THE UNIVERSITY OF MELBOURNE

Student ID

Semester 1 Assessment, 2016

School of Mathematics and Statistics

MAST30025 Linear Statistical Models

Writing time: 3 hours

Reading time: 15 minutes

This is NOT an open book exam

This paper consists of 8 pages (including this page)

Authorised materials:

- Scientific calculators are premitted, but not graphical calculators.
- One A4 double-sided handwritten sheet of notes.

Instructions to Students

- You must NOT remove this question paper at the conclusion of the examination.
- You should attempt all questions. Marks for individual questions are shown.
- The total number of marks available is 90.

Instructions to Invigilators

- Students must NOT remove this question paper at the conclusion of the examination.

This paper may be held in the Baillieu Library

This paper must not be removed from the examination room

Question 1 (9 marks)

(a) Let A be a square matrix and suppose that $A^k = A^{k+1}$ for some $k \geq 1$. Show that A is idempotent.

(b) Let X be an $n \times p$ matrix of full rank, where $n > p$. Show that $H = X(X^TX)^{-1}X^T$ is idempotent, and find its rank. (You may assume that H is symmetric.)

(c) Show that if a square matrix A is positive semidefinite, then its eigenvalues are non-negative.

Question 2 (10 marks) Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim MVN\left(\begin{bmatrix} a \\ -a \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \right),$$

where a is a constant.

(a) What is the distribution of $y_1 + y_2$?

(b) What is the distribution of $\frac{1}{2}\left(y_1^2 - 2y_1y_2 + y_2^2 + y_3^2\right)$?

(c) Suppose $a = 0$. For what values of c does

$$c\frac{y_1^2 - 2y_1y_2 + y_2^2 + y_3^2}{y_1^2 + 2y_1y_2 + y_2^2}$$

have an F distribution?

Question 3 (14 marks) Consider the full rank linear model, $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

(a) State the assumptions involved in fitting this model.

(b) Define the term BLUE (best linear unbiased estimator).

(c) Is it better to fit this model using the method of least squares or maximum likelihood estimation? Justify your answer.

(d) Define and explain the purpose of the leverage of a point.

(e) Explain the difference between a model relevance test and a model relevance test using a corrected sum of squares.

(f) When is a model with fewer explanatory variables more desirable or less desirable than a model with more explanatory variables?

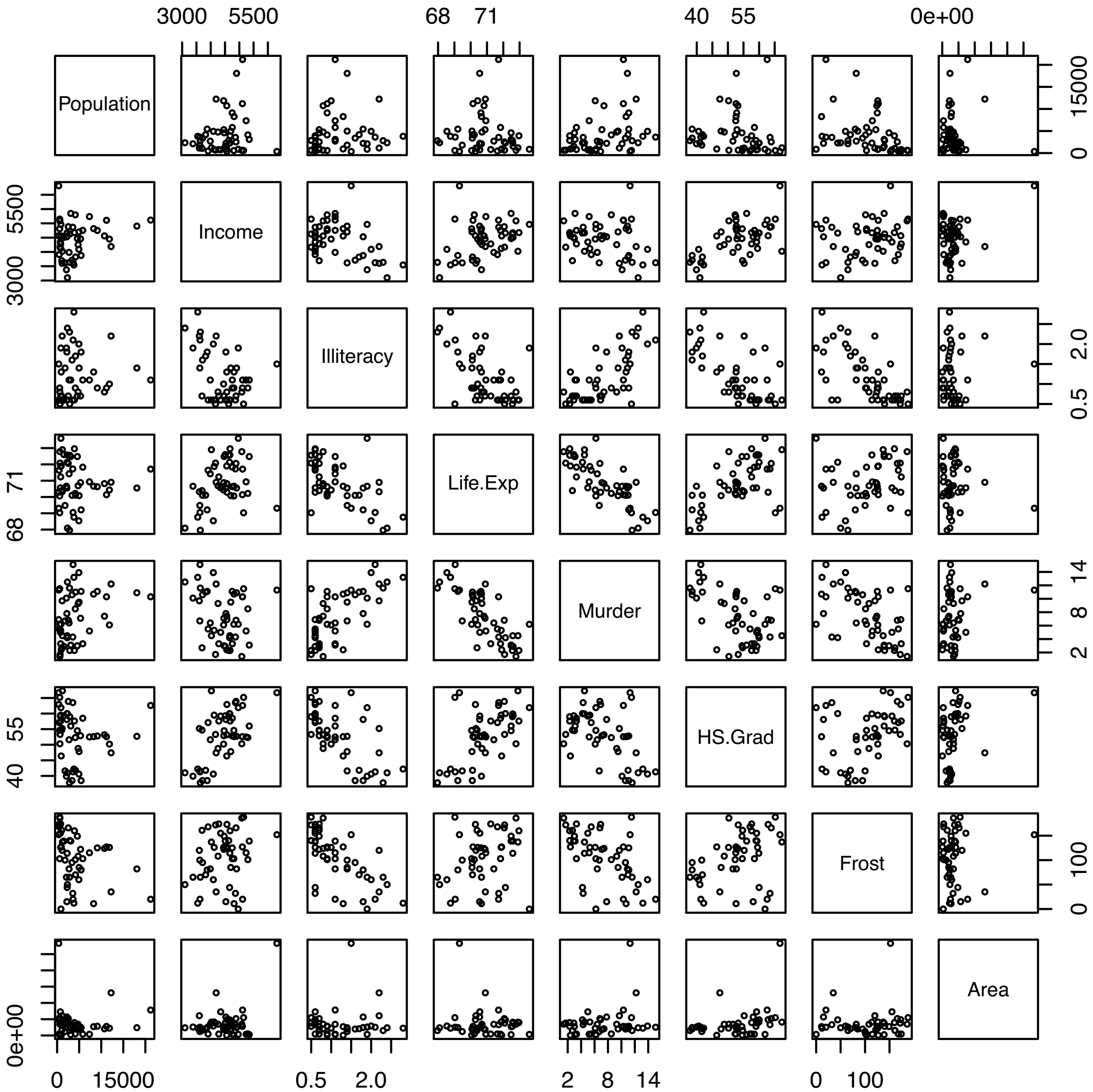
(g) Explain why the residual sum of squares SS_{Res} is not an appropriate goodness-of-fit measure for model selection.

