# Probability & Statistics - 2
## Assignment - 1

1. $X = 200$

$X \sim Poi(\theta)$

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

$P[X_1 < \theta < X_2] = 0.95$

$P[-1.96 < \theta < 1.96] = 0.95$

$\rightarrow$ 95% CI $= \theta \pm 1.96\sqrt{\theta}$

$= 200 \pm 1.96\sqrt{200}$

$= (172.281, 227.718)$

2. $X_i = 1, 2, \ldots n$ ; iid's RV

Mean $= \mu$

Variance $= \sigma^2$

$\overline{X} = \frac{\sum_1^n X_i}{n}$

$S^2 = \frac{1}{n-1} \sum_1^n (X_i - \overline{X})^2$

(i) for mean of $\overline{X}$

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

$= \frac{\sum[E(X_i)]}{n}$

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

$= \boxed{\mu}$

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

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

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

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

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

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

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

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

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

iii $X_i \sim Norm$

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

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

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

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

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

iv. $n = 10$

$\sigma^2 = 100$

$$P[50 < s^2 < 150] = P[s^2 < 150] - P[s^2 < 50]$$

$$= P\left[\frac{(n-1)s^2}{\sigma^2} < \frac{9 \times 150}{100}\right] - P\left[\frac{(n-1)s^2}{\sigma^2} < \frac{9 \times 50}{100}\right]$$

$$= P[\chi^2_9 < 13.5] - P[\chi^2_9 < 4.5]$$

$$= 0.8587 - 0.1245$$

$$= \underline{0.7342}$$

3. X : No. of claims

$X \sim Bin(4, p)$

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

$$L = \prod_{i=1}^{180} P(X = x)$$

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

$$= \left({}^4C_0 p^0 (1-p)^4\right)^{86} \cdot \left[{}^4C_1 p(1-p)^3\right]^{75} \cdot \left[{}^4C_2 p^2 (1-p)^2\right]^{16} \cdot \left[{}^4C_3 p^3 (1-p)\right]^2 \cdot \left[{}^4C_4 p^4\right]$$

$$L = \text{Constant} \times (1-p)^{603} p^{117}$$

$$\ln L = \text{Constant} + 603 \ln(1-p) + 117 \ln p$$

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

$$\Rightarrow 603\, p = 117(1-p)$$

$\rightarrow 720p = 117$

$\rightarrow \boxed{p = 0.1625}$

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

Ans $\hat{p} = 0.1625$

(ii)

| $x_i$ | $O_i$ | $E_i$ |
|---|---|---|
| 0 | 86 | $180 \times P(x=0) = 88.5547$ |
| 1 | 75 | $180 \times P(x=1) = 68.729$ |
| 2 | 16 | $180 \times P(x=2) = 20.0032$ |
| 3 | 2 | $180 \times P(x=3) = 2.5874$ |
| 4 | 1 | $180 \times P(x=4) = 0.1255$ |

Since $E_i$'s at $x_i = 3$ & $4$ are very small, we club the rows with the above to avoid erroneous values

Adjusted Table

| $O_i$ | $E_i$ | $\frac{(O_i - E_i)^2}{E_i}$ |
|---|---|---|
| 86 | 88.5547 | 0.0737 |
| 75 | 68.729 | 0.5722 |
| 19 | 22.7159 | 0.6078 |

$\chi^2_{cal} = \sum \frac{(O_i - E_i)^2}{E_i} = 1.2537$

$\chi^2_{tab} = \chi^2_1 = 3.841$

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

$\rightarrow$ We do not reject $H_0$ at 5% Level of Significance

$\Rightarrow$ No. of claims on each policy conform to the binomial model.

4.

