

# Homework #8

**See Canvas for the HW #8 assignment and due date.** Complete all of the following problems. Ideally, the theoretical problems should be answered in a Markdown cell directly underneath the question. If you don't know LaTeX/Markdown, you may submit separate handwritten solutions to the theoretical problems, but please see the [class scanning policy](#). Please do not turn in messy work. Computational problems should be completed in this notebook (using the R kernel). Computational questions may require code, plots, analysis, interpretation, etc. Working in small groups is allowed, but it is important that you make an effort to master the material and hand in your own work.

## A. Theoretical Problems

NOPE!

## B. Computational Problems

### B.1 Do remote students perform worse?

In spring 2019, STAT 5010 had two sections: an in-person section and an online/remote section. The Department of Applied Math is interested to know whether the in-person students performed better than the remote students. For this question, let  $\alpha = 0.05$ .

Each row of the dataset is a student from STAT 5010 in 2019. The variables in the dataset are:

- `modality`: 1 denotes in-person, 2 denotes remote.
- `score`: The final score in the class.

```
In [ ]: ##library(tidyverse)
library(ggplot2)

df = read.csv("InPersonRemote.csv")
df$modality = as.factor(df$modality)
summary(df)
dim(df)
```

|         | score  | modality |
|---------|--------|----------|
| Min.    | :74.64 | 1:13     |
| 1st Qu. | :79.15 | 2:19     |
| Median  | :80.69 |          |
| Mean    | :80.99 |          |
| 3rd Qu. | :82.04 |          |
| Max.    | :87.84 |          |

32 · 2

**B.1 (a) [10 points]** Explore the data to decide whether the normality and equal variance assumptions might be violated (see the "Tests for Differences in Means" section of the notes). If histograms don't help you determine whether the data come from a normally distributed population, you may want to investigate the R function `shapiro.test()`.

