
ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: “Big” Data Analytics
Proposal (or Final Report)

Aditya Sehgal
902905469

Problem Description:

For this project, I wish to perform a study on male and female height and shoe size. I wish to test the following hypothesis :

1. “mean male height is greater than mean female height”
2. “mean male shoe size is greater than mean female shoe size”

I will then attempt to determine if there is evidence to suggest a linear correlation between (i) male height and male shoe size, and (ii) female height and female shoe size.

I will use a 5% confidence level when performing tests.

Data Source:

For my data set, I will send out an anonymous form on a social media site such as Facebook, requesting my peers to enter their gender, height and shoe size. The samples can be thought of as random. I wish to collect more than 30 entries.

Method:

To test my hypothesis, I will perform a t-test (since population standard deviation is unknown to us) for the following:

H0: “mean male height is equal to mean female height”
H1: “mean male height is greater than mean female height”

H0: “mean male shoe size is equal to mean female shoe size”
H1: “mean male shoe size is greater than mean female shoe size”

Then finally, I will use linear regression to determine if there is enough evidence to support a linear correlation between (i) male height and male shoe size, and (ii) female height and female shoe size.