Institute of Actuaries of India

Subject CT3 – Probability & Mathematical Statistics

December 2018 Examination

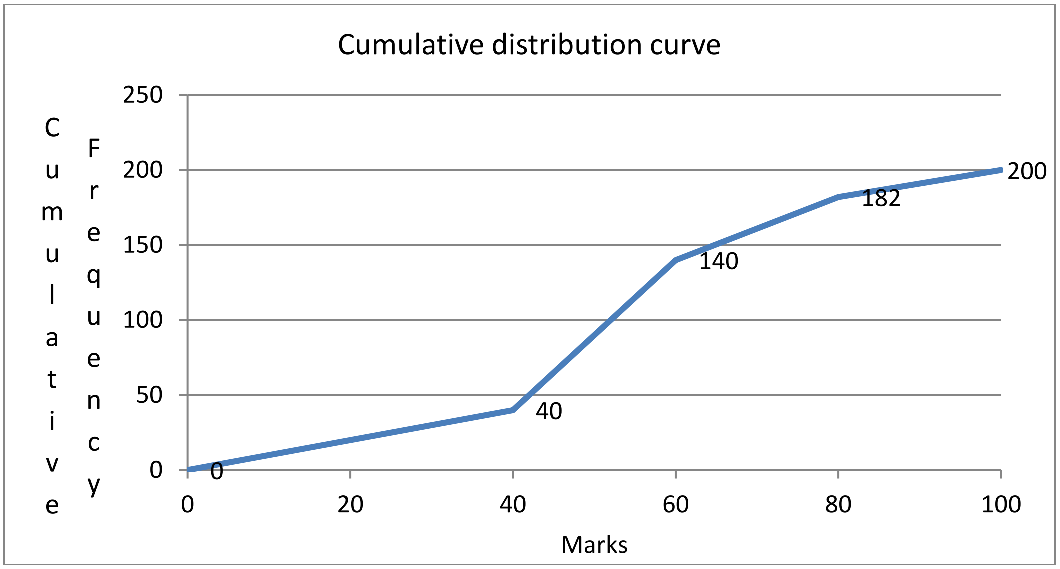
INDICATIVE SOLUTION

Introduction

The indicative solution has been written by the Examiners with the aim of helping candidates. The solutions given are only indicative. It is realized that there could be other points as valid answers and examiner have given credit for any alternative approach or interpretation which they consider to be reasonable.

Solution 1:

i)



[2]

ii) The median value splits the distribution into two equal halves, so that there are 100 observations below the median and 100 observations above the median. The median lies in the $40 \le X < 60$ interval. Using linear interpolation gives:

$$40 + \frac{60}{100}(60 - 40) = 52$$

Alternatively, using the $\frac{1}{2}(n+1)$ *formula we get:*

$$40 + \frac{60.5}{100}(60 - 40) = 52.1$$

If the values are distributed uniformly within each interval, the mean of each interval will equal the midpoint value. Therefore, the overall mean is:

$$\frac{40\,(20) + 100\,(50) + 42\,(70) + 18\,(90)}{200} = 51.8$$ [3]

[5 Marks]

Solution 2:

Total number of cases = 6^4 =1296

The various combinations for the sum being 19 and corresponding number of arrangements in each case are:

(6,6,6,1) -------> $\frac{4!}{3! * 1!}$ = 4

(6,6,5,2) -------> $\frac{4!}{2! * 1! * 1!}$ = 12

(6,6,4,3) -------> $\frac{4!}{2! * 1! * 1!}$ = 12

(6,5,5,3) -------> $\frac{4!}{2! * 1! * 1!}$ = 12

(6,5,4,4) -------> $\frac{4!}{2! * 1! * 1!}$ = 12

(5,5,5,4) -------> $\frac{4!}{3!*1!}$ = 4

Number of favorable cases are 4+12+12+12+12+4 =56

Hence the required probability is $\frac{56}{1296}$ = 0.04321 **[4 Marks]**

Solution 3:

i) We know that $\sum_{x=1}^{10} C\,(2x - 1) = 1$

This implies 2 C$\sum_{x=1}^{10} x$ − 10 C = 1; 2 C $\frac{10(11)}{2}$ − 10 C = 1;

