Revision Exercises 1

1. (a) Explain the difference between $\bar{x}$, $\bar{X}$ and $\mu$. 
   (b) Explain the difference between a standard error and a standard deviation. [5 marks]

2. To test the hypothesis that a coin is fair, it is tossed 100 times, resulting in 62 heads. Find the $P$-value for this result. What is your conclusion? [4 marks]

3. Show that if $X \overset{d}{=} R(0, \theta)$, then $X_{(n)}$ has cdf given by:
   
   $$F(x) = \left(\frac{x}{\theta}\right)^n \quad (0 < x < \theta)$$

   Hence show that
   
   $$\Pr\left( X_{(n)} > 0.05^{1/n}\theta \right) = 0.95$$

   and use this result to derive an upper 95% confidence limit for $\theta$ given a sample of $n = 12$ observations for which $x_{(12)} = 13.45$. [7 marks]

4. A statistic $T$ is such that $E(T) = \frac{n\theta}{n-1}$ and $\text{var}(T) = \frac{\theta^2}{n-2}$

   A random sample of $n = 10$ observations is obtained, for which the observed value of $T$ is found to be $t_{\text{obs}} = 3.45$. Find an unbiased estimate of $\theta$ and give its standard error. [5 marks]

5. (a) What is meant by:
   i. a 95% confidence interval for a parameter $\theta$?
   ii. a 95% prediction interval for a random variable $X$?

   (b) A random sample of $n = 25$ observations on $X \overset{d}{=} N(\theta, 1)$ gave $\bar{x} = 14.73$ and $s = 1.31$. Find:
   i. a 95% confidence interval for $\theta$;
   ii. a 95% prediction interval for $X$; [6 marks]

6. To test $H_0$: $X \overset{d}{=} N(0, 1)$ against $H_1$: $X \overset{d}{=} N(3, 2)$, the decision rule is to reject $H_0$ if $X > 3/2$. The observation $x = 2$ is obtained. Evaluate:
   (a) the size; (b) the power; (c) the $P$-value. [6 marks]

7. The likelihood function for a set of observations is given by
   
   $$L(\theta) = e^{-3\theta}(1 + \theta)^{36}$$

   Find the maximum likelihood estimate of $\theta$ and give an approximate value for its standard error. [5 marks]

8. A random sample of fifty observations on $X \overset{d}{=} \exp(\theta)$ produces the following data:

<table>
<thead>
<tr>
<th>interval</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 &lt; x &lt; 1</td>
<td>8</td>
</tr>
<tr>
<td>1 &lt; x &lt; 5</td>
<td>11</td>
</tr>
<tr>
<td>5 &lt; x &lt; 20</td>
<td>17</td>
</tr>
<tr>
<td>20 &lt; x &lt; 100</td>
<td>14</td>
</tr>
</tbody>
</table>

   (a) Represent this information using a histogram.
   (b) By approximating the sample cdf, obtain an approximate value for the sample median.
   (c) Give an approximate value for the sample mean, explaining your method.
   (d) Write down the likelihood function for this sample data. [8 marks]
9. A random sample of \( n \) observations is obtained on \( X \) which has pdf:
\[
f(x) = \theta x e^{-\frac{1}{2} \theta x^2} \quad (x > 0)
\]
Determine the form of the likelihood ratio test to test \( H_0: \theta = 1 \) against \( H_1: \theta > 1 \). [5 marks]

10. The following is a random sample on \( X \) which has cdf given by:
\[
F(x) = 1 - (1 + \frac{x}{\theta})^{-1} \quad (x > 0)
\]
0.12, 3.38, 3.46, 4.36, 4.42, 4.54, 5.66, 6.83, 8.30, 8.92, 9.35, 9.46,
9.51, 10.71, 10.96, 11.30, 12.53, 13.37, 13.83, 14.20, 18.02, 18.37,
18.51, 23.06, 33.67, 39.72, 53.13, 109.81, 138.66, 144.04, 313.39.

(a) Show that the sample median, \( \hat{C}_{0.5} \), \( \sim N(\theta, \frac{4\theta^2}{n}) \), and hence obtain an approximate 95% confidence interval for \( \theta \).
(b) Why is the sample mean of no use in estimation of \( \theta \)? [7 marks]

11. The following frequency table is a summary of a random sample of 100 observations on a discrete random variable \( M \):

<table>
<thead>
<tr>
<th>( k )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>freq(( M = k ))</td>
<td>11</td>
<td>42</td>
<td>27</td>
<td>20</td>
</tr>
</tbody>
</table>

The following frequency table is a summary of an independent random sample of 100 observations on a discrete random variable \( N \):

<table>
<thead>
<tr>
<th>( k )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>freq(( N = k ))</td>
<td>21</td>
<td>29</td>
<td>33</td>
<td>17</td>
</tr>
</tbody>
</table>

(a) Draw a suitable diagram to represent the two frequency distributions.
(b) Test the hypothesis that \( M \sim N \).
(c) Give an approximate 95% confidence interval for \( \mu_M - \mu_N \). [9 marks]

12. The precision of a measuring instrument is indicated by the sample standard deviation of a sample of standard measurements. A random sample of twenty such measurements produced a sample standard deviation of 0.417. Assuming the measurement errors are normally distributed, find a 95% confidence interval for the true standard deviation of the measuring instrument.

How could you check the assumption of normality? [5 marks]

13. Test the hypothesis that the populations from which the following random samples were independently obtained are identically distributed:

| sample 1 | 27, 31, 42, 37, 40, 43, 33 |
| sample 2 | 35, 32, 39, 44, 43, 40, 29 |

(a) using a \( t \)-test;
(b) using a rank-based test. [7 marks]
14. (a) If \( Z \sim Bi(80, 0.5) \) verify that \( \Pr(31 \leq Z \leq 49) \approx 0.967 \).

Hence find a 96.7\% confidence interval for the population median given the following sample:

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>23.9</td>
<td>29.4</td>
<td>29.7</td>
<td>30.4</td>
<td>30.7</td>
<td>30.9</td>
<td>32.1</td>
<td>32.2</td>
<td>32.4</td>
<td>32.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.4</td>
<td>33.6</td>
<td>34.1</td>
<td>34.6</td>
<td>34.8</td>
<td>34.9</td>
<td>35.1</td>
<td>35.8</td>
<td>35.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.9</td>
<td>36.1</td>
<td>36.2</td>
<td>36.2</td>
<td>36.4</td>
<td>36.4</td>
<td>36.7</td>
<td>37.2</td>
<td>37.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.6</td>
<td>37.9</td>
<td>38.3</td>
<td>38.5</td>
<td>39.2</td>
<td>39.4</td>
<td>39.5</td>
<td>39.7</td>
<td>40.1</td>
<td>40.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.2</td>
<td>40.3</td>
<td>40.3</td>
<td>41.0</td>
<td>41.5</td>
<td>41.7</td>
<td>41.8</td>
<td>41.9</td>
<td>42.2</td>
<td>42.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.7</td>
<td>42.8</td>
<td>42.9</td>
<td>42.9</td>
<td>43.0</td>
<td>43.2</td>
<td>43.3</td>
<td>43.4</td>
<td>43.8</td>
<td>44.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.0</td>
<td>44.2</td>
<td>44.8</td>
<td>45.2</td>
<td>45.2</td>
<td>45.4</td>
<td>45.5</td>
<td>45.6</td>
<td>45.9</td>
<td>46.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.2</td>
<td>47.8</td>
<td>47.8</td>
<td>48.0</td>
<td>48.5</td>
<td>48.7</td>
<td>49.1</td>
<td>49.8</td>
<td>50.5</td>
<td>53.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Consider a random sample of \( n \) observations on \( X \).

Let \( Z_q \) denote the number of observations in the sample less than \( c_q \), where \( c_q \) denotes the \( q \)-quantile of \( X \).

Specify the distribution of \( Z_q \), and explain how this result could be used to find a confidence interval for \( c_q \).  

[8 marks]

15. The following data are independent realisations of a random variable \( Y \) at pre-set values of a variable \( x \). It may be assumed that the observations are normally distributed with equal variances.

\[
\begin{array}{c|cccc}
  x & 0 & 1 & 2 & 3 \\
  y & 10 & 6 & 6 & 3 \\
   & 5 & 8 & 7 & 4 \\
   & 3 & 5 & 2 & 4 \\
\end{array}
\]

Use the Minitab output below to test the hypothesis that the regression of \( Y \) on \( x \) is a straight line.  