Question 4 (17 marks) In this question, we study a dataset of 50 US states. This dataset contains the variables:

- `Population`: population estimate as of July 1, 1975
- `Income`: per capita income (1974)
- `Illiteracy`: illiteracy (1970, percent of population)
- `Life.Exp`: life expectancy in years (1969–71)
- `Murder`: murder and non-negligent manslaughter rate per 100,000 population (1976)
- `HS.Grad`: percentage of high-school graduates (1970)
- `Frost`: mean number of days with minimum temperature below freezing (1931–1960) in capital or large city
- `Area`: land area in square miles

We use linear models to model life expectancy in terms of the other variables. The following R output is produced.

```
> data(state)
> statedata <- data.frame(state.x77, row.names=state.abb, check.names=TRUE)
> pairs(statedata, cex=0.5)
```




```
> statedata$Population <- log(statedata$Population)
> statedata$Area <- log(statedata$Area)
> nullmodel <- lm(Life.Exp ~ 1, data = statedata)
> fullmodel <- lm(Life.Exp ~ ., data = statedata)
> model <- step(fullmodel, scope = ~ .)

Start:  AIC=-23.6
Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
    Frost + Area

             Df Sum of Sq    RSS      AIC
- Income      1    0.0018 22.650 -25.5934
- Illiteracy  1    0.0556 22.704 -25.4746
```

```
- Area        1    0.2106 22.859 -25.1344
<none>                    22.648 -23.5973
- Frost       1    1.2374 23.886 -22.9374
- Population  1    1.8854 24.533 -21.5992
- HS.Grad     1    2.4375 25.086 -20.4864
- Murder      1   23.2760 45.924   9.7483

Step:  AIC=-25.59
Life.Exp ~ Population + Illiteracy + Murder + HS.Grad + Frost +
    Area

             Df Sum of Sq    RSS      AIC
- Illiteracy  1    0.0556 22.705 -27.4708
- Area        1    0.2197 22.870 -27.1107
<none>                    22.650 -25.5934
- Frost       1    1.2602 23.910 -24.8862
+ Income      1    0.0018 22.648 -23.5973
- Population  1    2.1909 24.841 -22.9768
- HS.Grad     1    4.0374 26.687 -19.3918
- Murder      1   24.2130 46.863   8.7601

Step:  AIC=-27.47
Life.Exp ~ Population + Murder + HS.Grad + Frost + Area

             Df Sum of Sq    RSS     AIC
- Area        1    0.2157 22.921 -28.998
<none>                    22.705 -27.471
+ Illiteracy  1    0.0556 22.650 -25.593
+ Income      1    0.0017 22.704 -25.475
- Population  1    2.2792 24.985 -24.688
- Frost       1    2.3760 25.082 -24.495
- HS.Grad     1    4.9491 27.655 -19.612
- Murder      1   29.2296 51.935  11.899

Step:  AIC=-29
Life.Exp ~ Population + Murder + HS.Grad + Frost

             Df Sum of Sq    RSS     AIC
<none>                    22.921 -28.998
+ Area        1     0.216 22.705 -27.471
+ Illiteracy  1     0.052 22.870 -27.111
+ Income      1     0.011 22.911 -27.021
- Frost       1     2.214 25.135 -26.387
- Population  1     2.450 25.372 -25.920
- HS.Grad     1     6.959 29.881 -17.741
- Murder      1    34.109 57.031  14.578
```

```
> summary(model)

Call:
lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,
    data = statedata)

Residuals:
     Min       1Q   Median       3Q      Max
-1.41760 -0.43880  0.02539  0.52066  1.63048

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 68.720810   1.416828  48.503  < 2e-16 ***
Population   0.246836   0.112539   2.193 0.033491 *
Murder      -0.290016   0.035440  -8.183 1.87e-10 ***
HS.Grad      0.054550   0.014758   3.696 0.000591 ***
Frost       -0.005174   0.002482  -2.085 0.042779 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7137 on 45 degrees of freedom
Multiple R-squared:  0.7404,        Adjusted R-squared:  0.7173
F-statistic: 32.09 on 4 and 45 DF,  p-value: 1.17e-12

> anova(nullmodel, model, fullmodel)

Analysis of Variance Table

Model 1: Life.Exp ~ 1
Model 2: Life.Exp ~ Population + Murder + HS.Grad + Frost
Model 3: Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
    Frost + Area
  Res.Df    RSS Df Sum of Sq       F    Pr(>F)
1     49 88.299
2     45 22.921  4    65.378 30.3101 6.901e-12 ***
3     42 22.648  3     0.273  0.1688    0.9168
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> signif(vcov(model), 6)

            (Intercept)   Population       Murder      HS.Grad        Frost
(Intercept)  2.00740000 -1.18811e-01 -1.98357e-02 -1.44506e-02 -1.42795e-03
Population  -0.11881100  1.26650e-02 -3.56651e-04  2.36109e-04  8.91432e-05
Murder      -0.01983570 -3.56651e-04  1.25601e-03  1.84375e-04  3.42863e-05
HS.Grad     -0.01445060  2.36109e-04  1.84375e-04  2.17798e-04 -3.18945e-06
Frost       -0.00142795  8.91432e-05  3.42863e-05 -3.18945e-06  6.15931e-06
```

(a) Why do we take a logarithmic transformation of population and area?

