Tutorial letter 204/2/2018

Applied Statistics II

STA2601

Semester 2

Department of Statistics

Solutions to Trial Examination

Define tomorrow.

UNISA university of south africa

Dear Student
This is the last tutorial letter for 2018 semester 2. I would like to take this opportunity again of wishing you well in the coming examination and I also wish you success in all your examinations.

Tutorial letters

You should have received the following tutorial letters:

Tutorial letter no.	Contents
101	General information and assignments.
102	Updated information.
103	Installation of SAS JMP 13.
104	Errata to tutorial letter 101
105	Trial paper.
201	Solutions to assignment 1.
202	Solutions to assignment 2.
203	Solutions to assignment 3.
204	Solutions to trial papers (this tutorial letter).

Some hints about the examination:

- For hypothesis testing always

(i) give the null hypothesis to be tested

(ii) calculate the test statistic to be used

(iii) give the critical region for rejection of the null hypothesis

(iv) make a decision (*reject/do not reject*)

(v) give your conclusion.

- Whenever you make a conclusion in hypothesis testing we never ever say "**we accept $\mathbf{H_0}$**." The two correct options are "**we do not reject $\mathbf{H_0}$**" or "**we reject $\mathbf{H_0}$**".

- Always show **ALL** workings and maintain **four decimal places**.

- Always specify the level of significance you have used in your decision. For example H_0 *is rejected at the* 5% *level of significance* / *we do not reject* H_0 *at the* 5% *level of significance*.

- Always determine and state the rejection criteria. For example if $F_{\text{table value}} = 3.49$. Reject H_0 if f is greater than 3.49.

- Use my presentation of the solutions as a model for what is expected from you.

Solutions of May/June 2018 Paper One Final Examination

QUESTION 1

(a) $t$distribution or the standardised normal distribution. (study guide page 30) (1)

(b) (i) $E(T) = \theta$ (study guide page 41) (1)

(ii) $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ (study guide page 50) (1)

(c) The steps are:

- Step 1: Find $L(\theta) = \prod_{i=1}^{n} f_X(X_i;\ \sigma)$
- Step 2: Find $Log\ L(\theta)$
- Step 3: Find $\dfrac{d \log L(\theta)}{d\theta}$, set it to zero and solve for θ.

(4)

[7]

QUESTION 2

(a) $E(X_1) = c_1\theta_1$; $E(X_2) = c_1\theta_1 + c_2\theta_2$

The least squares estimator is

$$
\begin{aligned}
Q(\theta_1;\theta_2) &= \sum_{i=1}^{2}\left[X_i - E(X_i)\right]^2 \\
\Longrightarrow Q(\theta_1;\theta_2) &= (X_1 - E(X_1))^2 + (X_2 - E(X_2))^2 \\
&= (X_1 - c_1\theta_1)^2 + (X_2 - c_1\theta_1 - c_2\theta_2)^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial Q(\theta_1;\theta_2)}{\partial\theta_1} &= 2(X_1 - c_1\theta_1)(-c_1) + 2(X_2 - c_1\theta_1 - c_2\theta_2)(-c_1) \\
&= -2c_1(X_1 - c_1\theta_1) - 2c_1(X_2 - c_1\theta_1 - c_2\theta_2)
\end{aligned}
$$

If we set $\dfrac{\partial Q(\theta_1;\theta_2)}{\partial \theta_1} = 0$. Now

$$\begin{aligned} 0 &= -2c_1X_1 + 2c_1^2\theta_1 - 2c_1X_2 + 2c_1^2\theta_1 + 2c_1c_2\theta_2 \\ &= -2c_1X_1 + 4c_1^2\theta_1 + 2c_1c_2\theta_2 - 2c_1X_2 \end{aligned}$$

Making θ_1 subject of the formula:

$$\begin{aligned} -4c_1^2\theta_1 &= -2c_1X_1 - 2c_1X_2 + 2c_1c_2\theta_2 \\ \widehat{\theta}_1 &= \frac{-c_1c_2\theta_2 + c_1X_1 + c_1X_2}{2c_1^2} \\ &= \frac{-c_2\theta_2 + X_1 + X_2}{2c_1} \text{.........................(1)} \end{aligned}$$

Making θ_2 subject of the formula:

$$\begin{aligned} -2c_1c_2\theta_2 &= 4c_1^2\theta_1 - 2c_1X_1 - 2c_1X_2 \\ \widehat{\theta}_2 &= \frac{-2c_1^2\theta_1 + c_1X_1 + c_1X_2}{c_1c_2} \\ &= \frac{-2c_1\theta_1 + X_1 + X_2}{c_2} \text{.........................(2)} \end{aligned}$$

$$\begin{aligned} \frac{\partial Q(\theta_1;\theta_2)}{\partial \theta_2} &= 2(X_2 - c_1\theta_1 - c_2\theta_2)(-c_2) \\ &= -2c_2(X_2 - c_1\theta_1 - c_2\theta_2) \\ &= -2c_2X_2 + 2c_1c_2\theta_1 + 2c_2^2\theta_2 \\ &= -2c_2X_2 + 2c_2^2\theta_2 + 2c_1c_2\theta_1 \end{aligned}$$

If we set $\dfrac{\partial Q(\theta_1;\theta_2)}{\partial \theta_2} = 0$. Now

$$\begin{aligned} 0 &= -2c_2X_2 + 2c_2^2\theta_2 + 2c_1c_2\theta_1 \\ &= -c_2X_2 + c_2^2\theta_2 + c_1c_2\theta_1 \end{aligned}$$

Making θ_1 subject of the formula:

$$\begin{aligned}
-c_1c_2\theta_1 &= -c_2X_2 + c_2^2\theta_2 \\
\widehat{\theta}_1 &= \frac{-c_2^2\theta_2 + c_2X_2}{c_1c_2} \\
&= \frac{-c_2\theta_2 + X_2}{c_1}\text{........................}(3)
\end{aligned}$$

Making θ_2 subject of the formula:

$$\begin{aligned}
-c_2^2\theta_2 &= -c_2X_2 + c_1c_2\theta_1 \\
\widehat{\theta}_2 &= \frac{c_2X_2 - c_1c_2\theta_1}{c_2^2} \\
&= \frac{X_2 - c_1\theta_1}{c_2}\text{........................}(4)
\end{aligned}$$

Finding θ_1 by equating equations 2 and 4:

$$\begin{aligned}
\frac{-2c_1\theta_1 + X_1 + X_2}{c_2} &= \frac{X_2 - c_1\theta_1}{c_2} \\
-2c_1\theta_1 + X_1 + X_2 &= X_2 - c_1\theta_1 \\
-2c_1\theta_1 + X_1 &= -c_1\theta_1 \\
X_1 &= -c_1\theta_1 + 2c_1\theta_1 \\
X_1 &= c_1\theta_1 \\
\Longrightarrow \widehat{\theta}_1 &= \frac{X_1}{c_1}
\end{aligned}$$

