UCI ISI BUDS 2022
Organizing Your Plan for a Data Analysis Lab

The overarching goal of this project is to explore whether there is any evidence suggestive of discrimination by sex in the employment of the faculty at a single university (University of Washington). To this end, salary data was obtained on all faculty members employed by the University during the 1995 academic year. You have been asked to provide an analysis of 1995 salaries with the primary goal of determining whether or not gender discrimination exists with respect to pay. Along with the 1995 salary the following additional variables were also collected:

ID = The anonymous identification number for the faculty member
GENDER = Gender of the faculty member (coded as M or F)
DEG = The highest degree obtained by the faculty member (PhD, Professional, Other)
FIELD = Field of research during 1995 (Arts, Professional, Other)
STARTYR = Year starting employment at the university
YEAR = Year of data collection (1995 for all)
RANK = Faculty rank as of 1995 (Assistant, Associate, Full)
ADMIN = Does faculty member hold an administrative position as of 1995? (0=No, 1=Yes)
SALARY = 1995 salary in US dollars

Realizing that the strongest generalization of analysis results comes when the statistical question is decided before looking at the data, in this exercise we will simply think about the goal of the analysis and ways in which we should statistically approach the problem. **You should answer the following questions without downloading and opening the dataset.**

1. First, consider the sampling scheme for the current project. What population will we be able to make inference about? Can you think of any way in which the sampling scheme could lead to misleading inference regarding gender discrimination at the university?

2. Recall that a confounder in the relationship between gender and salary must be causally related to salary and associated with gender. Thus we can begin looking for potential confounders by first considering those factors that may influence salary. List out those factors which you feel strongest influence salary in this setting and justify your choices (Note: You should not limit yourself to factors for which data has been collected).

3. Among those factors listed above, decide which might reasonably be associated with gender and justify your choices. These are the potential confounders you would ideally like to adjust for in your analysis.

4. List any factors that you a priori feel would be effect modifiers in the relationship between gender and salary. Justify your choice(s).

5. Using your answers to (b), (c), and (d), describe what available adjustment variables you would include in an analysis to answer the question of interest. Classify your adjustment variables potential confounders or possible effect modifiers.