Assignments

Please see the study guide to note which questions forms part of which assignment.

Assignment A: (Q1, 2, 3)

---

1. Consider the dataset CFCS which gives the answers of 15035 individuals to questions in the "Consideration of Future Consequences Scale" questionnaire. The codebook that accompanies the data is given as well as instructions to clean the data. Work with the data after you have cleaned it (13861 observations).

   (a) Calculate the sample covariance between Q8 and Q10. Can you explain the sign of the covariance (positive or negative)?

   (b) Calculate the sample correlation between Q8 and Q10.

   (c) Calculate the average and standard deviation for Q8 for the following four groups:
   Males, Age<=25
   Males, Age>25
   Females, Age<=25
   Females, Age>25

2. Suppose $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ is a 3×1 vector with jointly distributed random variables such that

$$E(\boldsymbol{X}) = \boldsymbol{\mu} \text{ and } \boldsymbol{\Sigma} = \boldsymbol{cov}(\boldsymbol{X}, \boldsymbol{X}') = \begin{pmatrix} 50 & 36 & 18 \\ 36 & 36 & 0 \\ 18 & 0 & 72 \end{pmatrix}.$$

   Consider the linear combinations

$$\boldsymbol{b}'\boldsymbol{X} = 2X_1 + 2X_2 - X_3,$$
$$\boldsymbol{c}'\boldsymbol{X} = X_1 - X_2 + 3X_3$$

   and

$$\boldsymbol{d}'\boldsymbol{X} = X_1 + X_3.$$

   Use PROC IML to answer the following:

   (a) Calculate $var \;\; \boldsymbol{b}'\boldsymbol{X})$.

   (b) Calculate $var(\boldsymbol{c}'\boldsymbol{X})$.

   (c) Calculate $cov(\boldsymbol{b}'\boldsymbol{X}, \boldsymbol{c}'\boldsymbol{X})$.

   (d) Calculate $cor(\boldsymbol{b}'\boldsymbol{X}, \boldsymbol{c}'\boldsymbol{X})$.

   (e) Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{b}' \\ \boldsymbol{c}' \end{pmatrix}$. Calculate $cov(\boldsymbol{AX}, (\boldsymbol{AX})')$.

   (f) Let $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{b} & \boldsymbol{c} & \boldsymbol{d} \end{pmatrix}$. Calculate $cov\left(\boldsymbol{B}'\boldsymbol{X}, (\boldsymbol{B}'\boldsymbol{X})'\right) = \begin{pmatrix} \varsigma_{11} & \varsigma_{12} & \varsigma_{13} \\ \varsigma_{21} & \varsigma_{22} & \varsigma_{23} \\ \varsigma_{31} & \varsigma_{32} & \varsigma_{33} \end{pmatrix}$ and use this to give:

      i. ς_{23}

      ii. $cov(X_1 + X_3, 2X_1 + 2X_2 - X_3)$.

   (g) Calculate the following for $\boldsymbol{\Sigma}$:

      i. The eigenvalues and normalized eigenvectors of $\boldsymbol{\Sigma}$.

      ii. Calculate $\boldsymbol{\Sigma}^{\frac{1}{2}}$, the symmetric square root of $\boldsymbol{\Sigma}$.

      iii. Use the eigenvalues of $\boldsymbol{\Sigma}$ to calculate $|\boldsymbol{\Sigma}|$ and $\text{tr}(\boldsymbol{\Sigma})$.

3. Consider the data for *Iris Setosa* in the iris flower data (Iris.xls). Let $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i4} \end{pmatrix}$, $i = 1, 2, \ldots, 50$ indicate the vectors of observations for petal width (PW), petal length (PL), sepal width (SW), and sepal length (SL) respectively and let

$$\boldsymbol{X} : 50 \times 4 = \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_{50}' \end{pmatrix}$$

be the observed data matrix. Use SAS IML to calculate the sample mean ($\overline{\boldsymbol{x}}$), sample covariance matrix

$$\left(\mathbf{S} = \frac{1}{49}\mathbf{X}'\left(\boldsymbol{I}_{50} - \frac{1}{50}\mathbf{1}_{50}\mathbf{1}_{50}'\right)\mathbf{X}\right)$$

and sample correlation matrix ($\boldsymbol{R}$) for *Iris Setosa*.

