1. Let \( \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \) be a random vector with a multivariate normal distribution, with mean \( \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \) and covariance matrix \( \begin{pmatrix} 16 & 6 & k \\ 6 & 9 & k \\ k & k & 16 \end{pmatrix} \).

(a) What is the marginal distribution of \( X \)?

(b) What is the correlation between \( X \) and \( Y \)?

(c) Let \( U = X + Y \) and \( V = X + Z \).

(i) Find the mean vector and covariance matrix of \( U \) and \( V \).

(ii) For what value of \( k \) are \( U \) and \( V \) independent? For this value of \( k \), what is the variance of \( V \)?
2 Let $X$ be a random variable with probability density function

$$f_X(x) = \begin{cases} \frac{x + 1}{2} & -1 < x < 1 \\ 0 & \text{otherwise}. \end{cases}$$

(a) Let $Y = \sin^{-1}(X)$. Find the probability density function of $Y$. (4 marks)

(b) Let $Z = X^2$. Find the probability density function of $Z$. (5 marks)

3 Let $S$ be the square $\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ and let $X$ and $Y$ be two random variables with joint probability density function given by

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & (x, y) \in S \\ 0 & \text{otherwise}. \end{cases}$$

(a) Find the marginal probability density function of $Y$. (3 marks)

(b) Find the conditional probability density function of $X$, given that $Y = y$, assuming $0 \leq y \leq 1$. (3 marks)

(c) Let $U = XY$ and $V = X/Y$. Find the joint probability density function of $U$ and $V$, stating carefully the values for which it is non-zero. (8 marks)

4 Let $x = x_1, x_2, \ldots, x_n$ be a random sample from an Exponential distribution with parameter $\lambda > 0$.

(a) Find the likelihood of $\lambda$ given the data $x$. (2 marks)

(b) Find the maximum likelihood estimate of $\lambda$ given the data $x$. (7 marks)

(c) Let $n = 2$, and let the observations be $x_1 = 1.73$ and $x_2 = 3.03$. By considering the difference between the log likelihood at $\lambda$ and at its maximum, discuss how consistent these data are with

(i) $\lambda = 0.6$;

(ii) $\lambda = 5$. (5 marks)
Consider the linear model

\[ y_1 = \beta_0 x_1 + \beta_1 + \epsilon_1 \]
\[ y_2 = -\beta_0 + \beta_1 \left( \frac{x_2}{2} \right) + \epsilon_2 \]
\[ y_3 = \beta_1 x_3 + \epsilon_3 \]

where the random errors \( \epsilon_i \) are i.i.d. \( \sim N(0, \sigma^2) \). The sample \((x_i, y_i)\) is \((1, 0), (2, 1), (1, 1)\).

(a) Write down the model in matrix form. 

(b) We wish to test \( H_0 : \beta_0 = \beta_1 \) versus \( H_a : \beta_0 \neq \beta_1 \). Perform the F-test and report the P value in the form \( P(F_{?, ?} > ?) \). (You need to fill in the ? marks)

Consider the simple linear regression model \( y_i = \beta_0 + \beta_1 x_i + \epsilon_i \) where \( \epsilon_i \) are i.i.d. \( \sim N(0, \sigma^2) \), \( i = 1, 2, \ldots, n \). We wish to test \( H_0 : \beta_1 = 0 \) versus \( H_a : \beta_1 \neq 0 \). There are 2 possible ways to test this: the t-test and the F-test.

(a) Note that the T statistic for the t-test is given by

\[ T = \frac{\hat{\beta}_1}{\text{estimated standard error of } \hat{\beta}_1}. \]

Show that the F statistic is the square of the T statistic. (7 marks)

Reminder: You may use without proof that the estimators for the simple linear regression model are \( \hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \) and \( \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \), where \( s_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \) and \( s_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \). We also know that

\[ (X^T X)^{-1} = \begin{pmatrix}
\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} & -\frac{\bar{x}}{s_{xx}} \\
-\frac{\bar{x}}{s_{xx}} & \frac{1}{s_{xx}}
\end{pmatrix} \]

(b) Explain why both tests would give the same P-value. (3 marks)
The one-way ANOVA model was used to test the effectiveness of 3 fertilizers on the growth of a certain plant. 9 specimens of the plant were divided randomly into 3 groups corresponding to the 3 different treatments. The growth of the plant over a 6 month period is reported below.

\[
\text{Trt1: 1, 2, 3} \quad \text{Trt2: 3, 4, 5} \quad \text{Trt3: 5, 4, 4}
\]

An analysis was carried out in R to see if the number of parameters could be reduced. The output is shown below.

```r
> growth<-c(1,2,3,4,5,5,4,4)
> trt<-as.factor(c("1","1","1","2","2","3","3","3"))
> trt23<-as.factor(c("a","a","a","b","b","b","b","b"))
> trt13<-as.factor(c("a","a","a","b","b","a","a","a"))
> trt12<-as.factor(c("a","a","a","a","a","a","b","b","b"))
> lmfull<-lm(growth~trt)
> lm23<-lm(growth~trt23)
> lm12<-lm(growth~trt12)
> lm13<-lm(growth~trt13)
> lmreduced<-lm(growth~1)

> anova(lm23,lmfull)
          Res.Df RSS Df SumofSq    F   Pr(> F)
1           7 4.8333
2           6 4.6667  1  0.16667 0.2143 0.6597

> anova(lm13,lmfull)
          Res.Df RSS Df SumofSq   F   Pr(> F)
1           7 12.8333
2           6 4.6667  1  8.16667 10.5 0.01768

> anova(lm12,lmfull)
          Res.Df RSS Df SumofSq   F   Pr(> F)
1           7 10.6667
2           6 4.6667  1   6.7143 0.0321

> anova(lmreduced,lm23)
          Res.Df RSS Df SumofSq   F   Pr(> F)
1           8 14.2222
2           7 4.8333  1  9.3889 13.598 0.007782

> anova(lmreduced,lm12)
          Res.Df RSS Df SumofSq   F   Pr(> F)
1           8 14.2222
2           7 10.667  1  3.5556 2.3333 0.1705
```
> anova(lmreduced, lm13)
  Res.Df RSS Df Sum.ofSq  F  Pr(> F)
 1  8 14.222
 2  7 12.833  1  1.3889 0.7576 0.4129

> anova(lmreduced, lmfull)
  Res.Df RSS Df Sum.ofSq  F  Pr(> F)
 1  8 14.2222
 2  6  4.6667  2  9.5556 6.1429 0.03533

(a) Write down clearly the models being considered. Draw a nested diagram to show the relationship between the models. 

(b) Use hypothesis testing to find the most suitable model. Use size 0.05 for all your tests.

A two-way ANOVA model was used to test the effectiveness of combinations of 3 fertilizers $F_1, F_2, F_3$ and 2 insecticides $I_1, I_2$ on the growth of a certain plant. 6 specimens of the plant were randomly assigned to the six possible combinations. The growth of the plant after six months is reported below.

<table>
<thead>
<tr>
<th>Fertilizer</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insecticide</td>
<td>$I_1$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>$I_2$</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

Sketch two suitable plots to illustrate the presence or absence of interaction between the two factors. What do the plots suggest?

End of Question Paper