If you do not have evidence that the data are non-normal, you can carry on as if the data are normal. If you're not sure whether the variances are equal, assume they aren't.

```
In [ ]: ###install.packages("car")
library(car)

# Histograms
ggplot(df, aes(x = score, fill = modality)) +
  geom_histogram(position = "identity", bins = 30, alpha = 0.7) +
  facet_wrap(~modality, scales = "free") +
  labs(title = "Histograms of Scores by Modality")

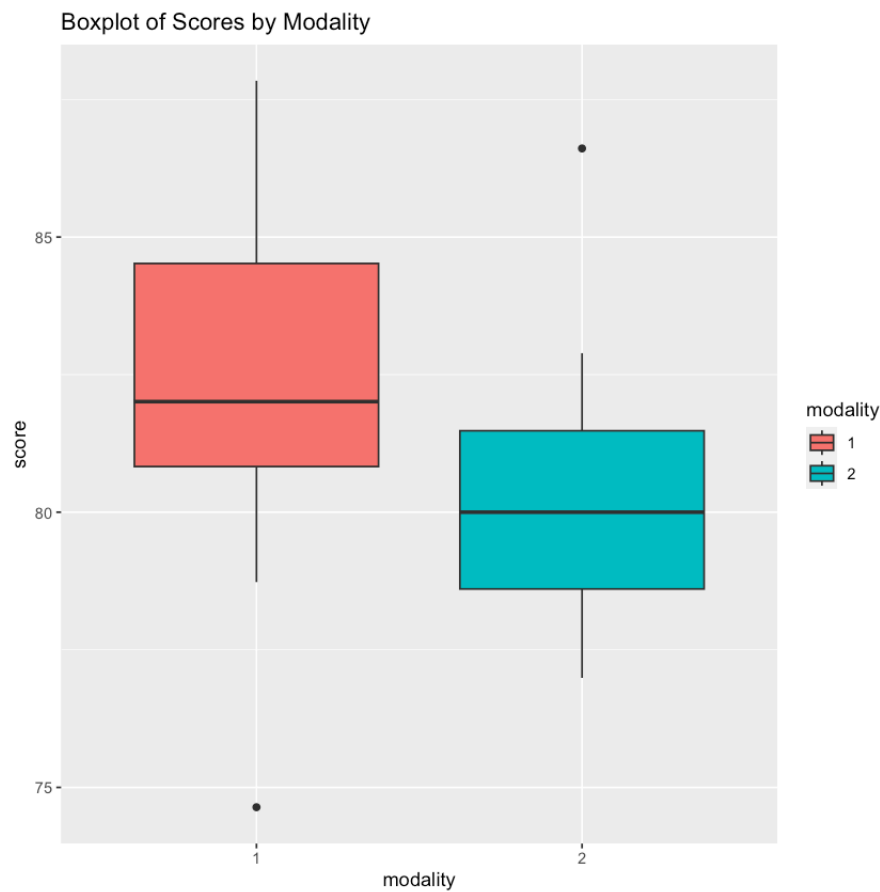
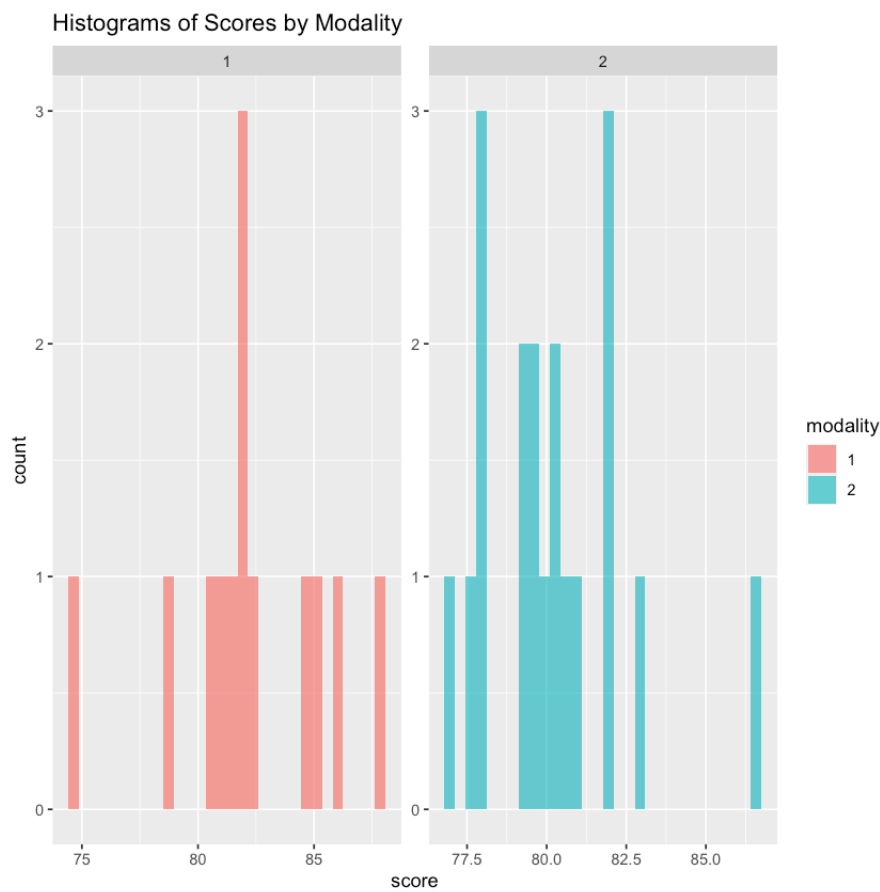
# Shapiro-Wilk test for normality
shapiro_test_in_person <- shapiro.test(df$score[df$modality == 1])
shapiro_test_remote <- shapiro.test(df$score[df$modality == 2])

# Print the results
cat("Shapiro-Wilk Test for Normality (In-person): p-value =", shapiro_test_in_person$p.value, "\n")
cat("Shapiro-Wilk Test for Normality (Remote): p-value =", shapiro_test_remote$p.value, "\n")

# Boxplot for comparing variances
ggplot(df, aes(x = modality, y = score, fill = modality)) +
  geom_boxplot() +
  labs(title = "Boxplot of Scores by Modality")
```

Shapiro-Wilk Test for Normality (In-person): p-value = 0.6599422

Shapiro-Wilk Test for Normality (Remote): p-value = 0.1042858



Both p-values are greater than 0.05, indicating that there is no strong evidence to reject the null hypothesis of normality. Therefore, based on the Shapiro-Wilk tests, you can

assume that the data from both the in-person and remote sections may come from normally distributed populations.