Assignment B: (Q4, 5, 6, 7)

4. Let $\boldsymbol{X}'$ be the random vector $\begin{pmatrix} X_1, & X_2, & X_3, & X_4 \end{pmatrix}$ with mean vector $\boldsymbol{\mu}_X' = \begin{pmatrix} 4, & 3, & 2, & 1 \end{pmatrix}$ and

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 7 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix}.$$

Partition $\boldsymbol{X}$ as

$$\begin{pmatrix} X_1 \\ X_2 \\ \hline X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}.$$

Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}.$$

Use PROC IML to calculate the following.

(a) $E(\boldsymbol{X}_2)$
(b) $E(\boldsymbol{A}\boldsymbol{X}_2)$
(c) $cov(\boldsymbol{X}_2, \boldsymbol{X}_2')$
(d) $cov(\boldsymbol{A}\boldsymbol{X}_2, (\boldsymbol{A}\boldsymbol{X}_2)')$
(e) $cov(\boldsymbol{X}_1, \boldsymbol{X}_2')$
(f) $cov(\boldsymbol{A}\boldsymbol{X}_1, (\boldsymbol{B}\boldsymbol{X}_2)')$

5. Write a SAS/IML program simulating the following theoretical fact:
Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(0,1)$ distribution. It is then known that $\overline{X}$ will follow a normal distribution, with mean 0 and standard deviation $\frac{1}{\sqrt{n}}$.
Accept that $\overline{X}$ follows a normal distribution. Demonstrate through simulation that the mean is 0 and the standard deviation is $\frac{1}{\sqrt{n}}$. Simulate 1000 samples, each of size 300. Do the simulation without using a do loop!
Give the theoretical and empirical values for the mean and standard deviation.

6. Work through Example A, Questions 1, 2 and 3.

7. Work through Example 5 and Example 6 in Section 2.2.

Assignment C: (Q8)

8. (a) Use PROC G3D and draw a 3 dimensi graph of a bivariate normal density function, that is for $\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv N_2\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right)$ for each of the $\boldsymbol{\Sigma}'s$ given below.

   Start out by defining a grid over the definition region of the density function in PROC IML. Continue by calculating the density function over the grid. Plot the density function using PROC G3D.

   Syntax for PROC G3D:

   ```
   proc g3d data=normal;
   plot x*y=fxy / side
   tilt=45 rotate=35;
   ```

   - 'side' produces a surface graph with a side wall.
   - 'tilt' specifies one or more angles $(0° - 90°)$ to tilt the graph toward you (default $= 70°$).
   - 'rotate' specifies one or more angles at which to rotate the $X - Y$ plane around the perpendicular Z axis (default $= 70°$).

   i. $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix}$

   ii. $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix}$

   iii. $\boldsymbol{\Sigma} = \begin{pmatrix} 4 & -4 \\ -4 & 6 \end{pmatrix}$

   (b) Calculate the correlation between X and Y for the three cases in 5(a) and comment on the association between the two random variables.

   (c) Refer to 5(a)(ii). Use numerical integration and determine the total volume under the density curve.

   (d) Within the context of 5(a)(ii), that is when $\mathbf{X} = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix}\right)$, draw a graph of a *slice* through the density function where $X = 2$. Use PROC GPLOT to draw the graph. Comment on the shape of the graph. Is it a density function, perhaps a normal density function?

   (e) Refer to 5(d). Use numerical integration and determine the area under the curve. How can the curve in 5(d)be transformed to ensure that it is a normal density curve.

Assignment D: (Q9, 10, 11, 12)

9. Use SAS/IML to rework the results of Example 11 and Example 12 in the notes.

Example 10 *Ficus is a genus of approximately 850 species of shrubs, trees, and vines in the family Moraceae, and are sometimes loosely referred to as "fig trees", or figs. The fruit that these trees bare are of vital cultural importance around the world and also serve as an extremely important food resource for wildlife. Consider the hypothetical example where* $p = 3$ *(this is also called "trivariate") about heights (in cm) of three different ficus species that are found in South Africa:*

$$\begin{aligned} X_1 &= \text{height of Ficus bizanae} \\ X_2 &= \text{height of Ficus burtt-davyi} \\ X_3 &= \text{height of Ficus tettensis} \end{aligned}$$

The interest in Ficus and modeling of its characteristics is important for understanding ecological advancement, and wildlife- as well as cultural conservation. The covariance matrix is given by

$$\boldsymbol{\Sigma} = (\sigma_{ij}) = \begin{pmatrix} 100 & 70 & 90 \\ 70 & 100 & 80 \\ 90 & 80 & 100 \end{pmatrix} .$$

