Tutorial Letter 202/1/2015

Applied Statistics II

STA2601

Semester 1

Department of Statistics

Solutions to Assignment 02

BAR CODE

QUESTION 1

$$n = 45, \qquad \sum_{i=1}^{45} X_i = 2\,835 \qquad \text{and} \qquad \overline{x} = \frac{\sum_{i=1}^{45} X_i}{n} = 63.$$

If we order the values from small to large we have:

43	44	45	47	48	49	51	51	52	53
53	54	55	55	56	56	57	58	58	59
60	61	61	62	64	64	65	66	66	67
67	69	70	70	71	73	74	76	79	79
81	84	85	86	91					

The corresponding $\left(X_i - \overline{X}\right)$ values are:

−20	−19	−18	−16	−15	−14	−12	−12	−11	−10
−10	−9	−8	−8	−7	−7	−6	−5	−5	−4
−3	−2	−2	−1	1	1	2	3	3	4
4	6	7	7	8	10	11	13	16	16
18	21	22	23	28					

The summary statistics are:

$$\sum_{i=1}^{45} |x_i - \overline{x}| = 448; \qquad \sum_{i=1}^{45} x_i^2 = 185\,121; \qquad \sum_{i=1}^{45} (x_i - \overline{x})^2 = 6\,516;$$

$$\frac{\sum_{i=1}^{45} (x_i - \overline{x})^2}{45} = 144.8 \qquad \sum_{i=1}^{45} (x_i - \overline{x})^3 = 33\,480; \qquad \sum_{i=1}^{45} (x_i - \overline{x})^4 = 2\,271\,828.$$

(a) (i) **Test for skewness:**

H_0 : The distribution is normal $\left(\Rightarrow \beta_1 = 0\right)$.

H_1 : $\beta_1 \neq 0$.

(Please note: The alternative must be two-sided. There is no indication of a one-sided test.)

The critical value is 0.558. Reject H_0 if $\beta_1 < -0.558$ or $\beta_1 > 0.558$ or $|\beta_1| > 0.558$.

$$\text{Now } \beta_1 = \frac{\frac{1}{n}\sum_{i=1}^{45}\left(X_i - \overline{X}\right)^3}{\left[\frac{1}{n}\sum_{i=1}^{45}\left(X_i - \overline{X}\right)^2\right]^{\frac{3}{2}}} = \frac{\frac{1}{45}(33\,480)}{\left[\frac{1}{45}(6\,516)\right]^{\frac{3}{2}}}$$

$$= \frac{744}{(144.8)^{\frac{3}{2}}}$$

$$= \frac{744}{1\,742.419982}$$

$$\approx 0.4270.$$

Since $-0.558 < 0.427 < 0.558$ we do not reject H_0 at the 10% level of significance level and conclude that the distribution is symmetric. (7)

(ii) **Test for kurtosis:**

We have to test:

H_0 : The distribution is normal $\left(\Rightarrow \beta_2 = 3\right)$.

H_1 : $\beta_2 \neq 3$.

Since we have a small sample, the test is based on A (page 113 in the study guide).

The size of the sample, $n = 45$, thus $n-1 = 44$. Since 44 is between 40 and 45, we need to interpolate the critical values.

From table C (page 114 in the study guide):

The upper 5% percentage point for A is

$$0.8540 + \frac{(44-40)}{45-40}(0.8508 - 0.8540) = 0.8540 + \frac{4}{5}(-0.0032) = 0.85144.$$

The lower 5% percentage point for A is

$$0.7470 + \frac{(44-40)}{45-40}(0.7496 - 0.7470) = 0.7470 + \frac{4}{5}(0.0026) = 0.74908.$$

We reject H_0 at the 10% significance level if $A <$ lower 5% point or $A >$ upper 5% point in table C.

The critical values are 0.7491 and 0.8514. Reject H_0 if $A < 0.7491$ or $A > 0.8514$.

