

Lab 4: Mini project (Gas Mileage)

1) Mini project overview

The project compares the gas mileages of US cars and Japanese cars. People may think that Japanese models have higher MPG, but we may need to test if this perception is correct or how statistically significant the difference is. This mini project uses formal statistical analysis procedures to address these questions.

Each student is expected to submit a report to compare the gas mileages of the US cars and Japanese cars based on the data given at the end of Lab 4 manual. Typically, a statistical problem is identified first and the importance is demonstrated. However, in this mini project, the problem and data sets are already provided. The students are expected to demonstrate the importance of the problem (the comparison of the gas mileage of the two types of cars) and carry the analysis on.

Section 4) below shows the sections of the report. The class did not cover materials needed for the whole project yet. Follow the instruction given in section 8) to submit the report.

2) Objectives of the lab

- To enhance the understanding of class material through systematic project work.
- To build up a basic capability to carry out full-scale statistical inference (estimation and test) projects.
- To enhance communication skills, including report writing skills, needed for most statistics projects and other industrial projects.
- To cultivate the ability to find meaningful problems in industry/workplace and clearly justify the importance of the problem.

3) Tasks to be done in the report

- Clearly describe the issue and the importance of the issue.
- Summarize the analyses you are to perform in the report, and explain why the analyses are needed. (This is not describing the result of the analysis; this addresses the analyses themselves, e.g., say if and why you want to perform a normality test, CI estimation or statistical test on variance, etc.)
- Perform the normality test, CI estimation and statistical test.
- Summarize the finding in a way end users can understand it easily along with the significance of the findings.

4) Report sections

Cover page

Table of contents, List of Tables and Figures

1. Introduction
 2. Importance
 3. Data Collection
 4. Data Presentation
 5. Normality Test of the Population
 6. Confidence Interval Estimation on the Gas Mileages
 7. Statistical Test on the Gas Mileages
 8. Conclusion and Recommendation
- Acknowledgement
References
Appendix

5) Section guideline

a. Cover page

- Title of the report (Should be clear, detailed and informative. Make two to three lines.)
- Name of the student and affiliation
- Date
- Copyright (specify your copyright here)

b. Table of Contents, List of Tables and Figures

This section identifies contents and organization of the document. It also shows list of tables and list of figures along with their captions. Note that appendix comes at the end of the report. Also note that acknowledgements, references and appendices do not have section numbers.

Page number one should start from section 1, not from the cover page or TOC page. It can be done following this way: (i) Go to the page to start numbering. (ii) Click Layout > Breaks > Next Page. (iii) Unclick “Link to Previous.” (iv) Double-click in the header or footer area (iv) Click Insert > Page number and continue (Or search for explanation using something like “Start page numbering later in your document in Word for Mac/Windows.”)

c. Section 1: Introduction

The aim of this section is to introduce the subject of the report to the reader. It is important to remember that the reader may be from a different branch of the discipline, and will require some orientation to the subject of your report. That has to be taken care of in this section.

The introduction contains the subject and purpose of the report or statement of the problem. This is where you briefly state why the report is being written and what the report is intended to achieve (objective of the study). Assume you are solving a real problem, e.g.

The objective of this study is to compare and test the performances of the engines of ABC Motors to those of WXY Motors for determining the proper price of the new motor scooters and advertisement strategy

This section also contains the background of the problem under study, describes how broad or how limited the treatment of the subject will be, the rationale of the study and an outline of the areas to be covered. This is where to bring out clearly your research questions and /or hypothesis, assumptions, limitations and delimitations and the general conclusion and/or the general recommendation. Pertinent terms in the report are also defined in this section.

If needed, put figures for the explanation. (However, the figure should very significantly contribute to the understanding of the problem. For example, do not put figures of any two cars saying (i) Japanese car (ii) US car. This can be waste of report space without adding much information.) For this assignment, refer to the internet for information if needed. Before writing this section, refer to the data file of gas mileage.

d. Section 2: Importance

Demonstrate the importance of the *problem* and the importance of *your research*. Note these two types of importance are different ones. Use some internet search to demonstrate the importance. After reading this part, a hypothetical funding agency should agree with the importance of the research. Write two paragraphs, each longer than six lines.