(b) Find the Akaike's Information Criterion for the model with variables `Population`, `Murder`, `Frost`, and `Area`.

(c) Write down the final fitted model (including any variable transforms used).

(d) Calculate the sample variance s^2 for the final model.

(e) Calculate a 95% confidence interval for $\beta_{Population} - \beta_{Murder}$. (The 97.5% critical value for a t distribution with 45 d.f. is 2.014.)

(f) What conclusions do you draw from the tests in the ANOVA table?

(g) If you were to perform an F test of $H_0 : \beta_{Frost} = 0$ in the final model, what would your F statistic and p-value be?

(h) Explain the F-statistic for the final model (last line of the `summary` call). Why is it different to the F-value in line 2 of the ANOVA table?

Question 5 (14 marks) Consider the general linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, which may be of full or less than full rank.

(a) Define the term estimable.

(b) Show that if $\mathbf{t}^T = \mathbf{t}^T(X^TX)^cX^TX$, then $\mathbf{t}^T\boldsymbol{\beta}$ is estimable.

(c) Show that in a one-factor model, all treatment contrasts are estimable.

(d) If $\mathbf{t}^T\boldsymbol{\beta}$ is estimable, derive the distribution of $\mathbf{t}^T\mathbf{b}$, where $\mathbf{b}$ is the least squares estimator of $\boldsymbol{\beta}$.

(e) If $\mathbf{t}^T\boldsymbol{\beta}$ is estimable, show that $\mathbf{t}^T\mathbf{b}$ is independent of the sample variance s^2.

Question 6 (12 marks) The nursing director at a private hospital wishes to compare the weekly number of complaints received against the nursing staff during the three daily shifts: first (7am–3pm), second (3pm–11pm) and third (11pm-7am). Her plan is to sample 17 weeks and select a shift at random from each week sampled, recording the number of complaints received during the selected shift.

The following data is collected:

		number of complaints	
	number of observations	mean	sample variance
shift 1	5	10	2
shift 2	6	9	4.8
shift 3	6	12	4.4

The data is analysed using a one-way classification model.

(a) What kind of experimental design is this?

(b) Calculate the sample variance s^2 for the linear model.

(c) Calculate a 95% prediction interval for the total number of complaints received in a day. (The 97.5% critical value of a t distribution with 14 d.f. is 2.145.) (*Hint:* You will need to modify the formula for a prediction interval.)

(d) Test the hypothesis that shift has no effect on the number of complaints. (The 95% critical value of an F distribution with 2 and 14 d.f. is 3.739.)

Question 7 (14 marks)

(a) Discuss when it is best to use a completely randomised design, complete block design, or Latin square design.

(b) For a complete block design, why do we fit an additive model and not an interaction model?

(c) Write down a design matrix and parameter vector for a balanced incomplete block design for a model with 3 treatments and 3 blocks, each of size 2.

(d) Calculate the reduced design matrix $X_{2|1}$ for this model.

(e) Do you expect the reduced normal equations for this model to have the same solution as the normal equations for a completely randomised design of 6 experimental units over 3 treatments?

End of Exam—Total Available Marks = 90.

THE UNIVERSITY OF MELBOURNE

Library Course Work Collections

Author/s:
Mathematics and Statistics

Title:
Linear Statistical Models, 2016 Semester 1, MAST30025

Date:
2016

Persistent Link:
http://hdl.handle.net/11343/127833

Question 1 (9 marks)

(a) Let A be a square matrix and suppose that $A^k = A^{k+1}$ for some $k \geq 1$. Show that A is idempotent.

(b) Let X be an $n \times p$ matrix of full rank, where $n > p$. Show that $H = X(X^TX)^{-1}X^T$ is idempotent, and find its rank. (You may assume that H is symmetric.)

(c) Show that if a square matrix A is positive semidefinite, then its eigenvalues are non-negative.

a) Let λ be eigenvalue of A with eigenvector x

$\Rightarrow Ax = \lambda x$

$\Rightarrow A^k x = A^{k-1}Ax = A^{k-1}\lambda x = \lambda A^{k-1}x = \lambda A^{k-2}Ax = \lambda A^{k-2}\lambda x = \lambda^2 A^{k-2}x = \lambda^k x$

Similarly, $A^{k+1}x = \lambda^{k+1}x$

Since $A^k = A^{k+1}$

$A^k x = A^{k+1}x$

$\Rightarrow \lambda^k x = \lambda^{k+1}x$

$\lambda^k(1-\lambda)x = 0$

$\Rightarrow \lambda^k = 0$ or $1-\lambda = 0$ since x is nonzero.

$\Rightarrow \lambda = 0$ or $\lambda = 1$

$\Rightarrow A$ is idempotent.

b) $H^2 = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H \Rightarrow$ idempotent.

