < 1.5\sigma^2) = P(S^2 < 150)$

$= P\left(9S^2/\sigma^2 < \frac{150 \times 9}{100}\right)$

$= P(\chi_9^2 < 13.5)$

$= 0.8587$

$P(S^2 < 0.5\sigma^2) = P(S^2 < 50)$

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

$= P(\chi_9^2 < 4.5)$

$= 0.1245$

$P(50 < S^2 < 150) = 0.8587 - 0.1245$

$= 0.7342$

3) Ans:-

(i) The likelihood function is:-

$L(P) = C\,[(1-p)^4]^{86}[4p(1-p)^3]^{75}[6p^2(1-p)^2]^{16}[4p^3(1-p)]^2[p^4]^1$

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

$= C(1-p)^{603}\,p^{117}$ ... c is constant

Taking log and diffrentiating w.r.t. p

$\frac{d[\ln(L)]}{dp} = -\frac{603}{(1-p)} + \frac{117}{p^2} < 0$

$\Rightarrow$ Maximum value for $\ln(L)$ is at $p = 0.1625$

(ii) Goodness of fit test:-

We are testing the following hypothese using a χ^2 goodness of fit test:-

H_0 = the probabilities confirm Bin(4, P)

H_1 = The probabilities do not confirm Bin(4, p)

- Using $\hat{p} = 0.1625$ from part (i), the probablities for this binomial distribution are:-

$P(x=0) = (1-p)^4 = 0.49197$

$P(x=1) = 4p(1-p)^3 = 0.38183$

$P(x=2) = 6p^2(1-p)^2 = 0.11113$

$P(x=3) = 4p^3(1-p) = 0.01437$

$P(x=4) = p^4 = 0.00070$

The expected values are 88.55, 68.73, 20.00, 2.59 and 0.13.

Combining the expected values less than 5 to third group, we get.

No. of Claims	Observed (O_i)	Expected (E_i)
0	86	88.55
1	75	68.73
2 or more	19	22.72

The degrees of freedom $= 3-1-1 = 1$

$$\chi^2 = \sum (O_i - E_i)^2 / E_i$$

$$= (86 - 88.55)^2 / 88.55 + (75 - 68.73)^2 / 68.73 + (19 - 22.72)^2 / 22.75$$

$$= 0.074 + 0.572 + 0.608$$

$$= 1.254$$

This is less than the 5% critical value of 3.841. We have insufficient evidence at 5% level to reject H_0. Hence, the model is a good fit.

4) Ans:-

(i) from 1st principle

$f(x) = \frac{\alpha\lambda^a}{(\lambda+x)^{a+1}}$; $x>0$, $\lambda>0$ and $\alpha>0$

$$\int_0^\infty \left((\beta^3/(x+\beta)^4 \right) dx = \left[-\frac{c}{3} \frac{\beta^3}{(x+\beta)^3} \right]_0^\infty = 1$$

$\because E(x) = 0.5$ & $Var(x) = (0.5)$

$\therefore$ $\underline{\underline{c = 3}}$

(ii) Equating mean to (0.5) gives -

$\bar{X}_n = \beta/2 \Rightarrow$ The moment estimator $\hat{\beta} = 2\bar{X}_n$

The mean square error or $\hat{\beta}$ -

$MSE[\hat{\beta}] = var[\hat{\beta}] + (Bias\,[\hat{\beta}])^2$

$E[\hat{\beta}] = E[2\bar{X}_n] = 2E[\bar{X}_n]$

$= 2E[X]$

$= 2(\beta/2)$

$= \beta$

and, $Bias = E[\hat{\beta}] - \beta = 0$

$\therefore$ This estimator is unbiased.

