

5A-ModelComparison

March 3, 2022

1 Lab 5A: Model Comparison

1.0.1 PSYC 193L: Science of Learning Data Science

SleepStudy Data from a study of sleep patterns for college students. The data were obtained from a sample of students who completed various cognitive tasks, a survey of their attitudes and habits, and kept a sleep diary to record time and quality of sleep over a two week period. [Source](#)

1.1 Part 0: Reminder about collaboration on lab assignments in PSYC 193L

We strongly believe in the value of collaborating with your peers for enhancing your learning experience in PSYC 193L. Being able to successfully collaborate with others is an important skill to have when you enter the workforce, and everyone can get better at collaboration with practice. However, for collaboration to be maximally valuable, we need to set some ground rules, building on the expectations laid out in the [course syllabus](#): Show Up, Try, Ask for Help When You Need It, Be Professional.

1.1.1 Guidelines

- Rotate responsibilities between group members.
- Choose a driver and a navigator.
- Discuss your thinking process openly with your group.
- Be supportive, respectful, and patient with one another.

Rotate responsibilities between group members. You will generally be working with the same people (from your discussion section) for each lab, and a new group of people for the next lab. Because you will be working with the same classmates for the next week or so, you will have the opportunity to share responsibilities with one another.

Choose a driver and a navigator. We suggest that one group member volunteer to act as the “driver” (and share their screen) while the other group members act as the “navigators.” Next time, it is a good idea to exchange roles, so that everyone gets a chance to act as the driver at least once, if possible.

Discuss your thinking process openly with your group. We suggest that you discuss the way you are thinking about each problem with your group. It is more important to us that you gain practice explaining your reasoning to yourself and to your peers than it is to simply state what you think the “right answer” is, without explaining your reasoning.

Be supportive, respectful, and patient with one another. Try to give everyone an opportunity to play both a leading and supporting role. If you feel relatively comfortable with R, we encourage you to proactively encourage other members of your group who feel the least confident about writing R code to take a leading role. If you feel less confident about your R skills, please know that you are not alone! With practice and persistence over the course of the coming weeks, you will find your skills improving!

Every student is still responsible for submitting their own lab assignments. Although you are encouraged to work together on these lab assignments, please remember that everyone is responsible for submitting their own lab assignments.

```
[ ]: ## Run this code to load the required packages
suppressMessages(suppressWarnings(suppressPackageStartupMessages({
  require(tidyverse)
  require(supernova)
  require(ggformula)
  require(mosaic)
  require(NHANES)
})))
```

1.1.2 Learning objectives

The purpose of this lab is to get practice using R to compare different linear models with their corresponding null models.

In **Lab 5A**, we compare the models based on the Analysis of Variance table.

In **Lab 5B**, we will study about Null hypothesis significance testing (NHST) and then apply NHST to the variables we studied in Lab 5A.

1.1.3 Load the Lock5withR package which has the SleepStudy dataset

```
[ ]: require(Lock5withR)
```

SleepStudy Dataset variable description:

A dataset with 253 observations on the following 27 variables.

- Gender 1=male, 0=female
- Sex Female or Male
- ClassYear Year in school, 1=first year, ..., 4=senior
- LarkOwl Early riser or night owl? Lark, Neither, or Owl
- NumEarlyClass Number of classes per week before 9 am
- earlyClass Indicator for any early classes
- EarlyClass Indicator for any early classes
- GPA Grade point average (0-4 scale)
- ClassesMissed Number of classes missed in a semester
- CognitionZscore Z-score on a test of cognitive skills
- PoorSleepQuality Measure of sleep quality (higher values are poorer sleep)
- DepressionScore Measure of degree of depression

