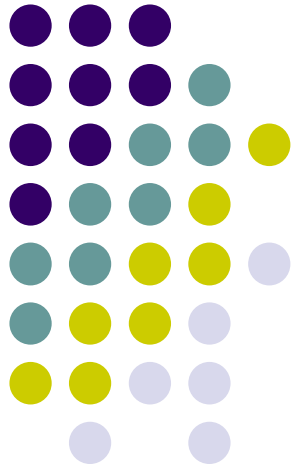
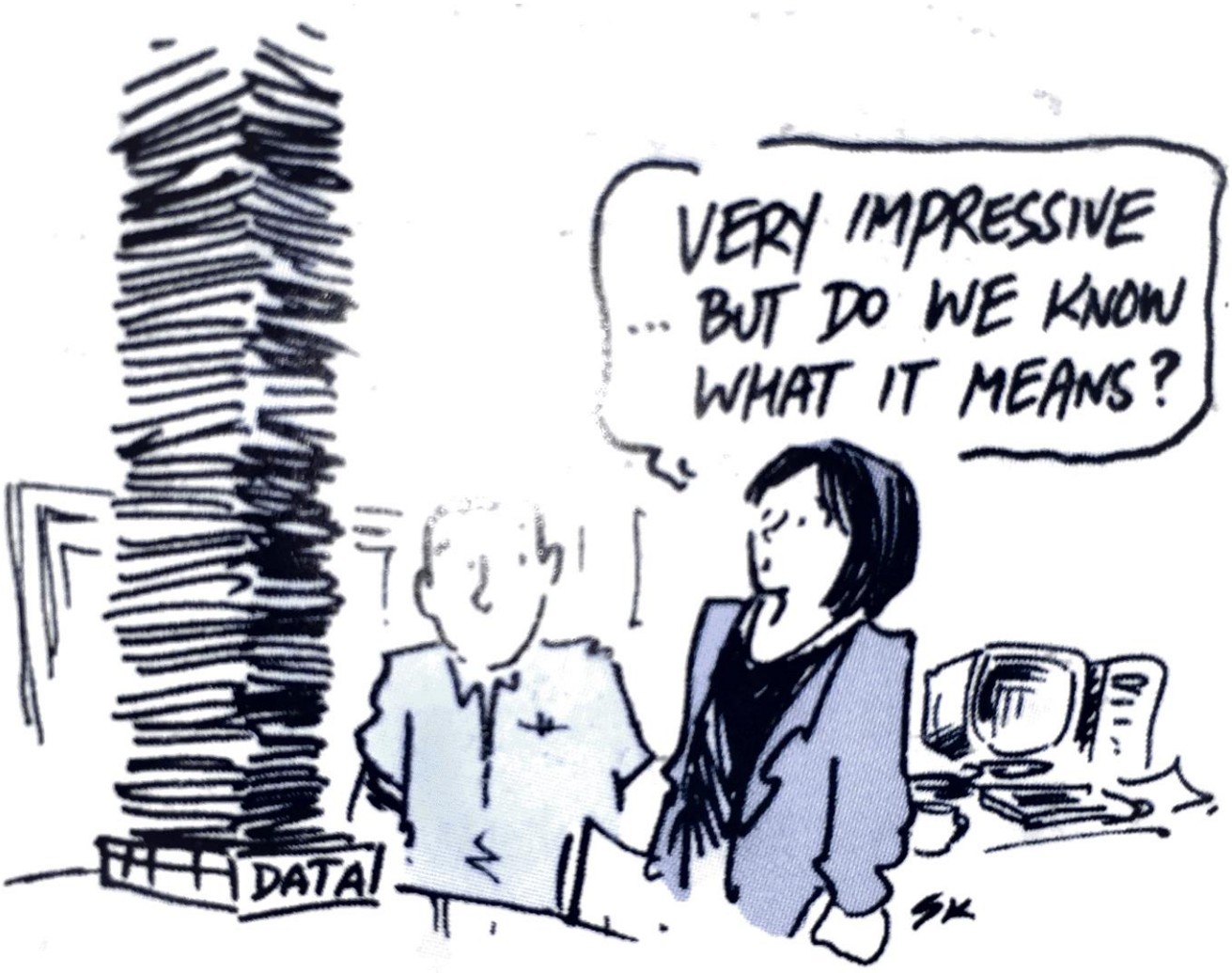
Normality Test

