# STA130H1S – Winter 2020

Week 5 Practice Problems - Sample Answers

L. Bolton & N. Moon

## Instructions

### How do I hand in these problems for the February 6th deadline?

Your complete .Rmd file that you create for *Questions 1 & 2* of these practice problems AND the resulting pdf (i.e., the one you 'Knit to PDF' from your .Rmd file) must be uploaded into a Quercus assignment (link: https://q.utoronto.ca/courses/138992/assignments/284429/) by 11:59PM, February 6th.

### What should I bring to tutorial on February 7th?

R output (e.g., output and explanations) for *Questions 1 & 2*. You can either bring a hardcopy or bring your laptop with the output.

## Tutorial Grading

Tutorial grades will be assigned according to the following marking scheme.

| | Mark |
|---|---|
| Completion of required problems (due on Quercus the day before your tutorial) | 1 |
| Attendance for the entire tutorial | 1 |
| In-class exercises | 4 |
| Total | 6 |

## Practice Problems

**[Question 1] A criminal court considers two opposing claims about a defendant: they are either innocent or guilty. In the Canadian legal system, the role of the prosecutor is to present convincing evidence that the defendant is not innocent. Lawyers for the defendant attempt to argue that the evidence is *not convincing* enough to rule out that the defendant could be innocent. If there is not enough evidence to convict the defendant and they are set free, the judge generally does not deliver a verdict of "innocent", but rather of "not guilty".**

**(a) If we look at the criminal trial example in the hypothesis test framework, which would be the null hypothesis and which the alternative?**

The null hypothesis would be that the defendant is innocent, and the alternative hypothesis would be that the defendant is guilty. Just like in the hypothesis testing framework, we are looking for enough convincing evidence against the null hypothesis (defendant is innocent).

**(b) What does a type 1 error mean in this context?**

If a judge/jury makes a type 1 error, it means that the defendant is really innocent ($H_0$ true) but wrongly convicted.

**(c) What does a type 2 error mean in this context?**

If a judge/jury makes a type 2 error, it means that the defendant is set free even though they were in fact guilty ($H_A$ true).

**[Question 2] (Adapted from "Biostatistics for the Biological and Health Sciences") Many students have had the unpleasant experience of panicking on a test because they found the first question very difficult. A study was conducted to study the relationship between the ordering of test questions and student anxiety. The following scores are measures of "test anxiety" (i.e. panic or blanking out), based on data from "Item Arrangement, Cognitive Entry Characteristics, Sex, and Test Anxiety as Predictors of Achievement in Examination Performance", by Klimko (2015). Note that higher scores indicate higher anxiety.**

```
test_ordering <- c(rep("easy_to_hard", 25), rep("hard_to_easy", 16));
anxiety_score <- c(24.64, 39.29, 16.32, 32.83, 28.02,
                   33.31, 20.60, 21.13, 26.69, 28.90,
                   26.43, 24.23, 7.10,  32.86, 21.06,
                   28.89, 28.71, 31.73, 30.02, 21.96,
                   25.49, 38.81, 27.85, 30.29, 30.72,
                   33.62, 34.02, 26.63, 30.26,
                   35.91, 26.68, 29.49, 35.32,
                   27.24, 32.34, 29.34, 33.53,
                   27.62, 42.91, 30.20, 32.54)
anxiety_data <- data.frame(test_ordering, anxiety_score)
glimpse(anxiety_data)
```

```
## Observations: 41
## Variables: 2
## $ test_ordering <fct> easy_to_hard, easy_to_hard, easy_to_hard, easy_to_hard,...
## $ anxiety_score <dbl> 24.64, 39.29, 16.32, 32.83, 28.02, 33.31, 20.60, 21.13,...
```

**(a) Construct boxplots of `anxiety_score` for each type of test. Write 2-3 sentences comparing the distributions of anxiety scores for the two types of test.**

```
anxiety_data %>% ggplot(aes(x=test_ordering, y=anxiety_score)) + geom_boxplot()
```


Students who write tests where the questions are ordered from difficult to easy tend to be more anxious than students whose tests start with the easiest questions. We can also observe that the level of anxiety is more variable among students who write tests beginning with easy questions, as both the range (maximum - minimum) and the interquartile range (IQR) of anxiety scores are larger than the corresponding quantities for students who wrote the tests beginning with the hardest questions.