- **AnxietyScore** Measure of amount of anxiety
- **StressScore** Measure of amount of stress
- **DepressionStatus** Coded depression score: normal, moderate, or severe
- **AnxietyStatus** Coded anxiety score: normal, moderate, or severe
- **Stress** Coded stress score: normal or high
- **DASScore** Combined score for depression, anxiety and stress
- **Happiness** Measure of degree of happiness
- **AlcoholUse** Self-reported: Abstain, Light, Moderate, or Heavy
- **Drinks** Number of alcoholic drinks per week
- **WeekdayBed** Average weekday bedtime (24.0=midnight)
- **WeekdayRise** Average weekday rise time (8.0=8 am)
- **WeekdaySleep** Average hours of sleep on weekdays
- **WeekendBed** Average weekend bedtime (24.0=midnight)
- **WeekendRise** Average weekend rise time (8.0=8 am)
- **WeekendSleep** Average weekend bedtime (24.0=midnight)
- **AverageSleep** Average hours of sleep for all days
- **allNighter** Had an all-nighter this semester? Yes or No
- **AllNighter** Had an all-nighter this semester? 0 or 1

1.2 Part 1: Preprocessing the dataset

1.1 Print the first 6 rows of the SleepStudy Dataset.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

1.2 Create a boxplot to visualize the relationship between **Stress** and **PoorSleepQuality**. **NOTE:** Use `gf_labs` to give the plot an appropriate x-axis label, y-axis label and title.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

1.3 Based on your visualization, what do you think is the relationship between **PoorSleepQuality** and **Stress**? Explain what features of the visualization justify your response.

YOUR ANSWER HERE

1.4 Create a scatterplot to visualize the relationship between **Happiness** and **PoorSleepQuality**. **NOTE:** Use `gf_labs` to give the plot an appropriate x-axis label, y-axis label and title.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

1.5 What do you think is the relationship between **PoorSleepQuality** and **Stress**? Explain what features of the visualization justify your response.

YOUR ANSWER HERE

Notice the **PoorSleepQuality** column values. The lower the value, the better the sleep quality. However, this way of encoding sleep quality may not be the most intuitive for many people. Hence, we will create a new column named **SleepQuality** such that the higher the value, the better the

sleep quality. This is also known as reverse coding. In the next question, we will walk through how to do this reverse coding step-by-step.

1.6 Recoding the PoorSleepQuality variable

1.6.1 Store the maximum PoorSleepQuality value as an R-object named **maxPSQ**. Similarly, Store the minimum PoorSleepQuality value as an R-object named **minPSQ**. *Hint: Try using the `min()` and `max()` functions.*

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

1.6.2 Now subtract all the PoorSleepQuality values from **maxPSQ** and then add **minPSQ**. Store the result as SleepQuality column within the SleepStudy dataframe.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

1.3 Part 2: Relationship between stress and sleep quality using model comparison

2.1 Fit a linear model called **stressModel** that predicts SleepQuality on the basis of Stress.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

2.2 Apply the **supernova** function to conduct an analysis of variance on **stressModel**.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

2.3 What can you tell from the **PRE** (Proportion Reduction in Error) value? How would you explain what this number means in plain language to someone who may not have yet taken PSYC 193L?

YOUR ANSWER HERE

2.4 What can you tell from the **F value**? What kind of information does the **F value** provide that may not be captured by **PRE** alone, and when might that be important to take into account?

YOUR ANSWER HERE

1.4 Part 3: Relationship between happiness and sleep quality using model comparison

3.1 Fit a linear model called **happinessModel** that predicts SleepQuality on the basis of Happiness.

```
[ ]: # your code here
fail() # No Answer - remove if you provide an answer
```

3.2 Apply the **supernova** function to conduct an analysis of variance on **happinessModel**.

```
[ ]: # your code here  
fail() # No Answer - remove if you provide an answer
```

3.3 What can you tell from the **PRE** value? How does it compare to the **PRE** value you found using the `stressModel` from Part 1?

YOUR ANSWER HERE

3.4 What can you tell from the **F value**? How might you use other columns of the table produced by calling the `supernova()` function to compute this **F value**?

YOUR ANSWER HERE

3.5 Create an appropriate plot to visualize the relationship between **Happiness** and **SleepQuality**. Also use `gf_lm()` to include a line visualizing the predictions of our best-fitting linear regression model.

```
[ ]: # your code here  
fail() # No Answer - remove if you provide an answer
```

2 All done? Great work!

Please add additional comments regarding either what you learned during this lab assignment so far, or aspects of it you found particularly challenging here.

INSERT YOUR RESPONSE HERE.