

### Homework 4

Due Date: April 17

There are 7 questions in this homework assignment. Submit the codes for all problems in a notebook file on canvas.

The female employees of a certain company threatened to sue the company's management for sex discrimination, claiming that female employees were paid less than male employees. In order to support their claim, the plaintiffs drew a simple random sample of 25 of the company's employees and obtained their salaries (in thousands of dollars). The data is given in a file named *salary.txt*.

The company officials counter that the difference in salaries, while real, is an historical artifact and no longer reflects the company's policies. In the "old days", officials claim, company management did have strong prejudice against female employees, but the new management does not suffer from such a bias. The reason that the salaries differ by gender, they argue, is simply that male employees have been there longer than female employees have, and the salary differences reflect longevity differences, rather than gender differences.

1. Calculate the maximum, standard deviation, and mean of salaries based on the gender. Interpret the results.
2. Create boxplots for salary for each gender. Interpret the results.
3. Calculate the percentage of salaries based on the gender and demonstrate it using a pie chart.
4. Show the scatter plot of salaries against years with the company for male and females employees. Interpret the graphs.
5. Plot the histogram of salaries (5 bins). Does it follow a normal distribution?
6. Show the regression plot between salaries and years with the company. Interpret the graph.
7. If you were the judge, would you rule in favor of the plaintiffs (females) or the defendant (company)?