SEX DISCRIMINATION PROBLEM

2. Study Design

Two following two questions about the study design are very important to answer the question about sex discrimination:

2.1 Does the study design enable us to prove sex discrimination?
2.2 Can we draw any inferences to populations based on the data?

2.1 Does the study design enable us to prove sex discrimination?

In order to support the discrimination claim we have to use the data to show that gender is the only *cause* of the observed disparity in starting salaries between males and females. Can statistical methods alone prove sex discrimination with the current study design? You will see that they cannot.

The case study is an example of an observational study because the sex of each of the 93 employees was not decided by the investigator. In other words, allocation of employees to the two gender groups (males, females) was not determined by any chance mechanism.

As the study is an observational study, we are not able to draw any causal conclusions from the statistical analysis alone. It is possible that some confounding variables are responsible for the disparity in the starting salaries. For example, graphical displays of the data in **Section 4** in the *Two-Sample Problems* module show that the males generally did have more years of education than the females, and this, not gender, may have been responsible for the observed differences in the starting salaries for males and females. Thus, the effect of gender cannot be separated from the effect of education.

In general, some confounding variables may not be recognized or measured and these variables consequently cannot be accounted for the observed disparity. Obviously, it would be difficult or even impossible to consider all variables affecting starting salary. Therefore, it may be possible to conclude that males tend to receive larger starting salaries than females, even after accounting for all available factors, and still not be possible to conclude, from the statistics alone, that this happens because they are males.

Although statistical analysis alone is not able to prove sex discrimination, it can be useful in a court of law to establish discrimination. In general, observational data are the only data likely to be available in discrimination cases.

Summarizing, both the study design and imperfect measures of individual qualification make it impossible to establish sex discrimination using statistics alone.

2.2 Can we draw any inferences to populations based on the data?

Notice that the 61 females and 32 males were not selected from any well-defined populations. Anyway, how would the populations be defined? Would they be defined as the populations of all *potential* male and female employees with the bank? In fact, there is no interest in the starting salaries of some larger population of individuals who were never hired. No random sampling was involved in the study. Moreover, the sex of each of these individuals was not randomly assigned. Hence, neither the random sampling nor the randomized experiment model can be used in this case.

Any plausible interpretation of the difference between the average starting salary of males and the average starting salary of females must be based on a fictitious chance model. One possible example is a model in which the employer assigns the starting salaries to the hired individuals at random. Then we can use statistical analysis to determine whether the observed difference between the average male and female salaries is likely assuming no sex discrimination.