The partial covariance matrix of the heights of the first two trees, if the influence of the third tree is eliminated, is:

$$\begin{aligned} \boldsymbol{\Sigma}_{11.2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \\ &= \begin{pmatrix} 100 & 70 \\ 70 & 100 \end{pmatrix} - \frac{1}{100}\begin{pmatrix} 90 \\ 80 \end{pmatrix}(90, 80) \\ &= \begin{pmatrix} 19 & -2 \\ -2 & 36 \end{pmatrix}. \end{aligned}$$

The partial correlation coefficient between the heights of the first two trees, if the influence of the height of the third tree is eliminated, then is:

$$\rho_{12.3} = \frac{-2}{\sqrt{19 \times 36}} = -0.076.$$

This analysis tells us for example that the heights between the first two trees are not particularly highly correlated, and thus gives insight to farmers/scientists/public that by seeing one tall Ficus might not mean any Ficus will be tall necessarily. This aids with plantation design and wildlife conservation - i.e. how "much" Ficus and Ficus leaves might be available for wildlife consumption as part of herd- and grazing planning.

Example 11 *Redo the above example, but calculate the partial correlation coefficient between the height of Ficus bizanae and Ficus tettensis, by eliminating the influence of Ficus burtt-davyi.*

Example 12 *Take note: the type of analysis above is crucial for understanding the relationships between, in this example, heights of shrubs/trees of the same genus but of different type. The elimination of the influence of one or more variables allows the scientist/practitioner to gain insight into the relationships that the remaining variables have. Think for example if your data is vast but the geographical area of your interest does not cater for Ficus bizanae. Then you can investigate the behaviour of the relationship of the remaining two types by disregarding the influence of Ficus bizanae!*

10. You are given the random vector $\mathbf{X}' = \begin{pmatrix} X_1, & X_2, & X_3, & X_4, & X_5 \end{pmatrix}$ with mean vector $\boldsymbol{\mu}' = \begin{pmatrix} 2, & -1, & 3, & 4, & 0 \end{pmatrix}$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & \frac{1}{2} & -\frac{1}{2} & -1 & 0 \\ \frac{1}{2} & 6 & 1 & 1 & -1 \\ -\frac{1}{2} & 1 & 4 & -1 & 0 \\ -1 & 1 & -1 & 3 & 0 \\ 0 & -1 & 0 & 0 & 2 \end{pmatrix}.$$

Use SAS/IML to calculate the following.

(a) Calculate ρ_{34}, the correlation between X_3 and X_4.

(b) Calculate $\rho_{34.25}$, the partial correlation between X_3 and X_4 controlling for X_2 and X_5.

(c) Calculate $E\left(\begin{pmatrix} X_3 \\ X_4 \end{pmatrix} \middle| \begin{pmatrix} X_2 \\ X_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$.

11. Work through Example A, Questions 4 and 5.

Note that if $\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{pmatrix}$ is the unbiased estimator for $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$, then it can be shown that an unbiased esimator for $\boldsymbol{\Sigma}_{11.2}$ is $\frac{n-1}{n-r-1}\left(\boldsymbol{S}_{11} - \boldsymbol{S}_{12}\boldsymbol{S}_{22}^{-1}\boldsymbol{S}_{21}\right)$ where n is the sample size and $\boldsymbol{S}_{22} : r \times r$.

12. Let $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}$ be a vector of random variables from a multivariate normal $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution with

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 1 \\ 1 \\ 3 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 9 & -1 & 2 & 0 \\ -1 & 7 & 3 & -1 \\ 2 & 3 & 13 & 3 \\ 0 & -1 & 3 & 9 \end{pmatrix}.$$

Generate a random sample of size 10 000 from this distribution and use the simulated data to calculate empirical values for the following.

(a) $P(\boldsymbol{X} < \boldsymbol{1}_4)$.
(b) $P(X_1 < 1, X_2 < 1, X_3 < 1)$.
(c) $P(X_1 + X_2 > X_3 + X_4)$.
(d) $var(X_4)$.
(e) $cov(X_1 + X_2, X_3 + X_4)$.

Assignment E: (Q13, 14)

13. Work through Example A, Questions 6 and 7.

14. The dataset FatherSons.txt gives the heights of a father and his two oldest sons in 137 randomly selected families. The variables are

$$\begin{aligned} X_1 &= \text{height of the first son} \\ X_2 &= \text{height of the second son} \\ X_3 &= \text{height of the father.} \end{aligned}$$

Calculate the maximum likelihood estimates for the following:

(a) The correlation between the first and second son.
(b) The correlation between the first and second son when controlling for the height of the father.
(c) The multiple correlation coefficient between the height of the first son and the heights of the father and second son.
(d) $E\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} | X_3 = 72\right)$.