That is, 100 C = 1; C= $\frac{1}{100}$ = 0.01 [1]

ii) E(X) = $\sum_{x=1}^{10} x\ P(X = x)$ = $\sum_{x=1}^{10} x\ \ 0.01\,(2x - 1)$

= 0.02 $\sum_{x=1}^{10} x^2$– 0.01 $\sum_{x=1}^{10} x$

= 0.02 $\frac{10*11*21}{6}$ −0.01 $\frac{10*11}{2}$

= 7.70 – 0.55 = 7.15 [3]

[4 Marks]

Solution 4:

i) Total persons served are $\frac{31\,(31+1)}{2} = 496$

31 December 2017 is Sunday, so we have 7, 14, 21 and 28 as Sundays in January 2018. Number of persons who were served on Sundays and got lunch and clothes is $70\ (= 7 + 14 + 21 + 28)$

Hence, the probability of a randomly selected person got lunch and clothes among all persons receiving charity is $\frac{70}{496} = 0.14113\ \approx 14\%$ [3]

ii) The probability of a randomly selected personwas served on third Sunday, given that the persongot lunch and clothes is

$$\frac{\left(\frac{21}{496}\right)}{\left(\frac{70}{496}\right)} = \frac{21}{70} = 0.3 = 30\%$$

[2]

[5 Marks]

Solution 5:

We know that:

$M_x(t) = E\,(e^{tX}) = \sum_x e^{tx}$ P (X=x)

It is given that

$M_x(t) = \sum_{k=0}^{\infty} \frac{e^{kt-1}}{k!}$

$M_x(t) = \sum_{k=0}^{\infty} \frac{e^{-1}}{k!} e^{kt}$

From the above, the co-efficient of e^{kt}, we get

P(X=3) $= \frac{e^{-1}}{3!} = \frac{1}{6e}$ =0.06131

[2 Marks]

Solution 6:

Let Y = $X_1 + X_2$

$M_y(t) = M_{X_1+X_2}(t) = E\left(e^{t(X_1+X_2)}\right)$

$= M_{X_1}(t)\, M_{X_2}(t) = \left(\frac{e^t+e^{2t}+e^{3t}+e^{4t}}{4}\right)\left(\frac{e^t+e^{2t}+e^{3t}+e^{4t}}{4}\right)$

$= \left(\frac{e^{2t}+2e^{3t}+3e^{4t}+4e^{5t}+3e^{6t}+2e^{7t}+e^{8t}}{16}\right)$

Hence$P(Y=2) = \frac{1}{16}$; $P(Y=3) = \frac{2}{16}$; $P(Y=4) = \frac{3}{16}$; $P(Y=5) = \frac{4}{16}$;

$P(Y=6) = \frac{3}{16}$; $P(Y=7) = \frac{2}{16}$; $P(Y=8) = \frac{1}{16}$

Or$P(Y=y) = \frac{4-|y-5|}{16}$ y=2,3,…,8

Alternate-1:

Using pgf:

$G_y(t) = G_{X_1+X_2}(t) = E\left(t^{(X_1+X_2)}\right)$

= $E(t^{X1})$* $E(t^{X2})$ = $G_{x1}(t) * G_{x2}(t)$ (Due to independence)

Now $G_{x1}(t) = E(t^{X1}) = \frac{t+t^2+t^3+t^4}{4}$

Similarly

$G_{x2}(t) = E(t^{X2}) = \frac{t+t^2+t^3+t^4}{4}$

Now

$G_{X_1+X_2}(t)$ = $E(t^{X1})$* $E(t^{X2})$

$= \left(\frac{t+t^2+t^3+t^4}{4}\right) * \left(\frac{t+t^2+t^3+t^4}{4}\right)$

$= \frac{t^2+2t^3+3t^4+4t^5+3t^6+2t^7+t^8}{16}$

P(Y=k) is co-efficient of t^k in $G_y(t)$ k=2,3,..,8 whre Y=X1+X2

Hence, $P(Y=2) = \frac{1}{16}$; $P(Y=3) = \frac{2}{16}$; $P(Y=4) = \frac{3}{16}$; $P(Y=5) = \frac{4}{16}$;

$P(Y=6)=\frac{3}{16};\ P(Y=7)=\frac{2}{16};\ P(Y=8)=\frac{1}{16}$

Alternate-2:

X1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
X2	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Y=X1+X2	2	3	4	5	3	4	5	6	4	5	6	7	5	6	7	8

P(Y=2) = P(X1=1 & X2=1) = P(X1=1)*P(X2=1) = $\frac{1}{4}*\frac{1}{4}=\frac{1}{16}$

P(Y=3) = P(X1=1 & X2=2) + P (X1=2 & X2=1)= P(X1=1)*P(X2=2) + P(X1=2)*P(X2=1)

= $\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}=\frac{\mathbf{2}}{\mathbf{16}}$

P(Y=4) = P(X1=1 &X2=3) + P (X1=2 & X2=2) + P (X1=3& X2=1)

= P(X1=1)*P(X2=3) + P(X1=2)*P(X2=2) +P(X1=3)*P(X2=1)

= $\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}=\frac{\mathbf{3}}{\mathbf{16}}$

P(Y=5) = P(X1=1 &X2=4) + P (X1=2 & X2=3) + P (X1=3& X2=2) + P (X1=4& X2=1)

= P(X1=1)*P(X2=4) + P(X1=2)*P(X2=3) +P(X1=3)*P(X2=2) +P(X1=4)*P(X2=1)

= $\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}=\frac{\mathbf{4}}{\mathbf{16}}$

P(Y=6) = P(X1=2&X2=4) + P (X1=3& X2=3) + P (X1=4& X2=2)

= P(X1=2)*P(X2=4) + P(X1=3)*P(X2=3) +P(X1=4)*P(X2=2)

= $\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}=\frac{\mathbf{3}}{\mathbf{16}}$

P(Y=7) = P(X1=3&X2=4) + P (X1=4 & X2=3)= P(X1=3)*P(X2=4) + P(X1=4)*P(X2=3)

= $\frac{1}{4}*\frac{1}{4}+\frac{1}{4}*\frac{1}{4}=\frac{\mathbf{2}}{\mathbf{16}}$

P(Y=8) = P(X1=4&X2=4) = P(X1=4)*P(X2=4) = $\frac{1}{4}*\frac{1}{4}=\frac{\mathbf{1}}{\mathbf{16}}$

[4 Marks]

Solution 7:

If Y and Z are independent then we have to show that P (Y=y, Z=z) = P(Y=y) P(Z=z), for all y and z

P(Y=y) = $\sum_{z=1}^{6} P(Y=y, Z=z)$

= P(Y=y,Z=y)+ $\sum_{z>y} P(Y=y, Z=z)$ = $\frac{1}{36}$ + (6-y) $\frac{2}{36}$ for y=1,2,…,6

P(Z=z) = $\sum_{y=1}^{6} P(Y=y, Z=z)$

= P(Y=z,Z=z)+ $\sum_{y<z} P(Y=y, Z=z)$ = $\frac{1}{36}$ + (z-1) $\frac{2}{36}$ for z=1,2,…,6

Since P(Y=1,Z=1)= $\frac{1}{36} \neq \left(\frac{11}{36}\right)\left(\frac{1}{36}\right)$
That means P(Y=1,Z= 1)≠P(Y=1)* P(Z=1)

Hence, Y and Z are not independent.

Alternatively
Probabilities:

Z=z		Y=y					
	y=1	y=2	y=3	y=4	y=5	y=6	**P(Z=z)**
z=1	**1/36**	**0**	**0**	**0**	**0**	**0**	**1/36**
z=2	**2/36**	**1/36**	**0**	**0**	**0**	**0**	**3/36**
z=3	**2/36**	**2/36**	**1/36**	**0**	**0**	**0**	**5/36**
z=4	**2/36**	**2/36**	**2/36**	**1/36**	**0**	**0**	**7/36**
z=5	**2/36**	**2/36**	**2/36**	**2/36**	**1/36**	**0**	**9/36**
z=6	**2/36**	**2/36**	**2/36**	**2/36**	**2/36**	**1/36**	**11/36**
P(Y=y)	**11/36**	**9/36**	**7/36**	**5/36**	**3/36**	**1/36**	**1**

Here for any pair P(Y=y,Z= z)≠ P(Y=y) P (Z=z) y=1,2,3,4 and z=1,2,3,4
eg
P(Y=1,Z= 1)= $\frac{1}{36}$
P(Y=1) = $\frac{11}{36}$ *and* P (Z=1) = $\frac{1}{36}$
P(Y=1,Z= 1)≠ P(Y=1)* P (Z=1)

Hence, Y and Z are not independent.