(i) PDF of random variable is given by:

$$f(x) = \frac{CB^3}{(x+B)^4} \quad ; x > 0,\ B > 0 \quad \text{—— (i)}$$

From the table book

This is similar to the pdf of the Pareto distribution given by:

$$f(x) = \frac{\alpha \lambda^{\alpha}}{(\lambda + x)^{\alpha+1}} \quad ; \quad \text{—— (ii)}$$

Comparing (i) & (ii)

$C = 3$

Mean of X:

$$E(X) = \frac{B}{C-1}$$

Ans $\boxed{E(X) = \frac{B}{2}}$

Variance of X:

$$V(X) = \frac{CB^2}{(C-1)^2(C-2)} = \frac{3B^2}{4 \cdot 1}$$

Ans $\boxed{V(X) = \frac{3B^2}{4}}$

(ii) Sample : $x_1, x_2, \ldots, x_n$

Using method of moments

$\Rightarrow E(X) = \frac{\sum_{i=1}^{n} x_i}{n}$

$\Rightarrow \frac{2B}{c-1} = \bar{x}_n$

$\Rightarrow \boxed{\hat{B} = 2\bar{x}_n}$

To verify ~~bias~~ unbiasedness :

$Bias = E[g(x)] - \theta$

$= E(\hat{B}) - B$

$= E(2\bar{x}_n) - B$

$= 2E(\bar{x}_n) - B$

$= 2E\left(\frac{\sum x_i}{n}\right) - B$

$= \frac{2\sum E(x_i)}{n} - B$

$= \frac{2}{n}\sum\left(\frac{B}{2}\right) - B$

$= \frac{2nB}{2n} - B$

$= 0$

$\Rightarrow$ Since $Bias = 0$

$\Rightarrow \hat{B}$ is an unbiased estimator of $B$

For consistency, we require the mean square error

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

$MSE = Var(\hat{B})$ $[Bias = 0]$

$= \cancel{Var}\ V(2\bar{X}_n)$

$= 4V\left(\frac{\sum X_i}{n}\right)$

$= \frac{4\sum[V(X_i)]}{n^2}$

$= \frac{4}{n^2}\sum\left(\frac{3B^2}{4}\right)$

$= \frac{3nB^2}{n^2}$

$= \frac{3B^2}{n}$

As $n \to \infty$, MSE tends to 0; hence $\hat{B}$ is consistent

iii. Estimators are of the form: $b\bar{X}_n$; $b \to$ const

$\rightarrow \hat{B} = b\bar{X}_n$

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

$Var(\hat{B}) = V(b\bar{X}_n) = b^2V(\bar{X}_n)$

$= \frac{b^2}{n^2}\sum V(X_i^2)$

$= \frac{b^2}{n^2}\left(\frac{n3B^2}{4}\right)$

$= \frac{3b^2B^2}{4n^4}$

$$E(\hat{B}) = E(b\bar{x})$$

$$= \frac{b}{n} \sum E(x_i)$$

$$= \frac{b}{n}\left[n \frac{B}{2}\right]$$

$$= \frac{bB}{2}$$

$$\text{Bias} = \frac{bB}{2} - B = \cancel{b(1} \frac{bB-2B}{2} = \frac{B(b-2)}{2}$$

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

$$= \frac{3b^2B^2}{4n} + \frac{B^2}{4}\left[b^2+4-4b\right]$$

$$= \frac{3b^2B^2 + B^2n\left[b^2+4-4b\right]}{4n}$$

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

$$\frac{d}{dB}[MSE] = \frac{B^2}{4n}\left[6b + 2nb - 4n\right] = 0$$

$$\Rightarrow 6b + 2nb - 4n = 0$$

$$\Rightarrow b(3+n) = 2n$$

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

$$\Rightarrow \boxed{b = \frac{2}{1+\frac{3}{n}}}$$

$$\frac{d^2(MSE)}{dB^2} = \frac{B^2}{4n}\left[6 + 2n\right] > 0$$

$$\Rightarrow \text{min at } \boxed{b = \frac{2}{1+\frac{3}{n}}}$$

iv. For unbiasedness

$$\text{Bias} = E(g(x)) - \theta$$
$$= E(\hat{B}) - B$$
$$= \frac{bB}{2} - B$$
$$= B\left[\frac{b-2}{2}\right]$$

