PRIOR AND POSTERIOR DISTRIBUTION

Q. The number of claims ~~in a week~~ received per day from a certain portfolio has a Poisson distribution with mean λ. The prior distribution of λ is as follows

λ	1	2	3
$P(\lambda)$	0.3	0.5	0.2

Given that 3 claims were received last day; determine the posterior distribution of λ and the mean.

<u>Solution</u>

Let X = Number of claims received per day

Here, $X \sim P(\lambda)$ so $P(X=x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$

Further, the prior distribution of λ is tabulated as below:

λ	1	2	3
$P(\lambda)$	0.3	0.5	0.2

Here, number of claims $X=3$. Thus, to obtain posterior distribut we must calculate $P(\lambda=1/X=3)$, $P(\lambda=2/X=3)$ and $P(\lambda=3/X=3)$

Now, probability of

$$P(\lambda=1/X=3) \propto P(\lambda=1) \cdot P(X=3/\lambda=1)$$

$$= 0.3 * \frac{e^{-1}\,1^{3}}{3!}$$

$$= 0.0183$$

$$P(\lambda=2/X=3) \propto P(\lambda=2) \cdot P(X=3/\lambda=2)$$

$$= 0.5 * \frac{e^{-2}\,2^{3}}{3!}$$

$$= 0.0902$$

$$P(\lambda=3/X=3) \propto P(\lambda=3) \cdot P(X=3/\lambda=3)$$

$$= 0.2 * \frac{e^{-3}\,3^{3}}{3!}$$

$$= 0.0448$$

Now,

$$P(X=3) = P(X=3, \lambda=1) + P(X=3, \lambda=2) + P(X=3, \lambda=3)$$
$$= 0.0183 + 0.0902 + 0.0448$$
$$= 0.1533$$

$\therefore$ The posterior distribution

$$P(\lambda=1/x=3) = \frac{0.0183}{0.1533} = \underline{0.1193}$$

$$P(\lambda=2/x=3) = \frac{0.0902}{0.1533} = \underline{0.5883}$$

$$P(\lambda=3/x=3) = \frac{0.0448}{0.1533} = \underline{0.2922}$$

Mean of posterior distribution

$$E(\lambda/x) = \sum \lambda \, P(\lambda/x)$$
$$= 1(0.1193) + 2(0.5883) + 3(0.2922)$$
$$= \underline{2.1725}$$

NON-PARAMETRIC TESTS

> MANN-WHITNEY U TEST

Q15. The nicotine contents of two brands of cigarettes, measured in milligrams was found to be as follows:

Brand A	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3		
Brand B	4.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2	1.9	5.4

If there is any significance difference between two brands of cigarettes. Use Mann-Whitney U-test.

Solution

Here, $n_1 = 8$, $n_2 = 10$, $\alpha = 0.05$ (supposed)

$H_0: Md_1 = Md_2$ i.e. there is no significant difference in nicotine contents of two brands of cigarettes.

$H_1: Md_1 \neq Md_2$ i.e. there is significant difference in nicotine contents of two brands of cigarettes (Two-tailed test)

$\therefore U_0 = \text{minimum of } \{23, 57\}$

$= 23$

Critical value : We have $\alpha = 0.05$ and $(n_1, n_2) = (8, 10)$

$\therefore U_{(8,10), 0.05}$ (two-tailed test) $= 17$

Decision : Since, $U_0 > U_{(8,10), 0.05}$, we do not reject H_0. Thus, there is no significant difference in nicotine contents of two brands of cigarettes.

Q17. Two independent random samples of unemployed men and women are drawn and the ages of 4 unemployed women and 5 unemployed men are recorded as follows:

Women	60	63	36	44	
Men	53	39	22	23	24

Do the data present sufficient evidence to conclude that there is a difference in the average age of unemployed men and women? Use Mann-Whitney U test at $\alpha = 0.05$.

Solution

Here, $n_1 = 4$

$n_2 = 5$

$\alpha = 0.05$

H_0: $Md_1 = Md_2$ i.e. there is no significant difference in average age of unemployed men and women.

