THE UNIVERSITY OF MELBOURNE

Student number	

Semester 1 Assessment, 2023

School of Mathematics and Statistics

MAST30025 Linear Statistical Models Assignment 3

Submission deadline: **Friday May 26, 5pm**

This assignment consists of 12 pages (including this page) with 5 questions and 46 total marks

Instructions to Students

Writing

- This assignment is worth 7% of your total mark.
- You may choose to either typeset your assignment in LaTeX, or handwrite and scan it to produce an electronic version.
- You may use R for this assignment, including the `lm` function unless otherwise specified. If you do, include your R commands and output.
- Write your answers on A4 paper. Page 1 should only have your student number, the subject code and the subject name. Write on one side of each sheet only. Each question should be on a new page. The question number must be written at the top of each page.

Scanning and Submitting

- Put the pages in question order and all the same way up. Use a scanning app to scan all pages to PDF. Scan directly from above. Crop pages to A4.
- Submit your scanned assignment as a single PDF file and carefully review the submission in Gradescope. Scan again and resubmit if necessary.

Question 1 (7 marks)

Let A be an $n \times p$ matrix with $n \geq p$.

(a) Show that $r(A^c A) = r(A)$.

(b) Show that $I - A(A^T A)^c A^T$ is idempotent.

(c) Show that $r(I - A(A^T A)^c A^T) = n - r(A)$.

(a)

$$r(A) = r(AA^c A) \leq r(A^c A) \leq r(A).$$

(b) This follows from the idempotency of $A(A^T A)^c A^T$ (proved in lectures) and the fact that if H is idempotent, then $I - H$ is also idempotent.

(c)

$$\begin{aligned} r(I - A(A^T A)^c A^T) &= tr(I - A(A^T A)^c A^T) \\ &= tr(I_n) - tr(A(A^T A)^c A^T) \\ &= n - r(A(A^T A)^c A^T) \\ &= n - r(A). \end{aligned}$$

Question 2 (11 marks)

We study the amount of rotting of a potato exposed to a variety of levels of oxygen, and a variety of temperatures. A small experiment is conducted and the following data obtained:

	Oxygen level		
Temperature	1	2	3
10	13	10	15
	11	4	2
	3	7	7
16	26	15	20
	19	22	24
	24	18	8

For this question, you may not use the `lm` or `ginv` functions in R.

(a) Plot an interaction plot of the data. Does there appear to be interaction?

(b) Fit an additive model, outputting your design matrix. Estimate the common variance.

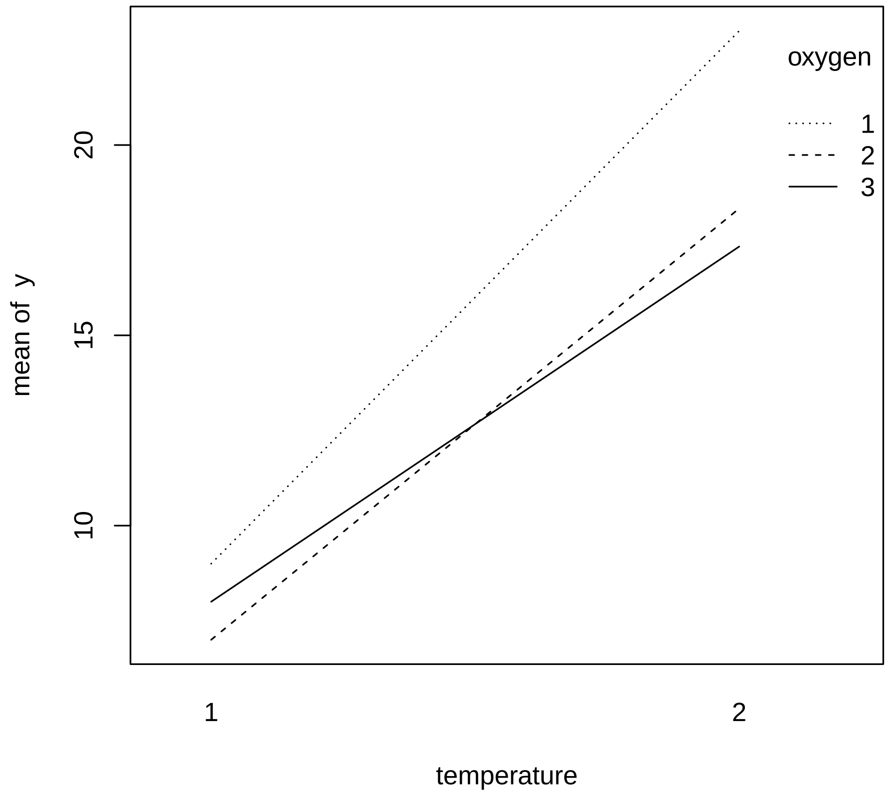
(c) Calculate a 95% confidence interval for the difference between the temperature effects.