At $b = \dfrac{2}{1+\frac{3}{n}}$

$$\text{Bias} = B\left[\frac{\frac{2n}{3+n} - 2}{2}\right]$$
$$= B\left[\frac{2n-6-2n}{2(n+3)}\right]$$
$$= \frac{-3B}{n+3}$$

Since Bias $< 0$, $\hat{B}$ is an underestimate of the true parameter value

$$MSE = Var(\hat{B}) + \text{Bias}^2$$
$$\cancel{= \frac{B^2}{4n}\left[3b^2 + \frac{2nb^2}{n+3} + 4n\left(1 - \frac{2nB}{n+3}\right)\right]} \quad \cancel{[\text{using (i)}]}$$

$$\cancel{= \frac{B^2}{4n}\left[\frac{b^2(3n+3+2n)}{n+3} + \frac{4n(n+3-2n)}{n+3}\right]}$$
$$\cancel{= \frac{B^2}{4n}\left[\frac{b^2(5n+3) + 4n(3-n)}{(n+3)}\right]}$$

$$= \frac{B^2}{4n}\left[3b^2 + nb^2 + 4n - 4nb\right] \quad [\text{using (i)}]$$

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

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

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

$$= \frac{B^2(4n^2+12n-4n^2)}{4n(n+3)}$$

$$MSE = \frac{3B^2}{n+3}$$

As $n \to \infty$, $MSE \to 0$, hence it is consistent.

5 (i) Let $x_1, x_2, x_3 \ldots x_n$ follow $U(a,b)$

$$\Rightarrow \bar{x} = \frac{\sum x_i}{n}$$

Using mom

$$E(x) = \frac{1}{2}(a+b) = \bar{x} \Rightarrow (a+b) = 2\bar{x} \quad \text{—(i)}$$

$$V(x) = s^2$$

$$\frac{1}{12}(b-a)^2 = s^2$$

Taking sq root on both sides

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

$$\Rightarrow b-a = 2\sqrt{3}s \quad \text{—(ii)}$$

Subtracting (ii) from (i)

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

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

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

(ii) For $\hat{b}$ :

~~Sub~~ Adding (i) & (ii)

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

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

iii Let the sample be : 5, 12, 20, 35, 50

$\bar{x} = 24.4$, $s = 18.147$

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

$\hat{a} = -7.018$

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

$\Rightarrow \hat{b} = 55.87$

Thus from the sample we observe that $\hat{a}$ is no where near the sample obs (5), as opposed to $\hat{b}$ which is still giving a better estimate of the sample value of b (50)

IV. Sample : $x_1, x_2, \dots, x_n$

$a = 0$

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

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

$L = \left(\frac{1}{b}\right)^n \quad [a = 0]$

$\ln L = -n \ln b$

$$\frac{d(\ln L)}{db} = -\frac{n}{b} \neq 0$$

Thus give $b = \infty$, which is not possible

Also ; $\frac{d^2(\ln L)}{db^2} = \frac{n}{b^2} > 0$ ; giving us a minima

Ans $\rightarrow \hat{b} = Max(x_i)$

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

(i) X : no. of hits in $1^{st}$ step
Y : no. of hits in $2^{nd}$ step

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

$$= P(X > 2/p=0.1) + P(Y > 4/p=0.1) P(X=0/p=0.1) + P(Y > 3/p=0.1) P(X=1/p=0.1) + P(Y > 2/p=0.1) P(X=2/p=0.1)$$

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

$$= 0.1433$$

$$= \underline{14.33\%}$$

ii. P ( rejecting $H_0$ ) when $p = 0.3$

$$= P(X > 2 / p = 0.3) + P(Y > 4 / p = 0.3)\, P(X = 0 / p = 0.3)$$
$$+ P(Y > 3 / p = 0.3)\, P(X = 1 / p = 0.3)$$
$$+ P(Y > 2 / p = 0.3)\, P(X = 2 / p = 0.3)$$

$$= (1 - 0.2528) + (1 - 0.07237)(0.0138)$$
$$+ (1 - 0.4925)(0.0850 - 0.0138)$$
$$+ (1 - 0.2528)(0.2528 - 0.0850)$$

$$= 0.9125$$
$$= \boxed{91.25\%}$$

iii. P ( type II error ) = 1 - P( type I error )

$$= 1 - 0.9125$$
$$= 0.0875$$
$$= \underline{8.75\%}$$

iv. $\hat{p} = \dfrac{40}{120} = \dfrac{1}{3} = 0.3333$

$$\rightarrow \bar{x} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Pivotal quantity : $\dfrac{\bar{x} - np}{\sqrt{\frac{p(1-p)}{n}}}$

