PS-2 Assignment

1) $X \sim Poi(\theta)$

$$\frac{X - \theta}{\sqrt{\theta}} \approx N(0,1)$$

$\hat{\theta} = 200$

CI for $\theta = \hat{\theta} \pm 1.96\left(\frac{\sqrt{\hat{\theta}}}{\sqrt{n}}\right)$ for 95%

$\theta = 200 \pm 1.96(\sqrt{200})$

$172.2814 < \theta < 227.7186$

2) ~~...~~ $x_i : i = 1, 2, \ldots n$

$\bar{x} = \sum_i^n x_i / n$ sample mean

$s^2 = \frac{1}{n-1} \sum_i^n (x_i - \bar{x})^2$ sample var.

$$E(\bar{x}) = E\left(\frac{\sum \bar{x}_i}{n}\right) = \frac{\sum_i^n (E(x_i))}{n}$$

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

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

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

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

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

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

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

$var(x) = E(X^2) - (E(X))^2$

$\sigma^2 = E(x_i^2) - \mu^2$

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

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

$= \sigma^2.$

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

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

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

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

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

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

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

(iv)

2) $P(X=0) = {}^4C_0\, p^0 q^4 = q^4$
$P(X=1) = {}^4C_1\, p^1 q^3 = 4pq^3$
$P(X=2) = 6\, p^2 q^2$
$P(X=3) = 4\, p^3 q$
$P(X=4) = p^4$

$$\therefore L = \text{constant} \times (q^4)^{86} \times (pq^3)^{75} \times (p^2q^2)^{16} \times (p^3q)^{2} \times (p^4)^1$$

$$= \text{constant} \times q^{603} \times p^{117}$$

$$\ln L = \ln c + 603 \ln q + 117 \ln p$$

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

$$0 = \frac{-603}{1-p} + \frac{117}{p}$$

$$603p = 117(1-p)$$

$$\frac{117}{720} = p$$

$$p = 0.1625$$

$P(X=0) = q^4 = (1-0.1625)^4 = 0.49197$
$P(X=1) = 4pq^3 = 0.38182$
$P(X=2) = 6p^2q^2 = 0.11113$
$P(X=3) = 0.014375$
$P(X=4) = 0.000697$

X	O_i	E_i	$(O_i - E_i)^2/E_i$
0	86	88.55	
1	75	68.727	
2	16	~~40.0068~~ 20.0034	
3	2	2.5875	
4	1	0.12546	
	180	180	

X	O_i	E_i	$(O_i - E_i)^2/E_i$
0	86	88.55	0.07343
1	75	68.7276	0.572 448
2	19	22.71636	0.60799

$\Sigma Z^2 = 1.253868$

$\Sigma Z^2 \sim \chi^2_{n-2}$

$\chi^2_{cal} = 1.253868$ $\quad n-3$

χ^2_{2} at 95% CI = 5.991

$\chi^2_{cal} < \chi^2_{tab}$

$\therefore$ Don't reject H_0

At 5% LOS, we don't reject H_0

At 10% LOS, we still don't reject H_0.

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

i) $f(x) = \dfrac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}$, $x>0$

$\therefore x \sim \text{Pareto}(c, \beta)$

$\therefore c = 3$

Mean $= E(X) = \dfrac{\lambda}{\alpha-1} = \dfrac{\beta}{c-1} = \dfrac{\beta}{3-1} = \dfrac{\beta}{2}$

$V(x) = \dfrac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)} = \dfrac{c\beta^2}{(c-1)^2(c-2)}$

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

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

$\dfrac{\hat{\beta}}{2} = \bar{X}_n$

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

$E(\hat{\beta}) = 2\bar{X}_n$

$E(\hat{\beta}) - \beta = 2\bar{X}_n - 2\bar{X}_n = 0$

MSE = Variance + Bias^2

$= \dfrac{3\beta^2}{4} + 0$

$= \dfrac{3\beta^2}{4} = \dfrac{3}{4}(2\bar{X}_n)^2$

$= 3(\bar{X}_n)^2 = \dfrac{3(\Sigma x_n)^2}{n^2}$

$\therefore$ As $n \to \infty$, MSE $\to \infty$

iii) Given :- set of estimator : $b\bar{X}_n$

b is constant

To prove :- $\frac{2}{(1+3/n)}$

Q.8 $\hat{\theta}$ is estimator of parameter θ.

i) unbiased estimator.

