Date ....
Page ....

1 If $x \sim$ Poisson (θ) and θ is large then the approximation

$$\frac{x-\theta}{\sqrt{\theta}} \sim \text{approx } N(0,1)$$

may be used to derive an approximation confidence interval for θ. You have obseled a single value of 200 from the distribution using the above approximation and a continuity correction, derive an approximatlate 96% symmetrical CI for θ.

Ans $\theta = 200$ (since a single value of θ is observed).

$x \sim$ Poission (200)

$x \sim N(200, 200)$.

$\frac{x-200}{\sqrt{200}} \sim z\,(0,1)$

$\bar{x} \pm z_{\alpha/2}\sqrt{\sigma}$

96% CI for λ = $200 \pm z_{5/2}\sqrt{200}$

$= 200 \pm 1.96\,(\sqrt{200})$.

$= 200 \pm 27.71858582$.

$= 200 - 27.71858582,\ 200 + 27.71858582$.

$= 172.2814,\ 227.7186$

2 Suppose $x_i : i = 1, 2, \ldots n$ are n independent and identically distributed R.V each with mean μ and variance σ^2. Let $\bar{x} = \sum_i^n x_i$ be the sample mean and $s^2 = \frac{1}{n-1}\sum_i^n (x_i - \bar{x})^2$ be the sample variance.

i) Show that the mean and variance $\bar{x}$ is μ and σ^2/n respecti

ii) Show that $E(s^2) = \sigma^2$

$$E(\bar{x}) = E\left(\frac{\sum x_i}{n}\right) = \frac{1}{n}E(\sum x_i) = \frac{1}{n}\sum E(x_i) = \frac{1}{n}\sum \bar{x} = \frac{n\mu}{n} = \mu.$$

$$V(\bar{x}) = V\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2}V(\sum x_i) = \frac{1}{n}\sum V(x_i) = \frac{1}{n}\sum\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\therefore \bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

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

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

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

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

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

By $E(x^2) = V(x) + [E(x)]^2$, $[E(x)]^2 = \bar{x}^2$, $V(x) = 6$

$$= \frac{1}{n-1}\left[\Sigma(u^2+6^2) - n\left(u^2 + \frac{6^2}{n}\right)\right]$$

$$= \frac{1}{n-1}\left[\cancel{nu^2} + n6^2 - \cancel{nu^2} - 6^2\right]$$

$$= \frac{6^2\,\cancel{(n-1)}}{\cancel{(n-1)}} = \underline{\underline{6^2}}$$

Hence proved.

Assume x_i's are from a normal population using distribution of $(n-1)s^2/6^2$ as chi square

iii) Show that variance of s^2 is $26^4\,(n-1)$

iv) Find the probability s^2 will fall between + or − 50% of its expected value when $n=10$ and $6^2 = 100$.

$$V\left(\frac{(n-1)s^2}{6^2}\right) = V\left(\chi_{n-1}\right)$$

$$\frac{(n-1)^2\,V(s^2)}{6^4} = 2(n-1) \text{ — formula.}$$

$$V(s^2) = \frac{26^4\,\cancel{(n-1)}}{(n-1)^{\cancel{2}}}$$

$$V(s^2) = \frac{26^4}{n-1} \text{ —— Hence proved.}$$

$6^2 = 100$

$P(s^2) = P(-50\% \overset{100}{\cancel{E(x)}} < s^2 < 50\% \overset{100}{\cancel{E(x)}})$

$= P(50 < s^2 < 150)$

$\frac{(n-1)s^2}{6^2} \sim \chi^2_{n-1} \quad (n-1 = 10-1 = 9)$

$\left(50 \times \frac{9}{100} < \frac{9s^2}{100} < 150 \times \frac{9}{100}\right)$

$(4.5 < y < 13.5) = P(y < 13.5) - P(y < 4.5)$

$= 0.8587 - 0.1245$

$= \underline{0.7342}$

Q3 An insurer believes that the distribution of number of claims on a particular type of policy is binomial with parameter = 4. A random sample of 180 policies revealed the following information.

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

i) Obtain the maximum likelyhood estimate of p.

