Data 557 Course Project

Introduction

In this group project, you will analyze faculty salary data from a US university to investigate potential differences between male and female faculty. The dataset includes information on faculty members employed at the university in 1995. Historical monthly salary data and other variables (sex, highest degree attained, rank, etc.) are also included. Through this analysis, you will demonstrate your ability to design and execute an applied statistics project, bringing together the concepts and methods we have covered in this course.

Acknowledgements: The dataset and questions were provided by Professor Scott Emerson.

Background

Differences in salaries between men and women at US colleges and universities are well documented. Because discrimination based on sex is illegal, other explanations for differences in salaries include differences in experience, degree attained, field, administrative responsibilities, productivities, etc.

There are a number of factors that influence faculty salaries, which may be confounded with differences in sex. Certain fields may attract higher salaries because of market demand of professionals in those areas. Salaries tend to increase with experience and when a faculty member is promoted from one rank to another. It is important to note that universities generally have strict timelines for promotion from Assistant to Associate Professor. Most faculty are considered for promotion in their sixth year and if not promoted, they will usually only be allowed to stay for one additional year. There is no timeline for promotion from Associate to full Professor. Productivity including research grant funding, number of papers published, teaching performance, and administrative duties can also affect salary.

In this project, we will examine salary data at the group level and explore whether there are differences in average salary between groups of men and women within an institution.

Questions of Interest

The overarching goal is to determine whether sex bias exists and to describe the magnitude and nature of the effect. We will focus on the following specific questions of interest:
1. Does sex bias exist at the university in the most current year available (1995)?

2. Has sex bias existed in the starting salaries of faculty members (salaries in the year hired)?
3. Has sex bias existed in granting salary increases between 1990 -1995?
4. Has sex bias existed in granting promotions from Associate to full Professor?

Dataset Description

The data to be used in this analysis consists of faculty members' salaries at a single US university. Data were collected on 1597 faculty members employed at the university in 1995 (medical school faculty were excluded). Monthly salary was determined for each faculty member for each year from 1976 through 1995. Other variables available include sex, highest degree attained, year of highest degree, field, year hired, rank, and administrative duties. Note that the latter two variables may change over time, but the others are constant over time.

The file salary.txt is in free field format with tabs separating the fields and can be downloaded from the class web page. Each record in the data file represents the information on salary and the other variables for a particular faculty member in a particular year (there are 19792 records). Missing data is denoted by NA. The first line of the file are the variable names.

The variable names and description are given below:
- *case* = case number
- *id* = identification number for the faculty member
- *sex* = M (male) or F (female)
- *deg* = highest degree attained: PhD, Prof (professional degree, eg, medicine or law), or Other (Master's or Bachelor's degree)
- *yrdeg* = year highest degree attained
- *field* = Arts (Arts and Humanities), Prof (professional school, ie, Business, Law, Engineering or Public Affairs), or Other
- *startyr* = year in which the faculty member was hired (2 digits)
- *year* = year (2 digits)
- *rank* = rank of the faculty member in this year: Assist (Assistant), Assoc (Associate), or Full (Full)
- *admin* = indicator of whether the faculty member had administrative duties (eg, department chair) in this year: 1 (yes), or 0 (no)
- *salary* = monthly salary of the faculty member in this year in dollars

Project Structure

You will be work in groups of approximately 5-6 students. While you are encouraged to divvy out tasks, each group member is expected to contribute equally to the project. Additionally, each group member is responsible for what is in the project, which means

that you should discuss and critically review each other's work. You will need to write a short "cover letter" at the end of the quarter, detailing everyone's contributions to the project.

There will be time given in class to discuss the project (and discuss with other groups), but you will need to work and collaborate on the project outside of class.

