````
---
title: "Stat 021 Homework 1"
author: "Tyler Soutendijk"
date: "Due: Fri, Sept 13"
output: pdf_document
---

```{r setup_pres, include=FALSE, echo=FALSE}
rm(list=ls())
library('tidyverse')                                        ##All of
this code here is called the "preamble"
library('gridExtra')
library('cowplot')

setwd("~/Desktop/Fall 2019 Courses/STAT")  ##fill in with your
own directory instead
```



<center>**Instructions:** A hard copy of your homework must be
handed in to me at the end of class on the due date or I must
have recieved via email a **pdf** version of your homework by
**noon** on the due date. If you are having trouble getting
your *.Rmd* file to compile, you need to get help with this
**before** the due date.

You are allowed to hand in **only one** late homework
assignment throughout the semester. If you need to hand in this
particular assignment late, you must let me know via email by
noon on the due date.

You are encouraged to study with your peers to help complete
the homework assignments but no copying is allowed. If I see
that two or more homework assignments are copied, all students
involved will recieve a grade of $0$ on that assignment and
will forfeit (perhaps retroactively) the opportunity to hand in
a late homework. </center>

***

**Q 1)** Download and install R and R Studio following the
instructions in class. Install the package *swirl()*
using the command "install.packages("swirl")". Once the package
````

1a celcius temp

is installed, call the package to your R session using the command "library("swirl"). Follow the instructions that pop up in your console. Select the course option "1: R Programming: The basics of programming in R" and complete the following lessons (about 20 minutes to complete each)

- 1: Basic Building Blocks

- 2: Workspace and Files

- 8: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the tutorial is covering something that you already know how to do, use the *skip()* command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (0 points)

Q 2) Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truely being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (5 points)

Solution:
My experimental study will investigate a population of college students at Swarthmore (n=100). The sample will be selected as the first 100 students to sign up that sleep on the standard Swarthmore housing spring mattresses. If eligable for the study, college students will be compensated for their time. The experimental study tests the correlation between sleeping on a memory foam mattress and alertness. A hypothesis would be that those who sleep on a memory foam mattress have greater alertness in the morning than those who sleep on a spring mattress. One bias that would be evident here is volunteer or self-selection bias from those who will opt to participate in the study because of the compensation. Another bias is for those with either pre-existing insomnia (sleeping issues) OR a

deficit in alertness. These biases will absolutely keep the sample from being truely random.

Once a sample is selected, treatment groups will be assigned by random number generation within a set of 100 numbers. If a student (basic observational unit) recieves an odd number then they are assigned to the control group. If a student receives an even number then they are assigned to the treatment group. The control group will sleep on a spring mattress that is masked as a foam mattress. The treatment group will sleep on a foam mattress. Both groups will spend the night on their mattresses and then tested for alertness through cognitive testing in the morning. The cognitive testing will involve exercises similar to the Acute Concussion Evaluation (ACE) test. Results will be recorded and a one-sided z-test will determine if there is ample evidence to favor a memory foam mattress to a spring mattress for alertness.

If you use **double asterisks** to surround a phrase or word to make it bold. Use *single asterisks* to make italics.

To start a new paragraph, make sure you leave enough line breaks between your text. To include mathematical expressions in a R Markdown document, use the same format as you would for a LaTex document and surround the equation with dollar signs like this: $\sum_{i=1}^{n}(y_{i}^{2}+\bar{y})^{2}=5$ for inline expressions and with double dollar signs for expressions centered on their own line such as

$$\sum_{i=1}^{n}(y_{i}^{2}+\bar{y})^{2}=5.$$

Q 3)
Access the data set called *sleep* in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (10 points)

1. Define a new variable called *group1.sleep* that includes only the values of the variable *extra* for individuals from group 1.

1. Perform a t-test on ``group1.sleep" to test of the extra hours slept by group 1 is smaller than or equal to 0.5 hours at an $\alpha = 0.1$ level. I.e. Test the hypothesis $H_0: \mu = 0.5$ vs. $H_1: \mu \neq 0.5$ at an $\alpha = 0.1$

significance level.

1. Report and interpret the 90% CI for the average extra hours of sleep for group 1.

1. Form a new categorical variable called *extra1.cat* that categorizes the variable *extra* into two groups, the first where extra hours slept is at least zero hours and the second where extra hours slept is less than zero hours. Print a table that counts the total number of observations in each group. (You may want to use the *ifelse()* function.)