For 90% one sided CI :

$$P\left[\frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}} < X_1\right] = 0.1$$

$$\rightarrow 90\% \text{ of CI} = \bar{x} - Z_{0.1}\sqrt{\frac{p(1-p)}{n}}$$

$$= 0.3333 - 1.2816\sqrt{\frac{0.3333(0.6667)}{120}}$$

$= \underline{0.2783}$

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

Since $\hat{p} = 0.3333 > p$ :- Continuity correction $\left(\frac{x - \frac{1}{2}}{n}\right)$

$$TS : \frac{\left[\frac{x - \frac{1}{2}}{n}\right] - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

$$z_{cal} = \frac{\frac{40 - 0.5}{120} - 0.278}{\sqrt{\frac{0.278(0.722)}{120}}}$$

$$= \frac{6.14}{\sqrt{\frac{(120)^2(0.200716)}{120}}} = 1.251$$

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

Since $Z_{cal} < Z_{tab}$
→ We do not reject $H_0$

vi. The (iv) suggests that there is a 10% chance that $p < 0.2782$ while in the hypothesis test $p$ could be less than or equal to 0.278 in 90% of the cases

vii $n_1 = 12$ $\sigma_1 = 54$
$\bar{X}_1 = 65$ $\sigma_2 = 70$
$\bar{X}_2 = 70$ $n_2 = 15$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Assumption: Population variance of both Sample are equal

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

$$S_p = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

$$= 4027.04$$

$$t_{cal} = \frac{(65 - 70)}{\sqrt{4027.04\left(\frac{1}{12} + \frac{1}{15}\right)}} = \frac{-5}{\sqrt{24.571}} = -0.2034$$

$$t_{tab} = (-2.060, 2060)$$

$\Rightarrow$ Since $t_{cal}$ lies in the 95% CI, we do not reject $H_0$

7. (i) Sample mean:

Private Hospitals $(\bar{x}_{pr}) = \frac{\sum x_i}{n} = \frac{315.5}{9} = 35.056$

Public Hospitals $(\bar{x}_{pu}) = \frac{270}{9} = 30$

Sample Variance:

$$S_{pr}^2 = \frac{\sum x_i^2 - n\bar{x}}{n-1} = 67.403$$

$$S_{pu}^2 = 64.3125$$

ii. For testing equal mean, we can use a 2-sided t-test assuming that the population variation of both is equal

To verify this assumption:

$H_0 : \sigma_1^2 = \sigma_2^2$ ; $H_1 : \sigma_1^2 \neq \sigma_2^2$

$$\frac{s^2_{pr}/s^2_{pu}}{\sigma^2_{pr}/\sigma^2_{pu}} \sim F_{n_{pr}-1, n_{pu}-1}$$

For 95% CI

$$\Rightarrow P\left[\frac{1}{x} < F_{8,8} < x\right] = 0.95$$

$$\Rightarrow P\left[\frac{1}{4.433} < \frac{67.403}{64.3125}\frac{\sigma^2_{1u}}{\sigma^2_{1r}} < 4.433\right]$$

$\rightarrow$ 95% of CI = (0.2152, 4.229)

Since at 5% LOS, CI includes 1, we do not reject $H_0$. Our assumption of equal population variance holds true in this case.