Project Deliverables

Along with several Milestones occurring approximately every two weeks throughout the quarter, your group will submit **two** of the following (note the different lists and their criteria):

List A (choose 0 or 1):

1. **Interactive Web Application** (e.g., Shiny, Flask): Allows others to explore the data and your findings dynamically.
2. **Detailed Simulation:** A simulation analysis that explores a statistical issue that arose while analyzing this data to help better understand the challenge. With this, you will write at least a 2-page document that describes the simulation methods and presents and describes the results.
3. **Case Study Analysis**: Write a detailed case study that explores a specific aspect of the data analysis (e.g., the impact of administrative duties on salary) that could be used as a teaching resource.
4. **Executive Summary**: A concise, no more than 2-page summary of the findings that would be aimed at university administrators or policymakers.

List B (choose 1 or 2):

1. **Oral Presentation** ("Job Talk" Style): A polished, 10–15-minute talk that you might give to colleagues or potential employers, highlighting your analysis and results. You will turn in the slides and present this to the instructor, TAs, and one other group that chooses a presentation or poster on March 12. Note that if you choose this option, your group will need to ask questions and provide feedback on another presentation or poster. *If you choose this option, choose something other than a poster for your other deliverable.*
2. **Journal Article**: Formal report similar to what you might see in an academic journal that summarizes the background, methods, results, and interpretations in a coherent narrative. The article should be between 10-20 pages, double-spaced.
3. **Blog Series**: A more conversational format, aiming to explain the project to a broader audience while still presenting rigorous findings. The series of blog posts should break the project down into smaller parts.
4. **Interactive Data Story** (e.g., R Markdown HTML): A step-by-step walk-through of your process, woven with visuals and discussion.
5. **Interactive Workshop Materials**: Create materials for a 1–2-hour workshop where you would teach others about the statistical methods used in the project. The workshop could include lecture materials (e.g., slides), hands-on activities, real-

time data analysis, review questions, and any solutions or discussion points. You should create a schedule or lesson plan of how you plan to structure the workshop and the amount of time spent on each.

6. **Poster**: Design a research poster that summarizes the project, suitable for presentation at an academic conference. You will turn in the poster and give a 10–15-minute presentation of this to the instructor, TAs, and one other group that chooses a presentation or poster on March 12. You will not need to print the poster—just present from a laptop. Note that if you choose this option, your group will need to ask questions and provide feedback on another presentation or poster. *If you choose this option, choose something other than an oral presentation for your other deliverable.*

For one of the deliverables, you can also propose something not from the above list, and you will need to get that approved by the instructor ahead of time. The deadline for getting approval on a deliverable not included in this list is **Wednesday, March 5**.

Milestones

Milestone 0: Project Preferences

Complete a short survey that will help us create groups. On the survey, you'll indicate:
- Work preferences (days vs evenings, weekdays vs weekends)
- Deliverable preferences
- R vs Python usage

We will do our best to match people based on preferences and interests, but cannot guarantee that there will be perfect agreement.

Milestone 1: Project Proposal

Submit a 1-2-page document that includes an outline of:
- The variables available and the roles you see them playing (e.g., outcome, predictor, confounder, …) in your analyses. What are some assumptions or considerations you have in deciding these roles?
- Classify the available measurements with respect to the type of measurement: qualitative (categorical) versus quantitative (numerical), unordered versus partially ordered versus ordered, discrete versus continuous, and interval versus ratio
- Which deliverables you are currently planning to produce (you can change these later)
- Roles and responsibilities for each group member (project management plan)

Milestone 2: Initial Descriptive Analysis

You will now be given access to the data and be asked to complete a descriptive analysis and small write-up. Details on what to include will be described in the Canvas assignment.

Milestone 3: Finalize Deliverables & Statistical Methods

Part 1: Finalize the two deliverables you want to go with.

Part 2: For each specific question, write out the statistical methods you plan to use (can be bullet points, can include notation). Details on what to include will be described in the Canvas assignment.

Grading

Milestone 1: 5 pts
Milestone 2: 10 pts
Milestone 3: 25 pts
Final deliverables and cover letter: 60 pts

Details regarding grading for each Milestone will be described in the Canvas assginment.