Finding θ_2 by equating equations 1 and 3:

$$\begin{aligned}
\frac{-c_2\theta_2 + X_1 + X_2}{2c_1} &= \frac{-c_2\theta_2 + X_2}{c_1} \\
-c_2\theta_2 + X_1 + X_2 &= -2c_2\theta_2 + 2X_2 \\
2c_2\theta_2 - c_2\theta_2 &= 2X_2 - X_1 - X_2 \\
c_2\theta_2 &= X_2 - X_1 \\
\Longrightarrow \widehat{\theta}_2 &= \frac{X_2 - X_1}{c_2}
\end{aligned}$$

OR

$E(X_1) = c_1\theta_1$ and $E(X_2) = c_1\theta_1 + c_2\theta_2$

In general $Q(\theta_1;\theta_2) = \sum_{i=1}^{n} [X_i - E(X_i]^2$

$$\begin{aligned}
\Longrightarrow Q(\theta_1;\theta_2) &= \sum_{i=1}^{2} [X_i - E(X_i]^2 \\
&= (X_1 - c_1\theta_1)^2 + (X_2 - c_1\theta_1 - c_2\theta_2)^2
\end{aligned}$$

To find the **least squares** estimators of θ_1 and θ_2, we need to minimize Q (see theorem 2.1).

Now

$$\begin{aligned}
\frac{\partial Q(\theta_1;\theta_2)}{\partial\theta_1} &= 2(X_1 - c_1\theta_1)(-c_1) + 2(X_2 - c_1\theta_1 - c_2\theta_2)(-c_1) \\
&= -2c_1(X_1 - c_1\theta_1 + X_2 - c_1\theta_1 - c_2\theta_2) \\
&= 2(-c_1X_1 + c_1^2\theta_1 - c_1X_2 + c_1^2\theta_1 + c_1c_2\theta_2) \\
&= 2(-c_1X_1 - c_1X_2 + 2c_1^2\theta_1 + c_1c_2\theta_2)
\end{aligned}$$

If we set $\frac{\partial Q(\theta_1;\theta_2)}{\partial\theta_1} = 0$

$$\begin{aligned}
\Longrightarrow 2(-c_1X_1 + 2c_1^2\theta_1 - c_1X_2 + c_1c_2\theta_2) &= 0 \\
\Longrightarrow 2(2c_1^2\theta_1 + c_1c_2\theta_2) &= 2c_1\,(X_1 + X_2) \\
\Longrightarrow 2c_1^2\theta_1 + c_1c_2\theta_2 &= c_1\,(X_1 + X_2) \,\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)
\end{aligned}$$

Similarly

$$\begin{aligned}
\frac{\partial Q(\theta_1;\theta_2)}{\partial\theta_2} &= 2(X_2 - c_1\theta_1 - c_2\theta_2)(-c_2) \\
&= -2c_2(X_2 - c_1\theta_1 - c_2\theta_2) \\
&= 2(-c_2X_2 + c_1c_2\theta_1 + c_2^2\theta_2)
\end{aligned}$$

If we set $\dfrac{\partial Q(\theta_1;\theta_2)}{\partial\theta_2} = 0$

$$\begin{aligned}
\Longrightarrow 2(-c_2X_2 + c_1c_2\theta_1 + c_2^2\theta_2) &= 0 \\
\Longrightarrow c_2^2\theta_2 + c_1c_2\theta_1 &= c_2X_2 \text{........................}(2)
\end{aligned}$$

From (1) (if we divide by c_1) it follows that

$$2c_1\theta_1 + c_2\theta_2 = X_1 + X_2 \text{........................}(3)$$

From (2) (if we divide by c_2) it follows that

$$c_2\theta_2 + c_1\theta_1 = X_2 \text{........................}(4)$$

If we subtract (4) from (3) we get
$$\begin{aligned}
2c_1\theta_1 + c_2\theta_2 - (c_2\theta_2 + c_1\theta_1) &= (X_1 + X_2) - X_2 \\
2c_1\theta_1 + c_2\theta_2 - c_2\theta_2 - c_1\theta_1 &= X_1 + X_2 - X_2 \\
c_1\theta_1 &= X_1 \\
\widehat{\theta}_1 &= \frac{X_1}{c_1}
\end{aligned}$$

Substituting the value of $\widehat{\theta}_1$ into (3) we get

$$\begin{aligned}
2c_1\theta_1 + c_2\theta_2 &= X_1 + X_2 \\
2c_1\left(\frac{X_1}{c_1}\right) + c_2\theta_2 &= X_1 + X_2 \\
2X_1 + c_2\theta_2 &= X_1 + X_2 \\
c_2\theta_2 &= X_1 + X_2 - 2X_1 \\
c_2\theta_2 &= X_2 - X_1 \\
\widehat{\theta}_2 &= \frac{X_2 - X_1}{c_2}
\end{aligned}$$

(7)

(b)

$$\begin{aligned} A_1 &= \frac{1}{n}\sum_{i=1}^{n}\left[X_i-\overline{X}\right]^2 \\ E(A_1) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\left[X_i-\overline{X}\right]^2\right) \\ &= E\left(\frac{\sigma^2}{n}\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2\right) \text{ (multiplying both sides by } \sigma^2\text{)} \end{aligned}$$

Note: $\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2 \sim \chi^2_{n-1} \Longrightarrow E\left(\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2\right) = n-1.$

$$\begin{aligned} \Longrightarrow E(A_1) &= \frac{\sigma^2}{n}E\left(\sum_{i=1}^{n}\left[\frac{X_i-\overline{X}}{\sigma}\right]^2\right) \\ &= \frac{\sigma^2}{n}\cdot(n-1) \\ &= \frac{(n-1)\,\sigma^2}{n} \end{aligned}$$

Or

$$\begin{aligned} E(A_1) &= \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2-2\sum_{i=1}^{n}X_i\overline{X}+\overline{X}^2\right) \\ &= \frac{1}{n}E\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}E\left(X_i^2\right)-E(\overline{X})^2 \end{aligned}$$

Now

$$\begin{aligned} Var(X_i) &= \left(E\left(X_i^2\right)-(E(X_i))^2\right) \\ \sigma^2 &= E\left(X_i^2\right)-\mu^2 \\ \Longrightarrow E\left(X_i^2\right) &= \sigma^2+\mu^2 \end{aligned}$$

$$\begin{aligned} Var(\overline{X}^2) &= E\left(\overline{X}^2\right)-\left(E(\overline{X})\right)^2 \\ \frac{\sigma^2}{n} &= E\left(\overline{X}^2\right)-\mu^2 \\ \Longrightarrow E\left(\overline{X}^2\right) &= \frac{\sigma^2}{n}+\mu^2 \end{aligned}$$

$$\begin{aligned}
\Longrightarrow E(A_1) &= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2+\mu^2\right)-\left(\frac{\sigma^2}{n}+\mu^2\right)\\
&= \frac{1}{n}\left(n\left(\sigma^2+\mu^2\right)-\frac{\sigma^2}{n}-\mu^2\right)\\
&= \sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2\\
&= \sigma^2-\frac{\sigma^2}{n}\\
&= \frac{n\sigma^2-\sigma^2}{n}\\
&= \frac{\sigma^2(n-1)}{n}
\end{aligned}$$

(5)

[12]

QUESTION 3

(a) No, data is slightly negatively skewed.