Now the value of the test statistic is

$$\begin{aligned}
A &= \frac{\frac{1}{n}\Sigma\left|X_i-\overline{X}\right|}{\sqrt{\frac{1}{n}\Sigma\left(X_i-\overline{X}\right)^2}} = \frac{\text{mean deviation}}{\text{standard deviation}} \\
&= \frac{\frac{1}{45}(448)}{\sqrt{\frac{1}{45}(6\,516)}} \\
&= \frac{9.955555556}{12.03328717} \\
&\approx 0.8273
\end{aligned}$$

Since $0.7491 < 0.8273 < 0.8514$, we do not reject H_0 at the 10% level of significance and conclude that the distribution does have the kurtosis of a normal distribution. (7)

(b) Yes, the distribution does originate from a normal distribution since it passed both tests. (1)

[15]

QUESTION 2

Let X denote the lifetime of a toy battery.

H_0 : X follows an exponential distribution with parameter λ.

$\lambda = \left(\frac{1}{E(X)} = \frac{1}{20} = 0.05\right)$

H_1 : X does not have an exponential distribution with $\lambda = 0.05$.

Before we can compute the expected frequency for each interval, we need to determine the probability π, for each interval a_i to b_i. We use the p.d.f. of the exponential distribution with $\lambda = 0.05$ to compute these probabilities.

$$\pi_i = \int_{a_i}^{b_i} \lambda e^{-\lambda x}dx = \left[-e^{-\lambda x}\right]_{a_i}^{b_i} = e^{-\lambda a_i} - e^{-\lambda b_i}$$

$$\begin{aligned}
\therefore P(0 < X \leq 5) &= e^{-0} - e^{-\frac{5}{20}} \\
&= 1 - e^{-0.25} \\
&= 1 - 0.77880 \\
&= 0.2212
\end{aligned}
\qquad
\begin{aligned}
P(5 < X \leq 10) &= e^{-\frac{5}{20}} - e^{-\frac{10}{20}} \\
&= e^{-0.25} - e^{-0.5} \\
&= 0.77880 - 0.60653 \\
&= 0.17227
\end{aligned}$$

$$\begin{aligned}
P(10 < X \leq 15) &= e^{-\frac{10}{20}} - e^{-\frac{15}{20}} \\
&= e^{-0.5} - e^{-0.75} \\
&= 0.60653 - 0.47237 \\
&= 0.13416
\end{aligned}
\qquad
\begin{aligned}
P(15 < X \leq 20) &= e^{-\frac{15}{20}} - e^{-\frac{20}{20}} \\
&= e^{-0.75} - e^{-1} \\
&= 0.47237 - 0.36788 \\
&= 0.10449
\end{aligned}$$

$$\begin{aligned} P(20 < X \leq 25) &= e^{-\frac{20}{20}} - e^{-\frac{25}{20}} \\ &= e^{-1} - e^{-1.25} \\ &= 0.36788 - 0.28650 \\ &= 0.08138 \end{aligned}$$

$$\begin{aligned} P(25 < X \leq 30) &= e^{-\frac{25}{20}} - e^{-\frac{30}{20}} \\ &= e^{-1.25} - e^{-1.5} \\ &= 0.28650 - 0.22313 \\ &= 0.06337 \end{aligned}$$

$$\begin{aligned} P(30 < X \leq 35) &= e^{-1.5} - e^{-1.75} \\ &= 0.22313 - 0.17377 \\ &= 0.04936 \end{aligned}$$

$$\begin{aligned} P(35 < X \leq \infty) &= e^{-1.75} - \lim_{b \longrightarrow \infty} e^{-b} \\ &= 0.17377 - 0 \\ &= 0.17377 \end{aligned}$$

We can now compute the expected number of batteries: $e_i = \pi_i \times 500$. We summarize the results in the following table:

Lifetime (in hours)	Number of batteries N_i	Expected number of batteries $e_i = \pi_i \times 500$
$t \leq 5$	108	110.6
$5 < t \leq 10$	107	86.135
$10 < t \leq 15$	90	67.08
$15 < t \leq 20$	50	52.245
$20 < t \leq 25$	45	40.69
$25 < t \leq 30$	43	31.685
$30 < t \leq 35$	32	24.68
$t > 35$	25	86.885
Total	500	500