$L = \prod_{i=1}^{180} P(x=0)^{86} \times P(x=1)^{75} \times P(x=2)^{16} \times P(x=3)^2 \times P(x=4)^1$

$L = \left({}^{4}C_0 (p)^0 q^4\right)^{86} \times \left({}^4C_1\, p^1 q^3\right)^{75} \times \left({}^4C_2\, p^2 q^2\right)^{16} \times \left({}^4C_3 (p^3)(1-p)^1\right)^2 \times \left({}^4C_4\, p^4\right)^1$

$L = [1\,(1-p)^4]^{86} \times \underline{4^{75}}(p^1(1-p)^3)^{75} \times \left[\left(\frac{4\times3}{2\times1}\right)^{16}(p^2)^{16}((1-p)^2)^{16}\right] \times$

$[\underline{4^2}(p^3)^2(1-p)^2] \times p^4$

$L = c \times (1-p)^{344} \times p^{75}(1-p)^{225} \times p^{32}(1-p)^{32} \times p^6(1-p)^2 \times p^4$

$L = c \times (1-p)^{603} \times p^{119}$ c is a constant.

taking log on both sides.

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

Differentiate with respect to p.

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

$\frac{d\ln(L)}{dp} = 0.$

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

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

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

$720p = 117$

$p = \frac{117}{720} = \underline{\underline{0.1625}}$.

ii Carry out a goodness fit test for the number of claim on each policy conforms to the binomial model.

χ^2 goodness of fit test:

H_0: probability conforms to be Bin(4, p) distribution

H_1: probability does not conforms to be Bin(4, p) distribution.

$\hat{p} = 0.1625.$

$P(X=0) = {}^4C_0\, p^x \times q^4 = 1 \times (1-p)^4 = (1-0.1625)^4 = 0.491971$

$P(X=1) = {}^4C_1\, p^1 \times q^3 = 4 \times p \times (1-p)^3 = 4(0.1625)(1-0.1625)^3 = 0.38182$

$P(X=2) = {}^4C_2\, p^2 q^2 = \frac{\cancel{4}^{2} \times 3}{\cancel{2} \times 1} p^2 (1-p)^2 = 6(0.1625)^2(1-0.1625)^2 = 0.1111$

$P(X=3) = {}^4C_3\, p^3 q^1 = 4\, p^3(1-p) = 4(0.1625)^3(1-0.1625) = 0.0143$

$P(X=4) = {}^4C_4\, p^4 q^0 = p^4 = (0.1625)^4 = 0.000697$

x_i	O_i	E_i
0	86	88.555
1	75	68.729
2	16	20.003
3	2	2.5885
4	1	0.126

Combining 2, 3 and 4 as E_i of 3 and 4 is < 5

Date ....
Page ....

x (No of claims)	Oi	Ei	$(O_i - E_i)^2 / E_i$
0	86	88.555	0.0737172
1	75	68.729	0.5721812
2+	19	22.717	0.6081828

degree of freedom = n-1 - no. of clubbing - no. of parameters estimated

$= 5-1-2-1$

$= 5-4 = \underline{1}$

$\chi^2_{cal} = \dfrac{\Sigma(O_i - E_i)^2}{E_i} = 0.0737172 + 0.5721812 + 0.6081828$

$= 1.254081215$

$\chi^2_{tab} = \chi^2_{1, 5\%} = 3.841$

$\chi^2_{tab} < \chi^2_{cal}$ $\therefore$ we donot reject Ho

4 The probability density function of a random variable X is given by $f(x) = \dfrac{c\beta^3}{(x+\beta)^4}$; $x > 0, \beta > 0$. where c is constant and β is a parameter

i) Determine the value of c and calculate the mean & variance of X as a function of β by using Formula and Tables for Actuarial Examination or otherwise

$f(x) = \dfrac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}$.

Question is on Pareto distribution (?)

The distribution belongs to Pareto distribution $E(X) = \dfrac{\lambda}{\alpha-1}$ $V(X) = \dfrac{\alpha\lambda^2}{(\alpha-1)^2(\alpha-2)}$

$\alpha + 1 = 4$ — By comparing $f(x)$.