Since H is also symmetric

$r(H) = tr(H) = tr(X(X^TX)^{-1}X^T) = tr(X^TX(X^TX)^{-1}) = tr(I_p) = p$

X^TX is $p \times p$ matrix.

c) Since for any x, $x^TAx \geq 0$

Choose x as eigenvector with corresponding eigenvalue λ.

$\Rightarrow x^TAx = x^T\lambda x = \lambda x^Tx = \lambda \sum_i x_i^2 \geq 0$

$x^Tx = [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_i x_i^2$

$\Rightarrow \lambda \geq 0$

Question 2 (10 marks) Let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \sim MVN\left(\begin{bmatrix} a \\ -a \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \right),$$

where a is a constant.

(a) What is the distribution of $y_1 + y_2$?

(b) What is the distribution of $\frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2 + y_3^2)$?

(c) Suppose $a = 0$. For what values of c does

$$c\frac{y_1^2 - 2y_1y_2 + y_2^2 + y_3^2}{y_1^2 + 2y_1y_2 + y_2^2}$$

have an F distribution?

$Ay \sim MVN(A\mu, AVA^T)$

a) $y_1+y_2 = [1\ 1\ 0]y \sim MVN\left([1\ 1\ 0]\begin{bmatrix} a \\ -a \\ 0 \end{bmatrix}, [1\ 1\ 0]\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right)$

$\sim MVN([0],[6])$

i.e. $y_1+y_2 \sim N(0,6)$.

b) $\frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2 + y_3^2)$

$= \frac{1}{2}[y_1\ y_2\ y_3]\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = y^T\left(\frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)y$

Let $A = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $V = var(y) = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, $\mu = E[y] = \begin{bmatrix} a \\ -a \\ 0 \end{bmatrix}$.

$AV = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

$= \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

AV is obviously symmetric

$(AV)^2 = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 2 & -2 & 0 \\ -2 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = AV$

$\Rightarrow AV$ is idempotent.

$\Rightarrow y^T A y \sim \chi^2(k = r(AV), \lambda = \frac{1}{2}\mu^T A \mu)$

$k = r(AV) = 2$ first 2 columns of AV are exact opposite.

$\lambda = \frac{1}{2}\mu^T A \mu = \frac{1}{2}[a \ \ -a \ \ 0] \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a \\ -a \\ 0 \end{bmatrix}$

$= \frac{1}{4}[2a \ \ -2a \ \ 0]\begin{bmatrix} a \\ -a \\ 0 \end{bmatrix}$

$= a^2$

$\Rightarrow \frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2 + y_3^2) \sim \chi^2(2, a^2)$

c) When $a=0$, $\frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2 + y_3^2) \sim \chi^2(2,0) \sim \chi^2(2)$, i.e. an ordinary χ^2.

We need $\alpha(y_1^2 + 2y_1y_2 + y_2^2)$ to be an independent ordinary χ^2 for some α.

From part a), $y_1 + y_2 \sim N(0, 6)$.

$\frac{y_1+y_2}{\sqrt{6}} \sim N(0,1) \Rightarrow \frac{(y_1+y_2)^2}{6} = \frac{1}{6}(y_1^2 + 2y_1y_2 + y_2^2) \sim \chi^2(1)$.

$\frac{1}{6}(y_1^2 + 2y_1y_2 + y_2^2) = y^T \frac{1}{6}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} y$

Let $B = \frac{1}{6}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$AVB = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}\frac{1}{6}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$= \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\frac{1}{6}\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \frac{1}{12}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0$

$\Rightarrow y^T A y$ and $y^T B y$ are independent.

$\Rightarrow \frac{y^T A y / 2}{y^T B y / 1} = \frac{\frac{1}{2}(y_1^2 - 2y_1y_2 + y_2^2 + y_3^2) \times \frac{1}{2}}{\frac{1}{6}(y_1^2 + 2y_1y_2 + y_2^2)} = \frac{3}{2}\,\frac{y_1^2 - 2y_1y_2 + y_2^2 + y_3^2}{y_1^2 + 2y_1y_2 + y_2^2} \sim F(2,1)$

$\Rightarrow c = \frac{3}{2}$.

Question 3 (14 marks) Consider the full rank linear model, $\mathbf{y} = X\beta + \varepsilon$.

(a) State the assumptions involved in fitting this model.

(b) Define the term BLUE (best linear unbiased estimator).

(c) Is it better to fit this model using the method of least squares or maximum likelihood estimation? Justify your answer.

(d) Define and explain the purpose of the leverage of a point.

(e) Explain the difference between a model relevance test and a model relevance test using a corrected sum of squares.

(f) When is a model with fewer explanatory variables more desirable or less desirable than a model with more explanatory variables?

(g) Explain why the residual sum of squares SS_{Res} is not an appropriate goodness-of-fit measure for model selection.

a) There is a linear relationship and $\varepsilon \sim MVN(0, \sigma^2 I)$

b) An unbiased estimator of $t^T\beta$ that has minimal variance among all such unbiased estimators.

c) LSE and MLE gives the same results of fitted parameters. However, LSE gives unbiased $\hat{\sigma}^2$ but MLE of σ^2 is biased.

d) Leverage of a point is its corresponding diagonal entry is the hat matrix $X(X^TX)^{-1}X^T$. It indicates extremeness of a point in the X space. It measures potential influence of a point to the fitted model.

e) Model relevance test tests whether all parameters, including intercept, are 0; Using corrected sum of squares will exclude intercept and test all parameters except intercept are 0.

f) More desirable if excessive variables are not significant and are useless in improving predictive power. Model with fewer variables is less likely to overfit.