The distribution is completely specified, so we use the test statistic

$$\begin{aligned} Y^2 &= \sum_{t=1}^{8} \frac{(N_i - e_i)^2}{e_i} \text{ which is approximately distributed } \chi^2_{k-1} \\ \therefore Y^2 &= \frac{(108 - 110.6)^2}{110.6} + \frac{(107 - 86.135)^2}{86.135} + \cdots + \frac{(25 - 86.885)^2}{86.885} \\ &= 0.0611 + 5.0543 + 7.8313 + 0.0965 + 0.4565 + 4.0407 + 2.1711 + 44.0784 \\ &= 63.7899 \end{aligned}$$

$\chi^2_{0.05;7} = 14.0671$. Reject H_0 if $Y^2 \geq 14.0671$.

Since $63.7989 > 14.0671$, we reject H_0 at the 5% level. We conclude that the data does not come from an exponential distribution with $\lambda = \dfrac{1}{20} = 0.05$.

[17]

QUESTION 3

(a) (i) Start the *JMP* program.

$>$ Enter *Attitude towards early retirement* in the first column and label it *Attitude towards early retirement*.

(make sure to change the scale to ordinal)

$>$ Enter *Index for working environment* in the second column and label it *Index for working environment*.

(make sure to change the scale to ordinal)

$>$ Enter the frequency in the third column and label it *Count*.

Your data should look like this.

Attitude towards early retirement	**Index for working environment**	**Count**
Good system	1 - 3	100
Mediocre system	1 - 3	40
Bad system	1 - 3	10
Good system	4 - 7	50
Mediocre system	4 - 7	90
Bad system	4 - 7	10
Good system	8 - 10	50
Mediocre system	8 - 10	70
Bad system	8 - 10	80

This is a chi-square test of association. To fit the model:

$>$ Choose Analyze$>$Fit Y by X with *Attitude towards early retirement* as X, Factor and *Index for working environment* as Y, Response and *Count* as $Freq$.

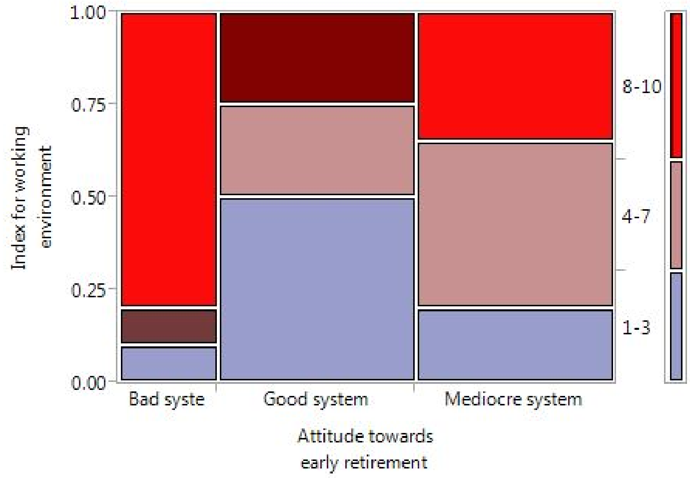
$>$ Click Ok.

Figure 1 shows the SAS JMP output with the Mosaic plot.

Contingency Analysis of Index for working environment By Attitude towards early retirement

Freq: Count

Mosaic Plot



Contingency Table

Index for working environment

Attitude towards early retirement	Count / Total % / Col % / Row %	1-3	4-7	8-10	
Bad system	Count	10	10	80	100
	Total %	2.00	2.00	16.00	20.00
	Col %	6.67	6.67	40.00	
	Row %	10.00	10.00	80.00	
Good system	Count	100	50	50	200
	Total %	20.00	10.00	10.00	40.00
	Col %	66.67	33.33	25.00	
	Row %	50.00	25.00	25.00	
Mediocre system	Count	40	90	70	200
	Total %	8.00	18.00	14.00	40.00
	Col %	26.67	60.00	35.00	
	Row %	20.00	45.00	35.00	
		150	150	200	500
		30.00	30.00	40.00	

Tests

N	DF	-LogLike	RSquare (U)
500	4	62.871890	0.1155

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	125.744	<.0001*
Pearson	129.167	<.0001*

Figure 1: Chi-square Test of Independence

(5)