Dr. Yazdan Asgari



2019

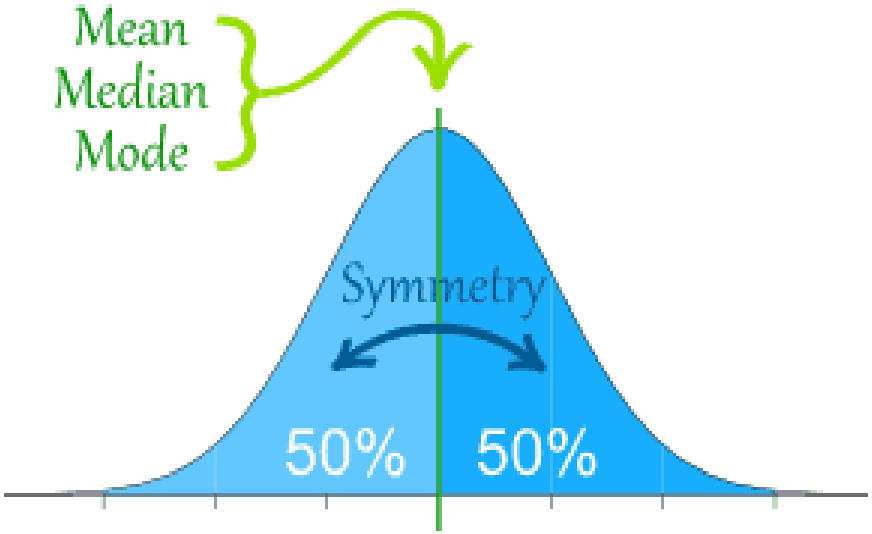
Descriptive Statistics - Goal



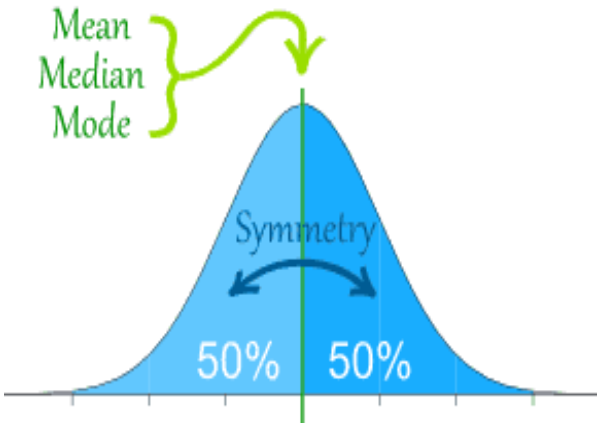
https://images.google.com/

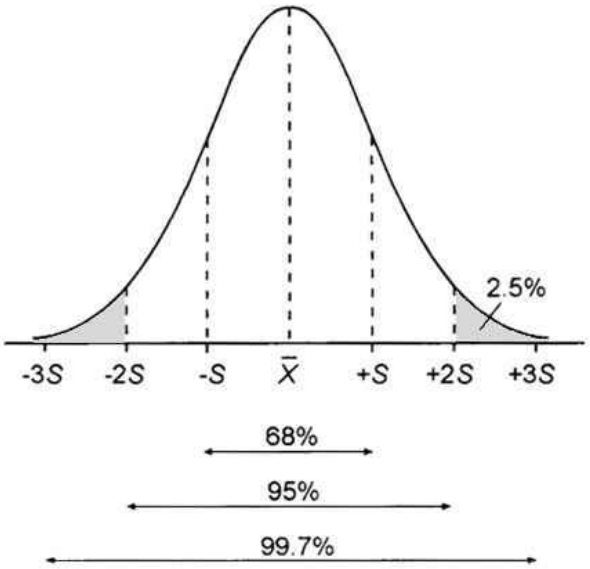
Normal Distribution

- It is a symmetric bell-shaped distribution
- where most of the observations cluster around the central peak
- and the probabilities for values further away from the mean taper off equally in both directions.