iii. $H_0 : \mu_{1r} = \mu_{pu}$ ; $H_1 : \mu_{1r} > \mu_{pu}$

This is a one-sided test

$\rightarrow$ Since popl^n var is unknown we use t-distribution

$$t_{cal} = \frac{(\overline{X}_{1r} - \overline{X}_{1u}) - (\mu_{pr} - \mu_{pu})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{5.056}{8.6756\sqrt{\frac{2}{9}}} = \frac{5.056 \times 3}{8.6756\sqrt{2}} = 1.2363$$

$t_{tab} = t_{16, 5\%} = 1.746$

Since $t_{cal} < t_{tab}$ ; we have insufficient evidence to reject $H_0$ at 5% level of significance

8. (i) An unbiased estimator ~~a~~ is an estimator whose mean value ($E(\hat{\theta})$) is equal to the true parameter value ($\theta$)

ii Bias is the measure of the diff b/w the expected value of the estimator & ~~it~~ the true parameter value

iii Mean square error is the second moment the estimator about the true parameter value i.e

Since ~~If~~ ~~Estimator~~ ~~is~~ ~~g(x)~~ $\hat{\theta} \to$ estimator
True parameter ~~is~~ is $\theta$

$\Rightarrow$ MSE $= E[(\overset{\hat{\theta}}{g(\cancel{x})} - \theta)^2]$

iv $E(\hat{\theta}) = \theta$ ; $E(\tilde{\theta}) > \theta$
Even though $\hat{\theta}$ has no bias, it could so happen that some of the estimate are too large & some too small but on average give a true value.
On the other hand $\check{\theta}$, the ~~a~~ biased estimator in this case could have a value that does not deviate much from true value & hence preferred over $\hat{\theta}$ whose values are all over the place

v As $n \to \infty$, MSE $\to 0$, which in this case would make either of the two estimators consistent

vi. 2 methods of estimating $\theta$:
Method of moments
MLE

9. (i) $t_k$ is defined as : $\frac{Z}{\sqrt{C/k}} \sim t_k$,

where $Z \sim N(0,1)$

$C \sim \chi^2_{k=n-1}$

$k$ denotes the degrees of freedom

ii Mean of $t_k = 0$

Variance $= \frac{k}{k-2}$, $k > 2$

iii. $n = 10$

~~Mean~~ $\bar{x} = 50$

$s^2 = 48.667$

(a) TS : $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

for 99% CI $= \bar{x} \pm t_{n-1, 0.5\%} \, S/\sqrt{n}$

$= 50 \pm 3.250 \times \sqrt{\frac{48.667}{10}}$

$= (42.83, 57.1697)$

(b) $\frac{\bar{x}-\mu}{S/\sqrt{n}} \sim N\left(0, \frac{k}{k-2}\right)$

Ts: $\frac{\bar{x}-\mu}{S/\sqrt{n}} \sim N\left(0, \frac{n-1}{n-1-2}\right)$

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

~~→ for 99% CI: $\bar{x} \pm Z_{0.5\%}$~~

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

$99\%\ CI = \bar{x} \pm Z_{0.5\%}\sqrt{\frac{9S^2}{7n}}$

$= 50 \pm 2.5758\sqrt{\frac{9(48.667)}{70}}$

Ans $99\%\ CI = (43.556, 56.443)$

10. X: No. of claims ~ Poi(3)

i. Type 1 error is the probability of rejecting Ho, when it is actually true

$H_0 = \mu = 3$; If $X < 3100$, $H_0$ → accepted

Type 1 error in this case is:

~~$P(\bar{X} > 3100 / \mu = 3) = 1 - P(\bar{X} < 3100 / \mu = 3)$~~

$\sum X_i \sim Poi(n\lambda) = \sum X_i \sim Poi(3 \times 1000)$

$= \sum X_i \sim Poi(3000)$

ii. For testing equal mean, we can use a 2-sided t-test assuming that the population variation of both is equal

To verify this assumption:
$H_0 : \sigma_1^2 = \sigma_2^2$ ; $H_1 : \sigma_1^2 \neq \sigma_2^2$

$\frac{s^2_{pr}/s^2_{pu}}{\sigma^2_{pr}/\sigma^2_{pu}} \sim F_{n_{pr}-1, n_{pu}-1}$