OR

Yes, if you fit a normal curve its almost symmetric. [Commenting of this plot is subjective]

(2)

(b) We need to pool, the first three and the last three classes since $e_i > 5$

$$\begin{aligned}
Y^2 &= \sum_{i=1}^{6}\frac{(N_i-\widehat{e_i})^2}{\widehat{e_i}}\\
&= \frac{(4-5)^2}{5}+\frac{(14-12)^2}{12}+\frac{(17-24)^2}{24}+\frac{(31-28)^2}{28}+\frac{(24-20)^2}{20}+\frac{(10-11)^2}{11}\\
&= 0.2+0.3333+2.0417+0.3214+0.8+0.0909\\
&= 3.7873
\end{aligned}$$

H_0 : The sample comes from a $n(20, 4)$ distribution
H_1 : The sample does not come from a $n(20, 4)$ distribution

$k = 6, r = 0,\ \chi^2_{\alpha;k-r-1} = \chi^2_{0.05;5} = 11.0705$. Reject H_0 if $Y^2 > 11.0705$.

Since $3.7873 < 11.0705$, we do not reject H_0 at the 5% level of significance and conclude that the data originates from a normal distribution.

(6)

(c)

$$\begin{aligned}
\beta_1 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^3}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{\frac{3}{2}}} \\
&= \frac{\frac{1}{100}(-282.3)}{\left[\frac{1}{100}(426.4)\right]^{\frac{3}{2}}} \\
&= \frac{-2.823}{[4.264]^{\frac{3}{2}}} \\
&= \frac{-2.823}{8.804927697} \\
&\approx -0.3206
\end{aligned}$$

$$\begin{aligned}
\beta_2 &= \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^4}{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]^{2}} \\
&= \frac{\frac{1}{100}(6\,958)}{\left[\frac{1}{100}(426.4)\right]^{2}} \\
&= \frac{69.58}{[4.264]^{2}} \\
&= \frac{69.58}{18.181696} \\
&\approx 3.8269
\end{aligned}$$

Testing for skewness:

We have to test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$

The critical value at 10% level is 0.389. We will reject H_0 if $\beta_1 < -0.389$ or if $\beta_1 > 0.389$ in other words if $|\beta_1| > 0.389$.

Since $|-0.3206| < 0.389$, we do not reject H_0 at the 10% level of significance and conclude that the sample has a skewness of a normal distribution

Testing for kurtosis:

We have to test $H_0 : \beta_2 = 3$ against $H_1 : \beta_2 \neq 3$

The critical values are 3.77 and 2.35. We will reject H_0 if $\beta_2 > 3.77$ or $\beta_2 < 2.35$ (two sided hypothesis).

Since $3.8269 > 3.77$, we reject H_0 in favour of H_1 at the 10% level of significance and conclude that the kurtosis of the sample is significantly different from the kurtosis of the normal population.

In conclusion, the sample failed one-test and hence we conclude that the sample is not from a normal population.

(14)

(d) In (b) the test was testing whether data was normally distributed with the parameters $\mu = 20$ and $\sigma^2 = 4$ while in (c) the test was just testing for normality. In this case the results were different as the second test showed that the data does not have the kurtosis of a normal distribution.

(3)

[25]

QUESTION 4

(a) H_0 : There is no relationship between cell phone model and degree of awareness.

H_1 : There is a relationship between cell phone model and degree of awareness..

For this 2×2 table for the exact test is

		Degree of awareness			
		Yes	**No**	**Row total**	
Cell phone model	**Method 1**	6	1	7	
	Method 2	1	$4(= x)$	5	$\leftarrow n$
	Column total	7	5	12	
			$\uparrow$ k		

Now $k = 5$, $n = 5$ and $x = 4$

In this case

$$\begin{aligned} P\left(X \geq x\right) &= 1 - P(X < x - 1) \\ P(X \geq 4) &= 1 - P(X \leq 3) \\ &= 1 - 0.955 \\ &= 0.045 \text{ (from table D study guide p131)} \end{aligned}$$

and $P(X \leq x) = P(X \leq 4) = 0.999$.

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

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

Since the test is two tailed we take the smaller of the two probabilities, i.e., we take 0.045. Since $0.045 > \frac{\alpha}{2} = 0.025$, we do not reject H_0 at the 5% level of significance and conclude that there is no association between cell phone model and degree of awareness, that is, customers are equally aware of models 1 and 2.

(9)

(b) (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

$$\begin{aligned} U &= \frac{1}{2}\log_e \frac{1+R}{1-R} \\ &= \frac{1}{2}\log_e \frac{1+0.2}{1-0.2} \\ &= \frac{1}{2}\log_e \frac{1.2}{0.8} \\ &= \frac{1}{2}\log_e 1.5 \\ &= \frac{1}{2} \times 0.405465108 \\ &\approx 0.2027 \end{aligned}$$

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

Now

$$\begin{aligned}
U - \frac{1.96}{\sqrt{n-3}} &< \eta < U + \frac{1.96}{\sqrt{n-3}} \\
0.2027 - \frac{1.96}{\sqrt{39-3}} &< \eta < 0.2027 + \frac{1.96}{\sqrt{39-3}} \\
0.2027 - \frac{1.96}{\sqrt{36}} &< \eta < 0.2027 + \frac{1.96}{\sqrt{36}} \\
0.2027 - 0.3267 &< \eta < 0.2027 + 0.3267 \\
-0.124 &< \eta < 0.5294
\end{aligned}$$

Now $\dfrac{e^{-0.124} - e^{0.124}}{e^{-0.124} + e^{0.124}} = \dfrac{0.8834 - 1.1320}{0.8834 + 1.1320} = \dfrac{-0.2486}{2.0154} \approx -0.1234 \approx -0.12$

and $\dfrac{e^{0.5294} - e^{-0.5294}}{e^{0.5294} + e^{-0.5294}} = \dfrac{1.6979 - 0.5890}{1.6979 + 0.5890} = \dfrac{1.1089}{2.2869} \approx 0.4849 \approx 0.48$

i.e., 95% confidence interval for ρ is $(-0.12; 0.48)$.

