Class 16

Sean Westwood

Research Assignment Instructions

This assignment will have you walk through the process of conducting a research study from start to finish. You will find a dataset, formulate research questions, draft hypotheses, and specify regression models to explore your proposed research questions.

This assignment should be approximately 1-2 single spaced pages.

You will work together, but you should each have an individual submission. The data, research questions, hypotheses and models can be common for all group members. You should

Why does this area merit consideration?

1. Choose a broad topic that interests you. It could be related to politics, education, economics, social behavior, or any area of social science. In no more than 200 words explain why this is an important part of social science and what you would want to learn with data.

Data

1. Find a publicly available dataset related to your chosen topic. Here are some sources where you can find datasets:
    - Kaggle
    - Data.gov
    - World Bank Data
    - ICPSR
    - Google Dataset Search
2. Download the data and look at what columns are present.

Research Questions

1. Review your dataset and identify key variables that align with your topic of interest.
2. Develop at least three research questions that you aim to answer. Your questions should be specific and measurable. For example:
    - How does level of education affect income?
    - What factors are associated with high levels of job satisfaction?
    - Is there a relationship between social media use and conspiratorial thinking?

Hypotheses

1. For each research question, draft a hypothesis that you will test. Your hypotheses should be clear and testable statements predicting the relationship between variables. For example:
    - Higher education levels are associated with higher income.
    - Job satisfaction is positively correlated with work-life balance.
    - Increased social media use is associated with higher levels of conspiratorial thinking.

Specifying Regression Models

Note: You do NOT have to run these regressions

1. Identify dependent variable(s) and independent variables for each hypothesis.
2. Write the regression equation for each hypothesis with Greek notation. For example:
    - Simple Linear Regression: $Y = \alpha + \beta_1 X + \mu$
    - Multiple Linear Regression: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \mu$
3. Also provide R code to run the corrsponding models.

Components

Include the following sections:

- Introduction: Briefly introduce your topic and dataset.
- Related Literature: Find at least publications that are related to your questions. Write a short (100 word) summary of what contributions to science each paper made.
- Research Questions: List your research questions.
- Hypotheses: Your hypotheses.
- Regression Models: Specify the regression models for each hypothesis.

Example **Topic**: Impact of Education on Income

Merit: Understanding the impact of education on income is crucial as it informs policy decisions aimed at reducing income inequality and promoting economic growth. It helps individuals make informed decisions about their educational pursuits and potential career paths, highlighting the long-term financial benefits of higher education. Additionally, it provides insights into how investments in education can contribute to overall societal well-being by fostering a more skilled and productive workforce.

Dataset: National Longitudinal Survey of Youth (NLSY)

Research Questions:

1. How does the number of years of education affect annual income?
2. Does the field of study influence income levels?
3. What is the impact of obtaining a higher degree (e.g., Master's, PhD) on income?

Hypotheses:

1. Higher years of education are associated with higher annual income.

2. Income levels vary significantly by field of study.
3. Obtaining a higher degree is positively correlated with increased income.

Regression Models:

Greek:

1. Income $= \alpha + \beta_1 \text{EducationYears} + \mu$
2. Income $= \alpha + \beta_1 \text{STEM} + \beta_2 \text{Arts} + \beta_3 \text{Business} + \mu$
3. Logistic Regression: income $= \alpha + \beta_1 \text{HigherDegree} + \beta_2 \text{Experience} + \mu$

R:

1. model1 <- lm(Income ~ EducationYears, data = data)
2. model2 <- lm(Income ~ STEM + Arts + Business, data = data)
3. model3 <- lm(income ~ HigherDegree + Experience, data = data)

Tips

1. You can find related research using Google Scholar: https://scholar.google.com/
2. You can enter equations with Google Docs with the following guide: https://support.google.com/docs/answer/160749?hl=en&co=GENIE.Platform%3DDesktop