For 95% CI
$\Rightarrow P\left[\frac{1}{x} < F_{8,8} < x\right] = 0.95$

$\Rightarrow P\left[\frac{1}{4.433} < \frac{67.403}{64.3125}\frac{\sigma^2_{pu}}{\sigma^2_{pr}} < 4.433\right]$

$\rightarrow$ 95% of CI = (0.2152, 4.229)
Since at 5% LOS, CI includes 1, we do not reject $H_0$. Our assumption of equal population variance holds true in this case.

iii. $H_0 : \mu_{pr} = \mu_{pu}$ ; $H_1 : \mu_{pr} > \mu_{pu}$
This is a one-sided test
$\rightarrow$ Since popl$^n$ var is unknown we use t-distribution

$$t_{cal} = \frac{(\bar{X}_{pr} - \bar{X}_{pu}) - (\mu_{pr} - \mu_{pu})}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{5.056}{8.6756\sqrt{\frac{2}{9}}} = \frac{5.056 \times 3}{8.6756\sqrt{2}} = 1.2363$$

$t_{tab} = t_{16, 5\%} = 1.746$

Since $t_{cal} < t_{tab}$ ; we have insufficient evidence to reject $H_0$ at 5% level of significance

Type 1 error in this case is

$$= P(X > 3100 / \mu = 3000)$$

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

$$= P(Z > 1.8257)$$

$$= 1 - P(Z < 1.8257)$$

$$= 0.0356$$

ii. Type 2 error is the probability of ~~rejecting~~ accepting $H_0$ when its supposed to be rejected.

iii Power of a test is the probability of rejecting $H_0$ when its false

$$\text{Power of test} = P(X > 3100 / \bar{X} \sim N(n\mu, n\mu))$$

$$= 1 - P(X < 3100 / \bar{X} \sim N(n\mu, n\mu))$$

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

iv. $\bar{X} \sim N\left(\mu, \frac{\mu}{n}\right)$

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

$$\rightarrow P\left[x_1 < \frac{\bar{x} - \mu}{\sqrt{\frac{\mu}{n}}} < x_2\right] = 0.99$$

$$\rightarrow P[-2.5758 < Z < 2.5758] = 0.99$$

$$\rightarrow 99\%\ CI = \frac{2900}{1000} \pm 2.5758\sqrt{\frac{2.9}{1000}}$$

Ans $99\%\ CI = (2.7613, 3.0387)$

11. $\sum x = 213200$

$n = 12$

$\sum x^2 = 4919860000$

$\bar{x} = 17,767$

$s = 10144$

$\Rightarrow \sum x_{new} = \sum x$ + correct value - incorrect value

$= 213200 + (27500 + 31500) - (11000 + 48000)$

$= 213200$

$$\bar{x}_{new} = \frac{\sum x_{new}}{n} = \frac{213200}{12} = 17767$$

$\sum x_{i_{new}}^2 = \sum x_i^2 - (11000^2 + 48000^2) + 27500^2 + 31500^2$

$= 4243360000$

$$s^2 = \frac{1}{n-1}\left(\sum x_{i\,new}^2 - n\bar{x}_{new}^2\right)$$

$$= \frac{1}{11}\left[4243360000 - (12 \times 17767^2)\right]$$

$s^2 = 41396776$

$\Rightarrow s = 6434$.

12. $X = X_1, X_2 \ldots X_n$

$\underline{x} \sim U(0, \theta)$

i. CRLB for variance of unbiased estimator of $\theta$ doesn't apply here as the range of the distribution involves the parameter, hence the derivative in the formula does not make sense.

