

# Lab 6: What Would Javi Do?

Javier E. Flores

April 1, 2019

## Introduction

At this point in the semester, we've learned of several statistical methods and have been able to apply them to real data. In past labs, I have provided guidance with respect to the appropriate approach for a given research question or dataset. For today's lab, such guidance will not be provided. As aspiring statisticians, it is imperative that you hone your understanding of the tools at your disposal and when they should be used. The table below summarizes all the methods we've learned throughout the semester thus far.

Data Scenario	Method
One-Sample Categorical Data	Confidence interval, one-sample z-test, randomization test for a single proportion
One-Sample Quantitative Data	Confidence interval, one-sample t-test, randomization test for a single mean
Two-Sample Categorical Data	Confidence interval, two-sample z-test, randomization test for a difference in proportions
Two-Sample Quantitative Data	Confidence interval, two-sample t-test, randomization test for a difference in means
Paired Quantitative Data	Confidence interval, paired t-test, randomization test for a single mean
Quantitative Data (Explanatory and Response)	Correlation coefficient, regression, randomization test for correlation coefficient or slope coefficient

In what follows, you will be presented with several research questions and the data necessary to address them. For each question, decide (as a group) upon the statistical approach you all feel should be used to answer the question at hand. Be sure to consider the assumptions of a particular approach in deciding which should be used. You may use StatKey or Minitab to implement the chosen approach. If your approach can be done by hand you may still use Statkey or Minitab, but you must include in your report the work that would be done if the approach was done manually. If you are unsure of what work I would expect to see for a given procedure, please be sure to ask me.

## MRI and Alzheimer's

The [Open Access Series of Imaging Studies \(OASIS\)](#) is a project which aims to make neuroimaging data freely available to the scientific community and interested public. In doing so, the sponsors of this project hope to facilitate future discoveries in basic and clinical neuroscience. For the following questions, we will analyze **longitudinal** (data occurring across time) MRI data supplied by the OASIS project.

This dataset (available [here](#)) consists of subjects aged 60 to 95. Each subject was scanned on two or more visits, separated by at least one year, but we will only consider the first and last visits of each subject. The following are a list of the contained variables and their brief description:

**Subject ID:** Subject identification

**MRI ID:** MRI Exam identification

**Group:** Dementia status

**Visit:** Visit number

**MR Delay:** MR delay time (contrast)

**M/F:** Biological sex

**Hand:** Handedness

**Age:** Age

**EDUC:** Years of education

**SES:** Socioeconomic status

**MMSE:** Mini mental state examination

**CDR:** Clinical dementia rating

**eTIV:** Estimated total intracranial volume

**nWBV:** Normalized whole brain volume

**ASF:** Atlas scaling factor

- Q1)** The Mini-Mental State Examination (MMSE) is a 30-point questionnaire used to measure cognitive impairment. This questionnaire is often used as a screener for dementia. At baseline, how different is the MMSE of subjects classified as demented from those classified as non-demented?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q2)** Typically, a MMSE score greater than or equal to 24 is indicative of normal cognition. At baseline, do subjects classified as demented fall below this threshold?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q3)** Aside from the classifications of "Demented" or "Nondemented", the data contain a third category, "Converted", which are individuals who initially did not have dementia but later developed it. If we (unrealistically) assume that each of these patients had a perfect MMSE at baseline (i.e. 30), we would expect to see greater than a 6 point decrease in MMSE by their last visit (in order to be consistent with their classification of "converted"). For the "Converted" patients, is the MMSE at baseline greater than the last visit MMSE by more than 6 points?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q4)** Is the progression of dementia, as measured by the change in MMSE from baseline to the last visit, different between those initially classified as having dementia compared to those who "converted"?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q5)** Normalized whole brain volume is widely considered to be a reliable method for quantifying neurodegeneration. Is the MMSE predictive of normalized whole brain volume?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

## Police Killings

These data were behind the FiveThirtyEight article, "[Where Police Have Killed Americans in 2015](#)". Data from the Guardian's "[The Counted](#)" database on police killings were linked to census data from the American Community Survey. The resulting dataset (which will be used to answer the following questions) may be accessed [here](#). You may read variable descriptions under the content section of [this page](#).

- Q6)** Recently, a jury in Pennsylvania acquitted a police officer who fatally shot an unarmed teenager in the back. How prevalent are police killings involving unarmed victims among all police killings?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

- Q7)** To describe the acquittal of the Pennsylvanian officer as well as the events surrounding the case, NPR wrote an article titled, "[Jury Acquits White Former Police Officer In Fatal Shooting of Unarmed Black Teen](#)". Is there a higher proportion of black unarmed deaths relative to white?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q8)** More generally, are the majority (i.e. more than half) of victims black?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q9)** The FiveThirtyEight article informed by these data states that "Police killings tend to take place in neighborhoods that are poorer and blacker than the U.S. as a whole." What proportion of killings are in black-majority neighborhoods whose census tracts are in the bottom 20% in terms of household income?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	

**Q10)** How different is this proportion from poor, majority hispanic communities?

Data Scenario	
Method chosen (include rationale)	
Hypotheses (if applicable)	
Statistic	
P-value (if applicable)	
Brief Interpretation	