

STAT 218 Winter 2024 Midterm Project

Team Member 1: NAME HERE Team Member 2: NAME HERE
Team Member 3: NAME HERE Team Member 4: NAME HERE
Team Member 5: NAME HERE

```
library(tidyverse)
library(openintro)
library(infer)
```

Directions: In this project, you will apply various statistical procedures covered in our course. Ensure to load the necessary R packages, including ‘openintro’ and ‘tidyverse’, render the `library()` function given above.

This code chunk below indicates how to load your data. If you use an OpenIntro data set, simply use `data()` function to load your data. If your data set is coming from any website as a .csv file, use `read_csv()` function to load it. Keep in mind that I provide a data set for paired samples *t*-test. Do not forget to load that data set too.

```
# Load your midterm data set here!
```

Part 1: Getting to Know Your Data Set

Question 1: Use the `glimpse()` function to introduce your data set(s). Write a short paragraph explaining key aspects of your data set(s).

```
# Type your codes here.
```

Type your answer here.

Question 2: Determine the number of variables and cases in the data set(s).

```
# Type your codes here
```

Type your answer here.

Question 3: Identify the types of variables in the data set(s) (nominal, ordinal, discrete, and continuous).

Type your answer here.

Question 4: Choose a numeric variable and visualize its distribution. Describe the graph.

```
# Type your codes here
```

Type your answer here.

Question 5: Choose a categorical variable and create a visualization. Describe the graph.

```
# Type your codes here
```

Type your answer here.

Question 6: Use the `geom_boxplot()` function to plot a graph showing the relationship between a single numerical and single categorical variable. Describe the graph and comment on IQRs, medians, and potential outliers.

```
# Type your codes here
```

Type your answer here.

Part 2: Statistical Inference

In this section, you'll perform several inferential statistics tests using your data set(s).

One Sample t -test

In this section, you are expected to write a case study based on your data set. In other words, choose a numeric variable and conduct a One Sample t -test. Show your work.

Question 7.1: Introduce your case here.

Type your case here.

Question 7.2: What is the research question of this study?

Type your answer here.

Question 7.3: What type of variable do we have in this study?

Type your answer here.

Question 7.4: List the assumptions required for this test and check the assumption of normality for this study. Interpret the assumptions overall.

- **Random Sampling:** Explain here.
- **Independence of Observation:** Explain here.
- **Normal Distribution:** Explain here.

```
# Type your codes here to check normality.
```

Interpret your outputs here.

Question 7.5. Use a one sample t -test to investigate your research hypothesis. Answer the questions given below. Show your work.

Question 7.5.1 Write out the null and alternative hypotheses in words, in the context of this study.

Type your answer here.

Question 7.5.2. Determine the (α) level.

Type your answer here.

Question 7.5.3. Calculate the t -statistic and explain what you see in the output.

```
# Type your codes here
```

Type your answer here.

Question 7.5.4. Draw a conclusion and write a conclusion statement.

Type your answer here.

Independent Samples t test

In this section, you are expected to write a case study based on your data set. Choose a numeric variable and a categorical variable (with 2 CATEGORIES) conduct Independent Samples t -test. Use 90% Confidence Interval. Show your work.

Question 8.1:Type your case here.

Type your case here

Question 8.2. What is the research question of this study?

Type your answer here.

Question 8.3. What type of variables do we have in this study?

Type your answer here.

Question 8.4. List the assumptions required for this test and check the assumption of normality for this study. Interpret the assumptions overall.

Random Sampling: Explain here.

Independence of Observation: Explain here.

Normal Distribution: Explain here.

```
# Type your codes here to check normality
```

Interpret the assumptions and type your reasoning.

Question 8.5. Use an independent samples t -test to investigate your research hypothesis. Answer the questions given below. Show your work.____

Question 8.5.1 Write out the null and the alternative hypothesis in words, in the context of this study:

Type your answer here.

Question 8.5.2. Determine the (α) level so that you can calculate 90% Confidence Interval.

Type your answer here.

Question 8.5.3. Calculate t statistic and explain what you see in the output.

```
# Type your codes here
```

Type your answer here.

Question 8.5.4. Explain what each function/argument does in this question.

`t_test`: Type your answer here.

`formula`: Type your answer here.

`order`: Type your answer here.

`alternative`: Type your answer here.

`conf.level`: Type your answer here.

`var.equal`: Type your answer here.

Question 8.5.5. Draw a conclusion and write a conclusion statement. Include your comment on 90% Confidence Interval.

Type your answer here.

Paired Samples t test

Example of a Case: Pollutants in a stream may accumulate or attenuate as water flows down the stream. In a study to monitor the accumulation and attenuation of fecal contamination in a stream running through cattle rangeland, monthly water specimens were collected at two locations along the stream over a period of 21 months.

The data set `stream` the total coliform count (MPN/100ml) for a water specimen.

Question 9.1. What is the research question of this study?

Type your answer here.

Question 9.2. What type of variables do we have in this study?

Type your answer here.

Question 9.3. Please list the assumptions required for this test and check the assumption of normality for this study. Interpret the assumptions overall.

Random Sampling: Explain here.

Independence of Observation: Explain here.

Normal Distribution: Explain here.

Type your overall reasoning here.

Question 9.4. Perform a paired samples t -test to assess whether the mean total coliform count is consistent across the two locations. Use the 5% significance level ($\alpha = 0.05$).

Type your answer here.

Question 9.5.1 Write out the null & the alternative hypothesis in words, in the context of this study:

Type your answer here.

Question 9.5.2. Determine (α) level (Find it within the case given above).

Type your answer here.

Question 9.5.3. Calculate t statistic and explain what you see in the output.

```
# load the data set given and perform the test.
```

Type your answer here.

Question 9.5.4: What is the difference between the two `t_test` functions in independent samples t test and paired samples t test?

Type your answer here.

Question 9.5.5: Draw conclusion and write a conclusion statement.

Type your answer here.

The Chi-Square for Goodness of Fit

In this section, you are expected to write a case study based on your data set. In other words, choose a categorical variable and conduct a The Chi-Square for Goodness of Fit test. Show your work.

Question 10.1 Introduce your case study here.

Type your case here.

Question 10.2. What is the research question of this study?

Type your answer here.

Question 10.3. What type of variable do we have in this study?

Type your answer here.

Question 10.4. Please list the assumptions required for this test.

Type your answer here.

Question 10.5.1 Write out the null & the alternative hypotheses:

Type your answer here.

Question 10.5.2. Determine the (α) level.

Type your answer here.

Question 10.5.3. Calculate Chi-square statistic and find the p-value and explain what you see in the output.

```
# Type your R Code here.
```

Type your answer here.

Question 10.5.4. Compute expected counts for each cell. Interpret the assumptions overall that you listed in Question 10.4.

```
# Type your R Code here.
```

Type your answer here. Interpret the assumptions overall and type your reasoning.

Question 10.5.5. Draw conclusion and write a conclusion statement.

Type your answer here.