

STTN316 Assignment 2

2021

Instructions:

- Complete all questions in this homework assignment.
 - You should spend about two hours a week for the next three weeks on this assignment (put aside at least 5 hours to complete it).
 - You must upload your answers to Assignment 2 - 2021-05-11 for 2021-05-31 on the STTN316 eFundi Assignment page.
 - You must submit your answers before 23h59 on the 31st of May, 2021.
 - Please write your name, surname, and student number on every page that you upload.
 - How to submit: Here are some options on how to submit this assignment (you'll need to pick one):
 - Type the answers out in Word using the equation editor (please note that this is very time consuming). Save the Word document and upload that Word document onto the STTN316 eFundi Assignment page. Use your student number as part of the file name.
 - Write out the answers on a sheet of paper and scan the pages using an App like CAMSCANNER on your phone to create a single PDF file. Save the PDF document and upload that PDF document onto the STTN316 eFundi Assignment page. Use your student number as part of the file name.
 - Write your answers on a sheet of paper, take photos of the assignment pages, and then upload the photos. Try to zip these photos together or create a combined PDF file from all the photos, and upload that single zip or PDF document onto the STTN316 eFundi Assignment page. If you cannot make the zip file or the PDF, then you can upload all the photos as separate files. Use your student number as part of the file names.
-

Question 1:

Consider the linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where \mathbf{Y} is a $(n \times 1)$ vector of observations on a dependent variable; \mathbf{X} is a $(n \times p)$ matrix of observations on non-stochastic independent variables; $\boldsymbol{\varepsilon}$ is a $(n \times 1)$ vector of unobserved disturbances with $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$, and $E(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) = \sigma^2$; and $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown coefficients.

1. Define the least squares residuals as $\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{b}$, where $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Show that $\mathbf{X}'\mathbf{e} = \mathbf{0}$. (1)
 2. Show that $\mathbf{e}'\widehat{\mathbf{Y}} = 0$. (1)
 3. Show that $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$, where $(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. (1)
-

Question 2:

Assume we are working with a model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{bmatrix},$$

$E(\boldsymbol{\varepsilon}) = \mathbf{0}_n$ and the covariance matrix of $\boldsymbol{\varepsilon}$ is $\boldsymbol{\Sigma} = \sigma^2\mathbf{V}$, where

$$\mathbf{V} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix},$$

with $0 \leq \rho \leq 1$. Recall that $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ and that we define $SSE = \mathbf{e}'\mathbf{e}$.

1. *Show* that $\boldsymbol{\Sigma}$ can be written in the form $\boldsymbol{\Sigma} = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}]$. (1)

2. *Show* that $E(MSE) = \sigma^2(1 - \rho) \leq \sigma^2$ (thus MSE underestimates σ^2 in this scenario). (4)

HINT:

- You may use some of the results that are stated in Chapter 5 of the notes to answer this question; you do not have to prove results that are already stated in the notes.
 - You may also use the following result: The sum of each of the rows of \mathbf{H} is equal to 1. Also, since \mathbf{H} is symmetric, we also have that the sum of each the columns of \mathbf{H} is equal to 1.
-

Question 3:

Throughout this module, we have assumed that $SSE/\sigma^2 \sim \chi_{n-p}^2$, however, we have never seen a proof for it! We now look at how to tackle this proof using matrix notation. We start by proving two simple theorems:

Theorem 1: Let \mathbf{Y} be an $(n \times 1)$ vector where $\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{I})$. Next, let \mathbf{A} be a $(n \times n)$ symmetric matrix such that we have the following two properties:

- \mathbf{A} is idempotent, and
- \mathbf{A} has rank r , i.e., $\text{rank}(\mathbf{A}) = r$.

We then have that $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi_r^2$.

Note that the two properties listed for \mathbf{A} imply that \mathbf{A} permits the following eigen-decomposition

$$\mathbf{A} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}',$$

where \mathbf{W} is the *orthogonal matrix* created from the eigen-vectors of \mathbf{A} (these eigen-vectors are denoted $\mathbf{w}_1, \dots, \mathbf{w}_n$ and are orthogonal to one another) and $\mathbf{\Lambda}$ is the $(n \times n)$ diagonal matrix formed from the eigen-values of \mathbf{A} , i.e., $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. However, since $\text{rank}(\mathbf{A}) = r$ we have $\lambda_j = 1, j = 1, 2, \dots, r$ and $\lambda_j = 0, j = r + 1, \dots, n$, which means that we have

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \lambda_{r+1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{(r \times r)} & \mathbf{0}_{(r \times (n-r))} \\ \mathbf{0}_{((n-r) \times r)} & \mathbf{0}_{((n-r) \times (n-r))} \end{bmatrix}$$

Using the above information, answer the following questions to prove the result:

1. Suppose we define $\mathbf{Z} = \mathbf{W}'\mathbf{Y}$. Determine $E(\mathbf{Z})$ and $\text{Var}(\mathbf{Z})$. (2)
2. Let Z_i denote the i^{th} value in the vector \mathbf{Z} . State the distribution of Z_i and Z_i^2 . (2)
3. Use the above two results and show that $\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi_r^2$. (3)

An extension of the previous theorem is to change the variance/covariance structure of the random vector.

