Statistics

Week 6 Recitation

ESD, SUTD

Term 5, 2017

Exercise: Inference for Two Samples

Two movie review websites (A and B) independently rate a variety of movies from 0 to 5. Martin noticed that one website seemed to provide in general higher ratings than the other one. He recorded ratings for 30 movies from both websites. Can you help him run statistical analyses to see if the observation is valid, on a significance level of 0.01?

The data are available on eDimension (*MovieRating.xlsx*).

Hint:

- Make plots to observe the data
- Should the ratings from website A and B be treated as two independent samples or matched pairs?

Exercise (Answer): Inference for Two Samples

Answer:

- Data should be treated as matched pair.
- The difference between each pair of ratings can be regarded as a new variable D
- H_0: $\mu_D = 0$
- H_1: $\mu_D \neq 0$
- p-value $= 1.27 \times 10^{-8} < 0.01$, reject the null hypothesis

Question: What would happen if you treat the two samples as independent?

- The variance of either rating A or rating B is much larger than the variance of D
- Won't be able to reject the null hypothesis that the means of rating A and B are the same
- $Var(\bar{A} - \bar{B}) = Var(\bar{A}) + Var(\bar{B})$ doesn't hold