**(b) Do these data support the claim that the median anxiety level is different for tests with questions ordered from easiest to hardest and tests with questions ordered from hardest to easiest?**

(i) State the hypotheses you are testing (be sure to define any parameters you refer to).

$H_0$ : The median anxiety scores for individuals who wrote the "easy to hard" tests and the "hard to easy" tests are equal to each other

vs

$H_A$ : The median anxiety scores for individuals who wrote the "easy to hard" tests and the "hard to easy" tests not equal to each other

$H_0 : M_{easy.to.hard} = M_{hard.to.easy}$ vs $H_A : M_{easy.to.hard} \neq M_{hard.to.easy}$

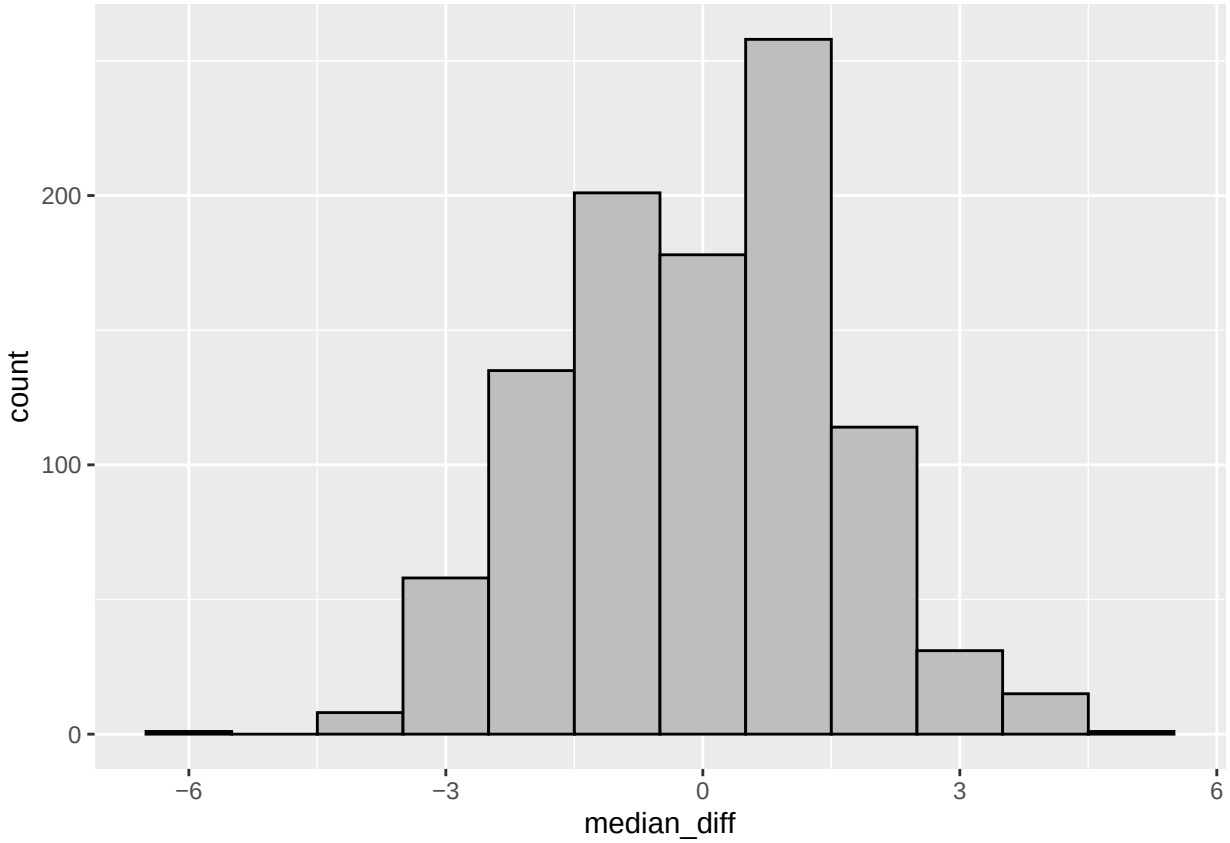
where $M_{hard.to.easy}$ is the median anxiety of score of students who wrote the test with questions ordered from hard to easy and $M_{easy.to.hard}$ is the median anxiety of score of students who wrote the test with questions ordered from easy to hard