OR alternatively

Using Table X we have

for $\eta = -0.1206 : \rho = 0.12$ and $\eta = -0.1307 : \rho = -0.13$

Using linear interpolation for $\eta = -0.124$

$$\begin{aligned}
\rho &= -0.12 + \frac{(-0.124 + 0.1206)}{(-0.1307 + 0.1206)}(-0.13 + 0.12) \\
&= -0.12 + \frac{-0.0034}{-0.0101} \times -0.01 \\
&= -0.12 - 0.003366336 \\
&= -0.12336336 \\
&\approx -0.12
\end{aligned}$$

for $\eta = 0.5230 : \rho = 0.48$ and $\eta = 0.5361 : \rho = 0.49$

Once more using linear interpolation for $\eta = 0.5294$

$$\begin{aligned}
\rho &= 0.48 + \frac{(0.5294 - 0.5230)}{(0.5361 - 0.5230)}(0.49 - 0.48) \\
&= 0.48 + \frac{0.0064}{0.0131} \times 0.01 \\
&= 0.48 + 0.004885496 \\
&= 0.484885496 \\
&\approx 0.48
\end{aligned}$$

Thus, the 95% confidence interval for ρ is $(-0.12; 0.48)$.

(8)

(ii) If 0 is contained in the interval then we do not reject $H_0 : \rho = 0$ at the 5% level of significance. (2)

[19]

QUESTION 5

(a) Yes, it is reasonable to assume that the two groups may be considered as independent groups because the blowout times of a tyre may not influence the blowout time tensile of another tyre. (2)

(b)

$$\begin{aligned} S_X^2 &= \frac{1}{n_1 - 1}\sum\left(X_{1j} - \overline{X}_2\right)^2 & S_Y^2 &= \frac{1}{n_2 - 1}\sum\left(Y_{2j} - \overline{Y}_2\right)^2 \\ &= \frac{1}{25 - 1}(11.0976) & &= \frac{1}{49 - 1}(11.5248) \\ &= \frac{1}{24}(11.0976) & &= \frac{1}{48}(11.5248) \\ &= 0.4624 & &= 0.2401 \end{aligned}$$

We have to test $H_0 : \sigma_X^2 = \sigma_Y^2$

against $H_1 : \sigma_X^2 \neq \sigma_Y^2$

The test statistic is

$$\begin{aligned} F &= \frac{\sigma_Y^2}{\sigma_X^2} \times \frac{S_X^2}{S_Y^2} \\ &= 1 \times \frac{0.4624}{0.2401} \\ &\approx 1.9259 \end{aligned}$$

The critical values is $F_{\alpha/2;n_1-1;n_2-1} = F_{0.025;24;48} \approx 1.94$ and $F_{1-\alpha/2;n_1-1;n_2-1} = \dfrac{1}{F_{\alpha/2;n_2-1;n_1-1}} = \dfrac{1}{F_{0.025;48;24}}$.

$$\begin{aligned} F_{0.025;48;24} &= 2.15 + \frac{8}{20}(2.08 - 2.15) \\ &= 2.15 + 0.4(-0.07) \\ &= 2.15 - 0.028 \\ &= 2.122 \end{aligned}$$

Now $\dfrac{1}{F_{0.025;48;24}} = \dfrac{1}{2.122} \approx 0.4713$. Reject H_0 if $F > 1.94$ or $F < 0.47$.

Since $0.47 < 1.9259 < 1.94$, we do not reject H_0 at the 5% level of significance and conclude that the variances are equal i.e. $\sigma_X^2 = \sigma_Y^2$.

(9)

(c) $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X < \mu_Y$

$$n_x = 25 \qquad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{83}{25} = 3.32 \qquad S_X^2 = 0.4624$$

$$n_y = 49 \qquad \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{196}{49} = 4 \qquad S_Y^2 = 0.2401$$

The test statistic is

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{nx} + \frac{1}{ny}}}$$

Now

$$\begin{aligned} S_p^2 &= \frac{(n_x - 1)S_X^2 + (n_y - 1)S_Y^2}{n_1 + n_2 - 2} \\ &= \frac{\sum (X_i - X)^2 + \sum \left(Y_i - \bar{Y}\right)^2}{25 + 49 - 2} \\ &= \frac{11.0976 + 11.5248}{25 + 49 - 2} \\ &= \frac{22.6224}{72} \\ &\approx 0.3142 \\ \Longrightarrow\ & S_{pooled} = \sqrt{0.3142} \approx 0.5605 \end{aligned}$$

Then

$$\begin{aligned}
T &= \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{S_p\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \\
&= \frac{(3.32-4)-(0)}{0.5605\sqrt{\frac{1}{25}+\frac{1}{49}}} \\
&= \frac{-0.68}{0.5605\sqrt{0.060408163}} \\
&= \frac{-0.68}{0.137760094} \\
&\approx -4.9361
\end{aligned}$$

The critical value is

$$\begin{aligned}
t_{\alpha;n_1+n_2-2} &= t_{0.05;72} \\
&= 1.671+\frac{12}{40}(1.660-1.671) \\
&= 1.671+0.3(-0.011) \\
&= 1.671-0.0033 \\
&\approx 1.668
\end{aligned}$$

Reject H_0 if $T < -1.668$.

Since $-4.9361 < -1.668$, we reject H_0 at the 5% level and conclude that the claim is true. The mean blowout time for tyres of Brand B is significantly higher than the mean blowout time for the tyres of Brand A., i.e., $\mu_X < \mu_Y$.

(7)

(d) In order to perform the test in (c) we assumed that:

- the observations in each sample are independent and also the two samples are mutually independent.
- the observations are normally distributed.
- the two population variances are equal.

(3)

[21]

QUESTION 6

(a) Consider the simple linear regression $\widehat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X$

Then

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum y_i\,(x_i - \bar{x})}{d^2} \\ &= \frac{180}{60} \\ &= 3\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1\overline{x} \\ &= \frac{276}{12} - 3\left(\frac{48}{12}\right) \\ &= 23 - 3\,(4) \\ &= 23 - 12 \\ &= 11\end{aligned}$$

The estimated regression line of y on x is $\widehat{\text{Yield}} = 11 + 3\text{Dose}$.