[5 Marks]

Solution 8:

Let W = $\frac{e^X}{1+e^X}$

The distribution function of W is
F(w) = P(W$\leq w$) = P$\left(\frac{e^X}{1+e^X} \leq w\right)$ = P$\left(\frac{1}{1+e^{-X}} \leq w\right)$ = P$\left(1 + e^{-X} \geq \frac{1}{w}\right)$

= P$\left(e^{-X} \geq \frac{1-w}{w}\right)$= P $\left(-\mathrm{X} \geq \ln\left(\frac{1-w}{w}\right)\right)$ =P $\left(\mathrm{X} \leq -\ln\left(\frac{1-w}{w}\right)\right)$=$\int_{-\infty}^{-\ln\left(\frac{1-w}{w}\right)} \frac{e^{-x}}{(1+e^{-x})^2}\,\mathrm{dx}$

Let e^{-x} = u => x = -ln u; dx= $\frac{-1}{u}$ du;for x =$-\infty$,u = ∞ and for x =$-\ln\left(\frac{1-w}{w}\right)$, u = $\frac{1-w}{w}$

F(w) = $\int_{\infty}^{\frac{1-w}{w}} \frac{u}{(1+u)^2}\left(\frac{-1}{u}\right)$ du = $\int_{\infty}^{\frac{1-w}{w}} \frac{-1}{(1+u)^2}$ du = $\left[\frac{1}{(1+u)}\right]_{\infty}^{\frac{1-w}{w}}$ = w

Hence, pdf of w is given by f(w) = $\begin{cases} 1 & if\ 0 < w < 1 \\ 0 & otherwise \end{cases}$

Alternatively

Let W = $\frac{e^X}{1+e^X}$

e^X = W *(1+e^X) = W+ W*e^X

e^X – W*e^X = W

e^X (1-W) = W=>e^X = $\frac{W}{1-W}$

e^{-X} = $\frac{1-W}{W}$

Hence X= ln ($\frac{W}{1-W}$)

Using the following, we can derive pdf of w

f(w) = f(x)* |$\frac{dx}{dw}$|

Now f(x) = $\frac{e^{-x}}{(1+e^{-x})^2}$for -∞<x <∞

f(x) = $\frac{\frac{(1-w)}{w}}{\left(1+\frac{1-w}{w}\right)^2}$for 0< w < 1

f(x) = w * (1-w)for 0< w < 1

$\frac{dx}{dw}$ = $\frac{1}{\frac{w}{(1-w)}}$ *[w *(-1)*$(1-w)^{-2}$ (-1) + $(1-w)^{-1}$ *(1)]

$\frac{dx}{dw}$ = $\frac{(1-w)}{w}$ *[$\frac{w}{(1-w)^2} + \frac{1}{(1-w)}$] = $\frac{(1-w)}{w}$ *[$\frac{w}{(1-w)^2} + \frac{(1-w)}{(1-w)^2}$] = $\frac{(1-w)}{w}$ *$\frac{1}{(1-w)^2}$

$\frac{dx}{dw}$ = $\frac{1}{w*(1-w)}$

Hence,

f(w) = f(x)* |$\frac{dx}{dw}$|

f(w) = w*(1-w)* |$\frac{1}{w*(1-w)}$| = 1 for 0<w<1

[5 Marks]

Solution 9:

Let X_i: Lifetime of the i[th] bulb installed; n: Light bulbs to be bought; S_n: Total lifetime of n bulbs; S_n= $X_1 + X_2 + X_3 + \ldots + X_n$

$E(S_n) =$ 3n; Var(S_n) $= \frac{n}{16}$

By Central Limit Theorem (CLT)$\frac{S_n - E(S_n)}{\frac{\sqrt{n}}{4}}$ ~ N (0, 1)

We need to find n such that P (S_n≥ 50) =0.95

P $\left(\frac{S_n-3n}{\frac{\sqrt{n}}{4}} \geq \frac{50-3n}{\frac{\sqrt{n}}{4}}\right)$ =P $\left(Z \geq \frac{200-12n}{\sqrt{n}}\right)$ = 0.95

1- P $\left(Z \leq \frac{200-12n}{\sqrt{n}}\right)$= 0.95 *OR*P $\left(Z \leq \frac{200-12n}{\sqrt{n}}\right) = 0.05$

$\frac{200-12n}{\sqrt{n}} = -1.6449$ from tables.