(ii) The sample size for good system is the same as mediocre system and almost twice that of bad system. The mosaic plot shows that the proportion of index 4–7 was almost twice of index 1 - 3 and one-half times of index 8 - 10 for mediocre systems. For good system the proportion of index 1- 3 is almost twice that of each of the indices 4-7 and 8 - 10 respectively. For bad system the proportion for index 8 - 10 is eight times that of each of the indices 1 - 3 and 4 - 7 respectively. Thus, the proportions across index for working environment are not the same as evidenced by horizontal lines which are not in alignment. The hypothesis of no association might be rejected. (3)

(iii) H_0: The variables *attitude towards early retirement* and *index for working environment* are independent factors.

H_1: The variables *attitude towards early retirement* and *index for working environment* are not independent factors.

(2)

(iv) The test statistic is $Y^2 = \sum_{k=1}^{k} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$ and the value is $Y^2 = 129.167$. (2)

(v)

		Index for working environment			**Total**
		1-3	**4-7**	**8-10**	
Attitude towards early retirement	**Good system**	100	50	50	**200**
	Mediocre system	40	90	70	**200**
	Bad system	10	10	80	**100**
Total		**150**	**150**	**200**	**500**

The expected values are:

$E_{ij} = \frac{\text{Row } i \times \text{Column } j}{\text{Grand total}}$

For

$$\begin{aligned} E_{11} &= \frac{\text{Row } 1 \times \text{Column } 1}{\text{Grand total}} \\ &= \frac{200 \times 150}{500} \\ &= 60 \end{aligned}$$

Expected values:					
		Index for working environment			**Total**
		1-3	**4-7**	**8-10**	
Attitude towards early retirement	**Good system**	60	60	80	**200**
	Mediocre system	60	60	80	**200**
	Bad system	30	30	40	**100**
Total		**150**	**150**	**200**	**500**

Then

$$
\begin{aligned}
Y^2 &= \sum_{k=1}^{k} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \\
\therefore Y^2 &= \frac{(100-60)^2}{60} + \frac{(50-60)^2}{60} + \cdots + \frac{(80-40)^2}{40} \\
&= 26.6667 + 1.6667 + 11.25 + 6.6667 + 15 + 1.25 + 13.3333 + 13.3333 + 40 \\
&= 129.1667
\end{aligned}
$$

(7)

(vi) Yes, the row percentages are different. There are, for bad system: 10%, 10%, 80%; for good system: 50%, 25%, 25% and for mediocre system: 20%, 45%, 35% for index groups *1 - 3,, 4 - 7* and *8 - 10* respectively. They are different and only a proper test could say whether it is significant or not. One might expect the null hypothesis to be rejected. (3)

(vii) The critical value is $\chi^2_{0.05;4} = 9.48773$. Since $129.167 > 9.48773$, we reject H_0 at the 5% level of significance and conclude that *attitude towards early retirement* and *index for working environment* are not independent factors.

Alternatively looking at Figure 1 we see that the p-value is $< 0.0001 << 0.05$, so we reject H_0 at the 5% level of significance and conclude that *attitude towards early retirement* and *index for working environment* are not independent factors. The p-value is highly significant and H_0 is also being rejected at the 1% level of significance.

(4)

(b) We want to test :

H_0 : $\pi_1 = 0.49; \quad \pi_2 = 0.38; \quad \pi_3 = 0.09; \quad \pi_4 = 0.04.$

H_1 : At least one of the proportions is different from the one specified above.

The expected values are $n\pi_i$:

	Observed frequency	**Expected frequency**
O	87	$170(0.49) = 83.3$
A	59	$170(0.38) = 64.6$
B	20	$170(0.09) = 15.3$
AB	4	$170(0.04) = 6.8$
Total	**170**	**170**

The test statistic is:

$$\begin{aligned} Y^2 &= \sum_{i=1}^{4} \frac{(N_i - e_i)^2}{e_i} \\ &= \frac{(87-83.3)^2}{83.3} + \frac{(59-64.6)^2}{64.6} + \frac{(20-15.3)^2}{15.3} + \frac{(4-6.8)^2}{6.8} \\ &= 0.1643 + 0.4854 + 1.4438 + 1.1529 \\ &= 3.2464 \end{aligned}$$

$Y^2 \sim \chi^2_{\alpha;k-1}$ (see study guide p.99) and we have $k-1 = 4-1 = 3$. Thus the critical value is $\chi^2_{0.05;3} = 7.81473$. Reject H_0 if $Y^2 \geq 7.81473$.