If $E(\hat{\theta}) - \theta = 0$ then $\hat{\theta}$ is an unbiased parameter.

ii) Bias is given by

$E(\hat{\theta}) - \theta$.

iii) MSE of estimator $\hat{\theta} = E(\hat{\theta} - E(\hat{\theta}))^2$

$=$ variance + bias^2

(iv) There is another estimator $\bar{\theta}$ of the same parameter θ such that $\hat{\theta}$ has no bias but higher MSE than $\bar{\theta}$ while $\bar{\theta}$ has positive bias. Estimator with smaller MSE is better, $\therefore$ $\bar{\theta}$ is a efficient estimator

v) An estimator is consistent when $n \to \infty$ and MSE $\to 0$.

vi) Method of moments

Maximum likelihood method

Q.7)

city	Public	private
1	24	30
2	45	54.5
3	29	30
4	33	40
5	20	28.5
6	40	36
7	26.5	30.5
8	25	30.5
9	27.5	35.5

$\Sigma x_{pu} = 270 \qquad \Sigma x_{pri} = 315.5$

$\bar{X}_{pu} = 30 \qquad \bar{X}_{pri} = 35.06$

$\Sigma x^2_{pu} = 8614.5 \qquad \Sigma x^2_{pri} = 11599.25$

$$S^2_{pu} = \frac{\Sigma x^2_{pu} - n\bar{x}^2_{pu}}{n-1} = \frac{8614.5 - 9(30)^2}{8}$$

$$= 64.31$$

$$S^2_{pr} = \frac{\Sigma x^2_{pr} - n\bar{x}^2_{pr}}{n-1} = \frac{11599.25 - 9(35.06)^2}{8}$$

$$= 67.42$$

ii) The primary assumption for testing equal ~~variances~~ mean is that both samples have equal population variances.

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

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

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

$$s_1^2/s_2^2 \sim F_{n_1-1,\, n_2-1}$$

$$\cancel{P\left(\frac{1}{F_{n_1-1,n_2-1}} < s_1^2/s_2^2 < F_{n_1-1,n_2-1}\right) = 0.95}$$

$$\cancel{P\left(\frac{1}{\;}\right.} \quad \frac{s_1^2}{s_2^2} = \frac{64.31}{67.42} = 0.9542$$

$F_{8,8}$ at $2.5\% = 0.2256$

$1/F_{8,8}$ at $2.5\% = 4.433$.

$\because$ 0.9542 lies between the values

$\therefore$ We don't reject the hypothesis.

$\therefore \sigma_1^2 = \sigma_2^2$ at 5% LOS.

iii) $H_0 \rightarrow \mu_1 = \mu_2$

$H_1 \rightarrow \mu_1 < \mu_2$

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

$$s_p^2 = \frac{((n-1)s_1^2) + ((n-1)s_2^2)}{n_1 + n_2 - 2} = \frac{(8 \times 64.31) + (8 \times 67.42)}{16}$$

$$= 65.86.$$

$$\therefore T.S. = \frac{(35.06 - 30) - (0)}{\sqrt{65.86\left(\frac{1}{9} + \frac{1}{9}\right)}}$$

$$= \frac{5.06(3)}{\sqrt{65.86 \times 2}}$$

$$= 1.32$$

$P(t_{16} > 1.32) = 20\%$

$\therefore$ we don't reject H_0

Q.6) i) P(Type I error) is probability of rejecting H_0 when H_0 is true.

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

$$P(\text{Type I error}) = (1-0.8891) + (0.2824)(1-0.9957) + (0.6590-0.2824)(1-0.9744) + (0.8891-0.6590)(1-0.8891)$$

$$= 0.14727 \approx 15\%$$

ii) Probability of rejecting the null hypothesis when $p = 0.3$.

$$= P(X \geq 3 \mid P = 0.3) + \cancel{[P(X \geq 3 \mid P = 0.3)]}$$
$$[P(X=0 \mid p=0.3) \cdot P(Y \geq 5 \mid p=0.3)] + [P(X=1 \mid p=0.3) \cdot P(Y \geq 4 \mid p=0.3)] + [P(X=2 \mid P=0.3) \cdot P(Y \geq 3 \mid P=0.3)]$$

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

$$= 0.91253 \approx 91\%$$

iii) P(Type II error) is the probability of accepting H_0 when H_0 is false (at $p=0.3$)

$$= 1 - 0.91253 = 0.08747 \approx 9\%$$

iv) 10 space agencies have developed missile 40 hits out of 120 independent missiles fired.