Less desirable if model with fewer variables has a poor fit comparing to model with more variables.

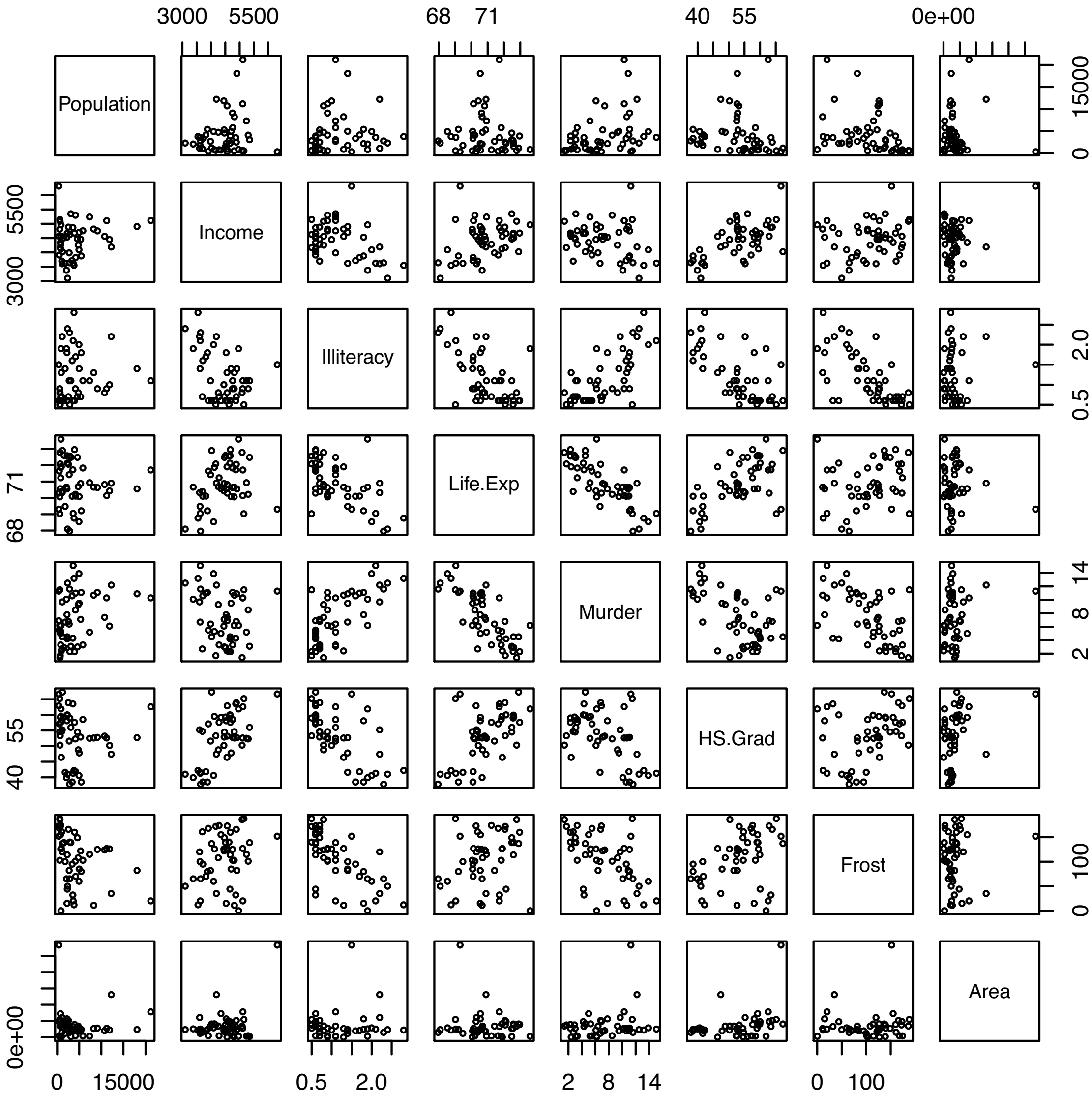
g) SSRes will always decrease when adding in new variables. It does not prevent overfitting.

Question 4 (17 marks) In this question, we study a dataset of 50 US states. This dataset contains the variables:

- `Population`: population estimate as of July 1, 1975
- `Income`: per capita income (1974)
- `Illiteracy`: illiteracy (1970, percent of population)
- `Life.Exp`: life expectancy in years (1969–71)
- `Murder`: murder and non-negligent manslaughter rate per 100,000 population (1976)
- `HS.Grad`: percentage of high-school graduates (1970)
- `Frost`: mean number of days with minimum temperature below freezing (1931–1960) in capital or large city
- `Area`: land area in square miles

