

Is it Normal?

In this lab we will be using Rstudio to investigate whether sample data follow a normal distribution. The data used in this lab are generated from a known distribution. There are 5 variables to consider. We will investigate the normality by looking at a histogram and a Q-Q plot. To begin, you will need to download the data set: IsitNormal.csv. Data sets can be found in Blackboard in the Data folder. You can rename the data set if you wish from the Import Dataset screen. In the following, the data set is called “Isitnormal”. The Rstudio Script that contains the commands is the file IsitnormalLab.R.

1. Upload the data set.
2. Create a histogram and Q-Q plot for each variable; x, y, z, w, h; .

```
> par(mfrow=c(1,2))  
> attach(Isitnormal) # ENTER as x, y, z, w, or h respectively  
> hist(x)  
# normal fit  
> qqnorm(x); qqline(x)
```

- a. Discuss whether or not the variable, x, appears to follow a normal distribution
- b. Discuss whether or not the variable, y, appears to follow a normal distribution
- c. Discuss whether or not the variable, z, appears to follow a normal distribution
- d. Discuss whether or not the variable, w, appears to follow a normal distribution
- e. Discuss whether or not the variable, h, appears to follow a normal distribution

3. ACTUAL DISTRIBUTION COMPARISON: For each variable compare your previous answer to the actual distribution. Discuss the reasons for a discrepancy if needed.

To generate the data, n data points were selected from the following distributions:

- a. ACTUAL x : t- distribution with $df=3$, $n=100$
- b. ACTUAL z : Exponential distribution with $\lambda=3$, $n=100$
- c. ACTUAL y: Standard Normal distribution, $n = 15$
- d. ACTUAL w: Standard Normal distribution, $n = 1000$
- e. ACTUAL h : Added 3 extreme outliers to w, $n = 1003$