

Case Study 2: California Department of Developmental Services (DDS)

Due date: Thursday, October 31, at 8:00 PM Ann Arbor time via Gradescope

Late submission policy: We offer a 1-hour grace period without any penalty. Submissions uploaded between 9:00 PM on 10/31/24 and 8:00 PM on 11/01/24 are subject to a 10% late penalty. No submissions are accepted after 8:00 PM on 11/01/24.

Purpose:

The overall purpose of this case study is to build your skills in conducting an applied statistical analysis in a real-world context. Specifically, this assignment is designed to (1) improve your statistical writing skills; (2) assess your proficiency with conducting and interpreting two independent sample t-tests; and (3) challenge you to make data-based arguments in a setting where confounding variables impact the overall results of statistical analysis.

Description:

Statistical analyses have played a key role in legal cases for decades. When the decision outcomes of a judge or jury are informed by statistical evidence, it is imperative the data have been properly analyzed. Failure to perform a sufficient analysis can lead to misunderstandings and misguided decisions that can have far-reaching consequences for a range of stakeholders.

Most states in the USA provide services and support to individuals with developmental disabilities (e.g., intellectual disabilities, cerebral palsy, autism, etc.) and their families. The State of California allocates funds via monthly support checks to developmentally disabled residents through the California Department of Developmental Services (DDS).

One of the responsibilities of DDS is to allocate funds that support over 250,000 developmentally disabled residents (referred to by the program as “consumers”). A team of researchers examined the mean annual expenditure on consumers by ethnicity and found that the mean annual expenditures for Hispanic consumers was approximately one-third ($\frac{1}{3}$) of the mean expenditures of White non-Hispanic consumers. This striking discrepancy prompted allegations of race-based discrimination against the California DDS. These allegations resulted in a formal suit being filed against the state of California, claiming that its DDS program was guilty of race-based discrimination.

State legislators have responded by enlisting the services of the Confound Law Firm to represent their interests in this matter. As a senior analyst at the Confound Law Firm, you have been assigned to defend the state of California in this case with the primary objective of assessing whether there is a valid basis for the prosecution’s claim and developing a strong defense. **The pivotal question before you, therefore, is whether there exists sufficient evidence to support the assertion of race-based discrimination within the operations of the California DDS. In other words, is a consumer’s ethnicity associated with the amount of money the state of California spends on their DDS services and support? Or is this relationship influenced by additional factors?**

The Assignment:

This case study has two parts: Part 1 - analyzing the DDS data, and Part 2 - writing up your results in a brief report. This assignment is worth 100 points.

1. Part 1: Data Analysis Task: (Completed during lab - Worth 40 points, see page 4 for rubric)

- Analysis of the *full* data set:
 - **Exploratory data analysis:** Perform a detailed exploratory data analysis (EDA) on the data collected. Create a graphical representation that illustrates the association between ethnicity and expenditures.
 - **Inferential analysis:** Conduct a t-test to confirm the association between ethnicity and expenditures.
- Exploratory data analysis of possible *confounding* factors:
 - Explore *age.group* as a potential confounding variable for the relationship between expenditures and ethnicity by creating a graphical representation of the *three* variables (ethnicity, expenditures, and *age.group*).
 - Explore *sex* as a potential confounding variable for the relationship between expenditures and ethnicity by creating a graphical representation of the *three* variables (ethnicity, expenditures, and *sex*).
 - Use these graphs to identify which variable appears to be a confounding variable (*age.group* or *sex*)
- Analysis of the *subset* data using the *identified* confounding variable:
 - Analysis of the *first subset*:
 - **Exploratory data analysis:** Using the appropriate subset, create a graphical representation that illustrates the association between ethnicity and expenditures for the first subset of data.
 - **Inferential analysis:** Conduct a t-test to assess whether or not an association exists between ethnicity and expenditures for the first subset of data.
 - Analysis of the *second subset*:
 - **Exploratory data analysis:** Using the appropriate subset, create a graphical representation that illustrates the association between ethnicity and expenditures for the second subset of data.
 - **Inferential analysis:** Conduct a t-test to assess whether or not an association exists between ethnicity and expenditures for the second subset of data.
- **NOTE:** We recommend checking in with your GSI to confirm your results, so that you can correct any errors / improve your analyses before proceeding to creating a written report.