(d) Test the hypothesis that oxygen level has no effect on rotting at the 5% significance level.

(e) Suppose we are interested in the effect of oxygen level only, but know that temperature affects the results, so we include it in our model. What type of design would this study be?

(a)
```
> n <- 18
> r <- 4
> y <- c(13,10,15,11,4,2,3,7,7,26,15,20,19,22,24,24,18,8)
> oxygen <- rep(1:3,6)
> temperature <- rep(1:2,each=9)
```

```
> interaction.plot(temperature, oxygen, y)
```



There does not appear to be much interaction.

(b)
```
> X <- matrix(0,n,6)
> X[,1] <- 1
> X[cbind(1:n,oxygen+1)] <- 1
> X[cbind(1:n,temperature+4)] <- 1
```

```
> X
      [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]    1    1    0    0    1    0
 [2,]    1    0    1    0    1    0
 [3,]    1    0    0    1    1    0
 [4,]    1    1    0    0    1    0
 [5,]    1    0    1    0    1    0
 [6,]    1    0    0    1    1    0
 [7,]    1    1    0    0    1    0
 [8,]    1    0    1    0    1    0
 [9,]    1    0    0    1    1    0
[10,]    1    1    0    0    0    1
[11,]    1    0    1    0    0    1
[12,]    1    0    0    1    0    1
[13,]    1    1    0    0    0    1
[14,]    1    0    1    0    0    1
[15,]    1    0    0    1    0    1
[16,]    1    1    0    0    0    1
[17,]    1    0    1    0    0    1
[18,]    1    0    0    1    0    1
> XtXc <- matrix(0,6,6)
> XtXc[2:5,2:5] <- solve((t(X)%*%X)[2:5,2:5])
> b <- XtXc%*%t(X)%*%y
> (s2 <- sum((y-X%*%b)^2)/(n-r))
[1] 26.12698
```

(c)

```
> tt <- c(0,0,0,0,1,-1)
> tt%*%b + c(-1,1)*qt(0.975,n-r)*sqrt(s2*t(tt)%*%XtXc%*%tt)
[1] -16.723555  -6.387556
```

(d)
```
> C <- matrix(c(0,1,-1,0,0,0,0,0,1,-1,0,0),2,6,byrow=T)
> Fstat <- (t(C%*%b) %*%
+             solve(C %*% XtXc %*% t(C)) %*% C%*%b)/2/s2
> pf(Fstat, 2, n-r, lower=F)
```

```
          [,1]
[1,] 0.4481124
```

We do not reject this hypothesis.

(e) It would be a complete block design.

Question 3 (5 marks)

Consider a linear model with only categorical predictors, written in matrix form as $\mathbf{y} = X_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$. Now suppose we add some continuous predictors, resulting in an expanded model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

Now consider a quantity $\mathbf{t}^T\boldsymbol{\beta}$, where $\mathbf{t}^T = [\mathbf{t}_1^T|\mathbf{t}_2^T]$ is partitioned according to the categorical and continuous predictors. Show that if $\mathbf{t}_1^T\boldsymbol{\beta}_1$ is estimable in the first model, then $\mathbf{t}^T\boldsymbol{\beta}$ is estimable in the second model.

If you write $X = [X_1|X_2]$, you may assume that $r(X) = r(X_1) + r(X_2)$.