Assignment F: (Q15, 16)

15. Suppose that the heights of married couples can be described by a bivariate normal distribution. The wives have a mean height of 169.7 cm and a standard deviation of 5.1 cm. The heights of husbands have a mean of 177.8 cm and a standard deviation of 6.3 cm. The correlation between the heights of husbands and wives is 0.68. Answer the following questions by making use of a simulation.

(a) What is the probability that for a randomly selected couple the wife is taller than her husband?
(b) Consider only couples where the husband is shorter than 175cm. What is the probability that for a randomly selected couple from this group the wife is taller than her husband?

16. The scores obtained by 87 students in three subtests are given in the dataset College.txt on ClickUP.

Let $X_1 =$ Social science and history score (entry in first column);
$X_2 =$ Verbal score (entry in second column);
$X_3 =$ Science score (entry in third column).

The summary statistics are given in the following SAS output:

```
                 The MEANS Procedure
Variable      N             Mean            Std Dev
---------------------------------------------------
social       87          526.586             76.211
verbal       87           54.690             11.227
science      87           25.126              4.807
---------------------------------------------------
```

Assume that $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right).$

Calculate the maximum likelihood estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

17. Work through Example B.

18. A shoe company evaluates new shoe models based on five criteria: style, comfort, stability, cushioning and durability, with each of the first four criteria evaluated on a scale of 1 to 10 and the durability criteria evaluated on the scale of 1 to 20. Based on the evaluations of 25 people about the company's latest prototype you have to determine whether the shoe is ready for release to the market. The goals expected from new products is that $\boldsymbol{\mu}$, the vector of average responses for each criteria, is equal to

$$\begin{pmatrix} \mu_{style} \\ \mu_{comfort} \\ \mu_{stability} \\ \mu_{cushioning} \\ \mu_{durability} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \\ 5 \\ 6 \\ 10 \end{pmatrix} = \boldsymbol{\mu}_0.$$

The observed data is given in the file shoes.xls on ClickUP.

Assume that the data comes from a multivariate normal distribution.

(a) Calculate the correlation coefficient between Comfort and Stability. Also calculate the partial correlation coefficient between Comfort and Stability controlling for Durability.

(b) Calculate the sample mean vector.

(c) The hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ must be tested on a 10% level of significance.

i. Calculate Hotellings T^2 statistic, that is $\dfrac{T^2}{N-1}$.

ii. Calculate the test statistics, that is $\dfrac{T^2}{N-1}\dfrac{N-p}{p}$.

iii. Calculate the critical value for the hypothesis and use it to test the hypothesis.

iv. Calculate the p-value for the hypothesis and use it to test the hypothesis.

v. Give a conclusion in terms of the problem.

19. **Exercise 6.14 from textbook: Linear Models in Statistics by AC Rencher and GB Schaalje.**

Use the data in Table 6.1 of the textbook to answer the following questions.
The data is is given on ClickUP under the datasets (Table6_1.xlsx)..

(a) Give a scatter plot of the data.

(b) Find $\widehat{\beta}_0$ and $\widehat{\beta}_1$ and interpret the value of $\widehat{\beta}_1$.

(c) Use the residuals to test the regression assumptions of equal variance and normality.

(d) Test $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 \neq 0$. Use $\alpha = 0.05$.

(e) Test $H_0 : \beta_0 = 0$ against the alternative $H_1 : \beta_0 \neq 0$. Use $\alpha = 0.05$.

(f) Calculate and interpret a 95% confidence interval for β_1.

(g) Give the total variation, explained variation and the unexplained variation.

(h) Give r^2 and interpret the value. Also show how r^2 can be calculated from the results in the previous question.

(i) Give MSE.

(j) Calculate a point estimate and 95% c.i. for the mean time until the next eruption for all eruptions that last 3 minutes.

(k) Calculate a point prediction and 95% c.i. for the time until the next eruption if the duration of a current eruption is 3 minutes.