We use linear models to model life expectancy in terms of the other variables. The following R output is produced.

```
> data(state)
> statedata <- data.frame(state.x77, row.names=state.abb, check.names=TRUE)
> pairs(statedata, cex=0.5)
```




```
> statedata$Population <- log(statedata$Population)
> statedata$Area <- log(statedata$Area)
> nullmodel <- lm(Life.Exp ~ 1, data = statedata)
> fullmodel <- lm(Life.Exp ~ ., data = statedata)
> model <- step(fullmodel, scope = ~ .)

Start:  AIC=-23.6
Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
    Frost + Area

             Df Sum of Sq    RSS      AIC
- Income      1    0.0018 22.650 -25.5934
- Illiteracy  1    0.0556 22.704 -25.4746
```

```
- Area         1    0.2106 22.859 -25.1344
<none>                     22.648 -23.5973
- Frost        1    1.2374 23.886 -22.9374
- Population   1    1.8854 24.533 -21.5992
- HS.Grad      1    2.4375 25.086 -20.4864
- Murder       1   23.2760 45.924   9.7483

Step:  AIC=-25.59
Life.Exp ~ Population + Illiteracy + Murder + HS.Grad + Frost +
    Area

             Df Sum of Sq    RSS      AIC
- Illiteracy  1    0.0556 22.705 -27.4708
- Area        1    0.2197 22.870 -27.1107
<none>                    22.650 -25.5934
- Frost       1    1.2602 23.910 -24.8862
+ Income      1    0.0018 22.648 -23.5973
- Population  1    2.1909 24.841 -22.9768
- HS.Grad     1    4.0374 26.687 -19.3918
- Murder      1   24.2130 46.863   8.7601

Step:  AIC=-27.47
Life.Exp ~ Population + Murder + HS.Grad + Frost + Area

             Df Sum of Sq    RSS     AIC
- Area        1    0.2157 22.921 -28.998
<none>                    22.705 -27.471
+ Illiteracy  1    0.0556 22.650 -25.593
+ Income      1    0.0017 22.704 -25.475
- Population  1    2.2792 24.985 -24.688
- Frost       1    2.3760 25.082 -24.495
- HS.Grad     1    4.9491 27.655 -19.612
- Murder      1   29.2296 51.935  11.899

Step:  AIC=-29
Life.Exp ~ Population + Murder + HS.Grad + Frost

             Df Sum of Sq    RSS     AIC
<none>                    22.921 -28.998
+ Area        1     0.216 22.705 -27.471
+ Illiteracy  1     0.052 22.870 -27.111
+ Income      1     0.011 22.911 -27.021
- Frost       1     2.214 25.135 -26.387
- Population  1     2.450 25.372 -25.920
- HS.Grad     1     6.959 29.881 -17.741
- Murder      1    34.109 57.031  14.578
```

```
> summary(model)

Call:
lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,
    data = statedata)

Residuals:
     Min       1Q   Median       3Q      Max
-1.41760 -0.43880  0.02539  0.52066  1.63048

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 68.720810   1.416828  48.503  < 2e-16 ***
Population   0.246836   0.112539   2.193 0.033491 *
Murder      -0.290016   0.035440  -8.183 1.87e-10 ***
HS.Grad      0.054550   0.014758   3.696 0.000591 ***
Frost       -0.005174   0.002482  -2.085 0.042779 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7137 on 45 degrees of freedom
Multiple R-squared:  0.7404,	Adjusted R-squared:  0.7173
F-statistic: 32.09 on 4 and 45 DF,  p-value: 1.17e-12

> anova(nullmodel, model, fullmodel)

Analysis of Variance Table

Model 1: Life.Exp ~ 1
Model 2: Life.Exp ~ Population + Murder + HS.Grad + Frost
Model 3: Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +
    Frost + Area
  Res.Df    RSS Df Sum of Sq       F    Pr(>F)
1     49 88.299
2     45 22.921  4    65.378 30.3101 6.901e-12 ***
3     42 22.648  3     0.273  0.1688    0.9168
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> signif(vcov(model), 6)

            (Intercept)   Population       Murder      HS.Grad        Frost
(Intercept)  2.00740000 -1.18811e-01 -1.98357e-02 -1.44506e-02 -1.42795e-03
Population  -0.11881100  1.26650e-02 -3.56651e-04  2.36109e-04  8.91432e-05
Murder      -0.01983570 -3.56651e-04  1.25601e-03  1.84375e-04  3.42863e-05
HS.Grad     -0.01445060  2.36109e-04  1.84375e-04  2.17798e-04 -3.18945e-06
Frost       -0.00142795  8.91432e-05  3.42863e-05 -3.18945e-06  6.15931e-06
```

(a) Why do we take a logarithmic transformation of population and area?

(b) Find the Akaike's Information Criterion for the model with variables `Population`, `Murder`, `Frost`, and `Area`.

(c) Write down the final fitted model (including any variable transforms used).

(d) Calculate the sample variance s^2 for the final model.

(e) Calculate a 95% confidence interval for $\beta_{Population} - \beta_{Murder}$. (The 97.5% critical value for a t distribution with 45 d.f. is 2.014.)

(f) What conclusions do you draw from the tests in the ANOVA table?

(g) If you were to perform an F test of $H_0 : \beta_{Frost} = 0$ in the final model, what would your F statistic and p-value be?