(7)

(b) The confidence interval is

$$\widehat{\beta}_1 \pm t_{\alpha/2;n-2} \times \frac{s}{d}$$

$\widehat{\beta}_1 = 3$ $\qquad t_{\alpha/2;n-2} = t_{0.025;10} = 2.228$

$d = \sqrt{60}$ $\qquad s = \sqrt{9} = 3$

The 95% confidence interval for $\widehat{\beta}_1$ is

$$\begin{array}{lcl} \widehat{\beta}_1 & \pm & t_{\alpha/2;n-2} \times \frac{s}{d} \\ 3 & \pm & 2.228 \times \frac{3}{\sqrt{60}} \\ 3 & \pm & 0.8629 \\ (3 - 0.8629 & ; & 3 + 0.8629 \\ (2.1371 & ; & 3.8629) \end{array}$$

(4)

(c) $x = 4$

The expected yield is

$$\begin{aligned}\widehat{\text{Yield}} &= 11 + 3\text{Dose} \\ &= 11 + 3(4) \\ &= 11 + 12 \\ &= 23\end{aligned}$$

(1)

(d) The confidence interval is $\left(\widehat{\beta}_0 + \widehat{\beta}_1 X\right) \pm t_{\alpha/2;n-2} \times S\sqrt{1 + \frac{1}{n} + \frac{\left(X - \overline{X}\right)^2}{d^2}}$.

Now

$$\begin{aligned}SE &= S\sqrt{1 + \frac{1}{n} + \frac{\left(X_i - \overline{X}\right)^2}{d^2}} \\ &= 3\sqrt{1 + \frac{1}{12} + \frac{(4-4)^2}{60}} \\ &= 3\sqrt{1 + \frac{1}{12}} \\ &= 3\sqrt{1.083333333} \\ &\approx 3.1225\end{aligned}$$

The 95% confidence interval for the expected yield of a new observation at $x = 4$ is

$$\begin{array}{lcl} \widehat{\beta}_0 + \widehat{\beta}_1 X & \pm & t_{\alpha/2;n-2} \times S\sqrt{1 + \frac{1}{n} + \frac{\left(X - \overline{X}\right)^2}{d^2}} \\ 23 & \pm & 2.228 \times 3.1225 \\ 23 & \pm & 6.9569 \\ (23 - 6.9569) & ; & 23 + 6.9569) \\ (16.0431 & ; & 29.9569) \end{array}$$

(4)

[16]

[100]

Solutions of May/June 2018 Paper Two Final Examination

QUESTION 1

(a) The statistic T is called an *unbiased estimator* for the parameter θ if $E(T) = \theta$. (study guide page 41) (1)

(b) The *significance level of a test* (probability of committing a Type I error) which is α, is the probability of rejecting the null hypothesis when in fact it is true, that is, α is called the significance level of t rejected| H_0 is true$) = \alpha$. (study guide page 56) (2)

(c) The *power of the test* is the probability, $1-\beta$, where $P($not rejecting $H_0 \mid H_1$ is true$) = \beta$ (study guide page 57) (2)

(d) The methods are least squares estimation and maximum likelihood estimation. (study guide page 44) (2)

(e) The methods are a chi-square goodness-of-fit test, test for skewness and kurtosis and using a normal quantile plot. (study guide page 119) (3)

[10]

QUESTION 2

$E(X_i) = \theta_1, \qquad i = 1, \ldots, (n-1) \qquad E(X_n) = \theta_1 + \theta_2.$

The least squares estimator is

$$\begin{aligned} Q(\theta) &= \sum_{i=1}^{n} (X_i - E(X_i))^2 \\ &= \sum_{i=1}^{n-1} (X_i - E(X_i))^2 + (X_n - E(X_n))^2 \\ &= \sum_{i=1}^{n-1} (X_i - \theta_1)^2 + (X_n - (\theta_1 + \theta_2))^2 \\ &= \sum_{i=1}^{n-1} (X_i - \theta_1)^2 + (X_n - \theta_1 - \theta_2)^2 \end{aligned}$$

$$\begin{aligned}
\frac{dQ}{d\theta_1} &= 2\sum_{i=1}^{n-1}(X_i-\theta_1)\times -1 + 2(X_n-\theta_1-\theta_2)\times -1 \\
&= -2\left(\sum_{i=1}^{n-1}(X_i-\theta_1) + X_n-\theta_1-\theta_2\right) \\
0 &= -2\left(\sum_{i=1}^{n-1}(X_i-\theta_1) + X_n-\theta_1-\theta_2\right) \\
0 &= \sum_{i=1}^{n-1}(X_i-\theta_1) + X_n-\theta_1-\theta_2 \\
0 &= \sum_{i=1}^{n-1}X_i - \sum_{i=1}^{n-1}\theta_1 + X_n-\theta_1-\theta_2 \\
0 &= \sum_{i=1}^{n-1}X_i - (n-1)\theta_1 - \theta_1 + X_n-\theta_2 \\
0 &= \sum_{i=1}^{n-1}X_i - n\theta_1 + \theta_1 - \theta_1 + X_n-\theta_2 \\
0 &= \sum_{i=1}^{n-1}X_i - n\theta_1 + X_n-\theta_2 \\
0 &= \sum_{i=1}^{n}X_i - n\theta_1 - \theta_2 \\
&\Longrightarrow \widehat{\theta}_1 = \frac{\sum_{i=1}^{n}X_i-\theta_2}{n}\text{..................(1)} \\
\widehat{\theta}_2 &= \sum_{i=1}^{n}X_i - n\theta_1\text{.........................(2)}
\end{aligned}$$

$$\begin{aligned}
\frac{dQ}{d\theta_2} &= 2(X_n-\theta_1-\theta_2)\times -1 \\
0 &= -2(X_n-\theta_1-\theta_2) \\
0 &= X_n-\theta_1-\theta_2 \\
&\Longrightarrow \theta_1 = X_n-\theta_2\text{......................(3)} \\
&\Longrightarrow \theta_2 = X_n-\theta_1\text{......................(4)}
\end{aligned}$$

Equating 2 and 4

$$X_n-\theta_1 = \sum_{i=1}^{n}X_i - n\theta_1$$

$$\begin{aligned}
n\theta_1 - \theta_1 &= \sum_{i=1}^{n} X_i - X_n \\
\theta_1 (n-1) &= \sum_{i=1}^{n-1} X_i + X_n - X_n \\
\theta_1 (n-1) &= \sum_{i=1}^{n-1} X_i \\
\theta_1 &= \frac{\sum_{i=1}^{n-1} X_i}{n-1} \\
\theta_1 &= \overline{X}
\end{aligned}$$

Equating 1 and 3

$$\begin{aligned}
\frac{\sum_{i=1}^{n} X_i - \theta_2}{n} &= X_n - \theta_2 \\
\sum_{i=1}^{n} X_i - \theta_2 &= nX_n - n\theta_2 \\
\sum_{i=1}^{n} X_i - nX_n &= \theta_2 - n\theta_2 \\
\sum_{i=1}^{n} X_i - nX_n &= \theta_2 (1-n) \\
nX_n - \sum_{i=1}^{n} X_i &= \theta_2 (n-1) \\
nX_n - X_n - \sum_{i=1}^{n-1} X_i &= \theta_2 (n-1) \\
X_n (n-1) - \sum_{i=1}^{n-1} X_i &= \theta_2 (n-1) \\
X_n - \frac{\sum_{i=1}^{n-1} X_i}{n-1} &= \theta_2 \\
X_n - \overline{X} &= \theta_2 \\
&\Longrightarrow \widehat{\theta}_2 = X_n - \overline{X}
\end{aligned}$$

[8]

QUESTION 3

(a) We have to test

H_0 : The observations come from a normal distribution.