$\alpha = 3$ $\quad \lambda = \beta$

$c = \alpha = 3$

$E(x) = \dfrac{\beta}{3-1} = \beta/2$

$V(x) = \dfrac{3\beta^2}{(3-1)^2(3-2)} = \dfrac{3\beta^2}{2^2 \cdot 1} = \dfrac{3\beta^2}{4}$

Date ....
Page ....

ii Show that the method of moments estimator of $\hat{\beta}$ is $2\bar{x}_n$ and verify the unbiasedness & consistency of the estimator.

$E(x) = \frac{\beta}{2}$ ___ from (i)

$\underline{2\bar{x}_n = \beta}$

$$\text{Bias} = E[\hat{\beta}] - \beta$$
$$= E(2\bar{x}_n) - \beta$$
$$= 2E(\bar{x}_n) - \beta$$
$$= 2 \times \frac{\beta}{2} - \beta$$
$$= \underline{0}.$$

By Mean Square error.

$MSE[\hat{\beta}] = V[\hat{\beta}] + (\text{Bias})^2 = 4\,Var[\bar{x}_n]$

∵ Bias = 0 ∴ The estimator is unbaised

$4\,Var(\bar{x}_n) = 4\,Var(\bar{x})/n = 4\,\overline{3/4\,\beta^2}^{\text{from(i)}}/n = \underline{3\beta^2/n}$

$\underline{MSE = 3\beta^2/n}$

As $n \to \infty$ $MSE \to 0$ ∴ estimator consistent.

iii Consider set of estimators of the form $b\bar{x}_n$ where b is constant. Show that value of b minizes the MSE of $b\bar{x}_n$ is $2/(1+3/n)$.

$$MSE(b\bar{x}_n) = Var(b\bar{x}_n) + [\text{Bias}(b\bar{x}_n)]^2$$
$$= b^2\,Var(\bar{x}_n) + (E(b\bar{x}_n) - \beta)^2$$
$$= \left[b^2 \times \frac{3}{4}\frac{\beta^2}{n}\right] + \left\{\left[b \times \frac{\beta}{2}\right] - \beta\right]^2$$
$$= \frac{3b^2\beta^2}{4n} + \left[\frac{\beta b}{2} - \beta\right]^2$$
$$= \frac{3b^2\beta^2}{4n} + \frac{\beta^2 b^2}{4} - \frac{2\beta^2 b}{2} + \beta^2$$

To minimise differentiate wrt

$$\frac{dMSE}{d\beta} = \frac{6b\beta^2}{4n} + \frac{2\beta^2 b}{4} - \frac{\cancel{2}\beta^2}{\cancel{2}}$$
$$= 3b\beta^2/2n + \beta^2 b/2 - \beta^2$$
$$= \frac{\beta^2}{n}\left[\frac{3b}{2} + \frac{bn}{2} - n\right]$$
$$= \beta^2/n\left[\frac{3b}{2} + n(b/2 - 1)\right]$$

$\frac{2n}{n+3} \longrightarrow$ minimum value

Differentiate wrt b

$$\frac{d^2MSE}{db^2} = \frac{\beta^2}{n}\left[\frac{3}{2} + \frac{n}{2}\right] - \left[\frac{3+n}{2n}\right]\beta^2$$

Date ....
Page...........

Q5i) Show that the MOM estimate for 'a' of a continuous uniform distribution $U(a,b)$ is $\hat{a} = \bar{x} - \sqrt{3}s$ where $\bar{x}$ is sample mean and s is sample standard deviation.

$E(x) = \frac{a+b}{2}$ $\qquad V(x) = \frac{1}{12}(b-a)^2$

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

$b = a - 2E(x) - a$

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

$3s^2 = (\bar{x}-a)^2$ $\qquad s^2 = \frac{4}{\cancel{12}\,3}(\bar{x}-a)^2$

square rooting both sides

$\sqrt{3}\,s = (\bar{x}-a)$ $\qquad s^2 = \frac{(\bar{x}-a)^2}{3}$

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

Hence proved.

ii) State a formula for $\hat{b}$ (method of moments estimate for 'b') in terms of sample mean and sample standard deviation.

$b = 2E(x) - a$

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

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

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

iii) Create a sample of 5 observations and use the sample to demonstrate potential weakness of method of moments estimati of 'a' and 'b' for $U(a,b)$.