(h) Explain the F-statistic for the final model (last line of the `summary` call). Why is it different to the F-value in line 2 of the ANOVA table?

a) Both have very skewed distributions.
Both do not have a linear relationship with respone Life.Exp.

b) -19.612.

c) Life.Exp = 68.72 + 0.25 log(Population) - 0.29 Murder
+ 0.055 HS.Grad - 0.0052 Frost

d) From summary(model): $0.7137^2 = 0.5094$

From stepwise: RSS = 22.921, df = 45 (From summary or calculated by n-p = 50-5)

$$s^2 = \frac{22.921}{45} = 0.5094$$

From anova, RSS = 22.921, Res. df = 45 $\Rightarrow s^2 = \frac{22.921}{45} = 0.5094$

e) CI for $t^T\beta$ is $t^Tb \pm t_{0.975} \cdot \sqrt{s^2 \cdot t^T(X^TX)^{-1}t} = t^Tb \pm t_{0.975} \cdot \sqrt{t^T \underbrace{var(b)}_{\downarrow \text{ vcov(model)}} t}$

Here $t^T = [0 \ 1 \ -1 \ 0 \ 0]$, let $C = var(b)$. i.e. vcov(model) in R code

$$t^T \cdot var(b) \cdot t = [0 \ 1 \ -1 \ 0 \ 0] \ C \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

$$= [C_{21}-C_{31} \quad C_{22}-C_{32} \quad C_{23}-C_{33} \quad C_{24}-C_{34} \quad C_{25}-C_{35}] \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= (C_{22}-C_{32}) - (C_{23}-C_{33}).$$

$\Rightarrow$ 95% CI of $\beta_{population} - \beta_{murder}$

$= (0.246836 - (-0.290016)) \pm 2.014 \times \sqrt{(0.0126650 - (-0.000356651)) - (-0.000356651 - 0.00125601)}$

$= (0.2932137, 0.7804903)$.

f) In the model selected by AIC, at least one variable is relevant. Other variables discarded by AIC are all not relevant.

g) F stats will be square of t-stats but p-value remains the same.

$F\text{-stats} = (-2.085)^2 = 4.35$, p-value $= 0.043$.

h) Summary uses SSRes for the selected model.

ANOVA uses SSRes for the full model.

Question 5 (14 marks) Consider the general linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, which may be of full or less than full rank.

(a) Define the term estimable.

(b) Show that if $\mathbf{t}^T = \mathbf{t}^T(X^TX)^cX^TX$, then $\mathbf{t}^T\boldsymbol{\beta}$ is estimable.

(c) Show that in a one-factor model, all treatment contrasts are estimable.

(d) If $\mathbf{t}^T\boldsymbol{\beta}$ is estimable, derive the distribution of $\mathbf{t}^T\mathbf{b}$, where $\mathbf{b}$ is the least squares estimator of $\boldsymbol{\beta}$.

(e) If $\mathbf{t}^T\boldsymbol{\beta}$ is estimable, show that $\mathbf{t}^T\mathbf{b}$ is independent of the sample variance s^2.

a) $t^T\beta$ is estimable if there exists matrix C s.t. $E[Cy] = t^T\beta$,

i.e. $t^T\beta$ has an unbiased linear estimator.

b) want $E[Cy] = CX\beta = t^T\beta \Rightarrow$ need $CX = t^T = t^T(X^TX)^cX^TX$

Choose $C = t^T(X^TX)^cX^T$

$$E[Cy] = E[t^T(X^TX)^cX^Ty] = t^T(X^TX)^cX^TE[y]$$
$$= t^T(X^TX)^cX^TX\beta$$
$$= t^T\beta$$

$\Rightarrow t^T\beta$ is estimable.

c) A treat contrast is $\sum_i \alpha_i \tau_i$ where $\sum_i \alpha_i = 0$.

Therefore $\sum_i \alpha_i\tau_i = \sum_i \alpha_i\tau_i + \mu\sum_i\alpha_i$

$$= \sum_i \alpha_i(\mu + \tau_i).$$

Since $\mu + \tau_i$ is element of $X\beta$,

$\sum_i \alpha_i(\mu+\tau_i)$ is a linear combination of $X\beta$, thus estimable.

d) $b = (X^TX)^cX^Ty$

$t^Tb = t^T(X^TX)^cX^Ty$, thus is MVN.

$E[t^Tb] = t^T(X^TX)^cX^TX\beta = t^T\beta$

$$var(t^Tb) = var(t^T(X^TX)^cX^Ty) = t^T(X^TX)^cX^T\sigma^2 I\left(t^T(X^TX)^cX^T\right)^T$$
$$= \sigma^2 t^T(X^TX)^cX^TX(X^TX)^c t$$
$$= \sigma^2 t^T(X^TX)^c t \quad \Rightarrow t^Tb \sim MVN(t^T\beta, \sigma^2 t^T(X^TX)^c t).$$

e) $s^2 = \frac{Y^T(I-H)Y}{n-r}$, where $H = X(X^TX)^c X^T$, $r = r(X)$.

$t^T b = t^T (X^TX)^c X^T Y$.

Let $A = \frac{1}{n-r}(I - X(X^TX)^c X^T)$, $B = t^T(X^TX)^c X^T$, $V = var(y) = \sigma^2 I$

$$BVA = \frac{\sigma^2}{n-r} t^T (X^TX)^c X^T \; I \; (I - X(X^TX)^c X^T)$$

$$= \frac{\sigma^2}{n-r}\left[t^T(X^TX)^c X^T - t^T(X^TX)^c X^T X (X^TX)^c X^T\right]$$

$$= \frac{\sigma^2}{n-r}\left[t^T(X^TX)^c X^T - t^T(X^TX)^c X^T\right]$$

$$= 0$$

$\Rightarrow$ independant.

Question 6 (12 marks) The nursing director at a private hospital wishes to compare the weekly number of complaints received against the nursing staff during the three daily shifts: first (7am–3pm), second (3pm–11pm) and third (11pm-7am). Her plan is to sample 17 weeks and select a shift at random from each week sampled, recording the number of complaints received during the selected shift.