Hint: Use Theorems 6.9 and 6.3. For any vectors $\mathbf{z}_1$ and $\mathbf{z}_2$, you can write

$$\left[\begin{array}{cc|c} X_1^TX_1 & X_1^TX_2 & X_1^TX_1\mathbf{z}_1 \\ X_2^TX_1 & X_2^TX_2 & X_2^TX_2\mathbf{z}_2 \end{array}\right] = \left[\begin{array}{cc} X_1^T & \mathbf{0} \\ \mathbf{0} & X_2^T \end{array}\right]\left[\begin{array}{cc|c} X_1 & X_2 & X_1\mathbf{z}_1 \\ X_1 & X_2 & X_2\mathbf{z}_2 \end{array}\right].$$

We have

$$X^TX = \left[\begin{array}{c|c} X_1^TX_1 & X_1^TX_2 \\ \hline X_2^TX_1 & X_2^TX_2 \end{array}\right].$$

We need to show that $X^TX\mathbf{z} = \mathbf{t}$ has a solution for $\mathbf{z}$. To do this we show that

$$r\left(\left[X^TX\middle|\mathbf{t}\right]\right) = r(X^TX) = r(X) = r(X_1) + r(X_2).$$

We first observe that since $\mathbf{t}_1^T\boldsymbol{\beta}_1$ is estimable in the first model, there exists a solution $\mathbf{z}_1$ to the system $X_1^TX_1\mathbf{z} = \mathbf{t}_1$. Likewise, there exists a solution $\mathbf{z}_2$ to the system $X_2^TX_2\mathbf{z} = \mathbf{t}_2$, since X_2 can be assumed to be of full rank.

Now, it is immediate that

$$r\left(\left[X^TX\middle|\mathbf{t}\right]\right) \geq r(X^TX).$$

To show the reverse inequality, we have

$$\begin{aligned} r\left(\left[X^TX\middle|\mathbf{t}\right]\right) &= r\left(\left[\begin{array}{cc|c} X_1^TX_1 & X_1^TX_2 & \mathbf{t}_1 \\ X_2^TX_1 & X_2^TX_2 & \mathbf{t}_2 \end{array}\right]\right) \\ &= r\left(\left[\begin{array}{cc|c} X_1^TX_1 & X_1^TX_2 & X_1^TX_1\mathbf{z}_1 \\ X_2^TX_1 & X_2^TX_2 & X_2^TX_2\mathbf{z}_2 \end{array}\right]\right) \\ &= r\left(\left[\begin{array}{cc} X_1^T & \mathbf{0} \\ \mathbf{0} & X_2^T \end{array}\right]\left[\begin{array}{cc|c} X_1 & X_2 & X_1\mathbf{z}_1 \\ X_1 & X_2 & X_2\mathbf{z}_2 \end{array}\right]\right) \\ &\leq r\left(\left[\begin{array}{cc} X_1^T & \mathbf{0} \\ \mathbf{0} & X_2^T \end{array}\right]\right) \\ &= r(X_1) + r(X_2). \end{aligned}$$

Thus the equality is proved and $\mathbf{t}^T\boldsymbol{\beta}$ is estimable in the full model.

Question 4 (11 marks)

An experiment is conducted to investigate the distance travelled by three different toy cars on a smooth surface. Each car starts from a 16-inch long ramp tilted at different angles (in degrees), and the travel distance (in meters) is recorded. The data is given in the file `toycars.csv`, available on the LMS.

(a) Plot the data, using different colours and/or symbols for each car. What do you observe?

(b) Test for the presence of interaction between the angle and the type of toy car.

