

# 5B-Hypothesis Testing

March 3, 2022

## 1 Lab 5B: Hypothesis Testing

### 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 dataset and apply preprocessing

```
[ ]: ### Load the `Lock5withR` package which has the `SleepStudy` dataset
require(Lock5withR)

## Preprocessing to add SleepQuality as a column to the dataframe
maxPSQ <- max(SleepStudy$PoorSleepQuality)
minPSQ <- min(SleepStudy$PoorSleepQuality)
SleepStudy$SleepQuality <- maxPSQ - SleepStudy$PoorSleepQuality + minPSQ
```

## 1.2 Part 1: Introduction to Null Hypothesis Significance Testing

### 1.2.1 Model comparison

In this class we have focused on the central concept of fitting models to explain variation in data. The question we are trying to answer is: **which model does the best job of explaining such variation, without being overly complicated?**

$$DATA = MODEL + ERROR$$

So far you have learned about a specific kind of model that is widely used in statistics – the **General Linear Model (GLM)**. We’ve focused in this class on two major use cases of the GLM:

1. To specify the **empty model** where use a single number to predict the mean of the outcome variable. The only **parameter** to estimate is the mean, represented by  $b_0$ . In the equation below,  $Y_i$  represents a specific observed value of the outcome variable and  $e_i$  represents the error, or the deviation, between the mean ( $b_0$ ) and the observed value of the outcome variable,  $Y_i$ , for the  $i$  – *th* member of the dataset.

$$Y_i = b_0 + e_i$$

2. To specify a **model using an explanatory variable** to predict the mean of the outcome variable. These models will require you to estimate more than a single parameter.
  - If your explanatory variable is **quantitative**, you will be estimating two parameters:  $b_0$  again, representing the **y-intercept** and  $b_1$ , representing the **slope** of the best-fitting line.

$$Y_i = b_0 + b_1X_i + e_i$$

- If your explanatory variable is **categorical**, you will be estimating  $(L - 1) + 1$  parameters, where  $L$  is the number of levels of your categorical variable. For example, if your explanatory variable consists of three different levels (or groups), then see below for what your model would look like. Below,  $b_0$  represents the “y-intercept” of your model, but  $Y_i$  and  $e_i$  have the same meaning as in the previous equation.  $b_1$  represents the “effect of belonging to level 1 relative to the baseline group” and  $b_2$  represents the “effect of belonging to level 2 relative to the baseline group.”

$$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$$

Once you’ve fit both an empty model and model using an explanatory variable, how do you know which is better? That is where the concept of **Proportion Reduction in Error (PRE)** and the **F value** come into play. We can use **PRE** to ask how much better the more complex model (which contains an explanatory variable, and thus has more parameters to estimate) does at reducing error relative to the simpler empty model. We can use the **F value** to measure how much error is reduced for each additional parameter we have to estimate (the more parameters, the more complex the model, the less appealing it is *unless* we gain a lot in terms of the error reduction).

When we pit two models against each other to see which one does a better job, taking into account the complexity of each model, this is called **model comparison** and is a powerful and general approach to using statistics to explain real-world data.

## 1.2.2 Null hypothesis significance testing

We have spent comparatively less time on fleshing out the concept of “hypothesis testing,” or more specifically, “**Null Hypothesis Significance Testing (NHST)**.” NHST is very commonly used to evaluate empirical claims in psychology so we thought we should spend some time on it. However it also a much less powerful and general approach than model comparison, which is why we did not spend more time on it. The question being asked when we conduct NHST is: **How unlikely are**

the data under the null hypothesis (empty model)? Here are the key steps involved when performing NHST:

**Step 1: Formulate your research hypothesis about the relationship between your outcome variable and explanatory variable.** For example, “Romulans and Vulcans have different average heights.”

**Step 2: Specify null and alternative hypotheses.** For example, my **null hypothesis** is that “Romulans and Vulcans have the same average heights. My **alternative hypothesis** is that Romulans and Vulcans have different average heights.”

**Step 3: Collect some data relevant to the hypothesis.** For example, imagine observing several star buses arrive and measuring the height of each passenger.

**Step 4: Compute a test statistic that is based on the null hypothesis being true.** For example, when comparing the average heights of Romulans and Vulcans, we might compute something called a **t-statistic** that reflects how different the mean heights are between groups in our sample (star bus), divided by the standard deviation in heights. Here is the formula for calculating the *t* statistic when comparing the means of two independent groups:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Here, the  $\bar{X}$ ’s represent the sample mean heights for each group of aliens, *S*’s represent the sample standard deviation for each group, and the *n*’s represent the number of aliens in each group. We can use built-in functions in *R* to compute this *t*-statistic, as well as other common “test statistics.”

**Step 5: Compute the probability of the observed value of that statistic assuming that the null hypothesis is true.** If the absolute value of the *t* statistic is **large**, this means that if the null hypothesis were true (that there is no “real” difference in heights), and any observed differences in our sample are just due to sampling variability rather than a “real” difference, the difference in mean heights between groups would be **pretty unlikely**.

On the other hand, if the absolute value of the *t* statistic is **small** or close to zero, this means that if the null hypothesis were true (that there is no “real” difference in heights), the observed difference is **pretty likely** (i.e., not that surprising).