Sample - 1, 2, 3, 4, 5

$\bar{x} = \frac{15}{5} = 3$ $\qquad s = 1.581134$

MOM estimates using above formulas:

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

$= 3 - \sqrt{3}(1.58114)$ $\qquad = 3 + \sqrt{3}(1.58114)$

$= 0.26139$ $\qquad = 5.73861$

Result?

Date ....
Page ....

iv Find the maximum likelihood estimate of b for $U(a,b)$ based on sample $x_1, x_2 \ldots x_n$ when $a=0$.

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

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

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

$L = \prod_{1}^{n} 1/b-a$

$L = \prod_{1}^{n} 1/b$

$L = 1/b^n$

$L = 1/b^n$ or $L = b^{-n}$

taking log on both sides

$\log L = -n \log b$

differentiate wrt a&b

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

$\hookrightarrow 0$

$0 = \frac{-n}{b}$ $\boxed{\cancel{b = -n}}$

6 ABC Space agency, responsible to protect a planet from asteriod collection, devloped new space to space to be loaded in satellites.

The agency ~~placed~~ planned missile trail in 2 steps for testing H_0 $p = 0.1$ verses H_1 $p > 0.1$, where p is the proportion of hits the missiles, each missile targets at similar asteroid.

At the first step 12 missiles will be fired. If ~~there~~ are 3 or more independent hits are observed among the first 12 missiles, H_0 is rejected, the study is terminated and no more missiles are fired.

Otherwise, another 12 missiles will be fired in 2^{nd} step. If a total of 5 or more independent hits are observed among 24 missiles fired in 2 steps, then H_0 is rejected.

i) Calculate the probability of Type 1 error from two step testing procedure.

Type I error → Rejecting H_0 when H_0 is true.

P(Type I error) = ?

let X be the No. of hits in the first step (12 missiles) and Y be the No. of hits in the 2nd step (12 missiles).

$p = 0.1$

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

[By continuity correction on the lower end random

Date ....
Page ....

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

$$= 0.1109 + 0.00121432 + 0.00964096 + 0.02551809$$

$$= 0.14727337$$

$\therefore$ There is a 14.727% Type 1 error.

ii Calculate the probability of rejecting the null hypothesis Ho when p=0.3

$$P(\text{Rejecting Ho}, p=0.3) = (X \geq 3) + P(X=0)P(Y \geq 5) + P(X=1)P(Y \geq 4) + P(X=2)P(Y \geq 3)$$

$$= (X \geq 2.5) + P(X=0)P(Y \geq 4.5) + P(X=1)P(Y \geq 3.5) + P(X=2)P(Y \geq 2.5)$$ By continuity correct

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

$$= 0.7472 + 0.00381294 + 0.0361340 + 0.1253806$$

$$= ~~0.9125271~~ \; 0.9125271$$

$\therefore$ There is 91.257% probability of rejecting Ho when $p = 0.3$.

iii Calculate the probability of Type 2 error when $p=0.3$

Type II Error $\rightarrow$ Accepting Ho when Ho is false.

$\therefore P(\text{rejecting Ho}) = 0.9125271$.

$$P(\text{Accepting Ho} \mid \text{Type II Error}) = 1 - 0.9125271$$

$$= 0.0874729$$

$\therefore$ There is a 8.74729% Type 2 error.

Suppose that 10 space agencies have developed missile similar to ABC space agency. They have performed only the first step trials and observed aggregate of 40 hits out of 120 independent missiles fired.

iv Calculate an approximate lower bond of 90% right-tailed confidence interval for 'p'

Date ....
Page ....

- 10 Space agencies observed aggregate of 40 ships and 120 fires.
- Assuming that sample comes from binomial

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

$n=120 \quad x=40$

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

since it is one side.

$$P\,\hat{p} = \hat{p} \pm Z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \hat{p} - Z_{10\%}\sqrt{\frac{0.3333(1-0.333)}{120}}$$

$= 0.33333 - (1.2816 \times 0.043033314829)$

$= 0.33333 - 0.05515128285$

$= 0.27816205$

$= \underline{\underline{0.2782}}$

v) Test the Hypothesis $H_0: p = 0.278$ against $H_1: p > 0.278$ at 10% level of significance.