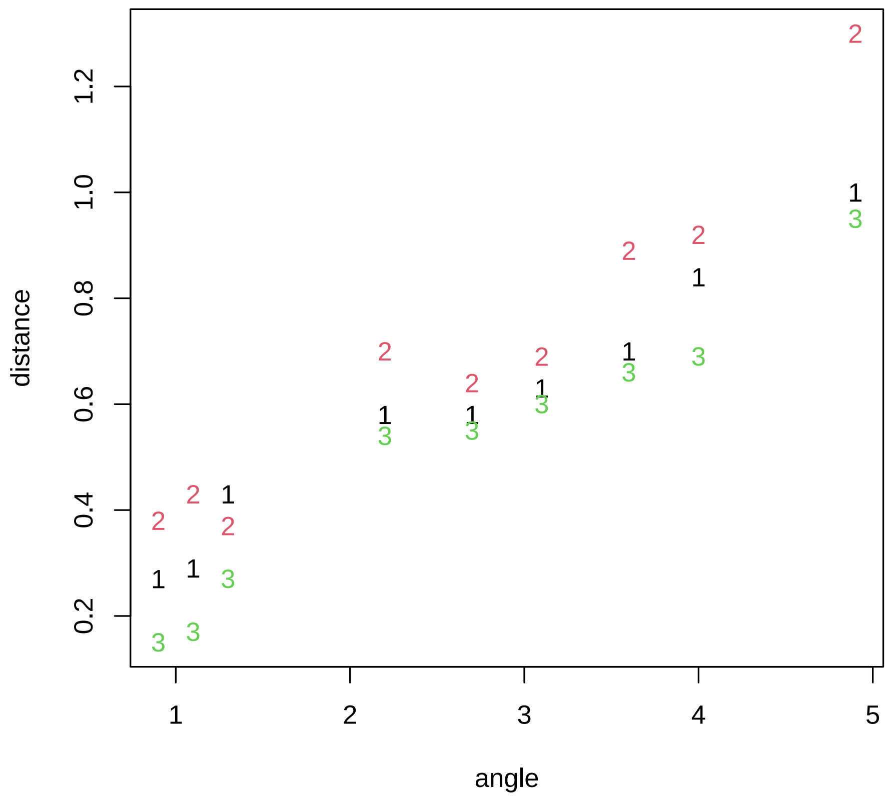
(c) Use backward elimination to select relevant variables for the data.

(d) For model from part (c), test the hypothesis that the type 1 toy car travels 0.05 meters more on average than the type 3 toy car, at the 5% significance level.

(e) In the full model with interaction, test the hypothesis that the distances travelled by the type 2 and type 3 toy cars are the same, at the 5% significance level.

(a)

```
> tcars <- read.csv('toycars.csv')
> plot(distance ~ angle, pch=as.character(car), col=car, data=tcars)
```



There appears to be a clear linear relationship between the distance and the angle. It also appears that the type of toy car has an effect on the distance it travelled, although this is less clear.

(b)
```
> tcars$car <- factor(tcars$car)
> amodel <- lm(distance ~ angle + car, data=tcars)
> imodel <- lm(distance ~ angle * car, data=tcars)
> anova(amodel, imodel)

Analysis of Variance Table

Model 1: distance ~ angle + car
Model 2: distance ~ angle * car
  Res.Df      RSS Df Sum of Sq      F Pr(>F)
1     23 0.105657
2     21 0.093271  2  0.012386 1.3944   0.27
```

Interaction is not significant.

(c)
```
> drop1(imodel, scope=~., test='F')

Single term deletions

Model:
distance ~ angle * car
          Df Sum of Sq     RSS      AIC  F value   Pr(>F)
<none>                 0.09327 -141.038
angle      1   0.44593 0.53920  -95.664 100.4023 1.87e-09 ***
car        2   0.01979 0.11307 -139.842   2.2284   0.1325
angle:car  2   0.01239 0.10566 -141.672   1.3944   0.2700
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> model2 <- lm(distance ~ angle + car, data=tcars)
> drop1(model2, scope=~., test='F')

Single term deletions

Model:
distance ~ angle + car
       Df Sum of Sq     RSS      AIC F value    Pr(>F)
<none>              0.10566 -141.672
angle   1   1.65108 1.75673  -67.774 359.416 1.547e-15 ***
car     2   0.16945 0.27511 -119.833  18.444 1.662e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The final model depends on both `angle` and `car`.

(d)
```
> library(car)
> C <- c(0,0,0,1)
> dst <- -0.05
> linearHypothesis(model2,C,dst)

Linear hypothesis test

Hypothesis:
car3 = - 0.05

Model 1: restricted model
Model 2: distance ~ angle + car

  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     24 0.11033
2     23 0.10566  1 0.0046722 1.0171 0.3237
```

We fail to reject the hypothesis at a 5% significance level.

(e)
```
> library(car)
> C <- matrix(c(0,0,1,-1,0,0,0,0,0,0,1,-1),2,6,byrow=T)
> dst <- c(0,0)
> linearHypothesis(imodel,C,dst)