[5 marks]
16. Suppose that the independent normally distributed random variables $Y_1$, $Y_2$, $Y_3$, $Y_4$ have means given by

$$E(Y_1) = \alpha + 2\beta, \quad E(Y_2) = 3\beta, \quad E(Y_3) = \alpha - \beta, \quad E(Y_4) = 2\alpha + \beta;$$

and equal variances, denoted by $\sigma^2$.

The following realisations are obtained:

$$y_1 = 4, \quad y_2 = -4, \quad y_3 = 5, \quad y_4 = 3$$

(a) Find the least squares estimates of $\alpha$, $\beta$ and $\sigma^2$.

(b) Find an estimate of $\eta = \alpha - \beta$ and its standard error.

(c) Find a 95% prediction interval for a future observation on $Y_3$. [9 marks]

17. (a) Explain the importance of randomisation and blocking in the design of experiments.

(b) The following results were obtained for a randomised block experiment for treatments $A$, $B$, $C$ and $D$ arranged in four blocks of four plots,

<table>
<thead>
<tr>
<th>block</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

A partial analysis of variance given below (in which some of the entries are replaced by asterisks).

<table>
<thead>
<tr>
<th>source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocks</td>
<td>**</td>
<td>*****</td>
<td>****</td>
</tr>
<tr>
<td>treatments</td>
<td>**</td>
<td>68.19</td>
<td>****</td>
</tr>
<tr>
<td>error</td>
<td>**</td>
<td>*****</td>
<td>1.28</td>
</tr>
<tr>
<td>total</td>
<td>**</td>
<td>202.94</td>
<td></td>
</tr>
</tbody>
</table>

Use this information to analyse these data, assuming independence and normality of the observations, in particular:

i. test in the significance of the treatment effects; and

ii. find a 95% confidence interval for $\tau_D - \tau_A$, the difference in the effects of treatment $D$ and treatment $A$. [9 marks]

18. Consider the general linear model:

$$\bar{Y} = A\hat{\theta} + \bar{E}, \quad \text{where} \quad E(\bar{E}) = 0 \quad \text{and} \quad D(\bar{E}) = \sigma^2 I$$

(a) Show that if $\hat{T}$ denotes the least squares estimator of $\hat{\theta}$ then $\hat{T}$ is such that

$$A'\hat{T} = A'\bar{Y}$$

where $M'$ denotes the transpose of $M$.

(b) Show that

$$A' A E(\hat{T}) = A' A \hat{\theta} \quad \text{and} \quad A' A D(\hat{T}) A' A = \sigma^2 A' A$$

Hence obtain expressions for the mean vector and the variance-covariance matrix of $T$ if $A' A$ is non-singular. [10 marks]
Revision Exercises 2

1. (a) What is meant by:
   i. a 95% confidence interval for a parameter $\theta$?
   ii. a 95% prediction interval for a random variable $X$?

   (b) A random sample of $n = 25$ observations on $X \sim N(\mu, \sigma^2)$ gave $\bar{x} = 14.73$ and $s = 2.81$. Find:
   i. a 95% confidence interval for $\mu$;
   ii. a 95% prediction interval for $X$;
   iii. a 95% confidence interval for $\sigma$.

2. (a) Suppose that $X$ has a logistic distribution $X \sim Lg(\theta, 1)$, so that $X$ has pdf given by
   $$ f_X(x) = \frac{e^{-(x-\theta)}}{[1 + e^{-(x-\theta)}]^2} $$
   For sampling on $X \sim Lg(\theta, 1)$, compare the sample mean and the sample median as estimators of $\theta$.

   (b) For sampling on $G(\theta)$ find the minimum variance bound for unbiased estimators of $\sigma^2 = (1 - \theta)/\theta^2$.

   (c) Estimators $T_1$ and $T_2$ are such that:
   $$ E(T_1) = \theta, \quad \text{var}(T_1) = \frac{2\theta^2}{n} \quad E(T_2) = \theta + \frac{\theta - 1}{n}, \quad \text{var}(T_2) = \frac{\theta^2}{n} $$
   Which of $T_1$ and $T_2$ is a better estimator of $\theta$? Give a detailed explanation of the reasons for your choice.

3. The likelihood function for an observed set of data is given by:
   $$ L(\theta) = e^{-10\theta} \theta^6 (1 - \theta)^3 \quad (0 < \theta < 1) $$
   (a) Find the maximum likelihood estimate of $\theta$.
   (b) Obtain an approximate value for the standard error of your estimate.

4. (a) If $X \sim R(0, \theta)$, and a random sample of $n$ observations is obtained on $X$,
   i. find the method of moments estimator of $\theta$;
   ii. show that the maximum likelihood estimate of $\theta$ is $X_{(n)}$.

   (b) Show that if $X \sim R(0, \theta)$, then $X_{(n)}$ has cdf given by:
   $$ F(x) = \left( \frac{x}{\theta} \right)^n \quad (0 < x < \theta) $$

   (c) Given the following sample on $X \sim R(0, \theta)$:
   2.56, 0.27, 1.23, 0.85, 2.59, 2.21, 1.33, 1.79, 0.52, 1.74;
   find a 95% confidence interval for $\theta$ based on $X_{(n)}$.

5. It is claimed that the failure time of brand X electronic components is greater than the failure time of brand Y components. To test this claim, the following measurements were obtained:

   | brand X | 47 | 73 | 52 | 72 | 26 | 89 | 77 | 113 | 63 | 138 |
   | brand Y | 20 | 39 | 87 | 39 | 62 | 143 | 27 | 55 | 72 | 41 |

   Carry out an appropriate test of this claim. Justify your choice of test.
6. Consider the random variable $X$ having pmf given by:

$$
\Pr(X = 0) = 1 - 3\theta, \quad \Pr(X = 1) = \theta, \quad \Pr(X = 2) = 2\theta
$$

Suppose a random sample of $n$ observations is obtained on $X$ yielding the following results:

freq($X = 0$) = 60, freq($X = 1$) = 16, freq($X = 2$) = 24

(a) Estimate $\theta$ using the method of moments and the method of maximum likelihood.

(b) Obtain the standard error for each estimator.

(c) Use the better estimate to obtain a 95% confidence interval for $\theta$.

(d) Test the goodness of fit of the model to the data. [12 marks]

7. (a) Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent random variables such that

$$Y_i \overset{d}{=} N(\beta x_i, \sigma^2) \quad i = 1, 2, \ldots, n$$

Using the results for the general linear model, show that the least squares estimator of $\beta$ is given by

$$\hat{B} = \frac{\sum x_i Y_i}{\sum x_i^2}$$

Show that

$$E(\hat{B}) = \beta \quad \text{and} \quad \text{var}(\hat{B}) = \frac{\sigma^2}{\sum x_i^2}$$

(b) Consider the following data set:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>9</td>
<td>16</td>
<td>26</td>
<td>39</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>20</td>
<td>31</td>
<td>45</td>
<td>59</td>
</tr>
</tbody>
</table>

i. Fit a straight line regression through the origin to these data.

ii. Test the goodness of fit of this model.

iii. Find a 95% confidence interval for $\beta$. [10 marks]

8. Determinations of the strength of a fibre after using three methods of treatment were as follows:

(1) control 29.8, 28.5, 28.7, 27.2
(2) treatment A 29.3, 31.5, 30.9, 30.1
(3) treatment B 31.2, 32.8, 32.3, 29.2

(a) Complete the following analysis of variance table derived from the above data:

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>between methods</td>
<td>1</td>
<td>16.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>within methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Test the hypothesis $H_0$: $\mu_1 = \mu_2 = \mu_3$ giving an approximate $P$-value.

(c) Give an estimate of the error variance.

(d) Find a 95% confidence interval for the effect of treatment A, i.e $\mu_2 - \mu_1$. [10 marks]
9. The random variables $Y_1, Y_2, \ldots, Y_6$ are independent and normally distributed with $\text{var}(Y_i) = \sigma^2$, $i = 1, 2, \ldots, 6$, and

\[
\begin{align*}
E(Y_1) &= \alpha, \quad E(Y_2) = \beta, \quad E(Y_3) = \alpha + \beta, \\
E(Y_4) &= \alpha - \beta, \quad E(Y_5) = 2\alpha + \beta, \quad E(Y_6) = \alpha + 2\beta.
\end{align*}
\]

The following observations were obtained:

\[
\begin{align*}
y_1 &= 4.1, \quad y_2 = 2.3, \quad y_3 = 6.7, \quad y_4 = 2.1, \quad y_5 = 10.4, \quad y_6 = 8.8.
\end{align*}
\]

(a) Write down the observational equations.