Normal Distribution

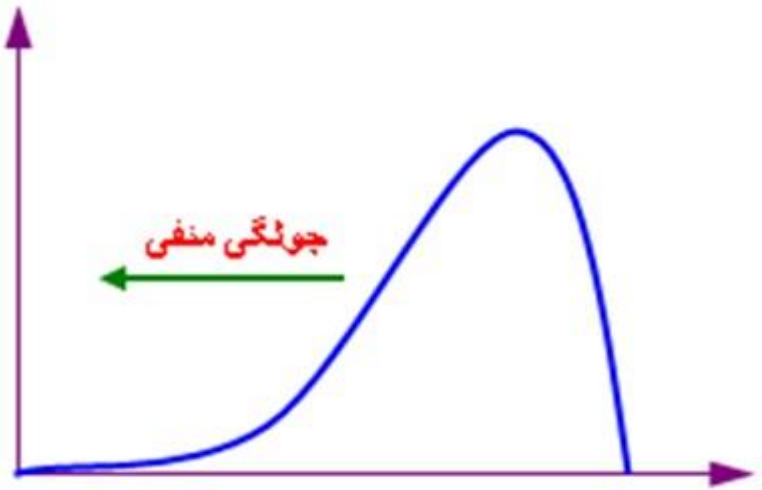


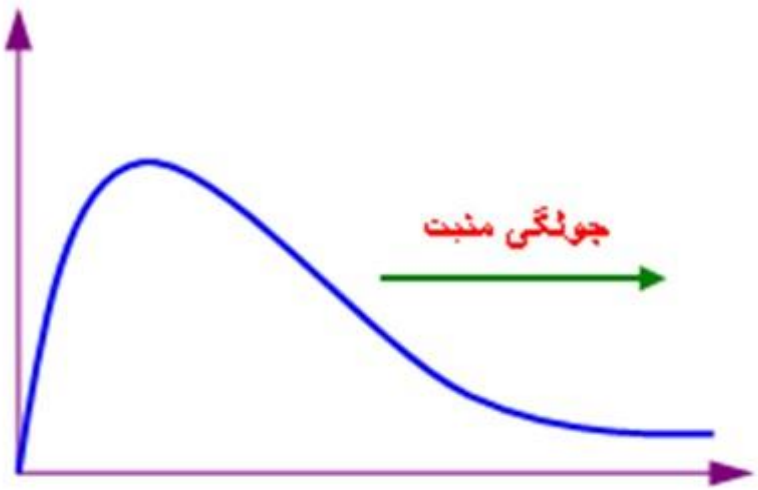


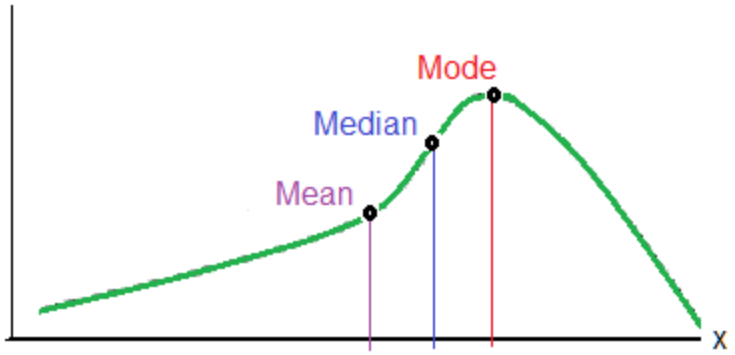
ویژگی های منحنی توزیع نرمال

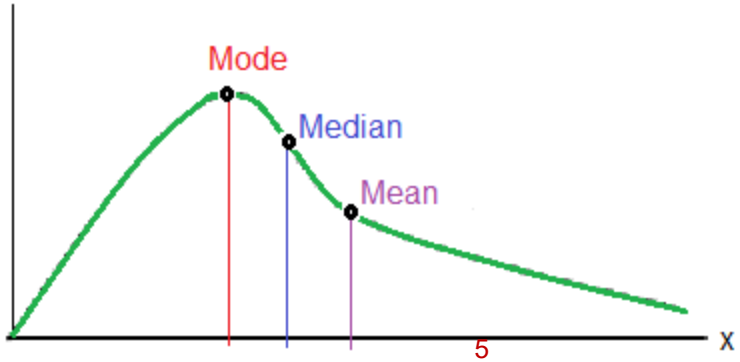
- زنگوله ای شکل است.
- حول میانگین متقارن است.
- سطح زیر منحنی نرمال برابر یک است.
- میانگین، میانه و نمای آن برابر است.
- پراکندگی حول میانگین، توسط انحراف معیار بیان می شود.
- تقریبا 68% منحنی نرمال با $\mu \pm s$ پوشش داده می شود.
- تقریبا 95% منحنی نرمال با $\mu \pm 2s$ پوشش داده می شود.
- تقریبا 99.7% منحنی نرمال با $\mu \pm 3s$ پوشش داده می شود