(ii) Look at the code below and write a few sentences explaining what the code inside the for loop is doing and why.

```
test_stat <- anxiety_data %>% group_by(test_ordering) %>%
  summarise(medians = median(anxiety_score)) %>%
  summarise(value = diff(medians))
test_stat <- as.numeric(test_stat)
test_stat
```

```
## [1] 3.28
```

```
set.seed(523)
repetitions <- 1000;
simulated_values <- rep(NA, repetitions)

for(i in 1:repetitions){
  simdata <- anxiety_data %>% mutate(test_ordering = sample(test_ordering))

  sim_value <- simdata %>% group_by(test_ordering) %>%
    summarise(medians = median(anxiety_score)) %>%
    summarise(value = diff(medians))

  simulated_values[i] <- as.numeric(sim_value)
}

sim <- tibble(median_diff = simulated_values)

sim %>% ggplot(aes(x=median_diff)) + geom_histogram(binwidth=1, color="black", fill="gray")
```




```
# Calculate p-value
num_more_extreme <- sim %>% filter(abs(median_diff) >= abs(test_stat)) %>% summarise(n())

p_value <- as.numeric(num_more_extreme / repetitions)
p_value
```

```
## [1] 0.037
```

The code inside the for loop simulates a single value of the difference in median anxiety scores which we could observe if there was no association between the type of test and anxiety scores. This is achieved by creating a new dataset `simdata` by randomly shuffling the test labels (i.e. "Easy to hard" and "Hard to easy") from the `anxiety_score` dataset, calculating the median anxiety score for each of these new "groups", and calculating the difference between these two medians. The resulting value is saved in the ith element of the `simulated_values` vector. This is repeated 1000 times, until we have simulated 1000 values of the difference in medians, assuming that the null hypothesis is true.

(iii) Write a few sentences summarizing your conclusions. Be sure to interpret the p-value carefully and to clearly address the research question.

Assuming that there is no difference in median anxiety score between students who write a test starting with the hardest questions and students who write a test starting with the easiest questions, the chance of seeing a difference in median anxiety scores of 3.28 or more is 0.037. This leads us to conclude that we have moderate evidence that the median anxiety score is different for students who wrote the two types of test, and the data suggest that students who begin with harder questions tend to be more anxious than students who begin with easier questions.

**(c) You will now conduct a hypothesis test to compare the mean anxiety scores of students who tests from easiest to hardest and those who wrote tests with questions ordered from hardest to easiest.**

(i) State the hypotheses you are testing (be sure to define any parameters you refer to).

$H_0$ : The mean anxiety scores for individuals who wrote the "easy to hard" tests and the "hard to easy" tests are equal to each other

vs

$H_A$ : The mean anxiety scores for individuals who wrote the "easy to hard" tests and the "hard to easy" tests not equal to each other

$H_0 : \mu_{easy.to.hard} = \mu_{hard.to.easy}$ vs $\mu_A : \mu_{easy.to.hard} \neq \mu_{hard.to.easy}$

where $\mu_{hard.to.easy}$ is the mean anxiety of score of students who wrote the test with questions ordered from hard to easy and $\mu_{easy.to.hard}$ is the mean anxiety of score of students who wrote the test with questions ordered from easy to hard

(ii) Calculate the test statistic.