(b) Write down the normal equations.

(c) Obtain the least squares estimates of $\alpha$ and $\beta$.

(d) Give an unbiased estimate of $\sigma^2$.

(e) Find a 95% confidence interval for $\alpha + \beta$.

[10 marks]

10. A randomised block experiment to compare four treatments ($T_1, T_2, T_3, T_4$) is carried out in six blocks ($B_1, B_2, \ldots, B_6$) of four plots/block with the following results:

<table>
<thead>
<tr>
<th></th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_3$</th>
<th>$B_4$</th>
<th>$B_5$</th>
<th>$B_6$</th>
<th>av</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_1$</td>
<td>40</td>
<td>52</td>
<td>60</td>
<td>26</td>
<td>28</td>
<td>34</td>
<td>40</td>
</tr>
<tr>
<td>$T_2$</td>
<td>46</td>
<td>54</td>
<td>60</td>
<td>32</td>
<td>32</td>
<td>40</td>
<td>44</td>
</tr>
<tr>
<td>$T_3$</td>
<td>46</td>
<td>52</td>
<td>64</td>
<td>30</td>
<td>33</td>
<td>39</td>
<td>44</td>
</tr>
<tr>
<td>$T_4$</td>
<td>45</td>
<td>51</td>
<td>63</td>
<td>29</td>
<td>32</td>
<td>38</td>
<td>43</td>
</tr>
</tbody>
</table>

Using MINITAB, these data were put into C1, the treatment index into C2 and the block index into C3. The following output was obtained:

MTB> TWOWAY C1 C2 C3

ANALYSIS OF VARIANCE

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>3</td>
<td>64.50</td>
<td>21.50</td>
</tr>
<tr>
<td>C3</td>
<td>5</td>
<td>3172.00</td>
<td>634.40</td>
</tr>
<tr>
<td>ERROR</td>
<td>15</td>
<td>32.00</td>
<td>2.13</td>
</tr>
<tr>
<td>TOTAL</td>
<td>23</td>
<td>3268.50</td>
<td></td>
</tr>
</tbody>
</table>

(a) Test the hypothesis that there is no difference between the treatments, specifying the assumptions on which this test is based. Do the assumptions appear to be valid?

(b) Give an estimate and 95% confidence interval for

i. the error variance;

ii. the treatment differences.

[8 marks]

11. The following data are observations on a random variable $Y$ at specified values of a variable $x$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>14, 15</td>
<td>25, 23</td>
<td>27, 29</td>
<td>33, 32</td>
<td>34, 35</td>
</tr>
</tbody>
</table>

With the $x$-values stored in C11 and the $y$-values stored in C10, the following MINITAB output was obtained:

MTB > regr c10 1 c11

The regression equation is

\[ C10 = 12.1 + 4.85 \times C11 \]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>Stdev</th>
<th>t-ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>12.150</td>
<td>1.651</td>
<td>7.36</td>
<td>0.000</td>
</tr>
<tr>
<td>C11</td>
<td>4.8500</td>
<td>0.4978</td>
<td>9.74</td>
<td>0.000</td>
</tr>
</tbody>
</table>

$s = 2.226$  
R-sq = 92.2%  
R-sq(adj) = 91.3%

Analysis of Variance
### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>470.45</td>
<td>470.45</td>
<td>94.92</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>8</td>
<td>39.65</td>
<td>4.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>510.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MTB > oneway c10 c11
Analysis of variance

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>C11</td>
<td>4</td>
<td>504.60</td>
<td>126.15</td>
<td>114.68</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>5</td>
<td>5.50</td>
<td>1.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>510.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### INDIVIDUAL 95 PCT CI’S FOR MEAN 
BASED ON POOLED STDEV

<table>
<thead>
<tr>
<th>Level</th>
<th>N</th>
<th>Mean</th>
<th>Stdev</th>
<th>Base on pooled stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>14.500</td>
<td>0.707</td>
<td>(---*)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>24.000</td>
<td>1.414</td>
<td>(<em>-</em>)</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>28.000</td>
<td>1.414</td>
<td>(-----)</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>32.500</td>
<td>0.707</td>
<td>(*)</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>34.500</td>
<td>0.707</td>
<td>(****)</td>
</tr>
</tbody>
</table>

POOLED STDEV = 1.049

It can be assumed that the observations are independent and normally distributed with equal variances.

(a) What is the sample correlation coefficient?
(b) Show that a quadratic regression provides a good fit, but that the linear regression does not.
(c) Find a 95% confidence interval for \( E(Y \mid x = 3) \):
   i. assuming a straight line regression model;
   ii. making no assumptions about the regression.  

[10 marks]
Revision Exercises 3

1. (a) Find an approximate 95% probability interval for the sample mean $\bar{U}$, obtained from a random sample of $n = 200$ observations on $U \overset{d}{=} \gamma(40, 2)$.

(b) A random sample of $n = 500$ observations is obtained on $V \overset{d}{=} \text{Pn}(2.4)$. Find an approximate 95% probability interval for the number of zeros in the sample.

(c) Find an approximate 95% probability interval for the sample median for a random sample of $n = 100$ observations on $W \overset{d}{=} \text{C}(10, 1)$, for which the pdf is given by:

$$f(w) = \frac{1}{\pi} \frac{1}{1 + (w - 10)^2}.$$  

(d) For a random sample of $n = 80$ on $X \overset{d}{=} \exp(0.1)$, find the probability that the sample maximum is greater than 50.

(e) For a random sample of $n = 40$ on $Y \overset{d}{=} \text{N}(\mu = 50, \sigma^2 = 100)$, find the probability that the sample standard deviation is greater than 11.4.  

2. The following is a random sample on $Y \overset{d}{=} \text{N}(\mu, \sigma^2)$:

\[
\begin{array}{cccccccccccc}
29.1 & 28.5 & 25.4 & 29.3 & 24.9 & 34.8 & 30.0 & 30.6 & 21.9 & 26.6 \\
23.7 & 28.6 & 29.7 & 36.7 & 31.7 & 21.9 & 29.4 & 22.2 & 26.6 & 31.0
\end{array}
\]

(a) i. Find a point estimate and a 95% confidence interval for $\mu$.

ii. Test the hypothesis $\mu = 30$.

(b) i. Find a point estimate and a 95% confidence interval for $\sigma$.

ii. Test the hypothesis $\sigma = 5$.

(c) Find a 95% prediction interval for a future observation on $Y$.  

3. A random sample of $n = 80$ observations is obtained on the discrete random variable $X$, with the following results:

\[
\begin{array}{cccc}
x & 0 & 1 & 2 & 3 \\
\text{frequency} & 29 & 21 & 16 & 14
\end{array}
\]

A proposed model for these data is as follows:

\[
\begin{array}{cccc}
x & 0 & 1 & 2 & 3 \\
\text{probability} & \theta & \theta(1-\theta) & \theta(1-\theta)^2 & (1-\theta)^3
\end{array}
\]

(a) Find the maximum likelihood estimate of $\theta$ under this model and approximate its standard error.

(b) Test the goodness of fit of the model.  

4. A random sample of $n$ observations is obtained on the random variable $Y$ which has a lognormal distribution, $\ell \text{N}(\theta, 1)$; the pdf of $Y$ is given by

$$f(y | \theta) = \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}(\ln y - \theta)^2} \quad (y > 0)$$

(a) Find the method of moments estimator of $\theta$, and derive an expression for its variance.

Note: You may use the results given in the summary notes without proof.

(b) Find the maximum likelihood estimator of $\theta$, and derive an expression for its variance.

(c) Comment on the consistency, unbiasedness and relative efficiency of these two estimators.  

5. Consider a two-state Markov chain, $\mathcal{X} = \{X_n, n = 0, 1, 2, \ldots\}$, with state space $\{0, 1\}$ and transition probability matrix

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

The following realisation of $\mathcal{X}$ is obtained:

00000 00101 00000 11110 00000 01110

(a) Find the maximum likelihood estimates of $\alpha$ and $\beta$ and their standard errors.

(b) Test the hypothesis that $\beta = 2\alpha$. [8 marks]

6. (a) To test $H_0: X \overset{d}{=} \text{N}(10, 4)$ against $H_1: X \overset{d}{=} \text{N}(15, 9)$, a sample of $n = 10$ observations is obtained on $X$ and the decision based on the sample mean. The observation $\bar{x} = 12.3$ is obtained. Specify the $P$-value for this test.