20. **Exercise 7.54 from textbook: Linear Models in Statistics by AC Rencher and GB Schaalje.**

21. **Continuation of Exercise 7.54 from textbook: Linear Models in Statistics by AC Rencher and GB Schaalje.**

Use the data in Table 7.4 to answer the following questions. Use a 5% level of significance.
The SAS dataset is given on ClickUP under the datasets (ex7_54.sas7bdat).
Consider the two models:

Model 1: $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

Model 2: $y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1^2 + \beta_5 x_2^2 + \beta_6 x_3^2 + \beta_7 x_1 x_2 + \beta_8 x_1 x_3 + \beta_9 x_2 x_3 + \varepsilon$

(a) Test for the overall significance of the Model 1. That is, test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

(b) For Model 1: Test $H_0 : \beta_i = 0$ for $i = 1, 2, 3$.

(c) For Model 1: Test $H_0 : 2\beta_1 = 2\beta_2 = \beta_3$.

(d) Consider the PROC REG or PROC GLM output from the second model. What can jou say about the overall significance of the model? Comment on the additional significance of each independent variable to the model given that the other independent variables are already in the model.

To create functions of the independent variables, say x_1^2, x_2^2 and x_3^2, the following syntax in SAS can be used:

```
data chemical;

input y1 x1 x2 x3;
x1x1=x1*x1; x2x2=x2*x2 x3x3=x3*x3;
cards;
41.5 162 23 3
33.8 162 23 8
```

;Alternatively, if the data has already been entered into the dataset chemical, the following syntax in SAS can be used:

```
data chem2;

set chemical;
x1x1=x1*x1; x2x2=x2*x2 x3x3=x3*x3;
```

22. This example comes from the textbook:
Daniel, W.W. (1974). *Biostatistics: A Foundation for Analysis in the Health Sciences*. New York: Wiley.

Blood sugar levels (mg/100g) were measured on 10 animals from each of five breeds. The results are presented in the table below.

Blood sugar levels (mg/100g) for 10 animals from each of five breeds (A - E)

A	B	C	D	E
124	111	117	104	142
116	101	142	128	139
101	130	121	130	133
118	108	123	103	120
118	127	121	121	127
120	129	148	119	149
110	122	141	106	150
127	103	122	107	149
106	122	139	107	120
130	127	125	115	116

A program is given on ClickUP under the datasets (Blood sugar.pdf) that can be used to enter the data into SAS.

(a) Give the summary statistics (n, mean and standard deviation) for the data as a whole as well as by breed.

(b) Give a boxplot of the data by breed.

(c) Answer the following questions by using PROC GLM with the statements

```
proc glm;
class breed;
model sugar=breed;
```

i. Test the hypothesis of equality of means for the five breeds.

ii. Do Tukey's test for pairwise comparisons of the five means and draw a conclusion from the results.

iii. Test the hypothesis $H_0 : \mu_A = \mu_B$.

iv. Test the hypothesis $H_0 : \dfrac{\mu_A + \mu_B + \mu_C}{3} = \dfrac{\mu_D + \mu_E}{2}$.

(d) Use PROC GLM to fit the following model to the data:

$$y = \mu + \tau_1 D_1 + \tau_2 D_2 + \tau_3 D_3 + \tau_4 D_4 + \varepsilon$$

where the values of the dummy variables D_1, D_2, D_3 and D_4 are as follows for the different breeds:

Breed	D_1	D_2	D_3	D_4
A	1	0	0	0
B	0	1	0	0
C	0	0	1	0
D	0	0	0	1
E	-1	-1	-1	-1

Use the model to answer the questions that follow.

i. Give and interpret the estimates for the one-way ANOVA model.

ii. Test the hypothesis $H_0 : \mu_D = \mu_E$.

iii. Test the hypothesis $H_0 : \dfrac{\mu_A + \mu_B}{2} = \mu_C$.

(e) Answer the following questions by using PROC GLM using the same dummy variables as in d.

i. Calculate the error sum of squares (SSE), the sum of squares for the model (SSR) and the total sum of squares (SST).

ii. Give the estimates of the parameters.

iii. Test the hypothesis $H_0 : \dfrac{\mu_A + \mu_B + \mu_C}{3} = \dfrac{\mu_D + \mu_E}{2}$.

23. A company studied the effects of three different types of promotions on the sales of a specific brand of crackers:

- Promotion 1 – The crackers were on their regular shelf, but free samples were given in the store.
- Promotion 2 - The crackers were on their regular shelf, but were given additional shelf space.
- Promotion 3 - The crackers were given special display shelves at the end of the aisle in addition to their regular shelf space.

The company selected 15 stores to participate in the study. Each store was randomly assigned one of the 3