Homework 5

Write your name here

19 May, 2022

Background: For this section, we will work with the final grades of students in a statistics class. We are interested in the effect of the hours of study and number of classes missed on the final grades of students in a statistics course. We have the grades (`grades`) of 100 students alongside the number of classes that they missed during a quarter (`classes_missed`) and the average number of hours that the student reported they spent studying on average during a week (`study_week`).

```
# Code 1: Load the data (The link can be found at the end of lecture 22)
#
# In this section you should write the code needed to load the data
# IMPORTANT: Remember to delete the comments inside the code
# chunks BEFORE submitting your homework.
```

1. Obtain the prediction and error of a the Null model for this study. What is the prediction of the Null model?

```
# Code 2: calculating the null model's predictions and error
#
# In this code chunk you can write the functions that you need in order
# to calculate the predictions and error of the null model.
```

ANS: [WRITE YOUR ANSWER HERE]

2. Obtain the predictions and error of a **simple linear regression** model that includes **only** the **hours of study** as predictor of a student's final grade. What are the values of the parameters in the model?

```
# Code 3: simple linear regression (hours of study)
#
# In this code chunk you can write the functions that you need to obtain the
# predictions and errors for the simple linear regression that includes only
# hours of study as a predictor.
```

ANS: [WRITE YOUR ANSWER HERE]

3. Obtain the predictions and error of a **simple linear regression** model that includes **only** the **number of classes missed** as predictor of a student's final grade. What are the values of the parameters in the model?

```
# Code 4: simple linear regression (hours of study)
#
# In this code chunk you can write the functions that you need to obtain the
# predictions and errors for the simple linear regression that includes only
# hours of study as a predictor.
```

ANS: [WRITE YOUR ANSWER HERE]

4. Obtain the predictions and errors of a **multiple linear regression** model that includes both **hours of study** and the **number of classes missed** as predictors of a students final grade. What are the

values of the parameters in the model?

```
# Code 5: multiple linear regression (hours of study and classes missed)
#
# In this code chunk you can write the functions needed to obtain the
# predictions and errors of a multiple linear regression that includes
# hours of study and classes missed as predictors.
```

ANS: [WRITE YOUR ANSWER HERE]

5. Someone has noticed that in our data set we seem to have 2 groups, a group of students that never missed a class (0 classes missed) and a group of students that missed at least one class. Create an indicator variable that assigns the value 1 to students that missed one or more classes and assigns the value 0 to students that missed no classes. What is the proportion of students that missed at least one class?

```
# Code 6: Creating a new indicator variable for students that missed a class
#
# In this code chunk you can write the functions that you need in order to
# create an indicator variable that takes the value 1 when the student missed
# one class or more and takes the value 0 if they missed no classes.
# (hint: how does the average of the indicator variable relate to the
# proportion?)
```

ANS: [WRITE YOUR ANSWER HERE]

6. Using your new categorical variable (the one you created in question 5) obtain the predictions and error of a **multiple linear regression** that includes the **hours of study** and the **indicator of at least one class missed** as predictors of a student's final grade. What are the values of the parameters in the model?

```
# Code 7: Calculating the predictions and errors of a multiple linear regression
# with hours of study and the indicator of at least one class missed
#
# In this code chunk you can write the functions that you need in order to
# calculate the predictions and errors of the multiple linear regression that
# includes the hours of study and the indicator of at least one class missed
# as predictors of final grade.
```

ANS: [WRITE YOUR ANSWER HERE]

7. Using your new categorical variable (the one you created in question 5) obtain the predictions and error of a **multiple linear regression** that includes the **hours of study**, the **indicator of at least one class missed** and an interaction between those variables as predictors of a student's final grade. What are the values of the parameters in the model?

```
# Code 8: Adding the predictions and error for a multiple linear regression that
# includes hours of study, the indicator of at least one class missed and
# the interaction of those variables as predictors
#
# In this code chunk you can write the functions that you need in order to
# obtain the predictions and errors of this multiple linear regression model
```

8. Calculate the Sum of Squared Errors of each of the 6 linear models that you have (Null, hours of study, classes missed, hours of study + classes missed, hours of study + indicator, hours of study + indicator + hours of study * indicator).

```
# Code 9: obtain the SSE of each of the 6 linear models that you have used.
#
```

```
# In this code chunk you can write the functions that you need in order to
# calculate the SSE for each linear model used before. (your code will be the
# answer you don't need to print the values)
```

9. Calculate the Mean SE of each of the 6 linear models that you have.

```
# Code 10: obtain the Mean SE of each of the 6 linear models that you have used.
#
# In this code chunk you can write the functions that you need in order to
# calculate the Mean SE for each linear model used before. (your code will be
# the answer you don't need to print the values)
```

10. Calculate the BIC of each of the 6 linear models that you have. What are the BIC values of each model (remember to indicate which value belongs to which model)?

```
# Code 11: obtain the BIC value for each of the 6 linear models.
#
# In this code chunk you can write the functions that you need in order to
# calculate the BIC of each of the 6 linear models.
```

ANS: [WRITE YOUR ANSWER HERE]

11. Based on the BIC values, which model would you select? Interpret the values of the parameters of the model that you selected in the context of the experiment.

ANS: [WRITE YOUR ANSWER HERE]

12. What is the proportion of the variability accounted for by the model?

```
# Code 13: obtain the proportion of variance accounted for (R^2) of the model
# that you choose on question 11.
#
# In this code chunk you can write the functions that you need in order to
# calculate the proportion of variance accounted for by the model that you
# selected.
```

ANS: [WRITE YOUR ANSWER HERE]

13. How would you summarize the results of your model comparison? (for example, what does the model that you selected tell us about our original research question?)

ANS: [WRITE YOUR ANSWER HERE]