Test whether $p > 0.278$ at 10% level of significance

$H_0\ p = 0.278 \quad H_1\ p > 0.278.$

$$\frac{x-np}{\sqrt{np(1-p)}} \sim N(0,1) \quad - \text{CLT}$$

$x = 40$ by continuity correction $x = 39.5$.

$$\frac{39.5 - 120(0.278)}{\sqrt{120(0.278)(1-0.278)}} = \frac{6.14}{4.907740824} = 1.251084811$$

One sided test then the value is less that Z_{10} i.e 1.2816

$Z_{cal} < Z_{tab}$ $\therefore$ We donot reject H_0

Date ....
Page ....

vi comment on your result of CI obtained and hypothesis testing in (v)

There hypothesis testing obtained in (v) suggests that we do reject Ho. Stating that $p = 0.278$ at 10% level of significance (one sided).

Apptitude tests were conducted by ABC Space agency at their 2 Institutes (1 & 2) that provide special training on missile technology. Test the scores shown in the table below.

	Institute 1	Institute 2
No. of trainees	12	15
Mean scores	65	70
Standard deviation	54	70

vii) Test the equality in mean scores of the population associated with 2 institutes. State any assumption made.

$H_0 : \mu_1 = \mu_2 \qquad H_1 \; \mu_1 \neq \mu_2.$

Assumption → sample comes from normal distribution & is independent.

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

$\bar{x}_1 = 65 \quad n_1 = 12 \quad \bar{x}_2 = 70 \quad n_2 = 15 \qquad \mu_1 - \mu_2 = 0.$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{1}{12+15-2}\left[(12-1)(54)^2 + (15-1)(70)^2\right]$$

$$= \frac{1}{25}\left[(11)(2916) + 14(4900)\right]$$

$$= \frac{1}{25}(32076 + 68600)$$

$$S_p^2 = 4027.04$$

$$S_p = 63.45896312$$

$$\frac{65-70-0}{63.4589\left(\sqrt{\frac{1}{12}+\frac{1}{15}}\right)}$$

$$\frac{-5}{63.4589\left(\sqrt{\frac{1}{27}}\right)}$$

$$\frac{-5}{12.21268315} = -0.4094104416 - t_{cal}$$

$t_{tab} = t_{27-2} = t_{25,\,2.5\%} = \pm 2.060$ $\because t_{tab} > t_{cal}$

$\therefore$ we do not reject Ho. at 5% level (two sided)

Q7 A study into the average claim (in Rs'000) per health ins policy was performed for the claims incurred in public and private hospitals. Data from some cities is given below.

	City 1 ~~Claim~~	City 2	City 3	City 4	City 5	City 6	City 7	C
Public	24	45	29	83	20	40	26.5	2
Private	30	54.5	30	40	28.5	36	30.5	3

i) Determine the sample mean & sample variance of the average claim size in both the type of hospital.

Public$^{(1)}$	Private$^{(2)}$
24	30
45	54.5
29	30
83	40
20	28.5
40	36
26.5	30.5
25	30.5.
27.5	35.5

~~$n = 8$~~

~~$\Sigma x_{public} = 242.5$~~

~~$\bar{x} = \frac{242.5}{8} = 30.3125$~~

~~$6\, s_1 = 8.514437027$~~

~~$6\, s_1^2 = 63.43359375$~~

$n = 9$

$\Sigma x_1 = 270$

$\bar{x} = \frac{270}{9} = 30$

$s_1 = 8.019507466$

$s_1^2 = 64.3125$

Date ....
Page ....

$n = 9$

$\Sigma x_2 = 315.5$

$\bar{x}_2 = 35.056$

$S_2 = 8.209919475$

$S_2^2 = 67.402778.$

ii) State the primary condition that needs to be true for testing equal means & verify whether that condition is satisfied in the above example.