We can use built-in functions in *R* to convert a *t*-statistic to an exact probability value (or **p-value**). Probabilities can only take on values between zero and one. Low *p*-values close to zero reflect unlikely outcomes, whereas high *p*-values closer to one reflect likely outcomes.

**Step 6: Assess the “statistical significance” of the result.** Under conventional applications of NHST in psychology, a *p*-value that is below 0.05 is considered **statistically significant**. A *p*-value that is above 0.05 is considered **not statistically significant**. Keep in mind that the word **significant** in this context does not mean the same thing as “important” or “meaningful.” There can be “statistically significant” effects that are not that meaningful, in the sense of making a noticeable, practical difference.

If an effect is “statistically significant,” then you would conclude that “**There is sufficient evidence to reject the null hypothesis.**” If an effect is **NOT** “statistically significant,” then you would conclude that “**There is insufficient evidence to reject the null hypothesis.**”

**Important:** When you have sufficient evidence to reject the null hypothesis, you can’t say that you’ve shown that your alternative hypothesis is true. Only that you’ve provided evidence against the null hypothesis explaining your data.

YOUR ANSWER HERE

**Hypotheses concerning the relationship between Stress and PoorSleepQuality** I hypothesize that \_\_\_\_\_.

My null hypothesis is: \_\_\_\_\_.

My alternative hypothesis is: \_\_\_\_\_.

YOUR ANSWER HERE

**Hypotheses concerning the relationship between Happiness and PoorSleepQuality** I hypothesize that \_\_\_\_\_.

My null hypothesis is: \_\_\_\_\_.

My alternative hypothesis is: \_\_\_\_\_.

### 1.3 Part 2: Testing relationship between stress and sleep quality using a hypothesis test

**2.1.0** Compute a t-statistic for the difference in sleep quality between stress groups. The function `t.test()` is available in R for performing t-tests. Let’s test it out first and examine the output. [Reference](#)

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

**2.1.1** Since `Stress` is a categorical variable, the correct way to use the function would be `t.test(y~x,data=yourData)`. Now, replace `y`, `x` and `yourData` to conduct a t-test that explains variation in `SleepQuality` with variation in `Stress`. Store the output in a R-object called `ttest1`.

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

**2.2** Let’s compute the p-value from the `ttest1` object step-by-step.

**2.2.1** We need to find out how to extract the p-value (or some other quantity of interest) from the output of the `t.test` function. For this function, the R help page has a detailed list of what the object returned by the function contains. A general method for a situation like this is to use the `names` function to find where the quantity of interest is.

Use the `names()` function to see the details of `ttest1` object

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

**2.2.2** Notice that the value we want is named ‘p.value’. To extract it, we can use the dollar sign notation (similar to what we do for a dataframe). Now extract the p.value from the `ttest1` object

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

**2.3** Assess the “statistical significance” of the result and decide whether to “reject” or to “fail to reject” the null hypothesis.

YOUR ANSWER HERE

**2.4** Look back at Lab 5A to see the results you obtained from fitting a linear model to analyze the relationship between stress and sleep quality. How do the results of your hypothesis test compare to those obtained in Lab 5A?

YOUR ANSWER HERE

## 1.4 Part 3: Testing relationship between happiness and sleep quality using a hypothesis test

**3.1** Since `Happiness` is a continuous variable, the correct way to use the function would be `t.test(yourData$y,yourData$x)`. Now, replace `y`, `x` and `yourData` to conduct a t-test that explains variation in `SleepQuality` with variation in `Happiness`. Store the output in a R-object called `ttest2`.

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

**3.2** Compute the p-value from the `ttest2` object (Similar to what you did in **2.2**)

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

**3.3** Assess the “statistical significance” of the result and decide whether to “reject” or to “fail to reject” the null hypothesis.

YOUR ANSWER HERE

**3.4** Look back at Lab 5A to see the results you obtained from fitting a linear model to analyze the relationship between happiness and sleep quality. How do the results of your hypothesis test compare to those obtained in Lab 5A?

YOUR ANSWER HERE

## 1.5 Part 4: Reflecting on the relationship between model comparison and hypothesis testing

**4.1** What are some general similarities between **model comparison** and **hypothesis testing**? What kinds of questions do they answer, and how do they answer those questions? Please be

detailed and concrete.

YOUR ANSWER HERE

**4.2** What are some differences between **model comparison** and **hypothesis testing**? Please be detailed and concrete. When do you think might these differences be important? Please explain your reasoning.

YOUR ANSWER HERE

### 1.5.1 LAB 5 Reflection

As part of each lab we ask that you look back at the three parts that you completed and write a short reflection about the experience with the lab.

- What parts did you find most challenging?
- What was the most interesting part?
- How does this lab connect to the CourseKata readings?
- What concepts did this lab cover?
- Is there anything else you would like us to know about your experience with the lab? (Note: this response is a large part of your grade on the overall lab, please take time to give us thoughtful feedback, as it will help us make the course better for future students!)

**WRITE YOUR REFLECTION HERE. PLEASE DO YOUR BEST TO ADDRESS EACH OF THE QUESTIONS ABOVE.**

Before submitting this lab via DataHub, remember to consult the pre-submission checklist on the course website.