1. Produce two boxplots for the variable *extra*, one corresponding to each group. Make sure each plot has a title. (You may find the function *grid.arrange()* in the package *gridExtra* useful for displaying two plots at once.)

Do me a favor and write your solutions to the different parts of Q 3 all in the same space (not between each bulletted list item). You can include a chunk of R code like this:

```
```{r myLabelForThisChunk, echo=TRUE, warning=FALSE}
 data(sleep)
 sleep %>% head()

# Question 1
group1.sleep <- sleep %>%
  filter (group == 1) %>%
  select(extra)

# Question 2
t.test(group1.sleep, alternative="less", mu=0.5,
conf.level=0.90)

# Question 3
# From the one sample t-test of group1.sleep, I don't have
enough evidence to reject the null hypothesis in favor of the
alternative hypothesis (p-value (0.6655) is not less than the
alpha level (0.10)). The 90 percent confidence interval shows
that the true mean is 90 percent likely to be located within
the interval (-Inf to 1.53).

# Question 4
group1.sleep <- group1.sleep %>%
  mutate (extra1.cat <-
```
```

```
ifelse(group1.sleep>=0,"Greater","Less"))
colnames(group1.sleep) <- c("extra", "extra1.cat")
table(group1.sleep$extra1.cat)

# Question 5
extra_hours_boxplot1 <- ggplot(data = group1.sleep,
aes(group1.sleep$extra1.cat, group1.sleep$extra)) +
geom_boxplot() + labs(title = "Extra Hours of Sleep", x =
"Greater or Less than Zero Hours?", y = "Number of Hours")
#extra_hours_boxplot2 <- ggplot(data = group1.sleep,
aes(x=group1.sleep$`extra hours`)) + geom_density() +
 #
geom_vline(xintercept=0, col="red")
 # I prefer the above formatting for my boxplot
extra_hours_boxplot1
```

Note that the deliminators for r code are **not** apostraphes but are the tick marks found in the upper left hand corner of your keyboard. You will never need to print out an entire data set for me in your homework, just the fist few rows using the *head()* function is fine.

Another note, the "echo=TRUE" and "warning=FALSE" options in your R code chunk are settings that will make the incorporation of your code into your document a lot neater. These options tell R to print the output of the code to your document and to not print any warning signs that may come up in the console, respectively.

To include a plot, I recomend the following options for your R code chunk:

````
```{r aDifferentCodeChunk, echo=TRUE, warning=FALSE, fig.height
= 4, fig.width = 10, fig.align = 'center'}
myPlot <- ggplot(sleep, aes(x=extra)) + geom_density() +

geom_vline(xintercept=0, col="red")
myPlot
```
````

In the code above, *aes()* is short for aesthetic which doesn't make a whole lot of sense to me, regardless, it is the function that enables you to define your x (and y) variable(s).

Q 4) Suppose $X \sim N(\mu,\sigma^2)$. Show that the random variable defined as $Z = \frac{Z-\mu}{\sigma}$ has $E[Z] = 0$ and $Var[Z]=1$. Show all of your steps (you may handwrite your answer to this question). (5 points)

Hint: Recall/look up some common properties of normal random variables and the rules of the expectation and variance operations.

For $E[Z]=0$

$$Z = \frac{X-\mu}{\sigma}$$

$$E[Z] = \frac{E[X]-E[\mu]}{E[\sigma]}$$

$X \sim N(\mu,\sigma^2)$, $E[\mu]=\mu$ & $E[\sigma]=\sigma$ because they are parameters

$$\therefore E[Z] = \frac{E[X]-\mu}{\sigma} \quad (*)$$

$E[X]=\mu$ because it is an estimate of μ

$$\therefore E[Z] = \frac{\mu-\mu}{\sigma}$$

$$E[Z] = \frac{0}{\sigma}$$

$$E[Z] = 0 \checkmark$$

For $Var[Z]=1$

$$Z = \frac{X-\mu}{\sigma}$$

$$var[Z] = \frac{var[X]-var[\mu]}{var[\sigma]}$$

$Var[X]=\sigma^2$
$var[\mu]=0$ w/ $X \sim N(\mu,\sigma^2)$
$var[\sigma]=\sigma^2$

$$var[Z] = \frac{\sigma^2-0}{\sigma^2}$$

$$var[Z] = 1 \checkmark$$