Math 537 Exam 1

July 17, 2024

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 marginal confidence intervals for the males and marginal confidence intervals 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 Wednesdays lecture, or deep dive into the textbook before you attempt this one)

e.) Ethan and Ranjosh are having an epic feud right now. Ethan 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 Ranjosh is equally insistent that this data came from a bivariate normal distribution with $\vec{\mu} = (21.5, 26)$ and $\Sigma = \begin{bmatrix} 9 & 8 \\ 8 & 16 \end{bmatrix}$

They've become so adamant about this issue that it's threatening to ruin their friendship. Based on the data, who do you think is right, Ethan or Ranjosh? (hint, likelihood ratios are our friends).

f.) Repeat question e.), only this time don't assume a covariance matrix for either Ethan or Ranjosh. How did your likelihood ratio change?

g.) Formally run your hypothesis test using a Hotelling's T-test (Think about your graph, how should you compute your T2 value)? Find a p-value using a chi-square distribution and again using a modified F-distribution. Did it really matter which distribution you used?