Looking at the box plots, since whiskers and box spreads are not similar, we can say that the variances are most likely not similar.

Let  $\mu_r$  represent the population mean of scores in the remote group. Let  $\mu_p$  represent the population mean of scores in the in-person group.

**B.1 (b) [10 points] Test whether there is statistical evidence that  $\mu_R \neq \mu_P$  at the  $\alpha = 0.05$  level. Use the appropriate test, based on the results from B.1(a). Code the test yourself (without using a built-in R function). Be sure to write out your null and alternative hypotheses and clearly state your decision.**

We'll assume that no other variables would be relevant to any difference in scores (but that's probably not true, and we learn how to deal with this kind of situation in STAT 5010!).

```
In [ ]: scores_remote <- df$score[df$modality == 2]
scores_in_person <- df$score[df$modality == 1]

# Sample statistics
mean_remote <- mean(scores_remote)
mean_in_person <- mean(scores_in_person)
sd_remote <- sd(scores_remote)
sd_in_person <- sd(scores_in_person)
n_remote <- length(scores_remote)
n_in_person <- length(scores_in_person)

# Welch's t-test statistic
t_stat <- (mean_remote - mean_in_person) / sqrt((sd_remote^2 / n_remote) + (

# Degrees of freedom
df_welch <- ((sd_remote^2 / n_remote + sd_in_person^2 / n_in_person)^2) / ((

# Two-tailed p-value
p_value <- 2 * pt(abs(t_stat), df_welch, lower.tail = FALSE)

# Print the results
cat("Welch's t-test Statistic:", t_stat, "\n")
cat("Degrees of Freedom (Welch):", df_welch, "\n")
cat("Two-tailed p-value:", p_value, "\n")

# Check if the null hypothesis is rejected
alpha <- 0.05
if (abs(t_stat) > qt(1 - alpha/2, df_welch)) {
  cat("Reject the null hypothesis.\n")
} else {
  cat("Fail to reject the null hypothesis.\n")
}
```

Welch's t-test Statistic: -1.849684  
Degrees of Freedom (Welch): 19.37781  
Two-tailed p-value: 0.07967117  
Fail to reject the null hypothesis.  
Degrees of Freedom (Welch): 19.37781  
Two-tailed p-value: 0.07967117  
Fail to reject the null hypothesis.

**B.1 (c) [10 points] Now find an R function that conducts the test that you used in B.1(b). Interpret the confidence interval given by that function.**

```
In [ ]: # Welch's t-test using t.test function
t_test_result <- t.test(scores_remote, scores_in_person, alternative = "two.

# Print the result
print(t_test_result)
```

Welch Two Sample t-test

data: scores\_remote and scores\_in\_person  
t = -1.8497, df = 19.378, p-value = 0.07967  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-4.1950236 0.2561572  
sample estimates:  
mean of x mean of y  
80.19211 82.16154

The p-value (0.07967) is greater than the significance level, suggesting that there is not enough evidence to reject the null hypothesis of equal means.

The 95% confidence interval for the difference in means includes zero, indicating that we cannot rule out the possibility of no significant difference between the two groups.

## B.2 Parametric tests without normality

Let  $X \sim \text{Pois}(\lambda)$ . We want to test the null hypothesis that  $\lambda \leq 15$ , and the alternative hypothesis is that  $\lambda > 15$ . Our sample size for this test is equal to 1. We will reject the null hypothesis if  $X$  is greater than a certain value  $c$ .