$Md_1 \neq Md_2$ i.e. there is significant difference in average age of unemployed men and women. (Two-tailed test)

Test statistic : Under H_0,

$U_0 = \text{minimum of } \{U_1, U_2\}$

where, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

Test statistic : Under H_0,

$$U_0 = \text{minimum of } \{U_1, U_2\}$$

where, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

Calculation

Brand A	Brand B	Combined Ranks of Brand A	Combined Ranks of Brand B
2.1	4.1	4	12
4.0	0.6	10.5	1
6.3	3.1	18	7
5.4	2.5	14.5	6
4.8	4.0	13	10.5
3.7	6.2	9	17
6.1	1.6	16	2
3.3	2.2	8	5
	1.9		3
	5.4		14.5
		$R_1 = 93$	$R_2 = 78$

Now,

$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$= 8 \times 10 + \frac{8 \times 9}{2} - 93$

$= 80 + 36 - 93$

$= 23$

$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

$= 8 \times 10 + \frac{10 \times 11}{2} - 78$

$= 80 + 55 - 78 = 57$

Calculation

Women	Men	Ranks of Women	Ranks of Men
		Combined	
60	53	8	7
63	39	9	5
36	22	4	1
44	23	6	2
	24		3
		$R_1 = 27$	$R_2 = 18$

Now,

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= 4\times5 + \frac{4\times5}{2} - 27$$

$$= 20 + 10 - 27$$

$$= 3$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$= 4\times5 + \frac{5\times6}{2} - 18$$

$$= 20 + 15 - 18$$

$$= 17$$

$\therefore U_0 = 3$

Critical value : We have, $\alpha = 0.05$ and $(n_1, n_2) = (4, 5)$.
$\therefore U_{(4,5),\,0.05}$ (two-tailed test) $= 1$

Decision : Since $U_0 > U_{(4,5),\,0.05}$, we do not reject H_0.
Thus, there is no significant difference in average age of unemployed men and women.

Q18. The following are the number of minutes it took a sample of 13 men and 12 women to complete the application form for a position.

Men	16.5	20.0	17.0	19.8	18.5	19.2	19.0	18.2	20.8	18.7	16.7	18.1	17.9
Women	18.6	17.8	18.3	16.6	20.5	16.3	19.3	18.4	19.7	18.8	19.9	17.6	

Use Mann Whitney U test at 0.05 level of significance to test the hypothesis that the two samples come from identical population against the alternative that two populations are not identical.

Solution

Here, $n_1 = 13$, $n_2 = 12$
& $\alpha = 0.05$

Here, we have large number of samples.

H_0 : $Md_1 = Md_2$ i.e. two samples come from identical population.

H_1 : $Md_1 \neq Md_2$ i.e. two samples do not come from identical population. (Two-tailed test)

Calculation

Men	Women	Ranks of Men (Combined)	Ranks of Women (Combined)
16.5	18.6	2	14
20.0	17.8	23	7
17.0	18.3	5	11
19.8	16.6	21	3
18.5	20.5	13	24
19.2	16.3	18	1
19.0	19.3	17	19
18.2	18.4	10	12
20.8	19.7	25	20
18.7	18.8	15	16
16.7	19.9	4	22
18.1	17.6	9	6
17.9		8	
		$R_1 = 170$	$R_2 = 155$

Now, $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

$= 13 \times 12 + \frac{13 \times 14}{2} - 170$

$= 77$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$= 13 \times 12 + \frac{12 \times 13}{2} - 155$$

$$= 79$$

$\therefore U_0 =$ minimum of $\{77, 79\}$

$= 77.$

<u>Test-Statistic</u>: Since the problem is large sample size so

$$Z = \frac{U_0 - E(U_0)}{\sqrt{Var(U_0)}}$$

Now, $E(U_0) = \frac{n_1 n_2}{2} = \frac{13 \times 12}{2} = 78$

and $Var(U_0) = \cancel{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$

$$= \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

$$= \frac{13 \times 12 \, (13 + 12 + 1)}{12}$$

$$= 338$$

$$\therefore Z = \frac{77 - 78}{\sqrt{338}}$$

$$= \frac{-1}{18.384}$$

$$= -0.054$$

$\therefore |Z_{calc}| = 0.054$

<u>Critical value</u>: We have $\alpha = 0.05$

$Z_{tab} = Z_{0.05}$ (two-tailed test) $= 1.960$

<u>Decision</u>: Since $|Z_{calc}| < Z_{tab}$, we do not reject H_0.

Thus, the two samples of observations come from an identical population.

MEDIAN TEST

[SMALL SAMPLE]

Q10. A quality controller wishes to determine whether there is a difference in outcome between two different tools of software I and II. The following data shows the outcome of two different tools. Can the controller conclude that a difference exists? Use median test at 5% level of significance.