Using the fact that the individual values of X_i are independent -

$Var[\hat{\beta}] = Var[2\bar{X}_n] = 4\,Var[\bar{X}_n] = 4\,Var[X]/n$

$= 4(3/4\,\beta^2)/n$

$= 3\beta^2/n$

Hence, $MSE[\hat{\beta}] = 3\beta^2/n + 0 = 3\beta^2/n$

This estimator is consistent scince MSE tends to zero as $n \to \infty$

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

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

$$= b^2 \frac{3}{4}\frac{\beta^2}{n} + (\beta(b/2 - 1))^2$$

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

Differentiating the MSE respect to b -

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

Setting this equal to 0 gives $b = 2n/(n+3)$

$$= 2/\left(1 + \frac{3}{n}\right)$$

$$\frac{d^2(MSE)}{db^2} = \frac{\beta^2}{n}\left(\frac{3}{2} + \frac{n}{2}\right) \quad \ldots \quad (+ve)$$

$\therefore$ a minimum value for the MSE is at $b = 2/(1+3/n)$.

(iv) It is seen in (ii) that $\hat{\beta} = 2\bar{X}_n$ is an unbiased estimator.
The estimator $b\bar{X}_n$ in (iii) when $b = \frac{2}{(1+3/n)}$ is -vely biased estimator as $b<2$ for $n \geq 1$.

As $n \to \infty$, $b \to 2$.
So, the estimator in (iii) is also consistent in veiw of (ii)

5) Ans :-

(i) $X \sim U(a,b)$

$\rightarrow E(X) = \frac{a+b}{2}$

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

$a = 2\bar{x} - b$ — (1)

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

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

$12s^2 = (b-a)^2$

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

$2\sqrt{3}\,s = 2\bar{x} - a - a$ .. $[b = 2\bar{x} - a]$

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

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

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

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

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

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

(iii) Sample of observations : 1, 2, 8, 50, 150

$\bar{x} = 42.5$

$s = 63.57$

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

$= -67.906$ $= 152.3064$

Hence we can conclude that value of $\hat{b}$ is approx close to sample value but $\hat{a}$ is very far from the true sample value, so it is the potential weathe of method of moments

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

$$L = \prod_{i=1}^{n} f(x_i)$$

$$= \left(\frac{1}{b}\right)^n$$

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

$$\frac{dl}{db} = \frac{-n}{b}$$

$\frac{dl}{db} = 0$, gives $b = \infty$ which is not possible.

Also,

$$\frac{d^2 l}{db^2} = \frac{n}{b^2}$$

$\frac{n}{b^2} > 0$, which we have a minimum likelihood estimate

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

6) Ans :-

(i) P(type I error) = P(rejecting H_0)

Let X be the no. of hits is first 12 missiles and Y be the no. of hits in step 12 missiles.

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

$$= (1-0.889) + [0.2824(1-0.9857) + (0.6590-0.2824)(1-0.9744) + (0.8891-0.6590)(1-0.8891)$$

$$= 0.14727 \approx 15\%$$

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

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

$$= (1-0.2528) + 0.0138\,(1-0.7237) + (0.0850-0.0138)(1-0.4925) + 10.2528 - 0.0850)(1-0.2528)$$

$$= 0.9125 \approx 91\%$$

(iii) Probability of the type II error

Accepting H_0 where H_0 is false

$$= 1 - 0.91253$$

$$= 0.08747$$

$$\approx 9\%$$

(iv) 10 space agencies is aggregate tried 120 missiles and recorded 40 hits.

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

$n = 120$, $X = 40$

$$\hat{p} = \frac{40}{120} = \frac{1}{3} = 0.3333$$

$Z_{10\%} = 1.2816$

lower bond of 90% ~~ritht~~ right tailed C.I. 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) $p > 0.278$ at 10% L.O.S.

$H_0 = p = 0.278$ vs $H_1 = p > 0.278$

Sample binomial model.

$$\frac{X - np_0}{\sqrt{np_0q_0}} \sim N(0,1) \text{ with continuity coved}^n$$

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

We don't have sufficied evidence to reject H_0 at 10% level of significance

(vi) Lower bond of C.I. implies that the ~~policy~~ is only of 10% change of $p \leq 0.2782$ whereas from the hypothesis test, 'p' could be less than or equal to 0.2780 with probablity of more than 10%

(vii) $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$

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

Given,

$\bar{X}_1 = 65$, $\bar{x}_2 = 70$

$S_1 = 54$, $S_2 = 70$

$n_1 = 12$, $n_2 = 15$

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

$$\frac{65-70}{65.459\sqrt{\frac{1}{12} + \frac{1}{15}}} = -0.2034$$

There is ± 2.060 $(= t_{25, 2.5\%})$

So we have insufficient evidence to reject H_0 at 5% level.