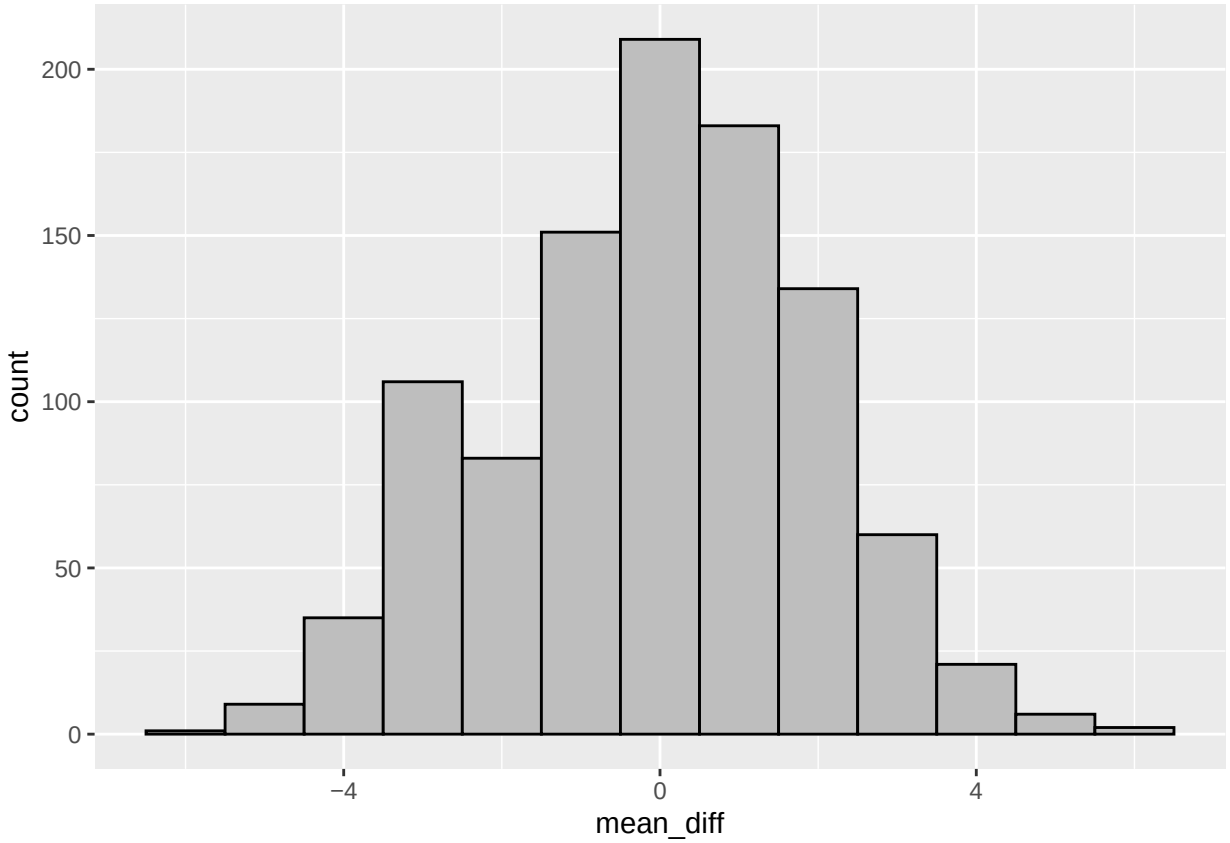
```
test_stat <- anxiety_data %>% group_by(test_ordering) %>%
  summarise(means = mean(anxiety_score)) %>%
  summarise(value = diff(means))
test_stat <- as.numeric(test_stat)
test_stat
```

```
## [1] 4.612925
```

(iii) Simulate data under the null hypothesis (use 1000 repetitions) and calculate the p-value.

```
set.seed(11) # Replace the number in the parentheses with the date of your birthday
#(suppose that my birthday is July 11th).
```

```
repetitions <- 1000;
simulated_values <- rep(NA, repetitions)

for(i in 1:repetitions){
  simdata <- anxiety_data %>% mutate(test_ordering = sample(test_ordering))

  sim_value <- simdata %>% group_by(test_ordering) %>%
    summarise(means = mean(anxiety_score)) %>%
    summarise(value = diff(means))

  simulated_values[i] <- as.numeric(sim_value)
}

sim <- tibble(mean_diff = simulated_values)

sim %>% ggplot(aes(x=mean_diff)) + geom_histogram(binwidth=1, color="black", fill="gray")
```




```
# Calculate p-value
num_more_extreme <- sim %>% filter(abs(mean_diff) >= abs(test_stat)) %>% summarise(n())

p_value <- as.numeric(num_more_extreme / repetitions)
p_value
```

```
## [1] 0.014
```

(iv) Write a few sentences summarising your conclusions. Do you have stronger evidence against the null hypothesis of no difference in median anxiety scores for the two groups, or more evidence against the null hypothesis of no difference in mean anxiety scores?