12 $(\sqrt{n})^2$-1.6449 $\sqrt{n}$ - 200 = 0

$\sqrt{n}$ = $\frac{1.6449+\sqrt{((-1.6449)^2)-4*12*(-200)}}{2*12}$ or = $\frac{1.6449-\sqrt{((-1.6449)^2)-4*12*(-200)}}{2*12}$

$\sqrt{n}$ = 4.15159 or -4.0145; n= 17.2357; So 18 bulbs should be bought.

[6 Marks]

Solution 10:

i) Given X~ N(μ,σ^2) where μ = 1000 hrs and σ = 240 hrs

Penalty for delay in each labour hour in excess of 1100 hours is Rs 1 lacs
If the minimum penalty paid by the company is 20 lacs, then the minimum delay in completing the construction of the airplane is 20 hours

Hence, it will take minimum of 1100+20 =1120 hours to complete the construction of the airplane.

We need to find
P(X≥1120) = P $\left(\mathrm{Z} \geq \frac{1120-1000}{240}\right)$ = P (Z≥0.5)

= 1- P(Z<0.5)= 1-0.69146 = 0.30854 [3]

ii) Let Y =$\sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(\mathrm{Xi}-1000)^2}{240^2} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{(\mathrm{Xi}-\mu)^2}{\sigma^2}$

Given: P (Y≤10) = 0.735

P$\left(\sum_{i=1}^{n} \frac{(Xi-\mu)^2}{\sigma^2} \leq 10\right)$ = 0.735

P ($\chi 2_n$ ≤ 10) = 0.735

From $\chi 2$ table, we get$n$ = 8

Hence n = 8 [4]

[7 Marks]

Solution 11:

i) $\mathrm{E[X]} = \frac{(\mathrm{a+b})}{2}$

$$\mathrm{b} = 2\mathrm{E[X]} - \mathrm{a} = 2\bar{x} - \mathrm{a}$$

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

$$\mathrm{s}^2 = \frac{(2\bar{x}-a-a)^2}{12} = \frac{(\bar{x}-a)^2}{3}$$

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

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

[3]

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

$\hat{b} = \bar{x} + \sqrt{3}s$ [1]

iii) Sample: 1, 2, 3, 4, 50

$\bar{x} = 12$; $s = 21.27$

Method of moments estimates using above formulae are:

$\hat{a} = 12 - \sqrt{3}(21.27) = -24.84; \hat{b} = 12 + \sqrt{3}(21.27) = 48.84$

For U (a, b),the probability of a sample point being less than 'a' or greater than 'b' is zero and we have a sample value 50 that is greater than our estimate of 'b'. This highlights a potential weakness of the method of moments.

[4]

iv) Likelihood for a sample of size n is$L(b) = \frac{1}{\mathrm{b}^n}$if $b \geq \max(x_i)$,otherwise L = 0

Differentiation with respect to b does not work because in the range of x depends on b

We must find b that maximizes$L(\mathrm{b})\ for\ \max(x_i)$ given.. We want $b$to be as small as possible subject to the constraint that b ≥ max (x_i).

Clearly the maximum is attained at b = max (x_i).

Hence $\hat{b}$ = max (x_i). [3]

[Marks 11]

Solution 12:

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

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

P(Type I error) = P (X≥ 3 | p = 0.1) + P (X = 0 | p = 0.1) P (Y ≥ 5 | p = 0.1) +
P (X = 1 | p = 0.1) P (Y ≥ 4 | p = 0.1) + P (X = 2 | p = 0.1) P (Y ≥ 3 | p = 0.1)

Using Actuarial Tables page 188 (probabilities for Binomial Distribution)

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

$$= 0.14727 \approx 15\%$$ [4]

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

= P (X ≥ 3 | p = 0.3) + P (X = 0 | p = 0.3) P (Y ≥ 5 | p = 0.3) +
P (X = 1 | p = 0.3) P (Y ≥ 4 | p = 0.3) + P (X = 2 | p = 0.3) P (Y ≥ 3 | p = 0.3)