Skewness

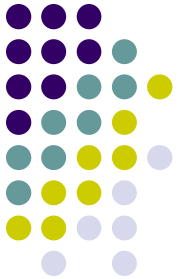






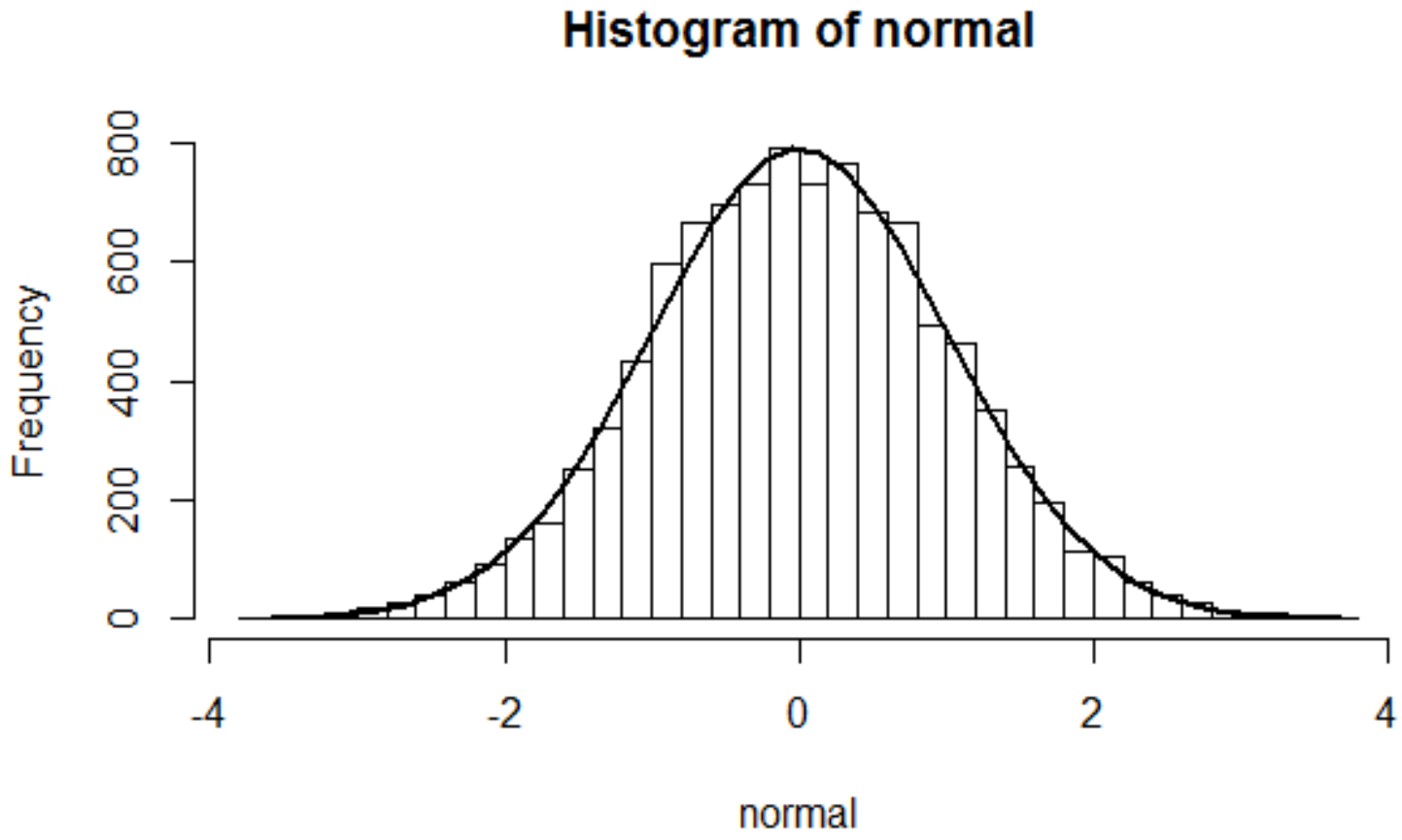


Checking normality in R

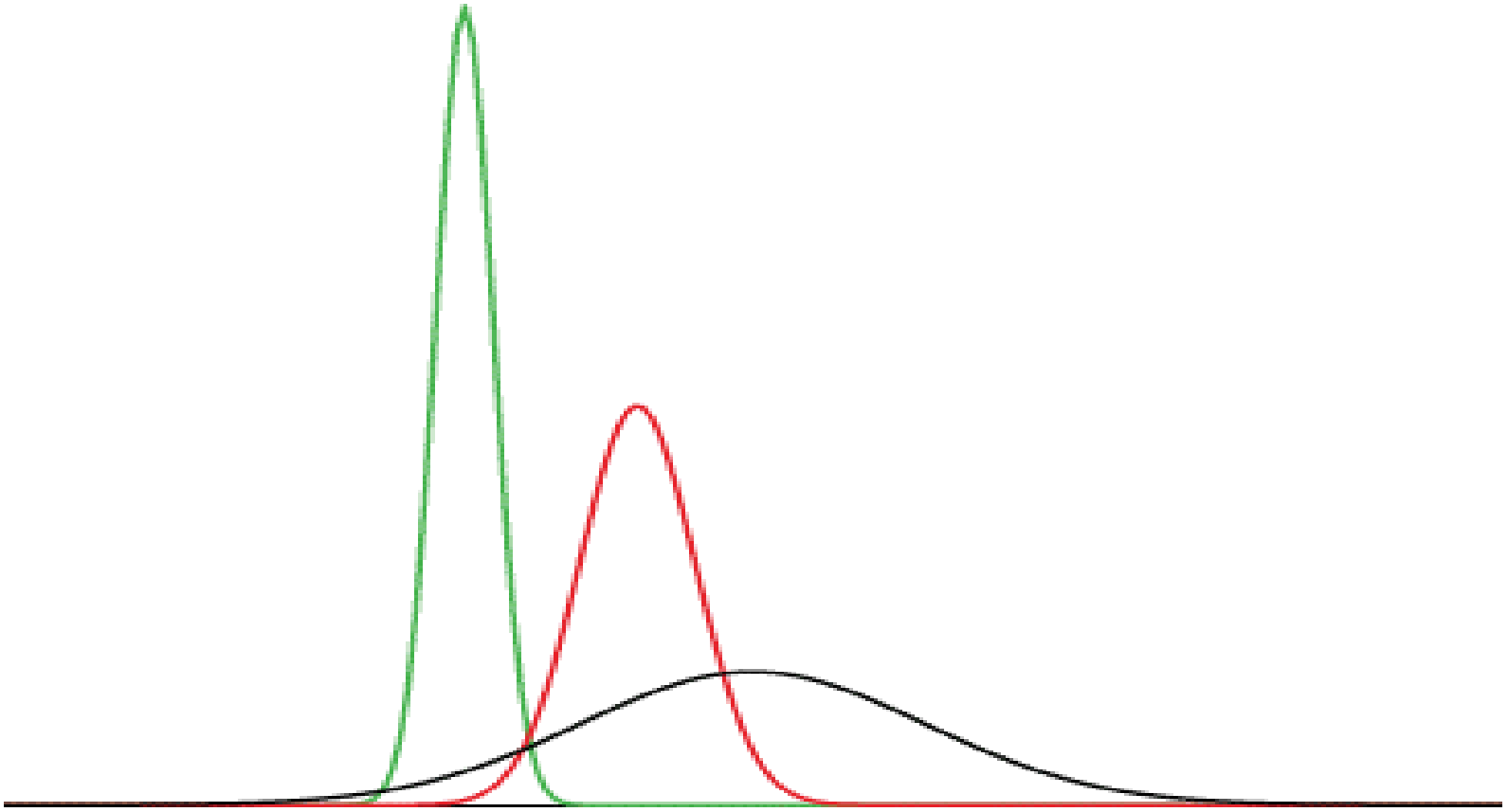


- Histogram
- Q-Q plot
- Normality tests / Goodness of fit tests
  - Kolmogorov-Smirnov test
  - Shapiro-Wilk test
  - Anderson–Darling test