Since the test statistic $Y^2 = 3.2464 < 7.81473$,we do not reject the null hypothesis at the 5% level. The postulate about the proportions seems to be correct. (11)

[37]

QUESTION 4

(a) H_0 : The lady has no discerning ability.

H_1 : The lady has discerning ability.

For this 2×2 table for the exact test is

		Lady says				
		Tea first	**Milk first**	**Total**		
Poured first	**Milk**	$5^* = x$	1	6	$\longleftarrow$	k
	Tea	1	5	6		
	Total	6 $\uparrow$ n	6	12	$\longrightarrow$	N

Now $k = 6$, $n = 6$ and $x = 5$

In this case

$$\begin{aligned} P(X \geq x) &= 1 - P(X < x-1) \\ P(X \geq 5) &= 1 - P(X \leq 4) \\ &= 1 - 0.96 \\ &= 0.04 \end{aligned}$$

and $P(X \leq x) = P(X \leq 5) = 0.999$.

We can only reject H_0 in favour of the two-sided alternative if x is too large or too small and if it represents a "rare event", in other words only if

$$P\left(X \le x\right) \le \frac{\alpha}{2} \text{ or if } P\left(X \ge x\right) \le \frac{\alpha}{2}$$

Since the test is two tailed we take the smaller of the two probabilities, i.e., we take 0.04. Since $0.04 > \frac{\alpha}{2} = 0.025$, we do not reject H_0 at the 5% level of significance and conclude that the lady has no discerning ability.

(7)

(b) $n = 10$ $\quad \sum X_i = 2\,013$ $\quad \sum X_i^2 = 407\,381$

$\sum X_i Y_i = 408\,745$ $\quad \sum Y_i = 2\,022$ $\quad \sum Y_i^2 = 410\,420$

$$\begin{aligned}
R &= \frac{\Sigma X_i Y_i - \dfrac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\sqrt{\left(\Sigma X_i^2 - \dfrac{(\Sigma X_i)^2}{n}\right)\left(\Sigma Y_i^2 - \dfrac{(\Sigma Y_i)^2}{n}\right)}} \\
&= \frac{408\,745 - \dfrac{(2\,013)\,(2\,022)}{10}}{\sqrt{\left(407\,381 - \dfrac{(2\,013)^2}{10}\right)\left(410\,420 - \dfrac{(2\,022)^2}{10}\right)}} \\
&= \frac{408\,745 - 407\,028.6}{\sqrt{(407\,381 - 405\,216.9)\,(410\,420 - 408\,848.4)}} \\
&= \frac{1\,716.4}{\sqrt{(2\,164.1)\,(1\,571.6)}} \\
&= \frac{1\,716.4}{\sqrt{3\,401\,099.56}} \\
&= \frac{1\,716.4}{1\,844.207027} \\
&\approx 0.9307
\end{aligned}$$

(i) The 95% confidence for η is

$$U - \frac{1.96}{\sqrt{n-3}} < \eta < U + \frac{1.96}{\sqrt{n-3}}$$

where $U = \frac{1}{2}\log_e \frac{1+R}{1-R}$ and $\eta = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho}$

Now

$$\begin{aligned}
U_1 &= \frac{1}{2}\log_e \frac{1+R}{1-R} \\
&= \frac{1}{2}\log_e \frac{1+0.9307}{1-0.9307} \\
&= \frac{1}{2}\log_e \frac{1.9307}{0.0693} \\
&= \frac{1}{2}\log_e 27.86002886 \\
&= \frac{1}{2} \times 3.327193004 \\
&\approx 1.6636
\end{aligned}$$