Linear hypothesis test

Hypothesis:
car2 - car3 = 0
angle:car2 - angle:car3 = 0

Model 1: restricted model
Model 2: distance ~ angle * car

  Res.Df      RSS Df Sum of Sq      F    Pr(>F)
1     23 0.265625
2     21 0.093271  2   0.17235 19.403 1.689e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We reject the hypothesis at a 5% significance level.

Question 5 (12 marks)

Consider the following model with both categorical and continuous predictors:

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij}, \quad i = 1, 2, \quad j = 1, 2, 3.$$

Here, we consider μ to be a nuisance parameter.
For this question, you may not use the `lm` function in R.

(a) State the reason why μ can be regarded as a nuisance parameter.

(b) Write down the matrices X_1 and X_2.

(c) Derive the reduced design matrix $X_{2|1}$.

(d) Suppose $(x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}) = (2, 4, 8, 7, 6, 4)$ and $(y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}) = (4, 2, 10, 8, 8, 12)$. Find a solution of the reduced normal equations.

(e) Find the corresponding solution for μ for your solution to part (d). Show directly that your solutions for (μ, τ_i, β) solve the full normal equations.

(a) The model is a less than full rank model, where the model for the ith treatment is a linear model with the intercept $\mu + \tau_i$ and the slope β. That is, μ does not really have a real physical meaning in the model, and is a nuisance parameter.

(b) $X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 1 & 0 & x_{11} \\ 1 & 0 & x_{12} \\ 1 & 0 & x_{13} \\ 0 & 1 & x_{21} \\ 0 & 1 & x_{22} \\ 0 & 1 & x_{23} \end{bmatrix}$.

(c) Let $\bar{x} = \frac{1}{6}\sum_{i,j} x_{ij}$. Then

$$H_1 = X_1(X_1^T X_1)^c X_1^T = \frac{1}{6}J_6,$$

$$X_{2|1} = (I_6 - H_1)X_2$$

$$= X_2 - \begin{bmatrix} 1/2 & 1/2 & \bar{x} \\ 1/2 & 1/2 & \bar{x} \\ 1/2 & 1/2 & \bar{x} \\ 1/2 & 1/2 & \bar{x} \\ 1/2 & 1/2 & \bar{x} \\ 1/2 & 1/2 & \bar{x} \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & x_{11} - \bar{x} \\ 1/2 & -1/2 & x_{12} - \bar{x} \\ 1/2 & -1/2 & x_{13} - \bar{x} \\ -1/2 & 1/2 & x_{21} - \bar{x} \\ -1/2 & 1/2 & x_{22} - \bar{x} \\ -1/2 & 1/2 & x_{23} - \bar{x} \end{bmatrix}.$$

(d)
```
> library(MASS)
> x <- c(2, 4, 8, 7, 6, 4)
> X21 <- matrix(c(rep(1/2,3), rep(-1/2,6), rep(1/2,3),
+                  x-mean(x)),6,3)
> y <- c(4, 2, 10, 8, 8,12)
> (rlse <- ginv(t(X21)%*%X21) %*% t(X21)%*%y)

           [,1]
[1,] -1.6857143
[2,]  1.6857143
[3,]  0.6285714
```

(e) From the proof of Theorem 8.2, we have $X_1^T X_1 \mathbf{b}_1 + X_1^T X_2 \mathbf{b}_2 = X_1^T \mathbf{y}$.

```
> X1 <- rep(1,6)
> X2 <- cbind(rep(1:0, each = 3), rep(0:1, each = 3), x)
> (m <- solve(t(X1)%*%X1, t(X1)%*%y - t(X1)%*%X2%*%rlse))

         [,1]
[1,] 4.085714

> X <- cbind(X1,X2)
> round(t(X)%*%X %*% rbind(m,rlse) - t(X)%*% y, 5)

   [,1]
X1    0
      0
      0
x     0
```

End of Assignment — Total Available Marks = 46