H_1 : The observations do not come from a normal distribution.

(2)

(b) $n = 30$ $\sum_{i=1}^{n} X_i = 41\,400,$

$$\widehat{\mu} = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{41\,400}{30} = 1\,380$$

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} \\ &= \frac{4\,623\,074}{30} \\ &= 154\,102.4667 \end{aligned}$$

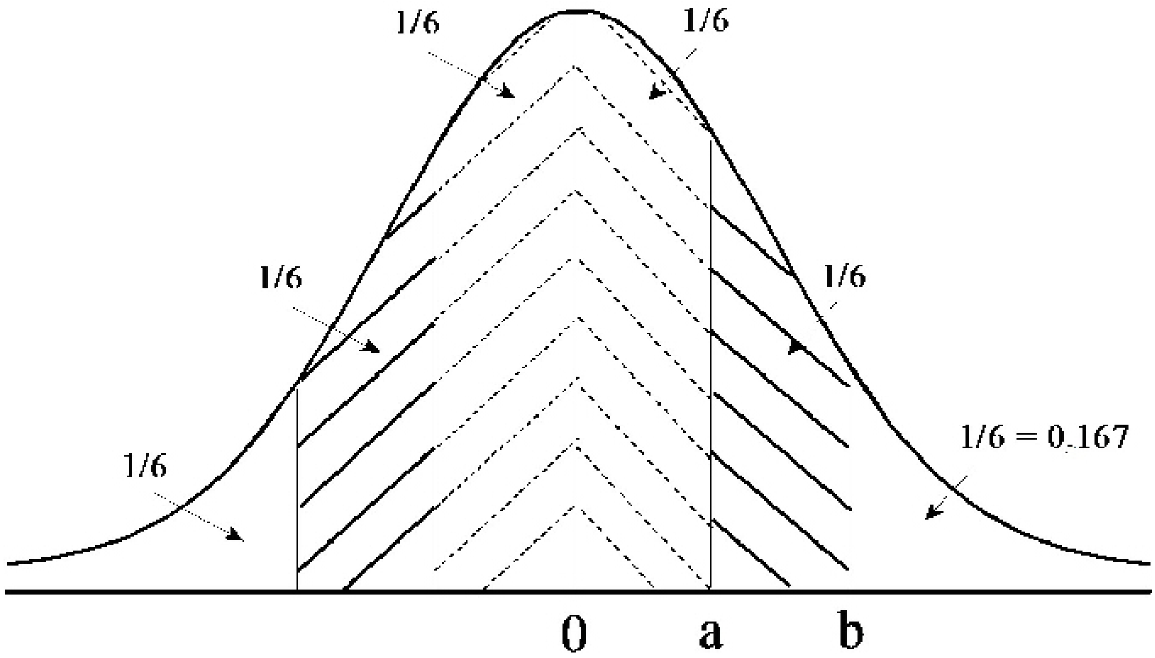
(5)

(c) If we divide the observations into 6 classes with equal expected frequencies, it means that $\pi_i = \frac{1}{6}$ for each interval $\Rightarrow n\pi_i = 5$.

The biggest problem is to *determine the interval limits* in terms of the X-scale such that each interval has a probability of $\frac{1}{6} = 0.167$.

We start with the standardised $n\,(0;\,1)$ scale (as always) and transform back to the X-scale by making use of

$$Z = \frac{X - \widehat{\mu}}{\widehat{\sigma}} = \frac{X - 1\,380}{\sqrt{154\,102.4667}} = \frac{X - 1\,380}{392.5589}.$$



From the first interval we have $P\,[Z \leq -b] = 0.167$ and we find b by working with the positive mirror image. i.e., $P(Z \geq b) = 0.167 \Longrightarrow P(Z \leq b) = 0.833$ (this probability is $1 - 0.167$).

From table II we get that $\Phi\,(0.966) = 0.833 \Rightarrow b = 0.966$.

The first interval is where $Z \leq -0.966$ [Note that P$(Z \leq -0.966) = 0.167$.

$$\begin{aligned}
\Longrightarrow \frac{X-\mu}{\sigma} = \frac{X - 1\,380}{392.5589} &\leq -0.966 \\
X - 1\,380 &\leq -0.966 \times 392.5589 \\
X - 1\,380 &\leq -379.2119 \\
X &\leq 1\,380 - 379.2119 \\
X &\leq 1\,000.7881 \\
X &\leq 1\,000.79
\end{aligned}$$

Hence the first interval is where $\infty \leq X \leq 1\,000.79$.

(4)

(d)

$$
\begin{aligned}
Y^2 &= \sum_{k=1}^{k} \frac{\left(\text{Observed} - \text{Expected}\right)^2}{\text{Expected}} \\
&= \frac{(6-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(5-5)^2}{5} \\
&= \frac{1}{5} + 0 + 0 + \frac{1}{5} + 0 + 0 \\
&= \frac{2}{5} \\
&= 0.4
\end{aligned}
$$

From table IV (Stoker) $\chi^2_{\alpha;k-1} = \chi^2_{0.10;5} = 9.23635$ and $\chi^2_{\alpha;k-r-1} = \chi^2_{0.10;3} = 6.25139$. Reject H_0 if $Y^2 > 6.25139$.

Since $0.4 < 6.25139$, we cannot reject H_0 at the 10% level of significance. We may conclude that the sample comes from a normal distribution.

FOR YOUR INFORMATION: For the degrees of freedom $k - r - 1$ where k is the number of classes and r is the number of estimated parameters.

Test	Value of r	Parameter unknown
$n\left(20, 5^2\right)$	$r = 0$	
$n\left(20, \sigma^2\right)$	$r = 1$	σ unknown
$n\left(\mu, 5^2\right)$	$r = 1$	μ unknown
$n\left(\mu, \sigma^2\right)$	$r = 2$	both μ and σ unknown

(6)

(e) (i) Testing $H_0 : \mu = 1\,300$ against $H_1 : \quad \mu > 1\,300$

<u>**Method I: Using the critical value approach**</u>:

From the output the test statistic is

$$
t = \frac{\sqrt{n}\left(\bar{x} - \mu_0\right)}{s} = \frac{\sqrt{30}\,(1\,380 - 1\,300)}{399.26976} \approx 1.0974
$$

The critical value is $t_{\alpha;n-1} = t_{0.025;29} = 2.045$.