7) Ans :-

(i) Public

$\bar{x}_1 = 30$

$S_1^2 = 64.3125$

Private

$\bar{x}_2 = 35.0555$

$S_2^2 = 67.40217$

(ii) 2sided t-test can be applied in case the samples come from population with equal variance.

We are testing $H_0 = \sigma_1^2 = \sigma_2^2$ Vs $H_1 = \sigma_1^2 \neq \sigma_2^2$

Test statistics is $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$

Value of static is $\dfrac{64.31}{67.42} = 0.9542$

$F_{8,8}$ values at 5% levels are 0.2256 & 4.433. Since the value of the statistics b/w the above value, we have sufficient evidence to reject the hypothesis and conclute the population variance are equal.

(iii) $H_0 = \mu_{Pr} = \mu_{Pu}$ & $H_1 \rightarrow \mu_{Pr} > \mu_{Pu}$

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

$$= \frac{35.0555 - 30}{8.1152 \sqrt{\frac{2}{9}}} = 1.32151$$

Tab value $= t_{16, 5\%} = 1.746$

Since, $T_{cal} < T_{tab}$ value

We have insuffcient evidence to reject H_0.

8) Ans:-

(i) Unbiased estimation

→ An unbiased estimation is an estimation where expected value is equal to the parameters.

→ Estimation is said to be unbaised if beat is equal to zero for all valeus of parameter θ.

(ii) Biased.

→ A biased estimation is one which delivers on estimate which is completely diffrent from the parameter to be estemated.

(iii) The mean squarer error of an estimate θ is defined by,

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

(iv) The estimater having less MSE is more conselent.

(v) They both would be considered to be consestent when MSE minimun such that when, $n \to \infty$; $MSE \to 0$.

(vi) (i) Method of moments.

(ii) Maximum likelyhood estimater.

10) Ans :-

$n = 1000$

$\lambda = 3$

$X \sim Poi(n\lambda)$

$X \sim Poi(300)$

(i) ~~P(Type I er~~

$P(\text{Type I error}) = P(\text{regret } H_0 / H_0 \text{ is true})$

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

using normal approx.

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

$= 1 - P(Z < 1.825)$

$= 3.39\%$

(ii) Type II error = event of accepting the hypothesis when is false

(iii) Power of list $= P(\text{regretting } H_0 / H_0 \text{ is false})$

$= P[X > 3100 / X \sim Poi(n\lambda) - N(n\lambda, n\lambda)]$

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

The value of power of that will depent on the value of parameter under after nature hypathesis

(iv) $\hat{\mu}$ is estimateion of μ, $\hat{\mu}$ followed $N(\mu, \hat{\mu})$

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

$$\therefore P\left(-2.578 < \frac{2.9 - \mu}{\sqrt{2.9/1000}} < 2.5758\right) = 0.88$$

Hence, the C.I. for μ is $(2.7613, 3.0387)$

11) Ans:-

$n = 12$, $\Sigma x = 213,200$

$\Sigma x^2 = 4,919,860,000$

$\bar{x} = 17,767$

$S = 10,144$

New sample standard deviation

$$\Sigma x^{2'} = 4,919,860,000 - (11,000)^2 - (48,000)^2 + (27500)^2 + (31,500)^2$$

$$= 4,243,360,000$$

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

$$= \frac{1}{11}\left(4,243,360,000 - 12(17,767)^2\right)$$

$$= 4,13,96,775.64$$

$$S' = 6434.032611$$

The sample mean remains the same after the indises of the ~~parameters~~ 2 permanent employees at the 2 value removed was outliers and the 2 values included write around the sample mean.

The sample standard devation reduces as the 2 values now included do not devate ~~very~~ much from the sample mean causing the variance to reduce and hence the reduction of the sample standard deviation.