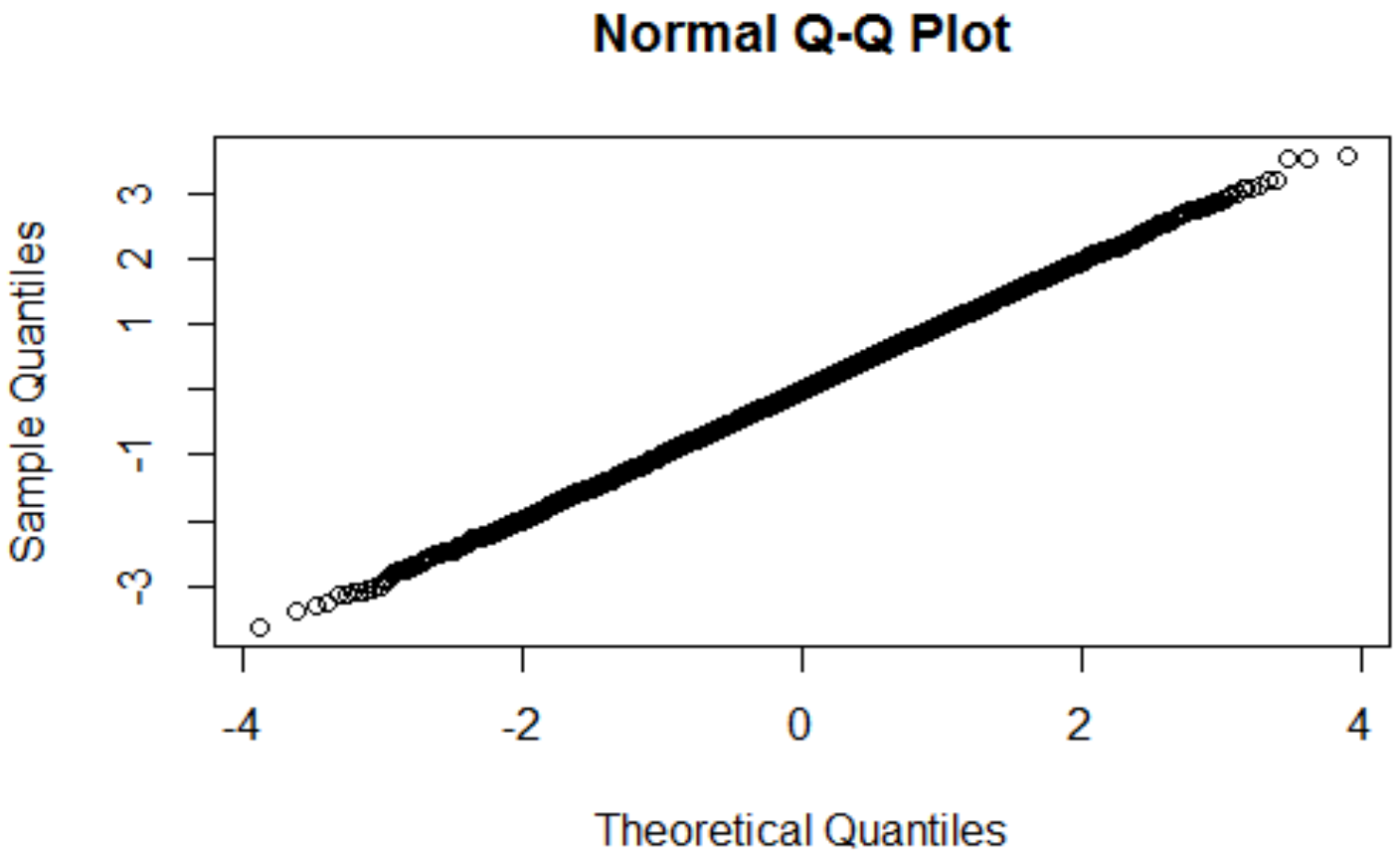
Histogram


Histogram of normal
Frequency
0
200
400
600
800
-4
-2
0
2
4
normal

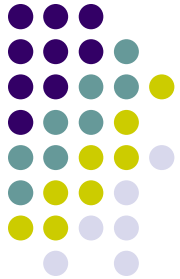
Histogram



Q-Q plot

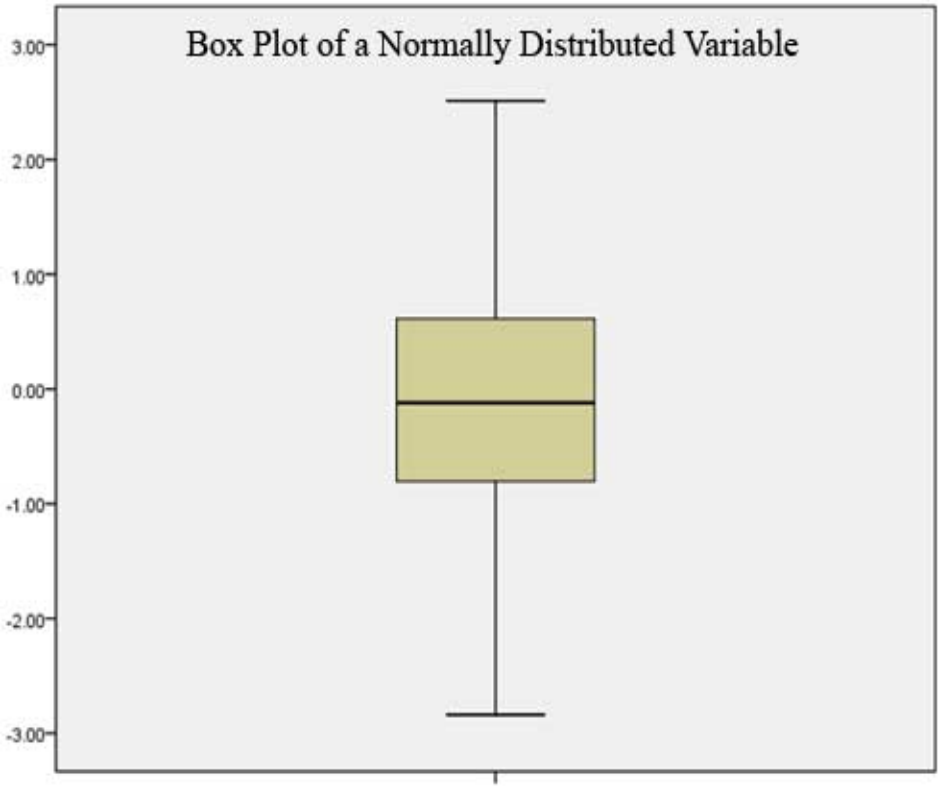

Normal Q-Q Plot
Sample Quantiles