We will reject H_0 if $T \geq 2.045$.

Since $1.0974 < 2.045$, we do not reject H_0 at the 2.5% level of significance and conclude that the mean service life is $1\,300$ days, i.e., $\mu = 1\,300$. Thus, the new design is not superior to the standard design with respect to mean service life.

Method II: Using the p-value approach

p-value $= 0.1407$. Since $0.1407 > 0.05$, we do not reject H_0 at the 2.5% level of significance and conclude that the mean service life is $1\,300$ days, i.e., $\mu = 1\,300$. Thus, the new design is not superior to the standard design with respect to mean service life.

(4)

(ii) The 95% two-sided confidence interval is computed as $\overline{X} \pm t_{0.025;29}\left(\frac{S}{\sqrt{n}}\right)$

$\overline{X} = 1\,380 \qquad S = 399.26976 \qquad t_{0.025;29} = 2.045$

Now

$$\begin{array}{lcl} \overline{X} & \pm & t_{\alpha/2;n-1} \times \frac{S}{\sqrt{n}} \\ 1\,380 & \pm & 2.045 \times \frac{399.26976}{\sqrt{30}} \\ 1\,380 & \pm & 149.0730 \\ (1\,380 - 149.0730 & ; & 1\,380 + 149.0730) \\ (1\,230.927 & & 1\,529.0730) \end{array}$$

We are 95% confident that the mean service life, μ, of batteries of the new design is $1\,230.9$ to $1\,529.1$

(5)

(iii) Let $U = \frac{\Sigma\left(X_i - \overline{X}\right)^2}{\sigma^2}$ then $U \sim \chi^2_{n-1}$ (unknown value of μ).. The 95% two-sided confidence interval for σ^2 now becomes $\left[\frac{\Sigma\left(X_i - \overline{X}\right)^2}{\chi^2_{\frac{1}{2}\alpha;n-1}} < \sigma^2 < \frac{\Sigma\left(X_i - \overline{X}\right)^2}{\chi^2_{1-\frac{1}{2}\alpha;n-1}}\right]$

$$\begin{array}{lclcl} \chi^2_{\frac{1}{2}\alpha;n-1} & = & \chi^2_{0.025;29} & = & 45.7222 \\ \chi^2_{1-\frac{1}{2}\alpha;n-1} & = & \chi^2_{0.975;29} & = & 16.0471 \end{array}$$

Thus, the 95% two-sided confidence interval for σ^2 now becomes

$$\left[\frac{\Sigma\left(X_i-\overline{X}\right)^2}{\chi^2_{\frac{1}{2}\alpha;n-1}}<\sigma^2<\frac{\Sigma\left(X_i-\overline{X}\right)^2}{\chi^2_{1-\frac{1}{2}\alpha;n-1}}\right]$$
$$\left[\frac{4\,623\,074}{45.7222}<\sigma^2<\frac{4\,623\,074}{16.0471}\right]$$
$$\left[101\,112.2387<\sigma^2<288\,094.0481\right]$$
$$\left[\sqrt{101\,112.2387}<\sigma<\sqrt{288\,094.0481}\right]$$
$$[317.98<\sigma<536.74]\,.$$

(5)

(iv) The assumption made was that the mean μ is unknown and hence the test statistic $U=\dfrac{\Sigma\left(X_i-\overline{X}\right)^2}{\sigma^2}$ was used at χ^2_{n-1}. (1)

[32]

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

$$P(X \leq x) \leq \frac{\alpha}{2} \text{ or if } P(X \geq x) \leq \frac{\alpha}{2}$$

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.

(9)

(b) (i) Yes. These are two measurements taken from the same observation. (1)

(ii) We have to test:

$H_0: \quad \mu_d = 0$ against

$H_1: \quad \mu_d < 0$

Method 1: Using the critical value approach

From the output the test statistics is

$$t = \frac{(\overline{x} - \mu_0)}{\frac{s}{\sqrt{n}}} = \frac{(-18.933 - 0)}{2.33068} \approx -8.12282$$

The critical value is $t_{\alpha;n-1} = t_{0.05;14} = 1.761$. Reject H_0 if $T \leq -1.761$

Since $-8.12282 < -1.761$, we reject H_0 in favour of H_1 at the 5% level of significance and conclude that $\mu_d < 0$, that is, the drug company's claim is true. Thus the drug lowers blood pressure.

Method II: Using the p-value approach

p-value $= 0.0001$. Since $0.0001 < 0.05$, we reject H_0 in favour of H_1 at the 5% level of significance and conclude that $\mu_d < 0$, that is, the drug company's claim is true. Thus the drug lowers blood pressure.

(5)

(c) We are testing $H_0 : \mu = 30$ against $H_1 : \mu \neq 30$.

The power of the test is a function of Φ which is defined as $\Phi = \frac{\delta}{\sqrt{2}}$

$$\begin{aligned}
\delta &= \frac{\sqrt{n}(\mu - \mu_0)}{\sigma} \\
&= \frac{\sqrt{10}(30 + \sqrt{2}\,\sigma - 30)}{\sigma} \\
&= \sqrt{10}\sqrt{2} \\
\Longrightarrow \Phi = \frac{\delta}{\sqrt{2}} &= \frac{\sqrt{10}\sqrt{2}}{\sqrt{2}} \\
&\approx 3.2
\end{aligned}$$

F**rom table F** we read of the power as 98% (i.e., $1 - \beta = 0.98$) (4)

(d) $H_0 : \rho_1 = \rho_2$ against $H_1 : \rho_1 < \rho_2$

$r_1 = 0.73$ $n_1 = 12$
$r_2 = 0.89$ $n_2 = 20$

$$\begin{aligned}
U_1 &= \frac{1}{2}\log_e \frac{1 + r_1}{1 - r_1} & U_2 &= \frac{1}{2}\log_e \frac{1 + r_2}{1 - r_2} \\
&= \frac{1}{2}\log_e \frac{1 + 0.73}{1 - 0.73} & &= \frac{1}{2}\log_e \frac{1 + 0.89}{1 - 0.89} \\
&= \frac{1}{2}\log_e \frac{1.73}{0.27} & &= \frac{1}{2}\log_e \frac{1.89}{0.11} \\
&= \frac{1}{2}\log_e 6.407407407 & &= \frac{1}{2}\log_e 17.18181818 \\
&\approx 0.9287 & &\approx 1.4219
\end{aligned}$$

(or just read the values for U_1 and U_2 from table X)

The test statistic is

$$\begin{aligned}
z &= \frac{U_1 - U_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \\
&= \frac{0.9287 - 1.4219}{\sqrt{\frac{1}{12 - 3} + \frac{1}{20 - 3}}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-0.4932}{\sqrt{\frac{1}{9}+\frac{1}{17}}} \\
&= \frac{-0.4932}{\sqrt{0.16993464}} \\
&= \frac{-0.4932}{0.412231294} \\
&\approx -1.1964
\end{aligned}
$$

$\alpha = 0.05$ and $Z_{0.05} = 1.645$. We reject H_0 if $Z < -1.645$.