e. Section 3: Data Collection

The required data is given to you at the end of this manual. However, pretend that you needed to collect this data by yourself (not from the Internet or from an ‘expert’). Discuss the data collection method and the sampling method you (hypothetically) chose (textbook, chapter 1). Use one paragraph to describe the data collection method chosen. Compare your choice with other possible alternatives, and justify yours. Also, describe the limitations of your data collection method. And make another paragraph that explains the sampling method chosen. Refer to the data collected by saying “The collected raw data is shown in Appendix.”

f. Section 4: Data Presentation

This is a section for descriptive data (textbook, chapter 2). Provide basic statistical summary such as sample size, mean, variance, range, and median. Also summarize the data using charts such as stem-and-leaf diagram, histogram, pie chart, etc. to show some characteristics of your choices. Try to get useful information from them. Guide the readers to easily follow your discussion in the text when you explain the tables and figures.

g. Section 5: Normality Test of the Population

As you know, a lot of the statistics analysis is based on the Normality of the populations. Test the normality of the two populations, US cars and Japanese cars. Use (i) the histogram + skewness + outlier test and (ii) probability plot test. In the case of probability plot, state if the population is normal or skewed, thick tailed

and/or thin tailed. Make clear conclusions about the Normality. If you can use Goodness-Of-Fit test (text ch. 11, p. 580), it is a plus. When you draw the probability plot using Minitab, do not include the 95% CI line in your figure. After that, even if the population is not Normal, assume it is normal and continue the analysis after saying this: **“Although the population is not normal, I continue the analysis assuming it is normal for the completion of the report.”**

h. Section 6: Confidence Interval Estimation on the Gas Mileages

Estimate the following (show calculations in a proper way). Add explanation why we are interested in these values. Show formulas and all important calculation steps in this section. Add narrative explanations to help the readers to follow the thinking process. Just doing mathematical calculation is not good.

- Difference of the population means of the two populations.
- Ratio of the population variances of the two populations.

i. Section 7: Hypothesis Test on the Gas Mileages

Perform these statistic tests (show calculations in a proper way), and explain the findings.

- a) Compare the two means.
- b) Compare the two variances.

Follow these steps for each test:

- a) Write the hypotheses clearly. In the hypotheses, use mathematical formulas. (See section 7 for examples.) Add narrative explanation on why you came up with these hypotheses.
- b) Write the test statistic formula along with the formula of degree of freedom.
- c) Calculate the Test Value.
- d) Find the P-value.
- e) Make a decision at the significance level of 5%.
- f) Summarize the decision, i.e., repeat the decision in more plain English for the end readers of the report. See the textbook for examples.

j. Section 8: Conclusion and Recommendation

The main purpose of this section is to state briefly the major inferences that can be drawn from the discussion, explain any conclusions that have resulted from the analysis of the data, cite and explain any shortcomings of the study and suggest a course of action based on the findings and conclusions.

This section must be based on information presented in the discussion. Each conclusion should be presented as a separate paragraph, with paragraphs numbered in sequence if possible for easy reference. Recommendations must follow logically from the conclusions and be supported both by the conclusions and by the data in the discussion. Comment on what should be done in future experiments and what you would recommend doing next. Many research papers often include directions for future work.

k. Acknowledgement

The purpose of this section is to show your appreciation to all those who took their time in helping you ensure that your work would turn out to be as effective as it has finally become. The contents of this section include expressions of acknowledgement to the support that you have received from family and friends, entities that financed your research if any, your sources of information/data.

l. References

The main objective of the reference section is to acknowledge the use of material from printed sources in the preparation of your report. General guidelines:

- Indicate the exact source of all quotations and/or results of previous work. When you use an idea from a book, a paper or a website, you need to “cite” that idea; otherwise, this action is considered plagiarism.
- There are common citation/reference styles (e.g. APA and MLA) that are used to cite other people’s work. It is advisable to use such styles and be consistent throughout the report.
- Avoid citations from sources such as Wikipedia.
- Common knowledge does not require a reference, e.g. the speed of light, the atomic weight of an element. If a new value for a commonly accepted quantity is cited, the source should be referenced.

- The references section may contain the author's name, title of book/journal, year published, publisher's name, city, ISBN number, page number. Bibliographic entries are listed alphabetically by the name of the author or by the first major work of the title.

Example:

In the early 1990s, many epidemiologic studies suggested that air pollution, even at the lower ambient air concentrations that had been achieved with regulations and control technology, was still associated with cardiopulmonary disease and mortality (Samet et al., 2000); especially the fine combustion-source pollution most commonly found near heavy traffic areas (Pope, 2000)...