ii. $\hat{\theta}(c) = cY$, $c \rightarrow$ constant; $Y = \max X_i$

Since $Y = \max X_i$

$F(Y) = F(X_{max}) = F(\max(x_1, x_2 \ldots x_n))$

$$= 1 - P(x \geq \max(x_1, x_2 \ldots, x_n))$$
$$= 1 - P(x > x_1)P(x > x_2) \ldots P(x > x_n)$$
$$= 1 - P(x \geq x)^n$$
$$= P(x < x)^n$$
$$= [F(x)]^n$$

$$F(y) = \left(\frac{x-0}{\theta-0}\right)^n = \left(\frac{x}{\theta}\right)^n = \left(\frac{y}{\theta}\right)^n$$

$$\Rightarrow f(y) = F'(y)$$
$$= \frac{d}{dy}\left(\frac{y}{\theta}\right)^n$$
$$= \frac{1}{\theta^n}\left[n y^{n-1}\right], \quad 0 < y < \theta$$

$$E[Y^k] = \int_0^\theta y^k f(y)\,dy$$
$$= \int_0^\theta y^k \left(\frac{n}{\theta^n} y^{n-1}\right) dy$$
$$= \frac{n}{\theta^n}\int_0^\theta y^{k+n-1}\,dy$$
$$= \frac{n}{\theta^n}\left[\frac{y^{n+k}}{k+n}\right]_0^\theta$$
$$= \frac{n\theta^{n+k}}{\theta^n(n+k)}$$
$$= \frac{n\cancel{\theta^n}\theta^k}{\cancel{\theta^n}(n+k)}$$

Ans $\boxed{E[Y^k] = \dfrac{n\theta^k}{n+k}}$ — (i)

iii. $\text{Bias}[\hat{\theta}(c)] = E(\hat{\theta}(c)) - \theta$
$$= E(cY) - \theta$$
$$= cE(Y) - \theta$$
$$= \text{Using (i) at } k = 1$$
$$\Rightarrow \text{Bias} = c\left[\frac{n\theta}{n+1}\right] - \theta$$
$$= \theta\left[\frac{cn}{n+1} - 1\right]$$

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

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

$$MSE = c^2\left[\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}\right] + \left[\theta\left(\frac{cn}{n+1} - 1\right)\right]^2$$
$$= \frac{c^2 n\theta^2}{n+2} - \frac{c^2n^2\theta^2}{(n+1)^2} + \theta^2\left[\frac{cn}{n+1} - 1\right]$$
$$= \frac{c^2 n\theta^2}{n+2} - \frac{c^2n^2\theta^2}{(n+1)^2} + \theta^2\left[\frac{c^2n^2}{(n+1)^2} + 1 - \frac{2cn}{n+1}\right]$$

Ans $\boxed{MSE = \dfrac{c^2 n\theta^2}{(n+2)} - \dfrac{c\,2n\theta^2}{(n+1)} + \theta^2}$

iv For $\theta(c)$ to become an unbiased estimator of $\theta$
$$\rightarrow \text{Bias} = 0$$
$$\rightarrow E(\hat{\theta}(c)) - \theta = 0$$

$\Rightarrow E(\hat{\theta}(c)) = \theta$

$\Rightarrow E(cY) = \theta$

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

$\Rightarrow c\left(\frac{n\theta}{n+1}\right) = \theta$

$\Rightarrow \boxed{C_u = \frac{n+1}{n}}$

v. $MSE = \frac{c^2 n\theta^2}{n+2} - \frac{c(2n\theta^2)}{n+1} + \theta^2$

$\frac{d(MSE)}{dc} = \frac{2Cn\theta^2}{n+2} - \frac{2n\theta^2}{n+1} = 0$

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

$\Rightarrow cn + c = n+2$

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

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

$\Rightarrow$ Minima at $c = \frac{n+2}{n+1}$

$\Rightarrow C_m = \frac{n+2}{n+1}$

vi. $\hat{\theta}(C_m)$ is preferred over $\hat{\theta}(C_u)$ as in the latter case it could so happen that some of the estimators are very large while other small, but on average give a true value.

In the case of the former, it could have a value that may not deviate much from the true value. As $n \to \infty$, both $\hat{\theta}(C_u)$ & $\hat{\theta}(C_m)$ become equal as $\hat{\theta}(C_u) = C_u \cdot Y = \left(\frac{n+1}{n}\right)Y \to Y$;
Similarly $\hat{\theta}(C_m) = C_m Y \to Y$