$n = 120 \quad \cancel{p}\, X = 40$

$$\hat{p} = \frac{X}{n} = \frac{40}{120} = 0.333$$

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

$$Z_{10\%} = 1.2816$$

$$\hat{p} - Z_{10\%}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.333 - 1.2816\left(\sqrt{\frac{0.333(1-0.333)}{120}}\right)$$

$$= 0.2782$$

$\therefore$ Lower bound of 90% right tailed confidence interval for p is ~~0.2~~ 0.2782

v) $H_0: p = 0.278$ at 10% LOS

$H_1: p > 0.278$

$$\frac{x - np}{\sqrt{npq}} \sim N(0,1)$$

$40 - 0.5 = 39.5$

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

$\because Z_{10\%} = 1.2816$

$\therefore Z_{tab} > Z_{cal}$

$\therefore$ we don't ~~at~~ reject H_0 at 10% LOS

vi) Lower bound of confidence interval shows $p \leq 0.2782$ at 10% LOS

Hypothesis testing shows $p \leq 0.278$ with more than 10% probability

This could be because of use of continuity correction

PAGE No.
DATE / /

vii) Test $\mu_1 = \mu_2$

Assume:- samples come from normal distribution with equal variance.

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

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

$n_1 = 12$ $n_2 = 15$

$S_1 = 54$ $S_2 = 70$

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

$$S_p = 63.459$$

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

$$\therefore \frac{(65-70)-(0)}{63.459\sqrt{\frac{1}{12}+\frac{1}{15}}} \sim t_{25}$$

$$\frac{-5}{63.459(0.38729)} = -0.2034$$

$t_{25, 2.5\%} = 2.060$

$\therefore$ We don't reject H_0 at 5% LOS

$\therefore \mu_1 - \mu_2 = 0$

PAGE No.	
DATE	/ /

9) i) variable $t_k \rightarrow \dfrac{N(0,1)}{\sqrt{\chi_k^2 / k}}$ k = d.o.f.

$N(0,1)$, χ_k^2 are r.v.

ii) Mean of $t_k = 0$

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

iii) a) $n = 10$, $s^2 = 48.667$, $\bar{x} = 50$

$\dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$

$t_9 = 3.25$

$\therefore$ CI for $\mu = 50 \pm 3.25 \sqrt{\dfrac{48.667}{10}}$

$\therefore$ CI for $\mu \in (50 - 7.169, 50 + 7.169)$

$\in (42.83, 57.169)$

b) $\cancel{t_{n-1} \sim \dfrac{N(0,1)}{\sqrt{\chi_{n-1}^2 / n-1}}}$ $t_9 \sim N\left(0, \sqrt{\dfrac{9}{7}}\right)$

$\cancel{t_9 \sim \dfrac{N(0,1)}{\sqrt{\chi_9^2 / 9}}}$

$\therefore \dfrac{\bar{X} - \mu}{s/\sqrt{n}} \sim N\left(0, \dfrac{3}{\sqrt{7}}\right)$

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

$Z_{0.5\%} = 2.58$

CI for $\mu \in \left(50 \pm 2.58 \sqrt{\dfrac{48.667}{10} \times \dfrac{9}{7}}\right)$

$\in (43.54, 56.45)$

10) i) Type 1 error: Rejecting Ho when it is true.

~~Let~~ X = total no. of claims

$\therefore X \sim Poi(n\mu) \sim Poi(3000)$

$H_0: \quad X \sim Poi(3000)$

$P(\text{Type I error}) = P(X > 3100$ when $X \sim Poi(3000)$

$X \sim N(3000, 3000)$

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

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

ii) Type II error: Accepting Ho when it is false.

iv) $\hat{\mu}$ is estimator of μ.

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

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

$\therefore Z_{0.5\%} = 2.58$

$\therefore P(-2.58 < Z < 2.58) = 0.99$

~~CI for μ ∈ (2.~~

$P\left(-2.58 < \frac{2.9 - \mu}{\sqrt{2.9/1000}} < 2.58\right) = 0.99$

$P(2.7613 < \mu < 3.0387) = 0.99$

CI $\mu \in (2.7613, 3.0387)$.

