PSTAT 122
Lab 7

Due on Thursday 5/22 at midnight

Please upload your Rmd file AND your pdf output file to Canvas.

Introduction

Today you will again be collecting data in class to investigate a question regarding how long you can hold your breath. This time, we will perform a 2^2 Factorial experiment, with the following two factors:

- Do 10 jumping jacks just prior to holding your breath (yes/no)
- Hold your arms straight up in the air while taking in the breath that you will hold (yes/no)
  - Your arms can come back down while you hold your breath

Data Collection

This time, each participant will record their own time using their own device. Please share devices as needed if anyone does not have a cellphone or other suitable timing device (e.g. a watch would be fine as well if it shows seconds). You can (and should) look at the amount of time during the whole duration of when you are holding your breath.

Anyone with a medical condition that prevents you from safely holding your breath for a prolonged period of time or doing either of the two treatment conditions may sit out and just watch.

Each of you should have an R session open for the purpose of randomly assigning yourself to a treatment group. This can be done as follows:

```
treatment <- c("control", "jumping jacks", "arms up", "both")
sample(treatment, 1)
```

Sidenote:

This is not the most ideal way to assign our treatment groups but we will do it this way for time and ease. The reason this is not the most ideal way is because it leaves the distinct possibility that the treatment groups will be unbalanced due to chance, or even have some combinations with no observations at all.

A better way to do it would be to predetermine the number of observations that each treatment combination will get, and distribute those out at random (similar to what is described below in the data collection option if you are doing it on your own). To do this in discussion sections would require knowing exactly how many people are going to show up to each discussion section, or coding it on the fly once everyone has shown up, but this is a bit too much to deal with in the timeframe of a 50 minute discussion section.

However, these are all things that should be discussed in your lab report (and your final project if relevant).

Times should be recorded at whatever level of precision your timing device gives (e.g. hundredths of a second). Then, all times should be compiled on the whiteboard in the room, on a table that looks something like this:

		Jumping Jacks	
		No	Yes
Arms Up	No		
	Yes		

All values should go into this table on the whiteboard, with multiple values per cell as appropriate. In this case, order of data collection is going to be very cumbersome to manage and record, so we will ignore that for this experiment (and also we aren't doing model checking for this lab since we haven't yet covered how to do that for factorial experiments).

In the unlikely event that any treatment group did not get any participants into it by random assignment, a volunteer should be obtained to switch into that treatment group. Any downsides of having done this should be discussed in your report.

Take a photo of the completed data on the whiteboard to include with your report.

If you did not attend a discussion section this week

You will be on your own to collect similar data by yourself, on yourself. In this case, you will obtain exactly 3 full replicates, for a total sample size of 12.

Thus, your random assignment can look like this:

```
treatments <- c(rep("control", 3), rep("arms up", 3),
                rep("jumping jacks", 3), rep("both", 3))

sample(treatments)
```

```
##  [1] "arms up"       "control"       "both"          "jumping jacks"
##  [5] "jumping jacks" "control"       "control"       "both"
##  [9] "both"          "arms up"       "arms up"       "jumping jacks"
```

This will give you a random order in which to assign the treatments to yourself, and you will collect your data in this order (the above is of course just an example order; yours may be different due to random chance and you should follow whatever output your code gives you – make sure to set a seed so that you don't lose the order and so that it can be included in your report!)

Sidenote: as mentioned above, this in fact would be the ideal way to do our random assignment for the in-class data collection – and if relevant and possible, this is how it should be done for your final projects

– but in the span of a 50-minute class session it would require a bit more time and effort than we want to expend.

Your report should explain in detail any advantages and disadvantages of what you have done here compared to what was done in the discussion sections, along with any specific issues that you ran into. If you have a medical condition that prevents you from safely holding your breath repeatedly, you may either enlist the help of a friend, or come up with something comparable to do on your own.

Part 1: Written Report

Now, we first enter all of these data into `R`. As always, it MUST be in the form of a rectangular dataframe! That is, one in which each row is an observation, and each column is a variable. The columns should be:

- time
- jumping jacks (coded as yes/no or 0/1)
- arms up (coded as yes/no or 0/1)

Your data may be entered into e.g. an Excel file and then imported into your R Markdown file, or you may input your data directly into R Markdown. Please either provide your Excel file or show your code for inputting your data.