**B.2 (a) [18 points] What value should we choose for  $c$  such that the probability of a type I error is at most 0.10? In the first cell, write out the relevant formulas. In the second (code cell), construct a table of values for  $c$  along with the corresponding type I errors. In the third, state  $c$ .**

1. Poisson Probability Mass Function (PMF):

$$P(X = k | \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

This formula represents the probability of observing k events when the average rate is lambda

## 2. Type I Error Definition:

The Type I error is the probability of rejecting a true null hypothesis. In this context, it is the probability of observing a value of X greater than a certain critical value c when the true rate is less than or equal to 15. Mathematically, it can be expressed as:

$$P(X > c | \lambda \leq 15)$$

The critical value (c) is chosen such that this probability is controlled to be at most a certain significance level alpha, which is typically set to 0.10 in this case:

$$P(X > c | \lambda \leq 15) \leq 0.10$$

```
In [ ]: lambda <- 15
alpha <- 0.10

# Function to calculate type I error for a given c
calculate_type_I_error <- function(critical_value, lambda) {
  p_value <- ppois(critical_value, lambda, lower.tail = FALSE)
  type_I_error <- 1 - p_value
  return(type_I_error)
}

# Construct a table of values for c along with type I errors
c_values <- seq(0, 20, by = 1)
type_I_errors <- sapply(c_values, calculate_type_I_error, lambda = lambda)

# Display the table
result_table <- data.frame(Critical_Value = c_values, Type_I_Error = type_I_errors)
head(result_table)

c <- qpois(1 - alpha, lambda, lower.tail = FALSE)

c
```

A data.frame: 6 x 2

|   | Critical_Value | Type_I_Error |
|---|----------------|--------------|
|   | <dbl>          | <dbl>        |
| 1 | 0              | 3.059023e-07 |
| 2 | 1              | 4.894437e-06 |
| 3 | 2              | 3.930845e-05 |
| 4 | 3              | 2.113785e-04 |
| 5 | 4              | 8.566412e-04 |
| 6 | 5              | 2.792429e-03 |

10

C = 10

**B.2 (b) [10 points]** If  $\lambda$  is actually equal to 20, what is the probability of making a type 2 error using the cut-off point from B.2 (a)? In the first cell, write the appropriate formula. In the second, calculate it using R .

$$\beta = P(X \leq c \mid \lambda = 20)$$

```
In [ ]: # Given values
lambda_true <- 20

# Using the critical value (c) from B.2 (a) to calculate Type II error
type_II_error <- ppois(c, lambda_true)

# Display the result
type_II_error
```

0.0108117188266527

## B.3 Coin Flipping and Power

Let  $X$  be the number of heads in  $n = 60$  tosses of a coin. Denote the probability of heads as  $p$ . In this example, we will explore properties of a test of whether the coin is biased toward heads.

**B.3(a) [10 points]** Explicitly write out the null and alternative hypotheses. Let the parameter space be all  $p \in (0, 1)$ .

Null Hypothesis ( $H_0$ ):  $H_0 : p = 0.5$

Alternative Hypothesis ( $H_1$ ):  $H_1 : p \neq 0.5$

The parameter space for ( $p$ ) is defined as ( $0 < p < 1$ ), indicating that ( $p$ ) can take any value between 0 and 1, exclusive.

**B.3(b) [8 points] Define the rejection region to be  $X \geq 37$ . Calculate the rate of type I error. (For this part, and all remaining parts, calculate probabilities exactly, i.e., without using the normal approximation).**

```
In [ ]: n <- 60 # number of coin tosses
p_null <- 0.5 # probability of heads under the null hypothesis
rejection_threshold <- 37 # rejection region

# Calculate Type I error
type_I_error <- 1 - pbinom(rejection_threshold - 1, n, p_null)

# Display the result
type_I_error
```