(b) The random variable $Y$ is exponentially distributed: $Y \overset{d}{=} \exp(\theta)$. To test the hypothesis $H_0: \theta = 0.5$ against $H_1: \theta = 0.2$, twenty independent observations are obtained on $Y$. Determine the likelihood ratio test with size 0.05, and find approximately the power of this test. [10 marks]

7. Consider the independent samples

| sample 1 | 27 | 34 | 39 | 40 | 43 |
| sample 2 | 41 | 44 | 52 | 93 |

(a) Draw dotplots of the two samples.

(b) Show that a two-sample t-test does not reject the null hypothesis of equal means.

(c) Show that using a rank-based test, the null hypothesis of equal locations is rejected.

(d) Explain the difference in the results of the tests. [10 marks]

8. Consider the following random sample of eighty observations on $T$, the time in hours between computer terminal breakdowns.

<table>
<thead>
<tr>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>13</th>
<th>14</th>
<th>23</th>
<th>26</th>
<th>30</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>40</td>
<td>41</td>
<td>53</td>
<td>59</td>
<td>60</td>
<td>66</td>
<td>77</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>91</td>
<td>94</td>
<td>99</td>
<td>100</td>
<td>104</td>
<td>117</td>
<td>120</td>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>124</td>
<td>125</td>
<td>126</td>
<td>128</td>
<td>128</td>
<td>131</td>
<td>132</td>
<td>133</td>
<td>134</td>
<td>139</td>
</tr>
<tr>
<td>145</td>
<td>147</td>
<td>154</td>
<td>156</td>
<td>162</td>
<td>166</td>
<td>174</td>
<td>184</td>
<td>185</td>
<td>187</td>
</tr>
<tr>
<td>188</td>
<td>190</td>
<td>190</td>
<td>191</td>
<td>193</td>
<td>193</td>
<td>199</td>
<td>200</td>
<td>203</td>
<td>208</td>
</tr>
<tr>
<td>223</td>
<td>245</td>
<td>249</td>
<td>261</td>
<td>277</td>
<td>317</td>
<td>319</td>
<td>328</td>
<td>329</td>
<td>332</td>
</tr>
<tr>
<td>349</td>
<td>366</td>
<td>374</td>
<td>417</td>
<td>432</td>
<td>462</td>
<td>533</td>
<td>539</td>
<td>549</td>
<td>885</td>
</tr>
</tbody>
</table>

(a) i. Find the sample median.

ii. Find a two-sided 95% confidence interval for $m$, making no assumptions about the underlying distribution.

(b) A possible model for these data is that $T$ has an exponential distribution.

i. How would you decide whether an exponential distribution was an appropriate model? Do not attempt to carry out any statistical procedures, just specify which procedure or procedures you would use.

ii. If you were to assume that the data were a random sample from an exponential distribution, describe (but do not perform) the method you would use to obtain a 95% confidence interval for $m$. [10 marks]

9. The following data are independent realisations of a random variable $Y$ at pre-set values of a variable $x$. It may be assumed that the observations are normally distributed with equal variances.

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>19, 16</td>
<td>14, 15</td>
<td>12, 10, 9</td>
<td>5</td>
<td>3, 6</td>
</tr>
</tbody>
</table>
(a) Use the Minitab output below (with $x$ in c1 and $y$ in c2) to test the hypothesis that the regression of $Y$ on $x$ is a straight line.

```
MTB > regr c2 1 c1
The regression equation is  $y = 17.6 - 3.35 x$
Predictor  Coef  StDev  T  P
Constant   17.600   0.875    20.1  0.000
$x$      -3.350   0.357    -9.37 0.000
$S = 1.599$  $R^2 = 91.6\%$  $R^2(adj) = 90.6\%$
```