Software I	24.0	16.7	22.8	19.8	18.9	
Software II	23.2	19.8	18.1	17.6	20.2	17.8

Solution

Here, $n_1 = 5$, $n_2 = 6$

and $\alpha = 0.05$

H_0: $Md_1 = Md_2$ i.e. there is no difference in outcome between two different tools of software I and II

H_1: $Md_1 \neq Md_2$ i.e. there is a difference in outcome between two different tools of software I and II. (Two-tailed test)

Test statistic: Under H_0, the test statistic is 'a'.

Calculation

Arranging observations of both samples in ascending order of magnitude.

16.7, 17.6, 17.8, 18.1, 18.9, 19.8, 19.8, 20.2, 22.8, 23.2, 24.0

$$\text{Median, } Md = \text{Value of } \left(\frac{n+1}{2}\right)^{th} \text{ term}; \quad n = n_1 + n_2$$

$$= \text{Value of } \left(\frac{11+1}{2}\right)^{th} \text{ term}$$

$$= \text{value of } 6^{th} \text{ term.}$$

$\therefore$ $Md = 19.8$

Now, a = Number of observations less than or equal to Md (19.8) in the first sample

= 3

Calculation of P-value

$$P_0 = P(A \geq a) \; ; \quad k = \frac{n_1 + n_2}{2} = \frac{5+6}{2} = 6 \text{ [Rounded up]}$$

$$= \sum_{a=3}^{6} \frac{{}^{n_1}C_a \; {}^{n_2}C_{k-a}}{{}^{n_1+n_2}C_k}$$

$$= \frac{1}{{}^{11}C_6} \left[{}^5C_3 \, {}^6C_3 + {}^5C_4 \, {}^6C_2 + {}^5C_5 \, {}^6C_1 + \underbrace{{}^5C_6}_{=0} \, {}^6C_0 \right]$$

$$= \frac{281}{462}$$

$$= 0.60$$

Decision : P-value $= 2P_0 = 2 \times 0.60 = 1.20$ ($\because$ The problem is two-tailed)

and $\alpha = 0.05$

Since, P-value $> \alpha$, we do not reject H_0. Thus, there is no difference in outcome between two different tools of Software I and II.

[LARGE SAMPLE]

Q8. The length of life in Kilowatt hours of some type of electric Neon tube and Helium tube made by two manufacturers were as follows

Tube	Length of life															
Ne	96	238	24	200	7	108	76	140	39	165	61	25	41	99		
He	11	125	47	20	34	101	25	68	17	59	178	30	83	75	45	28

Compare using Median test at 5% level of significance, the median lives of the electronic tubes made by manufacturers Neon and Helium.

Solution

$n_1 = 14,$

$n_2 = 16$

$\therefore n = 14 + 16 = 30$

$\alpha = 0.05$

Here, $n_1, n_2 > 10$ so the number of samples is large.

$H_0: Md_1 = Md_2$ i.e. the median lives Neon and Helium tubes made by two manufacturers are same

$H_1: Md_1 \neq Md_2$ i.e. the median lives of Neon and Helium tubes made by two manufacturers are different. (Two tailed test)

Test statistic : Under H_0,

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Calculation

Arranging observations of both samples in ascending order of magnitude :

7, 11, 17, 20, 24, 25, 25, 28, 30, 34, 39, 41, 45, 47, 59, 61, 68, 75, 76, 83, 96, 99, 101, 108, 125, 140, 165, 178, 200, 238

Median = Value of $\left(\frac{n+1}{2}\right)^{th}$ term

$= $ value of $\left(\frac{30+1}{2}\right)^{th}$ term

$= $ value of 15.5^{th} term

$\therefore$ Median $= \frac{59+61}{2} = 60$

2X2 Contingency Table

	No. of obs $\leq Md$	No of obs $> Md$	Total
Sample I	5 (a)	9 (b)	14 (a+b)
Sample II	10 (c)	6 (d)	16 (c+d)
Total	15 (a+c)	15 (b+d)	30 (a+b+c+d)

Now, $\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$

$$= \frac{30 \times (5\times6 - 10\times9)^2}{14\times16\times15\times15}$$

$$= 2.142$$

Critical value : We have $\alpha = 0.05$ and degree of freedom (df) $=1$

$$\therefore \chi^2_{tab} = \chi^2_{1,\,0.05} = 3.841$$

Decision : Since $\chi^2_{cal} < \chi^2_{tab}$, we do not reject H_0. Thus, the median lives of Neon and Helium tubes made by two manufacturers are same.