Math 537 Exam 1

June 5, 2020

Problem Statement

Download the dataset genderwage.csv . In this dataset you will find 62 observations. Each observation represents the average amount that a company in the manufacturing sector pays their non-management female employees and male employees respectively.

Once you have the data downloaded I would like you to perform the following tasks, with limited write-up and discussion on how you performed the task as well as what conclusions you've drawn when appropriate.

a.) Develop a statistical hypothesis test using a mean vector (no paired t-tests or simple linear regression models!) for whether or not you think that non-management males and females have the same average salary.

b.) Plot your data along with a 95% confidence region for where you believe the true average salary vector for males and females in the manufacturing sector is. Add your null hypothesis as a line to this plot. (If you're struggling with part a. I'd start with part b.)

c.) Separating the males and females, find a marginal confidence interval for the males and a marginal confidence interval for the females.

d.) Find simultaneous confidence intervals for the males and females in this dataset. (Note, you'll probably need to wait till after Tuesdays lecture, or deep dive into the textbook before you attempt this one)

e.) Mahalet and Rachel are having an epic feud right now. Mahalet insists that this data came from a bivariate normal distribution with $\vec{\mu} = (22.5, 24.5)$

and $\Sigma = \begin{bmatrix} 12 & 8 \\ 8 & 12 \end{bmatrix}$, while Rachel is extremely confident that this data came from a bivariate normal distribution with $\vec{\mu} = (21.5, 26)$ and $\Sigma = \begin{bmatrix} 9 & 8 \\ 8 & 16 \end{bmatrix}$

Based on the data, which side are you on, Rachels or Mahalets? (Hint, if you find yourself using the words 'likelihood ratio' and you've successfully engineered a likelihood ratio you're probably doing very well on this problem).

f.) Similarly to how you did on HW1 Problem 5. Load both of your variables onto the strongest eigen vector (the one with the largest eigen value). Pass μ_0 that you used for part a, through the same eigen rotation (lets call it v_0). Perform a simple t-test for whether or not the mean of your new information super-loaded variable could reasonably be equal to v_0. Compare your p-values from this new univariate test to the p-value from your bivariate test in part a.)