11) $\Sigma x = 213200$

$\Sigma x^2 = 4919860000$

$\bar{x} = 17767$

$s = 10144$

$T_1 = 11000$; $T_2 = 48000$

$P_1 = 27500$; $P_2 = 31500$

$\Sigma x_r = 213200 - 11000 - 48000 + 27500 + 31500$

$= 213200$

$\therefore \bar{x}_{revised} = \frac{213200}{12} = 17767$

$\Sigma x_r^2 = 4243360000$

$s_{rev}^2 = \cancel{3857} \left(\Sigma x_{rev}^2 - n\bar{x}_{rev}^2\right)/11$

$s_{rev}^2 = 41396775.64$

$s_{rev} = 6434.0326$

There is no change in sample mean However the standard deviation has reduced as the salary of permanent employees is closer to mean than the temporary employees.

12) $X_i \sim U(0, \theta)$

i) Cramer Rao lower bound: doesn't apply to this case.

ii) $g_Y(y) = \frac{n}{\theta^n} \cdot y^{n-1}, \quad 0 \le y < \theta$ — T.P.

$\hat{\theta}(c) = cY \qquad Y = \max_i X_i$

$$P[Y < y] = P[\max_i X_i < y]$$
$$= P\left[\bigcap_{i=1}^{n} (X_i < y)\right]$$
$$= \prod_{i=1}^{n} P(X_i < y]$$
$$= \prod_{i=1}^{n} \left[\int_0^y \frac{dx}{\theta}\right]$$
$$= \left(\frac{y}{\theta}\right)^n$$

p.d.f. of Y :-

$$g_Y(y) = \frac{d}{dy} P(Y < y)$$
$$= \frac{d}{dy}\left(\frac{y}{\theta}\right)^n = \frac{n\,y^{n-1}}{\theta^n}$$

ii) $E(y^k) = \dfrac{n\theta^k}{n+k}$

$$E(y^k) = \int_0^\theta y^k \, g_Y(y)$$

$$= \int_0^\theta y^k \frac{n\cancel{\theta} y^{n-1}}{\theta^n} \, dy$$

$$= \int_0^\theta \frac{n y^{n+k-1}}{\theta^n} \, dy$$

$$= \frac{n}{\theta^n}\left(\frac{y^{n+k}}{n+k}\right)_0^\theta$$

$$= \cancel{\frac{n}{\theta^n}\times\frac{\theta^{\theta+k}}{\theta+k}} \; \frac{n}{\theta^n}\frac{(\theta^{n+k})}{(n+k)}$$

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

iii) Biais $[\hat{\theta}(c)] = \left(\frac{cn}{n+1} - 1\right)\theta$

$$\text{Bias}[\hat{\theta}(c)] = E(\hat{\theta}(c)) - \theta$$
$$= E(cY) - \theta$$
$$= c(E(Y)) - \theta$$
$$= c\frac{n\theta}{n+1} - \theta$$
$$= \left(\frac{cn}{n+1} - 1\right)\theta$$

$$\text{MSE}[\hat{\theta}(c)] = E\left((\hat{\theta}(c) - \theta)^2\right)$$
$$= E\left((cY - \theta)^2\right)$$
$$= E(c^2Y^2 - 2cY\theta + \theta^2)$$
$$= c^2E(Y^2) - 2cE(Y)E(\theta) + E(\theta^2)$$
$$= c^2\frac{n\theta^2}{n+2} - \frac{2cn\theta E(\theta)}{n+1} + E(\theta^2)$$
$$= \frac{c^2n\theta^2}{n+2} - \frac{2cn\theta^2}{n+1} + \theta^2$$

PAGE No.
DATE / /

(iv) For $\hat{\theta}(c)$ to be unbiased.

$\text{Bias}[\hat{\theta}(c)] = 0$

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

$$\frac{cn}{n+1} = 1$$

$$c_u = \frac{n+1}{n}$$

(v) $MSE(\hat{\theta}(c))$ to be minimized

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

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

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

$$\frac{c}{n+2} = \frac{1}{n+1}$$

$$c_m = \frac{n+2}{n+1}.$$

$$\frac{d^2}{dc^2}(MSE(\hat{\theta}(c))) = \frac{2n\theta^2}{n+2} > 0$$

$\therefore$ MSE is minimum at $c = \frac{n+2}{n+1}$

PAGE No.

DATE / /

) To minimize error in estimation ~~[illegible]~~ $\hat{\theta}(Cm)$ is preferred over $\hat{\theta}(Cu)$

$\because$ MSE lower is preferred over lower bias.

As $n \to \infty$

~~[illegible]~~

$\hat{\theta}(Cu)$ and $\hat{\theta}(Cm)$ ~~as~~

become the same.