2. Part 2: Writing Task (Completed outside of lab - Worth 60 points, see page 5 for rubric)

Write a 700-1000 word report highlighting your findings. This report should focus on the central question of the trial: Do the observed DDS data support the claim that the California DDS is discriminating on the basis of race/ethnicity? In other words, is a consumer's ethnicity associated with the amount of money the state of California spends on their DDS services and support? Or is this relationship influenced by additional factors?

Your report must include the following:

- **Introduction:** Describe the purpose of the analysis and the importance of extensive data analysis that looks at all angles of the data. Include relevant summaries of the data that have been provided. (We encourage you to consider the prompt to Case Study 01, which outlines specifics often included in introductions of statistical reports, when writing this passage.)
- **Prosecution's anticipated analyses:** Provide the results of inferential analyses (e.g., a t-test) that you believe provides evidence *for* the claim that the DDS is engaged in racial discrimination (i.e., provides evidence *for* the claim that expenditures differ by ethnicity). State and interpret all components of the test and provide its conclusion. (Again, we recommend considering the prompt for Case Study 01 regarding how to interpret all testing components.)
- **Defense's substantive analyses:** Provide an analysis (including relevant exploratory analyses and inferential analyses) that you believe provides evidence *against* the claim that the DDS is engaged in racial discrimination. A successful analysis overview will explicitly provide...
 - A statement that clearly illustrates a full understanding of the role confounding factors play in studies.
 - A clear explanation with convincing arguments of why a third variable (either age.group or sex) was identified as a confounding variable. Be sure to comment on the patterns displayed in the appropriate graphical display.
 - A clear explanation of *how* the confounding variable will be factored into the analysis of the data.
 - After incorporating the confounding variable into the analysis of the data, provide an overview of the results. State and interpret all components of relevant tests and provide conclusions to each.
- **Conclusion:** a passage summarizing the key arguments you anticipate from the prosecution and the key evidence you will cite as defending legal counsel to refute their claims.

Criteria for Success:

Rubric for Data Analysis Task (40 points)

		Full data (10 points)	Confounding Variable (10 points)	Subset 1 (10 points)	Subset 2 (10 points)
Statistical Analysis	Proficient Correctly addresses all elements of the prompt, demonstrating expert understanding and interpretations of required concepts / skills.	10	10	10	10
	Emerging Correctly addresses most elements of the prompt, demonstrating solid understanding and interpretations of required concepts / skills. May include minor computational/conceptual mistakes.	6	6	6	6
	Needs Improvements Demonstrates limited understanding of statistical concepts / skills. Large errors or conceptual mistakes are present.	2	2	2	2
	Missing Completely missing any attempted analysis.	0	0	0	0

Rubric for Writing Task (60 points)

		Introduction (10 points)	Prosecution's Anticipated Analysis (20 points)	Defense's Substantive Analysis (25 points)	Conclusion (5 points)
Presentation	Proficient Section flows in a logical order, containing minimal grammatical or spelling errors. Writing is clear, creative, and informative.	5	5	5	5
	Emerging Section includes passages that are vague, unclear, or include substantial grammatical or spelling errors.	3	3	3	3
	Needs Improvements Section displays little logical order, suffering from many grammatical or spelling errors.	1	1	1	1
Statistical Reasoning	Exemplary Correctly addresses all elements of the prompt, demonstrating expert understanding and interpretations of required concepts / skills.	5	15	20	NA
	Proficient Correctly addresses all elements of the prompt, demonstrating solid understanding and interpretations of required concepts / skills. May include one or two minor conceptual mistakes.	4	12	17	NA
	Emerging Correctly addresses most elements of the prompt, demonstrating solid understanding and interpretations of required concepts / skills. May include more than two minor conceptual mistakes OR demonstrates a major misunderstanding in a core concept.	3	10	13	NA
	Needs Improvements Demonstrates very limited understanding of statistical concepts / skills. Includes more than one major misunderstanding in core concepts.	1	5	7	NA