**1.2 - Sex Discrimination in Employment – An Observational Study**

Did a bank discriminatorily pay higher starting salaries to men than to women? The data in Display 1.3 are the beginning salaries for all 32 male and 61 female skilled, entry-level clerical employees hired by the bank between 1969 and 1977. (Data from a file made public by the defense and described by H. V. Roberts, "Harris Trust and Savings Bank: An Analysis of Employee Compensation" (1979), Report 7946, Center for Mathematical Studies in Business and Economics, University of Chicago Graduate School of Business.)

| DISPLAY 1.3 | Starting salaries ($U.S.) for 32 male and 61 female clerical hires at a bank | | | | | | |
|---|---|---|---|---|---|---|---|

**Males**

| 4,620 | 5,700 | 6,000 |
|---|---|---|
| 5,040 | 6,000 | 6,000 |
| 5,100 | 6,000 | 6,000 |
| 5,100 | 6,000 | 6,300 |
| 5,220 | 6,000 | 6,600 |
| 5,400 | 6,000 | 6,600 |
| 5,400 | 6,000 | 6,600 |
| 5,400 | 6,000 | 6,840 |
| 5,400 | 6,000 | 6,900 |
| 5,400 | 6,000 | 6,900 |
|  | 6,000 | 8,100 |

**Females**

| 3,900 | 4,500 | 4,800 | 5,220 | 5,400 | 5,640 |
|---|---|---|---|---|---|
| 4,020 | 4,620 | 4,800 | 5,220 | 5,400 | 5,700 |
| 4,290 | 4,800 | 4,980 | 5,280 | 5,400 | 5,700 |
| 4,380 | 4,800 | 5,100 | 5,280 | 5,400 | 5,700 |
| 4,380 | 4,800 | 5,100 | 5,280 | 5,400 | 5,700 |
| 4,380 | 4,800 | 5,100 | 5,400 | 5,400 | 5,700 |
| 4,380 | 4,800 | 5,100 | 5,400 | 5,400 | 6,000 |
| 4,380 | 4,800 | 5,100 | 5,400 | 5,520 | 6,000 |
| 4,440 | 4,800 | 5,100 | 5,400 | 5,520 | 6,120 |
| 4,500 | 4,800 | 5,160 | 5,400 | 5,580 | 6,300 |
|  |  |  |  |  | 6,300 |

Answer the following questions about Case Study 1.2. Type your responses in a document and save as a pdf.  Submit this document under the Homework tab in Blackboard.  Submit your R* script *as a separate file* along with your pdf.  You may work together and use each other for help/advice/support but submit your work for this assignment.

1. If you worked with anyone, please give their first and last names. (Again, for this assignment please submit your own work!)
   Ans: I discussed with Sihyuan Han.

2. How is the *design* of this study like that in Case Study 1.1 from last time?  How does it differ? (Consider the *types* of variables measured on individuals and how values were "assigned" to individuals. See the displays on the next page and corresponding discussion from the text.)
   Ans: In Case Study 1.2, it has allocated two segments: male and female in the initial stage. Compared to Case study 1.1, it is randomly sampled and not classified at the beginning.

3. Using R find numerical and graphical summaries of this data. Use these to describe *the distribution* of the starting salaries for both males and females.
   (Please refer to the RMD file)

   a. Give and interpret the mean salary and standard deviation of salaries for females. Do this also for males.
   b. Give and interpret the median salary and the IQR of salaries for females. Do this also for males.
   c. Give a histogram of salaries for each group.
   d. Give side-by-side boxplots of salaries.
   e. Use a. to d. to describe the distribution of salaries for each group.
   f. Which group's salaries are larger, on average?
      Ans: Female group's salaries are larger since the sample size is bigger, but male's average salaries are larger, because 5956.875 (male's mean salaries) **>** 5138.852 (female's mean salaries)

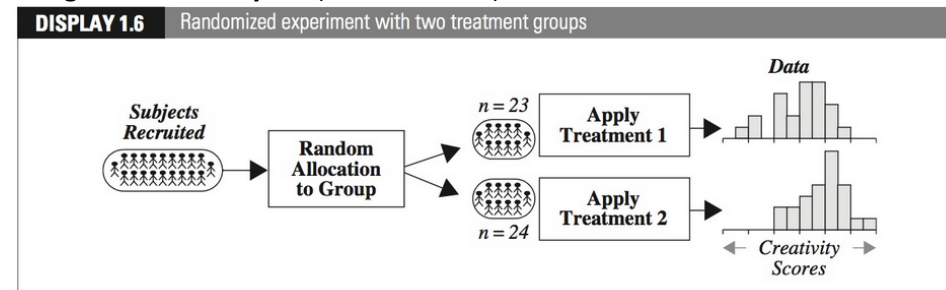*When I refer to R you can substitute your preferred software.

4. Did you get any unusual messages (warnings or errors) from R? If so, what? (You may not have!)
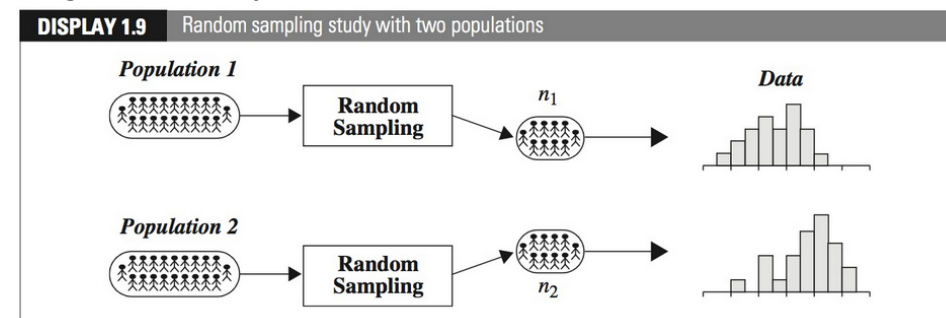   <mark>Ans:</mark> No errors on my R code
5. Your textbook gives the results of a fuller analysis (which we will get to eventually!) and state that there is "convincing evidence that the males, as a group, received larger starting salaries than the females," but that "statistics alone cannot address whether this difference is attributable to sex discrimination." Briefly explain their reasoning for drawing this conclusion.
   <mark>Ans</mark>: There are still some other factors we should consider, such as male may have had more years of previous experience.

**Design of Case Study 1.1** (from last time):



DISPLAY 1.6    Randomized experiment with two treatment groups

**Design of Case Study 1.2**:



DISPLAY 1.9    Random sampling study with two populations

**Drawing Conclusions from Statistical Studies**:

**Allocation of Units to Groups**

|  | **By Randomization** | **Not by Randomization** |
|---|---|---|
| **At Random** | A random sample is selected from one population; units are then randomly assigned to different treatment groups. | Random samples are selected from existing distinct populations. |
| **Not at Random** | A group of study units is found; units are then randomly assigned to treatment groups. | Collections of available units from distinct groups are examined. |

**Selection of Units**

*Inferences to the populations can be drawn*

*Causal inferences can be drawn*