

## **Introduction to Statistics Data Analysis Paper (UPDATED)**

Students are expected to complete an empirical paper demonstrating an understanding and application of the statistical tools covered in this course. The paper should be approximately 3-5 pages, double-spaced with 11 or 12 point font. The paper is due **Friday, April 27.**

Four Excel files are posted on Canvas in the Data subfolder in the Data Analysis Paper folder. The student should choose one of the datasets to analyze using both descriptive and inferential statistical tools. The student should fully describe the pertinent data and perform statistical analysis.

The topic choice is up to the student given the supplied data, but the paper must contain the following sections:

1. Introduction
2. Data description
3. Statistical analysis
4. Conclusion

### **Introduction**

This section should set up the topic of the paper, what data is being analyzed, what statistical tools will be used and how this analysis can be used to answer potential research questions. Examples of topics include:

- Do borrowers with lower income have more debt than those with higher incomes?
- What is the relationship between GDP and birth rates?
- Do European countries have higher or lower birthrates than countries elsewhere in the world?
- Is there a relationship between team batting average and winning percentage?
- Do schools with higher tuition have high enrollments as well?

Appropriate research questions will become clear as we progress through the course. I will also suggest additional topics throughout the beginning of the semester.

### **Data Description** [20 points]

This section should begin with a thorough verbal description of the data written in your own words. You should pick 3-6 variables from the dataset for this portion of the paper to discuss. This discussion should include the description/definitions of the variables of interest.

Students should report the usual descriptive/summary statistics for all variables in a properly labeled table in this section. (Usual summary statistics include: mean, median, min, max, standard deviation, and number of observations. You are welcome to include more at your discretion.) The descriptive statistics should also be interpreted in a meaningful way. Is there anything particularly interesting about these statistics? Is the standard deviation larger or smaller than the mean? Loosely describe the distribution of at least one of the variables. Mention if there are missing observations or outliers.

In this section, students should also provide at least one graphical representation of the data. This can be a histogram, pie chart, Pareto graph, etc. There should be a brief discussion of the graph. What does it show us? Why is it interesting?

Statistical Analysis [20 points]

In this section, students must perform **ONE** statistical inferential analysis based on their chosen data.

- Confidence interval
  - To receive full credit, the student must state confidence level, critical value, interval boundaries and fully interpret the interval in the variable context.
- One sample hypothesis test
  - Student must state null and alternative hypotheses, test type (z or t), level of significance and test statistic (or p-value). Additionally, they must state the conclusion and an interpretation of that conclusion.
- Two sample hypothesis test
  - Student must define the two samples, state null and alternative hypotheses, test type (paired t-test, separate variance test, etc.), level of significance and test statistic (or p-value). Additionally, they must state the conclusion and an interpretation of that conclusion.

Conclusion

This section of the paper should restate the topic, the data used to address the research question and the main statistical results. It should also include a sentence or two about further questions that could be addressed under the topic.

Other Notes & Requirements [10 points]

- There will be two mid-semester checkpoints where students will be asked to hand in intermediate work to garner instructor feedback. Failure to submit work for either part will result in a deduction of 5 points from the final paper grade.
  - Part I: Thursday, February 15
  - Part II: Tuesday, April 17 (**OPTIONAL – No points will be deducted if you do not submit Part II of the paper.**)
- Please use APA format for any citations.
- Use full sentences, complete paragraphs and proper grammar. Students will be graded on the quality of their writing.