Using Actuarial Tables page 188 (probabilities for Binomial Distribution)

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

$$= 0.91253 \approx 91\%$$ [3]

iii) P(type II error) is the probability of accepting H_0 when H_0 is false.
$= 1 - 0.91253 = 0.08747 \approx 9\%$(when p=0.3) [1]

iv) 10 Space agencies in aggregate fired 120 missiles and recorded 40 hits
Assuming that the sample comes from a binomial distribution, we know that the quantity

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

Here n = 120, X = 40. so$\hat{p} = \frac{40}{120} = \frac{1}{3} = 0.3333$

Using Actuarial Tables page162, $Z_{10\%} = 1.2816$
Lower bound of 90% right-tailed confidence interval for p is

$$\hat{p} - 1.2816\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.3333 - 1.2816\sqrt{\frac{0.3333(1-0.3333)}{120}} = 0.2782$$ [3]

v) Test whether p > 0.278 at 10% level of significance
H_0: P = 0.278 vs. H_1: p > 0.278
For one sample binomial model:

$\frac{X - np_0}{\sqrt{np_0 q_0}} \sim N(0,1)$with continuity correction.

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

We are carrying out a one-sided test. The value of the test statistic is less than

1.2816 (the upper 10% point of the N (0, 1) distribution) so we do not have sufficient evidence to reject H_0 at 10% level. [3]

vi) Lower bound of Confidence interval implies that there is only a 10% chance of 'p' ≤ 0.2782, whereas from the hypothesis test, 'p' could be less than or equal to 0.2780 with probability more than 10% (approx 10.5% corresponding to 1.2511).
The minor disconnect between Confidence interval and Hypothesis testing at the same level is due to

- use of sample proportion $\hat{p}$ to estimate population variance in calculating confidence interval, and
- applying continuity correction in hypothesis testing [3]

vii) Test whether there is a difference in the mean scores
We assume that the samples come from normal distributions with the same variance and that the samples are independent.

$H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$.

The pivotal quantity is: $\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

Given: $\overline{X}_1 = 65$; $\overline{X}_2 = 70$; $S_1 = 54$; $S_2 = 70$; $n_1 = 12$; $n_2 = 15$
The pooled variance is: $S_P^2 = \frac{1}{25}(11\ (2916) + 14\ (4900)) = 4027.04$

$$\frac{65-70-0}{63.459*\sqrt{\frac{1}{12}+\frac{1}{15}}} = -0.2034$$

This is within ± 2.060 $(= t_{25;2.5\%})$ So we have insufficient evidence to reject H_0 at the 5% level. Therefore, it is reasonable to conclude that there is no significant difference in the mean scores for the populations associated with the two Institutes. [4]

[21 Marks]

Solution 13:

i) Each Y_i has a $N(\alpha + \beta x_i, \sigma^2)$ distribution, so the joint likelihood function is:

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{y_i - \alpha - \beta x_i}{\sigma}\right)^2\right) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} exp\left(-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2}\right)$$

Taking logs: $\log L = -n log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 + constant$

Differentiating with respect to α and then with respect to β :

$$\frac{\partial \log L}{\partial \alpha} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)(-1)$$

$$\frac{\partial \log L}{\partial \beta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)(-x_i)$$

By setting $\frac{\partial \log L}{\partial \alpha}$ equal to zero we get $\sum_{i=1}^{n} y_i - n\widehat{\alpha} - \widehat{\beta}\sum_{i=1}^{n} x_i = 0$

By setting $\frac{\partial \log L}{\partial \beta}$ equal to zero we get $\sum_{i=1}^{n} x_i y_i - \widehat{\alpha}\sum_{i=1}^{n} x_i - \widehat{\beta}\sum_{i=1}^{n} {x_i}^2 = 0$

Solving the two equations: multiplying second equation with n and then subtracting first equation multiplied with $\sum_{i=1}^{n} x_i$ we get:

$n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i) - \widehat{\beta}(n\sum_{i=1}^{n} {x_i}^2 - (\sum_{i=1}^{n} x_i)^2) = 0$

$\widehat{\beta} = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} {x_i}^2) - (\sum_{i=1}^{n} x_i)^2} = \frac{S_{xy}}{S_{xx}}$; and