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>224.45</td>
<td>224.45</td>
<td>87.80</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual</td>
<td>8</td>
<td>20.45</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>244.90</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```
MTB > oneway c2 c1
Analysis of Variance for $y$
Source      DF | SS  | MS  | F    | P
-------------|-----|-----|------|----
$x$          | 4   | 225.40 | 56.35 | 14.45 | 0.006
Error        | 5   | 19.50  | 3.90  |      |    
Total        | 9   | 244.90 |      |      |    |
```

(b) Use the straight line regression model to obtain a 95% confidence interval for $E(Y \mid x = 5)$.

10. $P, Q, R$ and $S$ are four points on a line in the order given. Measurements are made on different segments with the following results:

- $PQ = 5$
- $QR = 8$
- $RS = 6$
- $PR = 10$
- $QS = 12$
- $PS = 15$

Let $\alpha, \beta, \gamma$ denote the lengths of the line segments $PQ, QR, RS$. Assume that the observations are unbiased with independent normally distributed errors having equal variances.

(a) Find the maximum likelihood estimates of $\alpha, \beta$ and $\gamma$; and obtain an unbiased estimate of the error variance.

The result $\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}^{-1} = \frac{1}{4} \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ may be useful.

(b) Obtain a 95% confidence interval for the length of $PS$.

(c) Test the hypothesis that $QR = RS$.

(d) Test the hypothesis that $PQ = QR = RS$. [12 marks]
11. The following results were obtained for a randomised block experiment for treatments \(A\), \(B\) and \(C\) arranged in six blocks of six plots,

<table>
<thead>
<tr>
<th>block 1</th>
<th>treatment (P)</th>
<th>treatment (Q)</th>
<th>treatment (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>block 2</td>
<td>21, 24</td>
<td>20, 14</td>
<td>15, 17</td>
</tr>
<tr>
<td>block 3</td>
<td>24, 27</td>
<td>25, 19</td>
<td>18, 21</td>
</tr>
<tr>
<td>block 4</td>
<td>27, 29</td>
<td>17, 26</td>
<td>13, 19</td>
</tr>
<tr>
<td>block 5</td>
<td>18, 23</td>
<td>12, 16</td>
<td>13, 17</td>
</tr>
<tr>
<td>block 6</td>
<td>17, 20</td>
<td>11, 15</td>
<td>13, 10</td>
</tr>
<tr>
<td>block 7</td>
<td>28, 21</td>
<td>22, 14</td>
<td>13, 17</td>
</tr>
</tbody>
</table>

A partial analysis of variance is given below (in which some of the entries are replaced by asterisks).

<table>
<thead>
<tr>
<th>source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocks</td>
<td>**</td>
<td>*****</td>
<td>*****</td>
</tr>
<tr>
<td>treatments</td>
<td>**</td>
<td>367.72</td>
<td>*****</td>
</tr>
<tr>
<td>error</td>
<td>**</td>
<td>*****</td>
<td>10.65</td>
</tr>
<tr>
<td>total</td>
<td>**</td>
<td>1033.89</td>
<td></td>
</tr>
</tbody>
</table>

Use this information to analyse these data, assuming independence and normality of the observations, in particular:

(a) test the significance of the treatment effects; and
(b) find a 95% confidence interval for the difference in the effects of treatment \(P\) and treatment \(C\).  

12. Consider the general linear model:

\[ Y = A\theta + E, \quad \text{where} \quad \text{E}(E) = 0 \quad \text{and} \quad \text{D}(E) = \sigma^2 I \]

Let \(T\) denote the least squares estimator of \(\theta\).

If \(A' A\) is non-singular, show that

(a) \(T = (A' A)^{-1} A' Y;\)
(b) \(\text{E}(T) = \theta, \text{D}(T) = \sigma^2 (A' A)^{-1};\)
(c) \(T\) is the best linear unbiased estimator of \(\theta.\)
Revision Exercises 4

1. A random sample on a continuous random variable $X$ yields the frequency data below:

<table>
<thead>
<tr>
<th>interval</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 &lt; x &lt; 40$</td>
<td>10</td>
</tr>
<tr>
<td>$40 &lt; x &lt; 50$</td>
<td>10</td>
</tr>
<tr>
<td>$50 &lt; x &lt; 60$</td>
<td>40</td>
</tr>
<tr>
<td>$60 &lt; x &lt; 70$</td>
<td>60</td>
</tr>
<tr>
<td>$70 &lt; x &lt; 80$</td>
<td>40</td>
</tr>
<tr>
<td>$80 &lt; x &lt; 100$</td>
<td>40</td>
</tr>
</tbody>
</table>

(a) Sketch a histogram for these data.

(b) Sketch the sample cdf and hence give approximate values for the sample median and the sample interquartile range.

(c) Estimate the population mean and standard deviation.

(d) Give an approximate 95% confidence interval for the population mean.

(e) Test the hypothesis that the population median is equal to 70 against a two-sided alternative. (12 marks)

2. (a) The diagram below is a 95% statistic-parameter diagram for the $P_n(\lambda)$ distribution. Describe briefly how such a diagram could be obtained.

(b) Use the statistic-parameter diagram:

i. to specify an approximate 95% probability interval for $X \overset{\text{d}}{=} P_n(20)$;

ii. to obtain a 95% confidence interval for $\theta$, based on the following random sample of $n = 20$ observations on $Y \overset{\text{d}}{=} P_n(\theta)
\begin{align*}
1, 0, 3, 0, 2, 0, 1, 0, 2, 2, 4, 0, 0, 1, 2, 2, 0, 0, 3, 1.
\end{align*}

(6 marks)
3. A random sample of fifteen observations is obtained from a population having a standard normal distribution, i.e. N(0, 1).

(a) i. Find the probability that the sample mean is greater than 0.3.
ii. Find the probability that the sample standard deviation is greater than 1.3.
iii. Find the probability that the largest observation is greater than 2.3.
iv. Find the probability that at least ten of the observations are positive.

A second, independent, random sample of ten observations is obtained from the same population.

(b) Specify the distributions of each of the following statistics:

\[
\bar{X}_1 - \bar{X}_2, \quad \frac{S^2_1}{S^2_2}, \quad 15\bar{X}^2_1 + 10\bar{X}^2_2, \quad 14S^2_1 + 9S^2_2
\]

where \( \bar{X}_1, \bar{X}_2, S^2_1 \) and \( S^2_2 \) denote the sample means and variances of the first and second samples.

(12 marks)

4. A random sample of \( n \) observations is obtained on \( X \) which has pdf and cdf given by

\[
f(x) = \frac{\theta}{(1 + x)^{\theta+1}} \quad (x > 0); \quad F(x) = 1 - \frac{1}{(1 + x)^{\theta}} \quad (x > 0).
\]

(a) Find the maximum likelihood estimate of \( \theta \) and an expression for its standard error.

(b) Let \( x_{(k)} \) denote the \( k \)th order statistic for this sample on \( X \); and define \( z_k = -\ln(1 - \frac{k}{n+1}) \). Explain how a plot of \( \ln(1 + x_{(k)}) \) against \( z_k \) yields: (i) a check of the distributional assumption; (ii) an estimate of \( \theta \).

Indicate in a rough sketch what such a plot might look like.

Comment on the possible advantages and/or disadvantages of this approach.

(c) Show that \( Y = \ln(1 + X) \overset{d}{=} \exp(\theta) \) and hence find an exact 95% confidence interval for \( \theta \) based on the following data:

\[
\begin{array}{ccccccccccc}
x & 0.93 & 0.19 & 0.41 & 12.22 & 1.62 & 0.46 & 1.37 & 0.01 & 0.29 & 1.22 \\
y & 0.66 & 0.17 & 0.35 & 2.58 & 0.96 & 0.38 & 0.86 & 0.01 & 0.25 & 0.80 \\
\end{array}
\]

(14 marks)

5. i. Independent estimators of \( \theta \) are such that

\[
E(\hat{\Theta}_1) = \theta, \quad \text{var}(\hat{\Theta}_1) = \sigma^2_1; \quad E(\hat{\Theta}_2) = \theta, \quad \text{var}(\hat{\Theta}_2) = \sigma^2_2.
\]

Specify the best estimator of \( \theta \) of the form \( a_1\hat{\Theta}_1 + a_2\hat{\Theta}_2 \), indicating what optimality criteria you have used.

ii. The following results are available:

\[
\hat{\theta}_1 = 2.43, \quad \text{se}(\hat{\theta}_1) = 0.34; \quad \hat{\theta}_2 = 2.96, \quad \text{se}(\hat{\theta}_2) = 0.72.
\]

Use these data to obtain a combined estimate of \( \theta \); and give its standard error.

iii. Describe how you would combine \( k \) estimates from independent sources: viz. \( \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k \) with standard errors \( \text{se}(\hat{\theta}_1), \text{se}(\hat{\theta}_2), \ldots, \text{se}(\hat{\theta}_k) \).

(9 marks)
6. The following data are from a random sample of $n = 67$ observations on $Y$, ordered from smallest to largest.

<table>
<thead>
<tr>
<th>0.35</th>
<th>0.48</th>
<th>0.50</th>
<th>0.54</th>
<th>0.54</th>
<th>0.56</th>
<th>0.59</th>
<th>0.69</th>
<th>0.71</th>
<th>0.73</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.73</td>
<td>0.74</td>
<td>0.78</td>
<td>0.80</td>
<td>0.80</td>
<td>0.82</td>
<td>0.84</td>
<td>0.85</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td>0.97</td>
<td>0.98</td>
<td>1.03</td>
<td>1.06</td>
<td>1.10</td>
<td>1.12</td>
<td>1.17</td>
<td>1.26</td>
<td>1.33</td>
<td>1.38</td>
</tr>
<tr>
<td>1.46</td>
<td>1.53</td>
<td>1.55</td>
<td>1.58</td>
<td>1.64</td>
<td>1.68</td>
<td>1.72</td>
<td>1.72</td>
<td>1.81</td>
<td></td>
</tr>
<tr>
<td>1.83</td>
<td>2.10</td>
<td>2.14</td>
<td>2.15</td>
<td>2.20</td>
<td>2.49</td>
<td>2.52</td>
<td>2.53</td>
<td>2.67</td>
<td>2.71</td>
</tr>
<tr>
<td>2.73</td>
<td>2.75</td>
<td>2.82</td>
<td>2.82</td>
<td>2.85</td>
<td>3.00</td>
<td>3.13</td>
<td>3.21</td>
<td>3.57</td>
<td>3.62</td>
</tr>
<tr>
<td>3.67</td>
<td>3.95</td>
<td>4.05</td>
<td>4.06</td>
<td>4.12</td>
<td>4.16</td>
<td>4.32</td>
<td>5.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Find the sample median and the sample quartiles and draw a boxplot for these data.

(b) If $Z \sim \text{Bi}(n = 67, p = 0.5)$, show that $\Pr(26 \leq Z \leq 41) \approx 0.95$ and hence find a 95% confidence interval for $m$, the median of $Y$.

(c) Explain why $(0.48, 4.32)$ gives a 94% prediction interval for $Y$.

(d) The distribution of $Y$ is actually lognormal. Use this information, and the descriptive statistics below, to obtain

i. a 95% confidence interval for $m$; and

ii. a 95% prediction interval for $Y$.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>67</td>
<td>1.866</td>
<td>1.182</td>
</tr>
<tr>
<td>$\ln y$</td>
<td>67</td>
<td>0.415</td>
<td>0.670</td>
</tr>
</tbody>
</table>

(13 marks)