The primary condition that needs to be true for testing equal means is that $\sigma_1^2 = \sigma_2^2$

Ho $\sigma_1^2 = \sigma_2^2$ v/s H_1 $\sigma_1^2 \neq \sigma_2^2$

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1,\, n-2}$$

$$F_{cal} = \frac{S_1^2}{S_2^2} = \frac{64.3125}{67.402778} = 0.9541520709.$$

$F_{tab} = \frac{1}{F_{8,8}}\ ;\ F_{8,8}$ [∵ Nothing is mentioned 95% CI is taken]

$\llcorner$ at 5%.

$= 0.2255808,\ 4.433.$

∵ $F_{tab} < F_{cal} < F_{tab}$ we do not reject Ho

∴ we can conclude that $\sigma_1^2 = \sigma_2^2$ (Variance is equal.

iii) Test whether the treatment in private hospital results higher claim size at 95% level.

Ho $\mu_1 = \mu_2$ $\quad H_1\ \mu_1 < \mu_2$

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

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

$$= \frac{64.3125(8) + 67.402778(8)}{\cancel{14}\ 8+8}$$

$$= 65.857639$$

$$Sp = 8.115271961$$

$$\frac{(35.056-30)-(0)}{8.115271961\sqrt{\frac{1}{9}+\frac{1}{9}}} = t_{cal}$$

$$= \frac{5.056}{8.115271961\sqrt{2/9}}$$

$$= 1.031631081$$

$\cancel{t_{16,5\%} = \pm 2.120}$ $\quad t_{16,5\%} = 1.746.$

~~$\therefore t_{tab} < t_{cal} < t_{tab}$~~ $\quad t_{tab} < t_{cal}$

~~$\therefore$ Do not reject H_0 $\therefore$~~ $\quad \therefore$ Reject H_0

and we can conclude that $\mu_1 \approx \mu_2$ (cost of claim in p... hospital is similar to public hospital).

Date ....
Page ....

8 If $\hat{\theta}$ is an estimator of parameter θ, answer the following

i) Define unbaised estimator - An unbaised estimator is the one that the sample estimator is equal to the population parameter. Example ($S^2 \to \sigma^2$), It is a disirable property.

ii) Define 'bias' - Bias is the over and underestimating the value of the parameters. $E(g(x)) \neq \mu$ (The parameter is not true).

iii Define Mean Square Error (MSE) of the estimator $\hat{\theta}$ - It calculate the range of fluctuation, Helps in Knowing the efficieny of the parameter. MSE is small then it is efficient and MSE is large then it is not efficient. $E[(g(x)-\theta)^2]$ = Variance + $Bias^2$.

There exist another estimator $\tilde{\theta}$ of the same parameter θ, such that $\hat{\theta}$ has no bias but higher MSE than $\tilde{\theta}$ while $\tilde{\theta}$ has a positive bias

iv) State giving reason, which estimator is "efficient"

$\tilde{\theta}$ is a more efficient than $\hat{\theta}$ because the Mean Square Error of $\hat{\theta}$ is higher than the MSE of $\tilde{\theta}$

v) When would either of the two estimates be termed as consistent?

$\hat{\theta}$ will be considered consistent because $\hat{\theta}$ does not have any bias whereas $\tilde{\theta}$ has a positive bias.

vi) Outline any 2 methods of estimation θ.

The two methods of estimating θ are as follows:-

1) MOM - Methods of Moments / MME

2) MLE - Maximum likelyhood Estimation.

Date ....
Page ....

Q9 i) Define the variable t_k used in t-test for sampling distrib of sample mean describing all symbols used

t_k can be defined as $\frac{N(0,1)}{\sqrt{\chi^2_k / k}}$

$k \to$ degrees of freedom

$N(0,1)$ and $\chi^2 k \to$ are independent.

ii) State the Mean and variance of t_k for $k>2$.

Mean for t_k where $k > \cancel{0}\,2 = 0$

and variance $= \frac{k}{k-2}$

iii A sample of 10 numbers from normal population has sample mean and sample variance as 50 and 48.667 respectively.
Determine the confidence interval for population mean at 99% confidence level.

a) Using the t-test tables

99% confidence Interval $= \overset{\bar{x}}{\cancel{50}} \pm t_{n-1, \frac{1\%}{2}} \left(\frac{\sigma}{\sqrt{n}}\right)$ $\quad \bar{x} = 50$

$\sigma^2 = 48.667$

$n = 10$

$= 50 \pm t_{9, 0.9\%} \left(\sqrt{\frac{48.667}{10}}\right)$

$= 50 \pm (3.25)(2.2060059836)$

$= 50 - 7.169694467, 50 + 7.1694467$

$= \boxed{42.83030553, 57.1694467}$

a) Assuming a normal distribution with parmeter as the results of part (ii) above.