$\widehat{\alpha} = \frac{\sum_{i=1}^{n} y_i - \widehat{\beta}\sum_{i=1}^{n} x_i}{n} = \bar{y} - \widehat{\beta}\bar{x}$ [5]

ii) It seems that temperature on land is related to distance from lake through an exponential function like Y = aexp (βX). Transforming this using W = log Y, we get a linear model $w_i = \alpha + \beta x_i + e_i$ where $\alpha = \log a$

Site (x)	1	2	3	4	5	6	7	8	9	10
Temp ^0C (y)	27.0	25.0	23.4	22.6	21.8	21.4	21.2	21.0	20.8	21.0
w = Log y	3.2958	3.2189	3.1527	3.1179	3.0819	3.0634	3.0540	3.0445	3.0350	3.0445

$\bar{x} = 5.5000;\ \bar{w} = 3.1109; S_{xw} = -2.1504;\ S_{xx} = 82.5000;\ S_{ww} = 0.0686$

$\beta = \frac{S_{xw}}{S_{xx}} = \frac{-2.1504}{82.5000} = -0.0261;\ \alpha = \bar{w} - \beta\bar{x} = 3.2542$; a = exp (α) = 25.8996

$\widehat{w} = 3.2542 - 0.0261\,x \Rightarrow \widehat{y} = 25.8996\exp(-0.0261\,x)$ [6]

iii) For $x = 4.5$; $\widehat{w} = 3.2542 - 0.0261\,(4.5) = 3.1369$
$\widehat{\sigma}^2 = \frac{1}{n-2}\left(S_{ww} - \frac{S_{xw}^2}{S_{xx}}\right) = \frac{1}{8}\left(0.0686 - \frac{-2.1504^2}{82.5000}\right) = 0.0016$

Standard error of the estimate $SE(\widehat{w}) = \sqrt{\sigma^2 \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2)}{S_{xx}}\right)}$

$= \sqrt{0.0016\{1 + \frac{1}{10} + \frac{(4.5-5.5)^2)}{82.5000}\}}$ = 0.041272

95% confidence interval for $\widehat{w} is \widehat{w} \pm t_{0.025,n-2}\, SE(\widehat{w})$

= $3.1369 \pm (2.306 \times 0.041272\,) =$ (3.04176, 3.23211)

Using $\hat{y} = \exp(\widehat{w})$, 95% confidence interval for $\hat{y}$ is $(20.94, 25.33)$

[5]

[16 Marks]

Solution 14:

The parameters μ and τ_i , i = 1, 2,...,k can be estimated using least squares by finding values for μ,τ_i , i = 1, 2,...,k such that:

$$q = \sum_i \sum_j \mathrm{e}_{\mathrm{ij}}^2 = \sum_i \sum_j \left(\mathrm{Y_{ij}} - \mu - \tau_\mathrm{i}\right)^2$$

is minimized.

Differentiating this partially with respect to μ and τ_i , i = 1, 2... k ,and equating to zero:

$$\frac{\partial q}{\partial \mu} = -\sum_i \sum_j 2\left(y_{ij} - \mu - \tau_i\right)$$

$$\sum_i \sum_j \left(y_{ij} - \mu - \tau_i\right) = 0;\; \sum_i \sum_j y_{ij} - n\mu - \sum_i n_i \tau_i = 0$$

Given that $\sum_i n_i \tau_i = 0$, so $\hat{\mu} = \frac{1}{n} \sum_i \sum_j y_{ij} = \bar{Y}_{..}$

$$\frac{\partial q}{\partial \tau_i} = -\sum_j 2\left(y_{ij} - \mu - \tau_i\right)$$

$$\sum_j \left(y_{ij} - \mu - \tau_i\right) = 0;\; \sum_j y_{ij} - n_i\mu - n_i\tau_i = 0$$

$$\hat{\tau}_i = \frac{1}{n_i} \sum_j y_{ij} - \mu = \bar{Y}_{i.} - \mu = \bar{Y}_{i.} - \bar{Y}_{..}$$

These results are just what we should expect. μ is the average for all treatments and we are estimating this using $\bar{Y}_{..}$ i.e. the average over all the data. τ_i represents the difference between the mean for treatment i and the overall mean, and we are estimating this based on the corresponding difference in the sample means.

[5 Marks]

****