Theorem 2: We now define \mathbf{X} as the $(n \times 1)$ vector where $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is defined as a positive definite matrix and so a non-singular matrix \mathbf{S} exists such that we are allowed to write $\mathbf{\Sigma} = \mathbf{S}\mathbf{S}'$. Next, let \mathbf{B} be a $(n \times n)$ symmetric matrix such that we have the following two properties:

- $\mathbf{B}\mathbf{\Sigma}$ is idempotent, and
- $\mathbf{B}\mathbf{\Sigma}$ has rank r , i.e., $\text{rank}(\mathbf{B}\mathbf{\Sigma}) = r$.

We then have that $\mathbf{X}'\mathbf{B}\mathbf{X} \sim \chi_r^2$.

Using the above information, answer the following questions to prove the result:

4. Show that if $\mathbf{B}\mathbf{\Sigma}$ is idempotent and $\mathbf{\Sigma}$ is positive definite (and thus non-singular), then it implies that $\mathbf{B}\mathbf{\Sigma}\mathbf{B} = \mathbf{B}$. (1)
5. Suppose we define $\mathbf{Y} = \mathbf{S}^{-1}\mathbf{X}$. Show that \mathbf{Y} has a $N_n(\mathbf{0}, \mathbf{I})$ distribution as required for Theorem 1. (2)
6. Since \mathbf{Y} satisfies the conditions of Theorem 1, show that $\mathbf{X}'\mathbf{B}\mathbf{X}$ can be written in the form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ (clearly define this new \mathbf{A} matrix using the other quantities introduced above). Determine if this \mathbf{A} matrix satisfies the two conditions stated in Theorem 1 for this type of matrix. (3)

HINT: You may use the following two results without proof:

- The trace of an idempotent matrix equals the rank of the matrix.
- If \mathbf{A} is $(m \times n)$ and \mathbf{B} is $(n \times m)$, then $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$.

7. Use the above results and Theorem 1 to confirm that $\mathbf{X}'\mathbf{B}\mathbf{X} \sim \chi_r^2$. (1)

Now, assume we have the normal error regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$, where the sample residuals are defined as $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ and where the error sum of squares is given by $SSE = \mathbf{e}'\mathbf{e} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$.

8. Set $\mathbf{r} = \mathbf{e}/\sigma$ so that $\mathbf{r}'\mathbf{r} = \mathbf{e}'\mathbf{e}/\sigma^2 = SSE/\sigma^2$. Use the results of Theorem 1 and 2 to show that $\mathbf{r}'\mathbf{r}$ follows a χ^2 distribution with $n - p$ degrees of freedom (thereby proving the result we originally wanted to show). (2)

HINT: You may once again use the following result without proof: The trace of an idempotent matrix equals the rank of the matrix.

Question 4:

Suppose we conduct an experiment where we study the amount time taken (Y) for a matrix multiplication operation to be completed when a large square matrix is multiplied by itself using specific computer code (run-time is measured in minutes). The variables that are used in the experiment as factors that potentially affect the run-time include the version of the compiler used to compile the code

$$X_1 = \begin{cases} \text{Compiler version 0.1.0} \\ \text{Compiler version 1.2.0} \\ \text{Compiler version 2.0.0} \end{cases},$$

as well as the variable X_2 which denotes the number of rows and columns of the square matrix used in the calculation (this variable can take on values anywhere between 1 000 and 1 000 0000).

NOTE: The same code is compiled on different versions of compilers and then the code is executed after it is compiled — it is thought that some versions of the compiler will perform better/worse for these kinds of matrix operations.

The experiment is conducted 9 times and the following results are obtained:

Y	X_1	X_2
12.5	Compiler v0.1.0	1.25×10^5
11.3	Compiler v0.1.0	0.90×10^5
10.4	Compiler v0.1.0	0.95×10^5
12.7	Compiler v1.2.0	1.25×10^5
10.3	Compiler v1.2.0	0.90×10^5
11.6	Compiler v1.2.0	0.95×10^5
12.0	Compiler v2.0.0	1.25×10^5
12.1	Compiler v2.0.0	0.90×10^5
12.5	Compiler v2.0.0	0.95×10^5

Interest lies in comparing the mean run-times of the different compilers to one another.