$X \sim N(50, 48.667)$

$99\% CI = \bar{x} \pm z_{\alpha/2} \left(\sigma / \sqrt{n}\right)$

$= 50 \pm z_{0.05} \left(\frac{48667}{\sqrt{50}}\right)$

$= 50 \pm 2.5758 \ (6.882553144$

$= (32.2719161, 67.728080 39)$

Date ....
Page ....

Q10 Number of claims in a year on an insurance policy is believed to follow a Poisson distribution. Claims on portfolio of 1000 such policies were observed for one year. It was suggested that the value of Poisson parameter is 3. If the observed number of claims in that one year is less than 3100 then the suggested value of 3 for Poisson distribution is accepted else rejected.

i) Define Type I error and estimate it for the above use.

Type I error → ~~rejected~~ rejecting Ho when it is true

$x \rightarrow$ No. of claims on the portfolio. $x \sim$ Poission $(n\mu)$ ie Poission(1000×3)

$x \sim$ Poission (3000)

$H_0 \rightarrow x \sim$ Poission (3000). $\quad x \sim$ Normal $(3000, 3000)$

P(rejecting Ho when Ho is true) $= P(x > 3100)$

$$P(x>3100) = P\left(\frac{x-\mu}{6} > \frac{3100-3000}{\sqrt{3000}}\right)$$

$$= P\left(z > 1.82574\right)$$

$$= 1 - P(z < 1.825)$$

$$= 1 - 0.96638$$

$$= \underline{\underline{3.362\%}}$$

ii) Define Type II error. - not rejecting Ho when it is false.

iii) Define a power of a test and determine the power of test in terms of μ in above case.

Power of a hypothesis test is a measure of how effective the test is at identifying a difference in populations if such a difference exists.

P(Reject Ho when Ho is false). $= P(x > 3100)$ $\quad x \sim Pois(n\mu) \sim N(n\mu, n\mu)$

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

Date ....
Page ....

iv If actual observed number of claims is 2900, determine the CI for poisson parameter at 99% confidence level.

$X \sim \text{Poisson}\left(\frac{2900}{1000}\right) \xrightarrow{\hat{u}} X \sim \text{Poisson}(2.9)$

$X \sim \text{Normal}\left(2.9, \frac{2.9}{1000}\right) \quad X \sim \text{Normal}\left(\hat{u}, \frac{\hat{u}}{n}\right)$.

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

$99\% \text{ CI} = \hat{u} \pm Z_{0.005}\left(\sqrt{\frac{u}{n}}\right)$

$= 2.9 \pm 2.5758\left(\sqrt{\frac{3}{1000}}\right)$

$= 2.9 \pm 0.1410823764$

$= \cancel{2.9}\ 2.075891764,\ 3.0410823764$

Q11 A random sample of basic monthly salary (x) of 12 employees of a multinational organization has been selected.
The summary of data is as follows:

$\Sigma x = 213,200$

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

Sample mean = 17,767

Sample standard deviation = 10144.

However later it was identified that two of these selected employees were temporary employees and need to be replaced with two permanent employees.
Calculate the sample mean and standard deviation of the revised sample if the ~~sample~~ salary of the two temporary employees is 11000 and 48,000 and salary of the 2 randomly selected permanent employees is 27500 and 31500.

$$\text{Revised Mean} = \frac{(213200 - 11000 - 48000 + 27500 + 31500)}{12} = \frac{213200}{12}$$

$$= \underline{\underline{17767}}$$

$$\text{Revised } \Sigma x^2 = 4919860000 - 121{,}000{,}000 - 2{,}304{,}000{,}000 + 756250000 + 992250000$$

$$= 4243360000$$

$$\text{Revised SD} = \sqrt{\frac{1}{N-1}\left(\Sigma x^2 - \frac{\bar{X}^2}{n}\right)}$$

$$= \sqrt{\frac{1}{11}\left(4243360000 - \frac{213200^2}{12}\right)}$$

$$= 6435.0367$$

No change in mean but reduction in SD from 10144 to 6435.0367

Date
Page

Q12 let $x_1, x_2 \dots x_n$ be a random sampled from uniform distribution over $(0, \theta)$, where θ is an unknow parameter (>0).

i) outline why the Cramér - Rao lower bound for the variance of unbaised estimators of θ does not appy in this case. Consider an estimator of $\hat{\theta}(L) = cY$ for some constant c where $Y = \max x_i$.