Since $-1.1964 > -1.645$, we do not reject H_0 at the 5% level of significance and conclude that $\rho_1 = \rho_2$, i.e., the correlation coefficient for population 1 is the same as that for population 2.

(6)

[25]

QUESTION 5

(a) We have to test:

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$, against $H_1 : \sigma_p^2 \neq \sigma_q^2$ for at least one $p \neq q$

Using the Levene's test (or Bartlett's test), p-value $= 0.0650$ (or p-value $= 0.4177$). Since $0.0650 > 0.05$ (or $0.4177 > 0.05$) $\Longrightarrow$ we can not reject H_0 at the 5% level of significance. The assumption of equal variances is not violated. The variances are equal, that is, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$.

Note: If student uses O'Brien[.5]; p-value $= 0.0378$ and thus we reject H_0 at the 5% level of significance. The assumption of equal variances is violated. At least one of the variances is different from the other.

(4)

(b) (i) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against
$H_1 : \mu_p \neq \mu_q$ for at least one $p \neq q$.

(ii) The test statistic is $F = \frac{MSTr}{MSE} \sim F_{k-1;n-k}$

From the output: Computations for ANOVA we see that $F = 23.3897$ which is highly significant with a p–value of $< 0.0001 << 0.05$. We reject H_0 in favour of H_1 at the 5% level of significance and conclude that there is a significant difference in the population mean yield among the four types of fertilizers, that is, $\mu_p \neq \mu_q$ for at least one $p \neq q$. The **mean yield** for the different types of fertilizers differ significantly (at any level of significance). This implies that $\mu_p \neq \mu_q$ for at least one pair $p \neq q$.

(4)

(c) Confidence intervals that include zero imply that the pairs of means are not significantly different from each other. Most pairs of means differ significantly except for the pair "*Weed-killer B* " and "*Weed-killer C*" . This is graphically confirmed by the "Means Diamonds" where we can see that "*Weed-killer B* " and "*Weed-Killer C*" have almost identical pictures and their two circles overlap to a large extent on the "All Pairs Tukey-Kramer" display.

From the output of the formal statistical test we see that the confidence interval for the difference of the mean yield ("*Weed-killer C*" - "*Weed-killer B*") $= (-0.93733 : 1.687332)$. It is the only interval which includes **zero** and implies we cannot reject $\mu_{Weed-killer\ B} = \mu_{Weed-killer\ C}$.

All the other intervals for the difference of the means are (positive value; positive value) which excludes zero and means we reject $\mu_p = \mu_q \Longrightarrow \mu_p \neq \mu_q$.The Abs(Dif) LSD of *Weed-killer B* and *Weed-killer C* is negative (-0.9373), they share the same letter B. However, the pairs $\mu_{Weed-killer\ A}$ and $\mu_{Control}$; $\mu_{Weed-killer\ C}$ and $\mu_{Control}$; $\mu_{Weed-killer\ A}$ and $\mu_{Weed-killer\ B}$; $\mu_{Weed-killer B}$ and $\mu_{Control}$ and $\mu_{Weed-killer\ A}$ and $\mu_{Weed-killer\ C}$ are significantly different since their confidence intervals do not include zero. Their Abs(Dif) LSD are positive. Thus conclude that $\mu_1 \neq \mu_2 = \mu_3$.

(3)

[11]

QUESTION 6

(a) Yes, There is a strong positive relationship. (2)

(b) $\hat{\beta}_0 = 28.6770$, $\hat{\beta}_1 = 7.9571$ and $\sigma^2 = 124.27$. (3)

(c) The least squares regression line is

$\widehat{Y} = 28.6770 + 7.9571x \Longrightarrow \widehat{\text{Number of dead cells}} = 28.6770 + 7.9571\text{Dose}.$ (1)

(d) Replace $X = 4$ in the regression equation in (c) then

$$\begin{aligned} Y &= \widehat{Y} = 28.6770 + 7.9571(4) \\ &= 28.677 + 31.8284 \\ &= 60.5054 \end{aligned}$$

$\therefore$ The expected number of dead cells when dose is 4 cubic cms is 61.

(1)

(e) We have to test $H_0: \ \beta_1 = 0$ against

$H_1: \ \beta_1 \neq 0.$

<u>**Method I: Using the critical value approach**</u>:

From the output:

$$\begin{aligned} T &= \frac{\hat{\beta}_1 - B_1}{s/d} \\ &= \frac{7.9571 - 0}{1.539453} \\ &\approx 5.17 \end{aligned}$$

$\alpha = 0.01$ $\alpha/2 = 0.005$ $t_{\alpha/2;n-2} = t_{0.005;18} = 2.878$. Reject H_0 if $T < -2.878$ or if $T > 2.878$ or if $|T|| > 2.878$.

Since $5.17 > 2.878$, we reject H_0 in favour of H_1 at the 1% level significance and conclude that $\beta_1 \neq 0$. This means that the regression line is significant to explain the variability in y. (Only when $\beta_1 = 0$, does it imply that regression is meaningless.)

<u>**Method II: Using the p-value approach**</u>

p-value $< 0.0001 << 0.05$. We reject H_0 in favour of H_1 at the 1% level of significance and conclude that $\beta_1 \neq 0$. This means that the regression line is significant to explain the variability in y. (Only when $\beta_1 = 0$, does it imply that regression is meaningless.)

(4)

(f) The confidence interval is

$$\widehat{\beta}_1 \pm t_{\alpha/2;n-2} \times \frac{s}{d}$$

$\widehat{\beta}_1 = 7.9571 \qquad t_{\alpha/2;n-2} = t_{0.005;18} = 2.878$

The 99% confidence interval for $\widehat{\beta}_1$ is

$$\begin{array}{lll} \widehat{\beta}_1 & \pm & t_{\alpha/2;n-2} \times \frac{s}{d} \\ 7.9571 & \pm & 2.878 \times 1.539453 \\ 7.9571 & \pm & 4.4305 \\ (7.9571 - 4.4305 & ; & 7.9571 + 4.4305 \\ (3.5266 & ; & 12.3876) \end{array}$$

(3)

[14]

[100]