Now

$$\begin{aligned}
U - \frac{1.96}{\sqrt{n-3}} &< \eta < U + \frac{1.96}{\sqrt{n-3}} \\
1.6636 - \frac{1.96}{\sqrt{10-3}} &< \eta < 1.6636 + \frac{1.96}{\sqrt{10-3}} \\
1.6636 - \frac{1.96}{\sqrt{7}} &< \eta < 1.6636 + \frac{1.96}{\sqrt{7}} \\
1.6636 - 0.7408 &< \eta < 1.6636 + 0.7408 \\
0.9228 &< \eta < 2.4044
\end{aligned}$$

Now $\frac{e^{0.9228} - e^{-0.9228}}{e^{0.9228} + e^{-0.9228}} = \frac{2.5163 - 0.3974}{2.5163 + 0.3974} = \frac{2.1189}{2.9137} \approx 0.727 \approx 0.73$

and $\frac{e^{2.4044} - e^{-2.4044}}{e^{2.4044} + e^{-2.4044}} = \frac{11.0718 - 0.0903}{11.0718 + 0.0903} = \frac{10.9815}{11.1621} \approx 0.984 \approx 0.98$

i.e., 95% confidence interval for ρ is $(0.73; 0.98)$.

OR alternatively

Using Table X we have

for $\eta = 0.9076 : \rho = 0.72$ and $\eta = 0.9287 : \rho = 0.73$

Using linear interpolation for $\eta = 0.9228$

$$\begin{aligned}
\rho &= 0.72 + \frac{(0.9228 - 0.9076)}{(0.9287 - 0.9076)}(0.73 - 0.72) \\
&= 0.72 + \frac{0.0152}{0.0211} \times 0.01 \\
&= 0.72 + 0.007203791 \\
&= 0.727203791 \\
&\approx 0.73
\end{aligned}$$

for $\eta = 2.3796 : \rho = 0.983$ and $\eta = 2.4101 : \rho = 0.984$

Once more using linear interpolation for $\eta = 2.4044$

$$\begin{aligned}
\rho &= 0.983 + \frac{(2.4044 - 2.3796)}{(2.4101 - 2.3796)}(0.984 - 0.983) \\
&= 0.983 + \frac{0.0248}{0.0305} \times 0.001 \\
&= 0.983 + 0.000813114 \\
&= 0.983813114 \\
&\approx 0.98
\end{aligned}$$

Thus, the 95% confidence interval for ρ is $(0.73; 0.98)$. (5)

(ii) $H_0 : \rho = 0.9$ against $H_1 : \rho > 0.9$

$n = 10 \qquad R = -0.9307$

$$\begin{aligned}
U &= \frac{1}{2}\log_e \frac{1+R}{1-R} & \eta &= \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} \\
&= \frac{1}{2}\log_e \frac{1+0.9307}{1-0.9307} & &= \frac{1}{2}\log_e \frac{1+0.9}{1-0.9} \\
&= \frac{1}{2}\log_e \frac{1.9307}{0.0693} & &= \frac{1}{2}\log_e \frac{1.9}{0.1} \\
&= \frac{1}{2}\log_e 27.86002886 & &= \frac{1}{2}\log_e 19 \\
&\approx 1.6636 & &\approx 1.4722
\end{aligned}$$

Note: You can read the value 0f 0.9 from Table X Stoker.

The test statistic is

$$\begin{aligned} Z &= \sqrt{n-3}(U-\eta) \\ &= \sqrt{10-3}(1.6636-1.4722) \\ &= \sqrt{7}\times(0.1914) \\ &\approx 0.5064 \end{aligned}$$

$\alpha = 0.05$, and $Z_{0.05} = 1.645$. Reject H_0 if $Z > 1.645$.

Since $0.5064 < 1.645$, we do not reject H_0 at the 5% level of significance and conclude that $\rho = 0.9$.

(6)

(c) (i) Let women be the first sample and men be the second sample.