Then, write a report following the general specifications laid out previously. Specifically for this case, you will have the following sections with these specifications:

Methods

- Start with a brief description of the aim of this lab
- Then give a description of how your data were collected, including how randomization was performed
- Describe any technical issues that may have come up, such as timer problems, or anything else.

Results

- Include an image of the whiteboard data.
- Then, present an appropriate graphical representation of your data and briefly describe what you see there.
- Include also a table of summary statistics within each treatment combination, such as:
  - sample mean
  - sample standard deviation
  - sample size
- Carefully write out your null and alternative hypotheses for your primary question of interest; if you use any mathematical symbols, carefully define what they represent
- Show your code for performing the hypothesis test. Make sure that any output from your test is not simply raw `R` output.
- Write your results in text, with a conclusion in the context of the question at hand.
- Carefully draw conclusions from the p-values of the individual coefficients, as discussed in Lecture 12.

Discussion

- Briefly summarize again what you did, and what you found or did not find.
- Describe any drawbacks of what you did, and of this study design in particular.
  - Specifically, as noted above, we did our randomization a little suboptimally.
  - Also describe any drawbacks of doing the timing on yourself.
  - If you were not in a discussion section and collected data on your own, a careful description of the drawbacks you ran into while doing that should be included here.

Other notes (same as before)

- All `R` code that performs anything beyond simply printing a table or a graph should be shown in the report. Note that this may make your narrative flow slightly less fluidly than it would if you did not show all of your code; that is ok and is exactly what we expect for this class.
- Again as noted previously, raw `R` **output** should not be shown anywhere in the lab report. This includes all `warnings` and `messages`; all such things should either be suppressed or otherwise dealt with so that they do not appear.

Part 2

For entering data as needed, you may do so either directly into your .Rmd file, or you may first enter them into a spreadsheet (such as Excel) and then import that into your .Rmd file.

1) Consider the data in Exercise 3.17 in our textbook:

3.17 A regional opera company has tried three approaches to solicit donations from 24 potential sponsors. The 24 potential sponsors were randomly divided into three groups of eight, and one approach was used for each group. The dollar amounts of the resulting contributions are shown in the following table.

Approach	Contributions (in $)							
1	1000	1500	1200	1800	1600	1100	1000	1250
2	1500	1800	2000	1200	2000	1700	1800	1900
3	900	1000	1200	1500	1200	1550	1000	1100

Enter these data into your R Markdown file as a rectangular data frame. Then,

- (a) Run `aov` on these data to investigate whether there are any differences in the mean contribution according to the approach. Report your p-value and conclusions at $\alpha = 0.05$.
- (b) Run `lm` on these data to investigate whether there are any differences in the mean contribution according to the approach. Confirm that your p-value from `lm` matches what you obtained from the ANOVA.

2) Consider the data from Exercise 6.14 in our textbook:

6.14 An article in the *AT&T Technical Journal* (March/April 1986, Vol. 65, pp. 39–50) describes the application of two-level factorial designs to integrated circuit manufacturing. A basic processing step is to grow an epitaxial layer on polished silicon wafers. The wafers mounted on a susceptor are positioned inside a bell jar, and chemical vapors are introduced. The susceptor is rotated, and heat is applied until the epitaxial layer is thick enough. An experiment was run using two factors: arsenic flow rate (A) and deposition time (B). Four replicates were run, and the epitaxial layer thickness was measured (μm). The data are shown in Table 6.1.

TABLE P6.1

The 2^2 Design for Problem 6.14

		Replicate					Factor Levels	
A	*B*	I	II	III	IV		Low (−)	High (+)
−	−	14.037	16.165	13.972	13.907	*A*	55%	59%
+	−	13.880	13.860	14.032	13.914			
−	+	14.821	14.757	14.843	14.878	*B*	Short	Long
+	+	14.888	14.921	14.415	14.932		(10 min)	(15 min)

Enter these data into your R Markdown file as a rectangular data frame. Set the "Low" level as the baseline level for each variable. Then,

- (a) Make two graphical representations of the data (refer back to Lecture 12), and comment briefly on what you observe.
- (b) Report the overall p-value from the linear model and write a brief statement of conclusions at $\alpha = 0.05$. Also evaluate the individual coefficient p-values according to what was discussed in Lecture 12.
- (c) Write down a regression equation, similar to what was seen in Lecture 13 Slide 26.
- (d) What is the estimated difference in epitaxial layer thickness attributed to Factor A, when Factor B is at the "high" level?