

IDS103: Statistical Intuitions and Applications

Final Assignment

Group Members:

1. Full Name _____ Student ID _____
2. Full Name _____ Student ID _____
3. Full Name _____ Student ID _____
4. Full Name _____ Student ID _____
5. Full Name _____ Student ID _____

Final Assignment Instructions:

1. You will work in groups; each group member must summarize their contributions at the end of their submitted work.
2. The goal of this assignment is to answer an overarching question: “what can you learn about a population from a sample dataset?”.
3. Write a succinct report by synthesizing different statistical concepts of this course.
4. You may use codes from previous assignments.
5. **Submission instructions:**
 - The primary source is a well-formatted PDF:
 - You may use this template.
 - All extra calculations, narrative, symbolic equations may be added in an appendix.
 - There is no page-limit for the appendix which may contain information such that the reader could replicate your calculations but should be as succinct as possible.
 - The secondary source is the .ipynb file along with the .csv file submitted in a single zip folder.
6. Make sure that your answers provide enough evidence in support of the following LOs in IDS-103 Statistical Intuition and Applications which will be assessed in the final assignment.

#Variables:	Identify and classify the relevant variables of a system, problem, or model.
#DescriptiveStats:	Calculate and interpret descriptive statistics appropriately.
#Correlation:	Apply and interpret measures of correlation; distinguish correlation and causation.
#Distributions:	Identify different types of distributions and make inferences based on samples from distributions appropriately.
#Visualizations:	Interpret, analyze, and create data visualizations.
#CompProgramDesign:	Generate working programs in a computer language that can solve computational problems; find and fix bugs that appear in them.
#InferentialStats:	Apply and interpret confidence intervals, statistical significance, and regression.

Objective:

The purpose of this assignment is to perform analysis of real survey data ...

1) Introduction

- Provide a description of your chosen data and outline the objectives of the report.
- Discuss the potential relationship between two selected quantitative variables.
- Present your chosen research question and explain the reasoning behind it.
- Focus primarily on providing a description in this section; calculations and technical details will go into subsections or appendix.

2) Methodology:

Describe the method you use in doing this research including the following:

- The source of your data.
- The process of your statistical analysis.
- Your approach to determining the potential correlation between the selected variables.

3) Results. Include histograms, scatter plots, regression line graphs, and numerical outcomes in the result section.**4) Discussion.** Provide a discussion or interpretation of the results, including if you accept or reject the hypothesis. Discuss if the results answer the research question. If necessary, provide alternative interpretation.**5) Conclusion.** Present the conclusions of your analyses, an evaluation of the study, and a discussion of any other issues you think relevant. Your conclusion is intended to assist the readers comprehend why your research should matter to them after they have read it. Your conclusion is not a summary of the main topics covered. It is a synthesis of key points; that is, a summary of your key findings and their significance in your study. One or two well-developed paragraphs is/are sufficient for your conclusion.**6) References.** You must list all references used, including internet resources.

Examples of Research Questions:

Each group in the class will work on only one dataset and your instructor will ensure that two groups are not working on the same dataset. Choose **one** from the following [datasets](#) in your [ipnyb file](#).

DataSet-1: Sleep Health and Lifestyle

- Do males and females differ in stress levels? (or physical activity or BMI category)
- Is there a statistically significant difference in the quality of sleep between doctors and nurses?
- What are the 95% and 99% confidence intervals for the duration of sleep for the different occupations for males in the population, assuming random sampling?
- Perform your hypothesis testing at a 5% significance level.

DataSet-2: Healthcare Insurance Expenses

- Do insurance charges differ between males and females?
- Is there a statistically significant difference in the insurance charges between smokers and nonsmokers?
- What are the 95% and 99% confidence intervals for the BMI of individuals in the different regions in the population, assuming random sampling?
- Perform your hypothesis testing at a 5% significance level.

DataSet-3: Online Education System – Review

- Do males and females engaged in online education differ on their average study time?
- Do males and females engaged in online education differ on their academic performance?
- Is there a statistically significant difference in the proportion of satisfied students engaged in online education between urban and rural areas?
- What are the 95% and 99% confidence intervals for each level of satisfaction in the population, assuming random sampling?
- Perform your hypothesis testing at a 5% significance level.

DataSet-4: Online Shop Customer Sales

- Is there a statistically significant difference in the [average] number of online purchases between males and females?
- Is there a statistically significant difference in the number of purchases between males and females in the population?
- What are the 95% and 99% confidence intervals for the total sales in the different payment methods?
- Perform your hypothesis testing at a 5% significance level.

DataSet-5: Airline Passenger Satisfaction

- Is there a statistically significant difference in the level of satisfaction with airline services between males and females?
- Is there a statistically significant difference in the proportion of satisfied passengers between Economy and Business classes?
- What are the 95% and 99% confidence intervals for the different flight services for females in the population, assuming random sampling?
- Perform your hypothesis testing at a 5% significance level.

Report Finalization:

In approximately 250 words, describe your specific role within the assignment. You need to cover the following points:

- Detail the parts of the research where you made significant contributions.
- Highlight any special skills, techniques, you applied to the project and their value.
- Explain how your work contributed to the overall goals of the assignment.
- Discuss any challenges you faced and describe how you helped overcome them.
- Describe how you collaborated with others.

Student's Full Name:

Summarize your contributions ...

Student's Full Name:

Summarize your contributions ...

Student's Full Name:

Summarize your contributions ...

Student's Full Name:

Summarize your contributions ...

Student's Full Name:

Summarize your contributions ...

Appendix:

You may provide here calculations, codes, snapshots, extra work (histograms, plots, analysis, technical details) which helped you to complete this report.