1. Describe one technique of dealing with the qualitative ‘compiler type’ variable. Carefully define any new variables created using this technique. (1)
2. Briefly explain why you chose this particular technique. (1)
3. You must now decide whether to use a model that incorporates interactions or not. Make a decision and fully motivate your answer. (1)
4. Provide an expression for the model you chose above. Write out the full form of the design matrix for the model you specified. (2)

5. Use your specified model and test the hypothesis that Compiler v0.1.0 produces shorter run times than Compiler v2.0.0 when calculating matrices of size 1.00×10^5 . Carefully state the hypothesis used and test the hypothesis at $\alpha = 0.05$. **(3)**
-

Question 5:

A study is conducted to determine how the attention span of small children is affected by various factors. The variables in the study are:

- Y : Attention span (in minutes),
- X_1 : Amount of sugary drinks consumed (in ml)
- X_2 : Age of child (in months).

The following multiple linear regression model is fit using data obtained from 20 children:

$$\hat{Y}_i = 38.24 - 0.084X_{i1} + 0.0217X_{i2}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.550983 & -0.000797 & -0.011573 \\ -0.000797 & 0.000003 & 0.000004 \\ -0.011573 & 0.000004 & 0.000397 \end{bmatrix}$$

Source	Mean Sum of Squares
Model	1208.284
Error	1.116

Sorted Leverages:

1	2	3	4	5
0.07181859	0.07422332	0.08665982	0.09028041	0.10633005
6	7	8	9	10
0.10730381	0.11410809	0.12050461	0.13903246	0.14909397
11	12	13	14	15
0.15657413	0.15823419	0.16164922	0.17701626	0.18215831
16	17	18	19	20
0.2055259	0.20660687	0.2071679	0.22180272	0.26390938

NOTE: For the questions below, accept that all model assumptions of the normal error multiple linear regression model are satisfied.

1. Using this information construct a 99% confidence interval for the expected response value (from a model based on these 2 predictor variables) for children that consume 100ml of sugary drinks and are 20 months old. (3)
2. Using this information construct a 99% prediction interval for a specific child that consumes 100ml of sugary drinks and is 20 months old. (1)
3. Interpret both intervals and, in particular, explain how the interpretation differs between these two intervals. In which circumstances would you recommend calculating the one over the other? (2)

4. Is the prediction interval calculated above a form of ‘extrapolation’ or ‘interpolation’? Calculate a value that will help motivate your answer. (1)
5. Test the hypothesis that the age of the child is a significant predictor in the model. State the hypothesis, calculate the test statistic, determine the critical value, and state your conclusion. (2)
-

Question 6:

Suppose that you have to find the best linear regression model for a response variable Y using four predictor variables X_1, X_2, X_3 and X_4 . Using an “all possible sub-sets” approach, you calculate the Mallows’s C_P, R^2 and R_{adj}^2 measures for all possible combinations of models that you could build using these three predictor variables and you tabulate the results below.

C_P	R^2	R_{adj}^2	Variables in model
138.7	0.675	0.645	X_4
142.5	0.666	0.636	X_2
2.7	0.979	0.974	X_1X_2
5.5	0.972	0.967	X_1X_4
3.0	0.982	0.976	$X_1X_2X_4$
3.0	0.982	0.976	$X_1X_2X_3$
5.0	0.982	0.974	$X_1X_2X_3X_4$

1. Use the above output and state which model you would consider using. Motivate your answer. **(2)**
2. What happens to the R_{adj}^2 measure as the number of variables added to the model becomes very large, i.e., $p \rightarrow \infty$. Does this result change when $n \rightarrow 1$ or $n \rightarrow \infty$? Motivate your answer. **(1)**
3. For fixed n, p and $SSTO$ values, indicate what the highest and lowest values R_{adj}^2 can assume? **(2)**

Question 7:

A research psychologist approaches you a research proposal. He wants to publish an article that details the effect that gender, weight, and income levels play in determining ‘Life Satisfaction’ (LS) of university students that graduated from South African universities in 2020. The LS variable is measured by having the participants take a questionnaire and produces an LS score between 0 and 100 (0 being the worst life satisfaction score and 100 being the best). The researcher wants to determine the relationship between the various levels of these factors and the expected value of the LS score for a population under study.

Write a short 100 to 300 word proposal that details your approach to addressing this research. Explain all necessary steps required to collect the necessary data, arrive at an appropriate model, and ensure that the model is adequate. In addition to taking the researcher’s problem statement into account, your proposal must also keep in mind real-world limitations of such studies. (4)