The Cramer's Rao lower Bond holds true with general conditions but for a range of distribution invowing parametes(in this case uniform distribution). Because the data is not continous.

ii] show that the probability density function of Y is given as

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

Hence show that

$E(Y^k) = \frac{n\theta^k}{n+k}$ for any non-negative real number K

$Y = \max X_i \qquad 0 < x < \theta$

$\therefore$ for $Y \quad 0 < Y < \theta$

$$P[Y < y] = P[\max x_i < y]$$
$$= P[\cap x_i < y]$$
$$= \prod_{i=1}^{n} x_i < y$$
$$= \prod_{i=1}^{n} \left[\int_0^y \frac{1}{\theta} dx\right]$$
$$= \prod_{i=1}^{n} \left(\frac{x}{\theta}\right)_0^y = \left(\frac{y}{\theta}\right)^n$$

$$f(x) = \frac{1}{b-a} = \frac{1}{\theta - 0} = \frac{1}{\theta}$$

$$g(y) = \frac{d}{dy} P[Y < y] = \frac{d}{dy}\left(\frac{y}{\theta}\right)^n = \frac{d}{dy}\frac{y^n}{\theta^n} = \frac{ny^{n-1}}{\theta^n}$$

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

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

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

$$= \frac{n}{\theta^n} \left[\frac{y^{n+k}}{n+k} \right]_0^\theta$$

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

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

$$= \frac{\cancel{\theta^n} \times \theta^k \, n}{n+k \, \cancel{\theta^n}}$$

$$= \frac{n\theta^k}{n+k} \quad \text{———} \; (1)$$

iii Show that bias & MSE of the estimator $\hat{\theta}(c)$ are given as follow

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

$$\text{Bias}[\hat{\theta}(c)] = E(\hat{\theta}c) - \theta$$
$$= E(cy) - \theta$$
$$= c\,E(y) - \theta \qquad k = 1 \text{ from (1)}.$$
$$= c \times \frac{n\theta}{n+1} - \theta$$
$$= \theta\left(\frac{cn}{n+1} - 1\right)$$

Hence proved.

$$\text{MSE} = \text{Variance} + \text{Bias}^2 \text{ or } E[(g-\theta)^2]$$
$$= E[(\hat{\theta}c - \theta)^2]$$
$$= E[(cy - \theta)^2]$$
$$= E[c^2y^2 - 2cy\theta + \theta^2]$$
$$= c^2E(y^2) - 2c\theta E(y) + \theta^2$$

— $k=2$, $k=1$ in (1)

Date ....
Page ....

$$c^2 \frac{n\theta^2}{n+2} - 2c\frac{n\theta^2}{n+1} + \theta^2 \quad \text{— ②}$$

— Hence proved

iv) Find the value of $c(=c_u)$ for which $\hat{\theta}(c)$ becomes an unbaised estimator of θ.

for unbaised estimator Bias $(\hat{\theta}(c)) = 0$.

$\therefore E(\hat{\theta}c) - \theta = 0$

$\cancel{c\,E(\theta) = 0}$ $\quad E(cy) - \theta = 0$

$$cE(y) - \theta = 0$$

$$c \times \frac{n\theta}{n+1} - \theta = 0.$$

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

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

$$cn = 1(n+1)$$

$$c = (n+1/n)$$

v) Find the value of $c(=c_m)$ for which mean square error of $\hat{\theta}(c)$ is minimised.

To minimse MSE $\frac{d}{dc}$MSE $= 0$.

from ②

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

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

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

$$\frac{c}{n+2} - \frac{1}{n+1} = 0$$

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

vi) - doubt/ask.