Assuming that there is no difference in mean anxiety score between students who write a test starting with the hardest questions and students who write a test starting with the easiest questions, the chance of seeing a difference in mean anxiety scores of 4.612925 or more is 0.014. This leads us to conclude that we have moderate evidence that the mean anxiety score is different for students who wrote the two types of test. We have stronger evidence against the hypothesis that the means are equal for the two types of tests than we do against the null hypothesis that the medians are equal.

**[Question 3] (Adapted from "Biostatistcs for the Biological and Health Sciences") The table below presents data from a random sample of passengers sitting in the front seat of cars involved in car crashes. Researchers are interested in whether the fatality rates (i.e. death rates) differ for passengers in cars with airbags and passengers in cars without airbags.**

| | Airbag available | No airbag available |
|---|---|---|
| Passenger Fatalities | 41 | 52 |
| Total number of Passengers | 11,541 | 9,853 |

**(a) Create a tidy data frame for this problem, using the R command `rep`. This function creates a vector which replicates its first argument the number of times indicated by its second argument. For example, the following command creates a vector with 5 elements, each of which is "hello". (Try it.)**

```
rep("hello", 5)
```

```
## [1] "hello" "hello" "hello" "hello" "hello"
```

Using the `rep` function, create a data frame for these data. View your data frame to verify that it is what you expect. Hint: Your data frame should have two variables, `group` (with levels: `airbag` and `no_airbag`) and `outcome` (with levels: `dead` and `alive`). Your data frame should also have 11,541+9,853=21,394 observations.

```
library(tidyverse)

data <- tibble(group=c(rep("airbag",11541),rep("no_airbag",9853)),
                        outcome=c(rep("dead",41), rep("alive",11541-41),
                                  rep("dead",52), rep("alive",9853-52)))
```

**(b) State appropriate hypotheses to compare the proportions of deaths in cars with and without airbags. Be sure to define any parameters you refer to in your hypotheses.**

$H_0 : p_{airbag} = p_{no.airbag}$ vs $H_A : p_{airbag} \neq p_{no.airbag}$, where $p_{airbag}$ is the proportion of deaths of passengers in cars with airbags and $p_{no.airbag}$ is the proportion of deaths of passengers in cars with no airbags

**(c) Carry out a hypothesis test for the hypotheses stated in part (b).**

```
set.seed(523) # Replace the number in the parentheses with the 1st, 3rd, and 5th
# digits in your student number.
```

```
repetitions <- 1000
simulated_stats <- rep(NA, repetitions)

n_airbag <- data %>% filter(group=="airbag") %>% summarise(n())
n_no_airbag <- data %>% filter(group=="no_airbag") %>% summarise(n())

# calculate the test statistic
airbag_deaths <- data %>%
  filter(outcome=="dead" & group=="airbag") %>%
  summarize(n())
noairbag_deaths <- data %>%
  filter(outcome=="dead" & group=="no_airbag") %>%
  summarize(n())
```

```
test_stat <- as.numeric(airbag_deaths/n_airbag - noairbag_deaths/n_no_airbag)
test_stat
```

```
## [1] -0.001725029
```

```
for (i in 1:repetitions){
  simdata <- data %>% mutate(outcome = sample(outcome))

  airbag_deaths <- simdata %>%
    filter(outcome=="dead" & group=="airbag") %>%
    summarize(n())

  noairbag_deaths <- simdata %>%
    filter(outcome=="dead" & group=="no_airbag") %>%
    summarize(n())

  p_diff <- as.numeric(airbag_deaths/n_airbag - noairbag_deaths/n_no_airbag)

  simulated_stats[i] <- p_diff
}

sim <- tibble(p_diff=simulated_stats)
sim %>%
  filter(p_diff >= abs(test_stat) | p_diff <= -1*abs(test_stat)) %>%
  summarise(p_value = n() / repetitions)
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.069
```