0.0462304905364621

**B.3(c) [14 points] Construct a grid of values for  $p$  under the alternative hypothesis. For each value of  $p$  in this grid, calculate the power of this test (remember that statistical power is the probability of rejecting the null hypothesis while an alternative hypothesis is true). Make a plot of the power ( $y$ -axis) vs  $p$  ( $x$ -axis). Interpret/comment on this plot. In particular, as  $p$  gets further from the  $p$  specified under the null hypothesis, what happens to the power?**

```
In [ ]: # Grid of values for p under the alternative hypothesis
p_values <- seq(0.1, 0.9, by = 0.01)

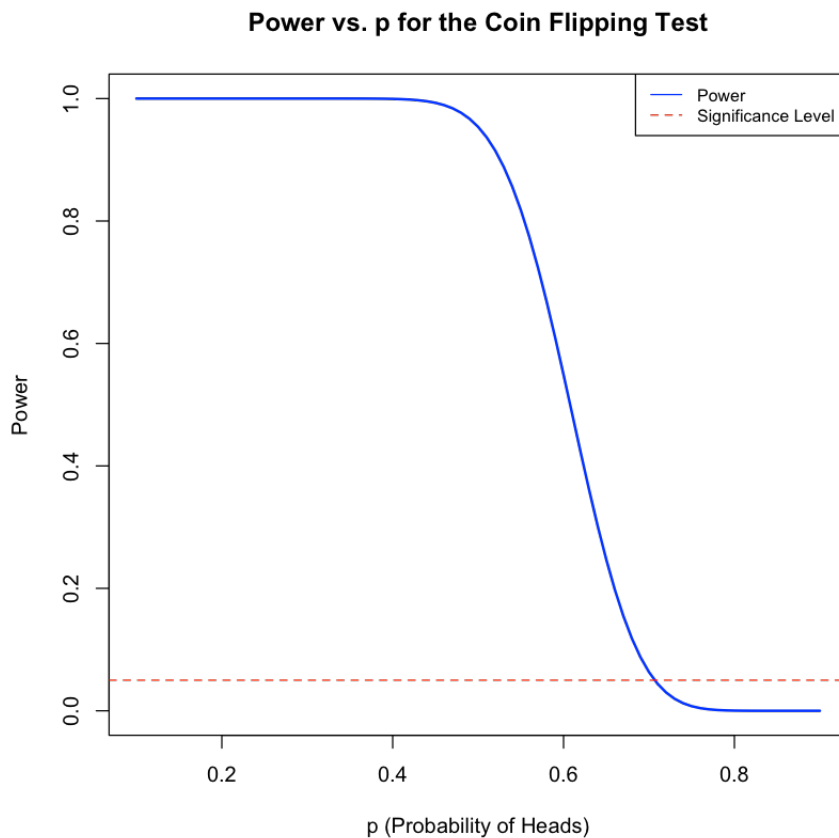
# Calculate power for each value of p
power_values <- pbinom(rejection_threshold - 1, n, p_values)

# Plot power vs. p
plot(p_values, power_values, type = "l", col = "blue", lwd = 2,
     xlab = "p (Probability of Heads)", ylab = "Power",
     main = "Power vs. p for the Coin Flipping Test",
     ylim = c(0, 1))

# Add a horizontal line for the significance level (Type I error)
abline(h = 0.05, col = "red", lty = 2)

# Add legend
legend("topright", legend = c("Power", "Significance Level"), col = c("blue"
```





The power is higher when the true probability is far from the null hypothesis value, indicating a better ability to reject the null hypothesis when it is false.

As  $p$  deviates further from the  $p_0$  specified under the null hypothesis ( $p_0 = 0.5$ ), the power of the test increases.

**B.3(d) [10 points] Suppose that you flip the coin and get 27 heads out of 60. Use this information to test your hypotheses from part B.3(a). State your conclusion.**

```
In [ ]: # Given values
n <- 60 # number of coin tosses
X <- 27 # observed number of heads

# Calculate the observed proportion
p_hat <- X / n

# Null hypothesis value
p_null <- 0.5

# Rejection region
rejection_threshold <- 37

# Test the hypothesis
if (X >= rejection_threshold) {
  conclusion <- "Reject H0"
} else {
```

```
    conclusion <- "Do not reject H0"  
  }  
  
  # Display the results  
  p_hat  
  conclusion
```

0.45

'Do not reject H0'

We do not reject H0 because since  $p_{\text{hat}}$  is less than  $37/60$