7. For a random sample of $n$ observations, the test statistic $T$ is such that $T \approx N(\theta, \theta/n)$. It is required to use $T$ to test $H_0$: $\theta = 4$ against the alternative $H_1$: $\theta > 4$.

i. Determine a test of size 0.05 based on $T$ for a sample of 25 observations.

ii. Find the power of this test when $\theta = 5$.

iii. Explain how you would go about finding a test of $H_0$ against $H_1$ with size 0.05 and power 0.90 when $\theta = 5$.

(6 marks)

8. Two procedures ($S$ and $G$) for sintering copper are compared by testing each procedure on each of ten different types of powder. The measurement of interest is the porosity of the test specimen. The results are given in the following table:

<table>
<thead>
<tr>
<th>powder</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>procedure $S$</td>
<td>21</td>
<td>27</td>
<td>28</td>
<td>22</td>
<td>26</td>
<td>19</td>
<td>21</td>
<td>26</td>
<td>28</td>
<td>32</td>
</tr>
<tr>
<td>procedure $G$</td>
<td>23</td>
<td>26</td>
<td>21</td>
<td>24</td>
<td>25</td>
<td>16</td>
<td>18</td>
<td>22</td>
<td>23</td>
<td>33</td>
</tr>
</tbody>
</table>

Test the hypothesis that the procedures have the same effect on the porosity level (i) using a $t$-test; (ii) using a rank-based test.

Let $\delta = \text{(effect of procedure } S - \text{effect of procedure } G)$. Find an approximate 95% confidence interval for $\delta$.

(8 marks)

9. Fifteen individuals are randomly divided into three groups of five. The first group is given treatment $A$, the second treatment $B$ and the third group is a control group and is given no treatment. The results obtained were reported as follows:

| group 1 (treatment $A$) | 46 | 43 | 45 | 44 | 38 |
| group 2 (treatment $B$) | 31 | 35 | 39 | 32 | 33 |
| group 3 (control group) | 28 | 22 | 29 | 54 | 27 |

The following analysis was presented:

MTB > oneway 'y' 'group'
Analysis of Variance for y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>2</td>
<td>356.8</td>
<td>178.4</td>
<td>3.00</td>
<td>0.088</td>
</tr>
<tr>
<td>Error</td>
<td>12</td>
<td>712.8</td>
<td>59.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
10. (a) Sketch a scatter plot for a sample of about twenty points such that \( r \approx -0.6 \).
(b) Compute the correlation coefficient for the following sample on (X,Y): (1,2), (2,1), (3,3), (4,5), (5,4).
(c) A random sample of fifty observations on (X,Y) gives sample correlation coefficient \( r = 0.4 \). Assuming the data are bivariate normally distributed, give a 95% confidence interval for \( \rho \). Is this evidence that X and Y are not independent? Explain.
(d) Give an example of a situation in which the correlation coefficient \( r \) and the rank correlation coefficient \( r' \) are quite different in magnitude.

11. (a) Suppose that \( Y_1, Y_2, \ldots, Y_n \) are independent random variables which are such that \( Y_i \overset{d}{=} \text{Pn}(\beta x_i) \) where \( x_1, x_2, \ldots, x_n \) are given constants. Find the maximum likelihood estimator of \( \beta \) and show that it is the MVB estimator.
For the following sample, find \( \hat{\beta} \) and \( se(\hat{\beta}) \).
\[
\begin{array}{ccccccc}
  x & 1.4 & 2.3 & 3.6 & 4.2 & 4.9 & 5.5 & 7.0 & 7.8 \\
  y & 4 & 3 & 5 & 2 & 7 & 8 & 9 & 7 \\
\end{array}
\]
(b) Suppose that \( \text{E}(Y) = A\hat{\beta} \) and \( \text{D}(Y) = \sigma^2 K \), where \( K \) is a diagonal matrix: \( K = \text{diag}(k_1, \ldots, k_n) \).
Let \( Y_w = K^{-1/2}Y \). Show that \( \text{D}(Y_w) = \sigma^2 I \).
Show that applying the method of least squares to \( Y_w \) gives an estimator \( \hat{\Theta}_w \) such that \( A'K^{-1}A\hat{\Theta}_w = A'K^{-1}Y_w \).
Let \( \hat{\Theta}_u \) be such that \( A'\hat{\Theta}_u = A'Y \). Which estimator is better? Why?

12. The treatments T₁, T₂ and T₃ were each applied at five levels of the variable x. The following yields (\( y_{ij} \)) were obtained:

<table>
<thead>
<tr>
<th>Total</th>
<th>14</th>
<th>1069.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>43.20</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>34.00</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>32.00</td>
</tr>
</tbody>
</table>

Individual 95% CIs For Mean Based on Pooled StDev

Pooled StDev = 7.707

Since \( P=0.088 \), we conclude that there is no significant difference between the treatments.
Do you agree with this analysis? If so, explain why, and give a 95% confidence interval for the effect of treatment A, i.e. the difference between group 1 and group 3. If not, explain why, and propose an alternative analysis indicating the likely outcome of your analysis. (6 marks)

For these data: \( n = 15 \), \( \Sigma x = 0 \), \( \Sigma y = 180 \), \( \Sigma x^2 = 30 \), \( \Sigma xy = 60 \), \( \Sigma y^2 = 2448 \).

(a) Fit the model \( \text{E}(Y_{ij}) = \alpha_i + \beta x_j \), \( \text{var}(Y_{ij}) = \sigma^2 \); i.e., estimate the parameters in the model.
Note: For this model, the mean yields are as given in the following table:

<table>
<thead>
<tr>
<th>x = -2</th>
<th>x = -1</th>
<th>x = 0</th>
<th>x = 1</th>
<th>x = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>T₁</td>
<td>( \alpha_1 - 2\beta )</td>
<td>( \alpha_1 - \beta )</td>
<td>( \alpha_1 + \beta )</td>
<td>( \alpha_1 + 2\beta )</td>
</tr>
<tr>
<td>T₂</td>
<td>( \alpha_2 - 2\beta )</td>
<td>( \alpha_2 - \beta )</td>
<td>( \alpha_2 + \beta )</td>
<td>( \alpha_2 + 2\beta )</td>
</tr>
<tr>
<td>T₃</td>
<td>( \alpha_3 - 2\beta )</td>
<td>( \alpha_3 - \beta )</td>
<td>( \alpha_3 + \beta )</td>
<td>( \alpha_3 + 2\beta )</td>
</tr>
</tbody>
</table>

(b) Find a 95% confidence interval for \( \beta \).
(c) Find a 95% confidence interval for the mean yield with treatment 3 and \( x = 2 \). (10 marks)
13. (a) A particular data set consists of a total of \( n = 25 \) observations. These observations can be assumed to have been obtained from independent and normally distributed random variables with equal variances. For these data \( \sum y_i^2 = 31546 \).

A general model \( H_1 \) with eight essential parameters is fitted to these data. It is found that the sum of squares for this model, \( SS(H_1) = \hat{\theta}'A'y = 31410 \).

The model \( H_0 \) is a special case of \( H_1 \) with five essential parameters. Fitting this model gives \( SS(H_0) = \hat{\theta}'A_0'y = 31320 \).

Use this information to test the goodness of fit of the model \( H_0 \).

(b) A randomised block experiment to compare three treatments was conducted in eight blocks with six plots per block. Each treatment was replicated twice in each block. It can be assumed that the observations are independent and normally distributed with equal variances.

i. Complete the following ANOVA table and hence test the significance of the treatment effects.

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocks</td>
<td>...</td>
<td>154</td>
</tr>
<tr>
<td>treatments</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>error</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>total</td>
<td>...</td>
<td>254</td>
</tr>
</tbody>
</table>

ii. Find a 95% confidence interval for the error variance.

iii. It is required to estimate the difference in effects of treatment 3 and treatment 1. This is done by comparing the average yield with treatment 1 and the average yield with treatment 3. The resulting estimate is 2.3. Find its standard error and hence obtain a 95% confidence interval for this difference. (12 marks)

Revision Exercises 5

1. (a) A random sample of fifteen observations is obtained on \( X \sim \text{Pn}(4.5) \). Find the probability that the sample contains at least two 1s.

(b) A random sample of 10 observations is obtained on \( W \), where \( W \) has cdf

\[
F(w) = 1 - (w + 1)^{-1} \quad (w > 0),
\]

Use tables of the Binomial pmf to obtain numerical values for

i. \( \text{Pr}(W_{(1)} > 1) \);

ii. \( \text{Pr}(W_{(10)} < 9) \). (8 marks)

2. Suppose that \( Z \) has a standard Logistic distribution, so that \( Z \) has cdf and pdf given by:

\[
F(z) = \frac{e^z}{1 + e^z} \quad \text{and} \quad f(z) = \frac{e^z}{(1 + e^z)^2}.
\]

i. Show that the \( q \)-quantile of \( Z \) is given by \( \ell(q) = \ln\left(\frac{q}{1-q}\right) \).

A random sample of \( n \) observations on \( Y \) is obtained. A Logistic model for \( Y \) is proposed: i.e. that \( Y \sim \theta + \phi Z \). To check on this model, the points \( \{(\ell(k/n+1), y(k)), k = 1, 2, \ldots, n\} \) are plotted.

ii. Explain how this plot provides

a. a check on the proposed model; and

b. estimates of \( \theta \) and \( \phi \).

iii. If the model is correct, indicate in a rough sketch what the plot might look like, showing the estimates of \( \theta \) and \( \phi \) on your sketch.

iv. For a sample of \( n = 100 \) on \( Y \), specify the approximate distribution of the sample median. (12 marks)
3. (a) If \( U_n = \frac{1}{n} \text{Pn}(ne^\theta) \), show that \( \ln U_n \) is a consistent estimator of \( \theta \).

(b) A random sample of \( n = 10 \) observations is obtained on \( X \overset{d}{=} \text{Pn}(\lambda) \) for which \( \sum_{i=1}^{10} x_i = 24 \). Use the following MINITAB output to deduce a 95% confidence interval for \( \lambda \).

MTB > pone 10 24
*ERROR* number of successes must be <= number of trials.
MTB > pone 100 24
Sample X N Sample p 95.0% CI
1 24 100 0.240000 (0.160225, 0.335735)
MTB > pone 1000 24
Sample X N Sample p 95.0% CI
1 24 1000 0.024000 (0.015436, 0.035501)
MTB > pone 10000 24
Sample X N Sample p 95.0% CI
1 24 10000 0.002400 (0.001538, 0.003569)
MTB > pone 100000 24
Sample X N Sample p 95.0% CI
1 24 100000 0.000240 (0.000154, 0.000357)

(c) Suppose that \( T = \theta Z \), where \( Z \) is a continuous random variable with selected quantiles as given in the table below:

<table>
<thead>
<tr>
<th>( q )</th>
<th>0.025</th>
<th>0.05</th>
<th>0.10</th>
<th>0.90</th>
<th>0.95</th>
<th>0.975</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_q(Z) )</td>
<td>0.25</td>
<td>0.4</td>
<td>0.6</td>
<td>0.75</td>
<td>0.8</td>
<td>0.875</td>
</tr>
</tbody>
</table>

If \( t_{\text{obs}} = 60 \),

i. find a (two-sided) 95% confidence interval for \( \theta \);

ii. find a (one-sided) 95% lower confidence limit for \( \theta \).

(d) If \( X \overset{d}{=} \text{Bi}(n, p) \) and \( n = 100 \), \( x_{\text{obs}} = 20 \), find an approximate 95% confidence interval for \( p \).

(e) A random sample of \( n = 14 \) observations from a normal population with variance \( \sigma^2 \) gives sample variance \( s^2 = 5.0 \). Find a 95% confidence interval for \( \sigma^2 \).

(14 marks)

4. (a) The table below gives the \( \text{Pn}(\lambda) \) cdf for \( \lambda = 4, 5, 6, 7, 8, 9, 10 \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0183</td>
<td>0.0067</td>
<td>0.0025</td>
<td>0.0009</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>0.0916</td>
<td>0.0404</td>
<td>0.0174</td>
<td>0.0073</td>
<td>0.0030</td>
<td>0.0012</td>
<td>0.0005</td>
</tr>
<tr>
<td>2</td>
<td>0.2381</td>
<td>0.1247</td>
<td>0.0620</td>
<td>0.0296</td>
<td>0.0138</td>
<td>0.0062</td>
<td>0.0028</td>
</tr>
<tr>
<td>3</td>
<td>0.4335</td>
<td>0.2650</td>
<td>0.1512</td>
<td>0.0818</td>
<td>0.0424</td>
<td>0.0212</td>
<td>0.0103</td>
</tr>
<tr>
<td>4</td>
<td>0.6288</td>
<td>0.4405</td>
<td>0.2851</td>
<td>0.1730</td>
<td>0.0996</td>
<td>0.0550</td>
<td>0.0293</td>
</tr>
<tr>
<td>5</td>
<td>0.7851</td>
<td>0.6160</td>
<td>0.4457</td>
<td>0.3007</td>
<td>0.1912</td>
<td>0.1157</td>
<td>0.0671</td>
</tr>
<tr>
<td>6</td>
<td>0.8893</td>
<td>0.7622</td>
<td>0.6063</td>
<td>0.4497</td>
<td>0.3134</td>
<td>0.2068</td>
<td>0.1301</td>
</tr>
<tr>
<td>7</td>
<td>0.9489</td>
<td>0.8666</td>
<td>0.7440</td>
<td>0.5987</td>
<td>0.4530</td>
<td>0.3239</td>
<td>0.2202</td>
</tr>
<tr>
<td>8</td>
<td>0.9786</td>
<td>0.9319</td>
<td>0.8472</td>
<td>0.7291</td>
<td>0.5925</td>
<td>0.4557</td>
<td>0.3328</td>
</tr>
<tr>
<td>9</td>
<td>0.9919</td>
<td>0.9682</td>
<td>0.9161</td>
<td>0.8305</td>
<td>0.7166</td>
<td>0.5874</td>
<td>0.4579</td>
</tr>
<tr>
<td>10</td>
<td>0.9972</td>
<td>0.9863</td>
<td>0.9574</td>
<td>0.9015</td>
<td>0.8159</td>
<td>0.7060</td>
<td>0.5830</td>
</tr>
<tr>
<td>11</td>
<td>0.9991</td>
<td>0.9945</td>
<td>0.9799</td>
<td>0.9467</td>
<td>0.8881</td>
<td>0.8030</td>
<td>0.6968</td>
</tr>
<tr>
<td>12</td>
<td>0.9997</td>
<td>0.9980</td>
<td>0.9912</td>
<td>0.9730</td>
<td>0.9362</td>
<td>0.8758</td>
<td>0.7916</td>
</tr>
<tr>
<td>13</td>
<td>0.9999</td>
<td>0.9993</td>
<td>0.9964</td>
<td>0.9872</td>
<td>0.9658</td>
<td>0.9261</td>
<td>0.8645</td>
</tr>
<tr>
<td>14</td>
<td>1.0000</td>
<td>0.9998</td>
<td>0.9986</td>
<td>0.9943</td>
<td>0.9627</td>
<td>0.9585</td>
<td>0.9166</td>
</tr>
<tr>
<td>15</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9995</td>
<td>0.9976</td>
<td>0.9918</td>
<td>0.9780</td>
<td>0.9513</td>
</tr>
<tr>
<td>16</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9998</td>
<td>0.9990</td>
<td>0.9963</td>
<td>0.9889</td>
<td>0.9730</td>
</tr>
<tr>
<td>17</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9996</td>
<td>0.9984</td>
<td>0.9947</td>
<td>0.9857</td>
</tr>
<tr>
<td>18</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9993</td>
<td>0.9976</td>
<td>0.9928</td>
</tr>
<tr>
<td>19</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9997</td>
<td>0.9989</td>
<td>0.9965</td>
</tr>
<tr>
<td>20</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9999</td>
<td>0.9996</td>
<td>0.9984</td>
</tr>
</tbody>
</table>

It is assumed that \( X \overset{d}{=} \text{Pn}(\theta) \). To test \( H_0: \theta = 4 \) against the alternative that \( \theta = 10 \), the decision rule is to reject \( H_0 \) if \( X \geq 8 \). The observation \( x_{\text{obs}} = 9 \) is obtained.

i. Specify the size, the power and the \( P \)-value.

ii. Explain why the exact (one-sided) 95% lower confidence limit for \( \theta \) is between 4 and 5.

iii. Use an appropriate normal approximation to specify an approximate (one-sided) 95% lower confidence limit for \( \theta \).
(b) The random variable $Y$ has pdf

$$f(y; \theta) = \theta e^{-\theta y} \quad (y > 0).$$

Derive the form of the likelihood ratio test of $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$, where $\theta_1 > \theta_0$. Is this test uniformly most powerful? Explain. (11 marks)

5. (a) Suppose that, based on a random sample of $n$ observations, the estimator $T_n$ is such that $E(T_n) = \theta + \frac{1}{n}$ and $\text{var}(T_n) = \frac{1}{n}(\theta^2 + 1)$. It is also known that the expected information contained in one observation $i(\theta) = \frac{1}{\theta^2}$. Is $T_n$ consistent? unbiased? asymptotically unbiased? efficient? asymptotically efficient? Give brief explanations for your answers.

(b) A random sample of $n = 18$ observations on $X$ gives sample mean and variance $\bar{x} = 21.0$ and $s^2 = 45.0$. If $E(X) = 1 + 4\theta$ and $\text{var}(X) = 2\theta^2$, where $\theta > 0$, find the method of moments estimate of $\theta$ and its standard error.

(c) The likelihood function for data set $D$ is given by

$$L(\theta) = k \theta^{12} e^{-4\theta} e^{-\frac{1}{2} \theta^2} \quad (\theta > 0).$$

Find the maximum likelihood estimate $\hat{\theta}$ and its standard error $\text{se}(\hat{\theta})$. Specify an approximate 95% confidence interval for $\theta$.

State the asymptotic results used to obtain your approximations. (13 marks)

6. Independent samples are obtained from normal populations $X_1 \overset{d}{=} N(\mu_1, \sigma_1^2)$ and $X_2 \overset{d}{=} N(\mu_2, \sigma_2^2)$, with the following results:

- $n_1 = 8; \quad \bar{x}_1 = 80, \quad s_1^2 = 40$;
- $n_2 = 8; \quad \bar{x}_2 = 50, \quad s_2^2 = 32$.

i. Find a 95% confidence interval for $\sigma_1/\sigma_2$; and verify that the hypothesis $\sigma_1^2 = \sigma_2^2$ would be accepted.

ii. Specify the pooled variance estimate based on both samples; and state its distribution under the assumption that $\sigma_1^2 = \sigma_2^2$.

iii. Using the pooled variance estimate, obtain a 95% confidence interval for $\mu_1 - \mu_2$, and specify the value of the $t$-statistic used to test $\mu_1 = \mu_2$.

iv. Give values for the entries indicated by asterisks (** in the one-way analysis of variance table below for these data:

<table>
<thead>
<tr>
<th>source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>between samples</td>
<td>**</td>
<td>..</td>
<td>..</td>
<td>**</td>
</tr>
<tr>
<td>within samples</td>
<td>**</td>
<td>..</td>
<td>..</td>
<td>**</td>
</tr>
<tr>
<td>total</td>
<td>..</td>
<td>..</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(12 marks)

7. Four treatments, labelled [00], [01], [10] and [11] are each applied to ten experimental units with the following results:

<table>
<thead>
<tr>
<th>trt</th>
<th>av</th>
<th>sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>[00]</td>
<td>-1</td>
<td>-10</td>
</tr>
<tr>
<td>[01]</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>[10]</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

i. Given that $\sum y^2 = 244$, complete the following analysis of variance table and hence test for the significance of the differences between treatment effects.

(10 marks)
source                  df    SS    MS   F   P
---                     --    ---  ---  --- ---
treatments             .     .    .    .   .
error                  .     .    .    .   .
total                  .     .    .    .   .

ii. When the analysis is done in MINITAB, the following additional output is generated:

Individual 95% CIs For Mean
Based on Pooled StDev
n  mean   sd  --------+---------+---------+---------+--
1  10 -1.0 2.0  (-------*-------)
2  10  1.0 1.9  (-------*-------)
3  10  2.0 1.8  (-------*--------)
4  10  2.0 2.2  (-------*--------)
--------+---------+---------+---------+--
-1.6  0.0  1.6  3.2

Fisher's pairwise comparisons
Family error rate = 0.197;
Individual error rate = 0.050
1 2 3
2 -3.814 -0.186
3 -4.814 -2.814 -1.186 0.814
4 -4.814 -2.814 -1.814 1.814

Give a brief explanation of this output; in particular explain the meaning of “family error rate”.

iii. Define a contrast that compares the average yield for treatments [10] and [11] with the average yield for treatments [00] and [01]. Compute the sum of squares corresponding to this contrast and hence or otherwise test the significance of this contrast.

8. The following data set is a random sample of 400 observations from a population with distribution D.

The sample has been re-ordered by magnitude from smallest to largest.

i. Construct a boxplot for this data set and briefly describe the main characteristics of the data set.

ii. If \( Z \sim Bi(400,0.5) \), verify that \( Pr(180 \leq Z \leq 220) \approx 0.960 \). Hence find a 96% confidence interval for the median of the population \( D \).

iii. Give rough values for the sample mean and the sample standard deviation of the above sample. Explain briefly how you obtained these values. Use these rough values to specify an approximate 95% confidence interval for the mean of the population \( D \). Comment on the differences between this interval and the confidence interval obtained in (b).

iv. Use this sample to specify an approximate 95% prediction interval for a future observation from the population \( D \).
9. You have a sample of ten observations and your client wants you to examine the hypothesis that the population median is equal to 30. Just to make sure of the result you carry out several tests. The data and the output from three tests are given in the MINITAB output below:

```
data x: 13 14 16 18 19 23 25 26 31 44

MTB > stest 30 x Sign test of median = 30 versus .neq. 30
   n   below  equal  above  P-value  median
x 10     8      0      2  0.109   21.00

MTB > wtest 30 x Signed rank sum test of median = 30 versus .neq. 30
   Wilcoxon estimated
   n   n* statistic  P-value  median
x 10  10     8.5  0.059  22.00

MTB > ttest 30 x t-test of mu = 30 vs mu .neq. 30
   n  Mean  StDev  t  P-Value
x 10  22.90  9.36 -2.40  0.040
```

But now you have a problem: the tests indicate different conclusions. Summarise what the output is saying in a line or two. Can you reconcile the results of the tests? What are you going to tell the client?

The following MINITAB output, which gives the corresponding confidence intervals, may be used in your answer.

```
MTB > sint x Sign confidence interval
   achieved
   n  median  confidence  conf interval
x 10  21.00    97.9     ( 14.0, 31.0)

MTB > wint x Wilcoxon signed rank confidence interval
   estimated  achieved
   n  median  confidence  conf interval
x 10  22.0     94.7     ( 16.0, 30.0)

MTB > tint x t confidence interval
   n  Mean  StDev  95% conf interval
x 10  22.90  9.36     ( 16.2, 29.6)
```

(6 marks)

10. (a) Four independent random variables \( Y_1, Y_2, Y_3 \) and \( Y_4 \) have means \( \alpha, \beta, \alpha + \beta \) and \( \alpha - \beta \) respectively and variances equal to \( \sigma^2 \). Show that the least squares estimator of \( \alpha \) is given by \( \hat{\alpha} = \frac{1}{2}(Y_1 + Y_3 + Y_4) \). Verify that \( \hat{\alpha} \) is unbiased for \( \alpha \) and find an expression for its variance.

(b) What does ‘LS is BLUE’ mean?

(c) A particular data set consists of a total of \( n = 26 \) observations. These observations can be assumed to have been obtained from independent and normally distributed random variables with equal variances. For these data \( \Sigma y_i^2 = 2584 \).

A general model \( H_1 \) with six essential parameters is fitted to these data. It is found that the sum of squares for this model, \( SS(H_1) = \hat{\theta}_1 A_1 y = 2544 \).

The model \( H_0 \) is a special case of \( H_1 \) with two essential parameters. Fitting this model gives \( SS(H_0) = \hat{\theta'}_0 A_0 y = 2512 \).

Use this information to test the goodness of fit of the model \( H_0 \). (10 marks)
11. (a) Give an outline derivation (algebraic or geometric or other) of the normal equations \( A'\hat{A} = A'y \) for the least squares estimate of \( \theta \) based on observation of \( Y \) where \( E(Y) = \hat{A} \theta \) and \( D(Y) = \sigma^2 I \).

(b) The normal equations for a linear model for a sample of \( n = 15 \) independent observations each with variance \( \sigma^2 \) are:

\[
\begin{bmatrix}
10 & 0 & 0 \\
0 & 7 & 2 \\
0 & 2 & 2
\end{bmatrix}
\begin{bmatrix}
\hat{\alpha} \\
\hat{\beta} \\
\hat{\gamma}
\end{bmatrix}
= 
\begin{bmatrix}
30 \\
16 \\
6
\end{bmatrix}
\]

i. Obtain estimates of \( \alpha \), \( \beta \) and \( \gamma \).

ii. Given that \( y'y = 158 \) obtain an estimate of \( \sigma^2 \).

iii. Find \( se(\hat{\alpha}) \) and \( se(\hat{\beta} + \hat{\gamma}) \).

(12 marks)

12. (a) A randomised block experiment to compare four treatments is to be conducted in five blocks with eight plots per block. Each treatment is to be replicated twice in each block. Explain in detail how the treatments would be assigned to the plots.

(b) The experiment described in (a) is carried out, resulting in the analysis of variance given below. For the analysis, it can be assumed that the observations are independent and normally distributed with equal variances.

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>blocks</td>
<td>...</td>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>treatments</td>
<td>...</td>
<td>.....</td>
<td>...</td>
</tr>
<tr>
<td>error</td>
<td>...</td>
<td>.....</td>
<td>20</td>
</tr>
<tr>
<td>total</td>
<td>...</td>
<td>2440</td>
<td></td>
</tr>
</tbody>
</table>

i. Complete the ANOVA table and hence test the significance of the treatment effects.

ii. Given that the difference in average yields with treatment 1 and treatment 2 is 10.0, derive a 95% confidence interval for the difference in effects of treatments 1 and 2. (10 marks)