References

Samet, J. M., F. Dominici, F. C. Curriero, I. Coursac, and S. L. Zeger (2000). Fine Particulate Air Pollution and Mortality in 20 Us Cities, 1987-1994, *New England Journal Of Medicine*. 343:1742-1749.

Pope, C. A. (2000). Review: Epidemiological Basis for Particulate Air Pollution Health Standards, *Aerosol Science And Technology*. 32:4-14.

In this report, use at least one on-line reference, one book reference (except for any statistics book) and one journal article reference. Use the IEEE method of citation and referencing of work (this is one of the allowable scientific referencing style in journals). The following URL contains the IEEE referencing style requirements: <http://www.ieee.org/documents/ieeecitationref.pdf>. Reference does not have section numbering.

m. Appendix

Put the raw data here. This is a section for large tables, lengthy mathematical derivations, computer codes and formulas. (In this project, show your formulas in the main text, not in this appendix.) If a table is small, put that in the chapter referencing. When data is large, it is placed in this section. Appendices can also contain program listings, drawings, technical specifications, or other detailed explanations of some aspects of your report. Appendix does not have section numbering.

6) Comments on Formatting

Formatting is a very important part of this lab. Follow these instructions carefully.

- Put a table caption on the top of a table. Put a figure caption on the bottom of the figure.
- Make a detailed figure and table captions. A two-line caption is very welcome.
- Remove the legend if there is only one series in a figure.
- Do not keep the figure title in the figure. That should be given to the caption of the figure.
- Give a good scaling of the x- and y-axes of figure.
- Make the letters in tables large. Reduce row and column widths if a space is not used.
- Use proper citation format for references.

7) More tips

- Include at least one figure and one table in the report.
- The actual calculation can be done by MINITAB, but the main formulas and calculation steps should be shown in the report.
- Give numbers to sections and sub-sections.
- All figures, tables and references should be referenced in the text.
- Use mathematical expression, not plain English, to tell the hypothesis to be tested except when testing Good of Fitness or independence. E.g. $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$, where μ_1 and μ_2 are population means of the UWM students and Marquette students, respectively. (Wrong example) H_0 : mean of UWM students is the same as the mean of Marquette students, H_1 : not equal.
- Be clear about the difference between population and sample. We are testing and estimating the population parameters and not sample values such as mean.
(Correct example) $H_0: \mu = 3$ $H_1: \mu < 3$
(Wrong example) $H_0: \bar{X} = 3$ $H_1: \bar{X} < 3$

8) Assignment

Part 1: Complete the report except for sections 5, 6, 7, and 8.

Upload the completed project report on D2L of your LAB (not lecture) by midnight March 26.

This report is carries the same weight as the report of one other lab.

Part 2: Complete sections 5, 6, 7, and 8 and insert these parts to the part 1 report

Upload the completed project report on D2L of your LAB (not lecture) by midnight May 5.

This report is carries the same weight as the report of one other lab.

For the grading of part 2 assignment, the report sections covered in part 1 will *not* be read nor graded again.

9) Data

Use this data set for the analysis. Ask TA for an electronic copy of this data set so that you do not have to type the data again.

MPG Data		MPG Data		MPG Data		MPG Data	
U.S.A. Vehicles	Japanese Vehicles	U.S.A. Vehicles	Japanese Vehicles	U.S.A. Vehicles	Japanese Vehicles	U.S.A. Vehicles	Japanese Vehicles
18	24	14	34	11	32	16	36
15	27	14	26	13	47	11	38
18	27	14	30	12	41	12	32
16	25	12	22	13	45	13	31
17	31	13	22	15	34	18	36
15	35	13	33	13	33	10	24
14	24	18	39	13	24	11	26
14	19	22	36	14	32	9	29
14	28	19	28	14	32	17	37
15	23	18	27	22	39	19	28
15	27	23	21	28	35	18	19
14	20	26	24	13	32	14	25
15	22	25	30	14	37	13	24
14	18	20	34	13	38	12	37
22	20	21	32	14	34	28	24
18	31	13	38	15	34	25	24
21	32	14	37	12	32	19	33
21	31	15	30	13	33	16	33
10	32	14	31	13	32	17	32
23	32	18	38	18	34		