$R_1 = -0.939 \qquad n_1 = 20$
$R_2 = -0.783 \qquad n_2 = 30$

$$\begin{aligned} U_1 &= \frac{1}{2}\log_e \frac{1+R_1}{1-R_1} & U_2 &= \frac{1}{2}\log_e \frac{1+R_2}{1-R_2} \\ &= \frac{1}{2}\log_e \frac{1-0.939}{1+0.939} & &= \frac{1}{2}\log_e \frac{1-0.783}{1+0.783} \\ &= \frac{1}{2}\log_e \frac{0.061}{1.939} & &= \frac{1}{2}\log_e \frac{0.217}{1.783} \\ &= \frac{1}{2}\log_e 0.031459515 & &= \frac{1}{2}\log_e 0.121704991 \\ &\approx -1.7295 & &\approx -1.0531 \end{aligned}$$

Now $\dfrac{U_1 - U_2}{\sqrt{\dfrac{1}{n_1-3}+\dfrac{1}{n_2-3}}} \sim n\,(0;\ 1)$.

The 95% confidence for $\eta_1 - \eta_2$ is

$$P\left(-1.96 \leq \frac{U_1 - U_2 - (\eta_1 - \eta_2)}{\sqrt{\dfrac{1}{n_1-3}+\dfrac{1}{n_2-3}}} \leq 1.96\right) = 0.95$$

Then $(U_1 - U_2) - 1.96\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}} \leq \eta_1 - \eta_2 \leq (U_1 - U_2) = 1.96\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}$

Thus, $U_1 - U_2 = -1.7295 + 1.0531 = -0.6764$

Now

$$\begin{aligned}
-0.6764 - 1.96\sqrt{\frac{1}{17} + \frac{1}{27}} &\leq \eta_1 - \eta_2 \leq -0.6764 + 1.96\sqrt{\frac{1}{17} + \frac{1}{27}} \\
-0.6764 - 1.96\sqrt{0.095860566} &\leq \eta_1 - \eta_2 \leq -0.6764 + 1.96\sqrt{0.095860566} \\
-0.6764 - 0.6068 &\leq \eta_1 - \eta_2 \leq -0.6764 + 0.6068 \\
-1.2832 &\leq \eta_1 - \eta_2 \leq -0.0696
\end{aligned}$$

Now $\frac{e^{-1.2832} - e^{1.2832}}{e^{-1.2832} + e^{1.2832}} = \frac{0.2771 - 3.6082}{0.2771 + 3.6082} = \frac{-3.3311}{3.8853} \approx -0.86$

and $\frac{e^{-0.0696} - e^{0.0696}}{e^{-0.0696} + e^{0.0696}} = \frac{0.9328 - 1.0721}{0.9328 + 1.0721} = \frac{-0.1393}{2.0049} \approx -0.07$

i.e., 95% confidence interval for $\rho_1 - \rho_2$ is $(-0.86; -0.07)$.

OR alternatively

Using Table X we have

for $\eta = -1.2562 : \rho = -0.85$ and $\eta = -1.2933 : \rho = -0.86$

Using linear interpolation for $\eta = -1.2832$

$$\begin{aligned}
\rho &= -0.85 + \frac{(-1.2832 + 1.2562)}{(-1.2933 + 1.2562)}(-0.86 + 0.85) \\
&= -0.85 + \frac{(-0.027)}{(-0.0371)} \times -0.01 \\
&= -0.85 - 0.007277628 \\
&\approx -0.86
\end{aligned}$$

for $\eta = -0.0601 : \rho = -0.06$ and $\eta = -0.0701 : \rho = -0.07$

Once more using linear interpolation for $\eta = -0.0696$

$$\begin{aligned}
\rho &= -0.06 + \frac{(-0.0696 + 0.0601)}{(-0.0701 + 0.0601)}(-0.07 + 0.06) \\
&= -0.06 + \frac{(-0.0095)}{(-0.01)} \times -0.01 \\
&= -0.06 - 0.0095 \\
&\approx -0.07
\end{aligned}$$

Thus, the 95% confidence interval for $\rho_1 - \rho_2$ is $(-0.86; -0.07)$.

(10)

(ii) $H_0 : \rho_1 = \rho_2$ against $H_1 : \rho_1 \neq \rho_2$

Since 0 is not contained in the interval $(-0.86; -0.07)$, we reject H_0 at the 5% level of significance and conclude that $\rho_1 \neq \rho_{2,}$ that is, the two correlations are significantly different from each other. (3)

[31]

[**Total Marks: 100]**