In our sample data, we the proportion of deaths was slightly lower in cars with airbags than in car without airbags (difference of 0.0017). Assuming there is no difference in the proportion of passenger deaths in cars with and without airbags, the probabability of seeing a difference at least as extreme as this is 0.06. In other words, we have weak evidence that the proportion of deaths for passengers of cars with airbags is different from the proportion of deaths for passengers of cars with no airbags.

**(d) Based on your answer in part (c), would you reject the null hypothesis at the 0.05 confidence level?**

In part (c), we found that the p-value was larger than 0.05, so we would not reject the null hypothesis that the proportions of deaths are equal for passengers of cars with and without airbags.

**(e) Based on your answer in part (d), what kind of error did you possibly make?**

In part (d), we failed to reject the null hypothesis. We do not know if this is the truth about reality, though. If in reality the proportion of deaths is truly different for passengers of cars with and without airbags, we would have made a Type 2 error.

**[Question 4] In class we've talked about two kinds of hypothesis tests. In the first kind (week 4) we talked about how to test whether a proportion is equal to a specific value, with hypotheses of the form: $H_0 : p = p_0$ vs $H_A : p \neq p_0$. In this week's class (week 5), we talked about how to test if there is a difference between two groups (e.g. a difference in the means of two groups, the medians of two groups, or proportions of two groups). A test for the difference between the means of two groups takes the form: $H_0 : \mu_1 = \mu_2$ vs $H_A : \mu_1 \neq \mu_2$.**

For each of the following scenarios, state appropriate hypotheses $H_0$ and $H_A$. Be sure to carefully define any parameters you refer to.

**(a) A university professor wants to learn about factors which affect student learning. She records which students attend all 10 weekly tutorials and which students miss one or more tutorials, and is interested in determining if there is an association between this and scores on the final exam.**

Since the researcher is interested in comparing values of a numerical variable (exam scores) across two groups, a two-sample test comparing the mean (or median) final exam scores for the two groups of students (attend all tutorials vs don't attend all tutorials) would be appropriate.

$H_0 : \mu_{all} = \mu_{some}$ vs $H_A : \mu_{all} \neq \mu_{some}$, where $\mu_{all}$ is the mean exam score for students who attend all tutorials and $\mu_{some}$ is the mean exam score for students who attend only some of the tutorials.

**(b) A health survey asked individuals to report the number of times they exercised each week. Researchers were interested in determining if the proportion of individuals who exercised at least 3 times per week differed between people who drink coffee every day and people who do not drink coffee every day.**

$H_0 : p_{coffee} = p_{nocoffee}$ vs $H_A : p_{coffee} \neq p_{nocoffee}$ where $p_{coffee}$ is the proportion of daily coffee drinkers who exercise at least 3 times per week and $p_{nocoffee}$ is the proportion of non-daily coffee drinkers who exercise at least 3 times per week.

**(c) A study was conducted to examine whether the sex of a baby is related to whether or not the baby's mother smoked while she was pregnant.**

$H_0 : p_{smoking} = p_{nonsmoking}$ vs $H_A : p_{smoking} \neq p_{nonsmoking}$, where $p_{smoking}$ is the proportion of male babies born to smoking mothers and $p_{nonsmoking}$ is the proportion of male babies born to non-smoking mothers. Under the null hypothesis that there is no relationship between a baby's sex and whether or not the mother smoked, there should be no difference between $p_{smoking}$ and $p_{nonsmoking}$. Note that there are other correct ways to formulate these hypotheses.

**(d) Based on results from a survey of graduates from the University of Toronto, we would like to compare the median salaries of graduates of statistics programs and graduates of computer science programs.**

$H_0 : M_{stat} = M_{CS}$ vs $H_A : M_{stat} \neq M_{CS}$ where $M_{stat}$ is the median salary for graduates of statistics programs and $M_{CS}$ is the median salary for graduates of computer science programs.