The following data is collected:

		number of complaints	
	number of observations	mean	sample variance
shift 1	5	10	2
shift 2	6	9	4.8
shift 3	6	12	4.4

The data is analysed using a one-way classification model.

(a) What kind of experimental design is this?

(b) Calculate the sample variance s^2 for the linear model.

(c) Calculate a 95% prediction interval for the total number of complaints received in a day. (The 97.5% critical value of a t distribution with 14 d.f. is 2.145.) (*Hint:* You will need to modify the formula for a prediction interval.)

(d) Test the hypothesis that shift has no effect on the number of complaints. (The 95% critical value of an F distribution with 2 and 14 d.f. is 3.739.)

a) CRD.

b) Within each shift, $s_i^2 = \frac{\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)}{n_i - 1}$ → part of SSRes, since $\bar{y}_i$ is prediction of y_{ij}

$$b = \begin{bmatrix} 0 \\ \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \end{bmatrix}, \quad xb = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_1 \\ \vdots \\ \bar{y}_2 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_3 \\ \bar{y}_3 \\ \vdots \end{bmatrix}$$

$$SSRes = 2\times(5-1) + 4.8\times(6-1) + 4.4\times(6-1) = 54$$

$$s^2 = \frac{54}{\underbrace{(5-1)+(6-1)+(6-1)}_{\text{same as } n-r}} = 3.86.$$

c)

$$b = \begin{bmatrix} 0 \\ 10 \\ 9 \\ 12 \end{bmatrix} \qquad (X^TX)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & \frac{1}{6} \end{bmatrix}$$

Here we estimate interval of $(\mu+\tau_1)+(\mu+\tau_2)+(\mu+\tau_3)$

$\Rightarrow$ construct $t = [3\ 1\ 1\ 1]$.

95% PI is

$$t^Tb \pm t_{0.975} \cdot \sqrt{s^2 \cdot [3 + t^T(X^TX)^c t]}$$

$$= (10+9+12) \pm 2.145 \times \sqrt{3.86 \times \left(3 + \frac{1}{5} + \frac{1}{6} + \frac{1}{6}\right)} = 31 \pm 7.92$$

$$= (23.08, 38.92)$$

d) $H_0: \tau_1 = \tau_2 = \tau_3$

Construct $C = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$

$$\text{F-stats} = \frac{(Cb)^T [C(X^TX)^c C^T]^{-1} (Cb)/2}{s^2}$$

$C(X^TX)^c C^T$

$$= \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \frac{1}{5} & -\frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{6} & -\frac{1}{6} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{5} + \frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{3} \end{bmatrix} = \frac{1}{30} \begin{bmatrix} 11 & -5 \\ -5 & 10 \end{bmatrix}$$

$$\Rightarrow [C(X^TX)^c C^T]^{-1} = \frac{30}{110 - 25} \begin{bmatrix} 10 & 5 \\ 5 & 11 \end{bmatrix} = \frac{6}{17} \begin{bmatrix} 10 & 5 \\ 5 & 11 \end{bmatrix}$$

$$\text{F-stats} = \frac{\begin{bmatrix} 1 & -3 \end{bmatrix} \left(\frac{6}{17} \begin{bmatrix} 10 & 5 \\ 5 & 11 \end{bmatrix} \right) \begin{bmatrix} 1 \\ -3 \end{bmatrix}}{2 \times 3.86}$$

$$= 0.0457 \begin{bmatrix} 1 & -3 \end{bmatrix} \begin{bmatrix} 10 & 5 \\ 5 & 11 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

$$= 0.0457 \begin{bmatrix} -5 & -28 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

$$= 3.6103 < 3.739$$

$\Rightarrow$ do not reject H_0. Therefore, effect of shift is not significant at 5% level.

Question 7 (14 marks)

(a) Discuss when it is best to use a completely randomised design, complete block design, or Latin square design.

(b) For a complete block design, why do we fit an additive model and not an interaction model?

(c) Write down a design matrix and parameter vector for a balanced incomplete block design for a model with 3 treatments and 3 blocks, each of size 2.

(d) Calculate the reduced design matrix $X_{2|1}$ for this model.

(e) Do you expect the reduced normal equations for this model to have the same solution as the normal equations for a completely randomised design of 6 experimental units over 3 treatments?

a) CRD: one treatment without any known confounding factor.
CBD: one confounding factor known to have some effect.
Latin square: two confounding factors. CBD in this case will have to treat all pairs of confounding factors as blocks, thus impractical and not feasible.

b) Purpose of blocking is to remove interaction between blocks and treatments.

c)

3 block with size 2

each treatment apply to 2 different blocks.

$$\beta = \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix}, \quad X = \begin{bmatrix} \underbrace{\begin{matrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{matrix}}_{X_1} & \underbrace{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}}_{X_2} \end{bmatrix}$$

each block has two different treatments.

d)
$$X_{2|1} = [I - H_1] X_2$$
$$= [I - X_1 (X_1^T X_1)^c X_1^T] X_2$$

$$(X_1^T X_1)^c = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$$

like one-way classification.

$$H_1 = X_1(X_1^T X_1)^c X_1^T = \begin{bmatrix} 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

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

$$= \begin{bmatrix} 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 & -0.5 & 0.5 \end{bmatrix} X_2$$

$$= \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

e) No. Blocks are not orthogonal to the treatments in BIBD.