-3
-1
1
2
3
-4
-2
0
2
4
Theoretical Quantiles

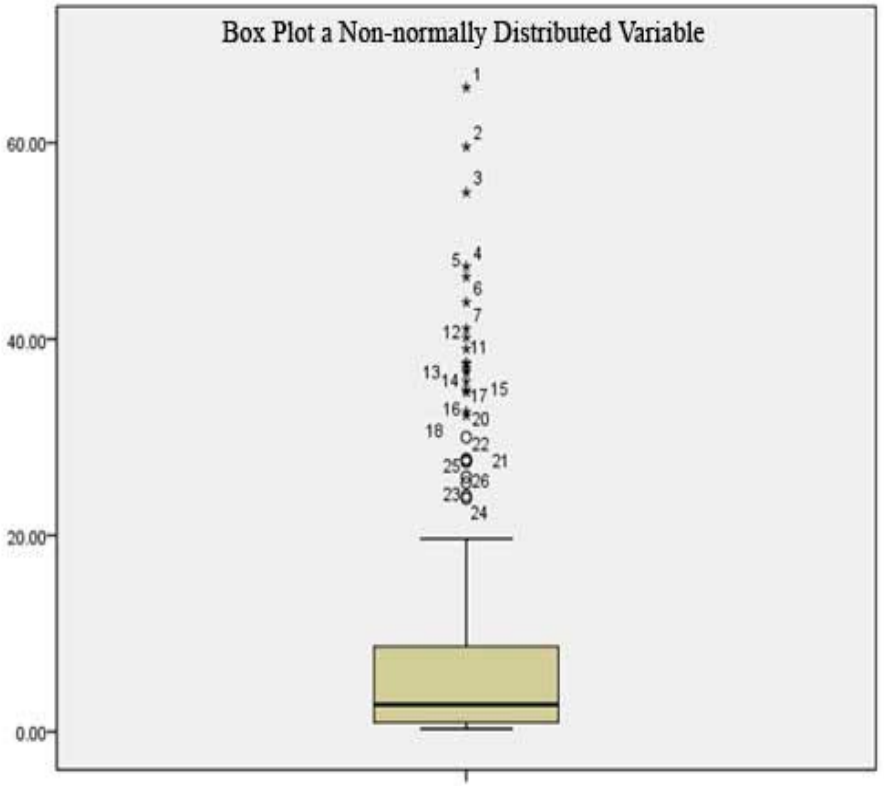
Shapiro-Wilk test



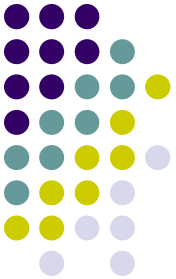
- The S.W. test examines the NULL hypothesis that "*the samples came from a Normal distribution*".
- This means that if your *p-value* <= *0.05*, then you would reject the NULL hypothesis that the samples came from a Normal distribution. Therefore, If *p-value* > *0.05*, normality can be assumed.
- In R: sample size must be between 3 and 5000

Other Methods: Box Plots

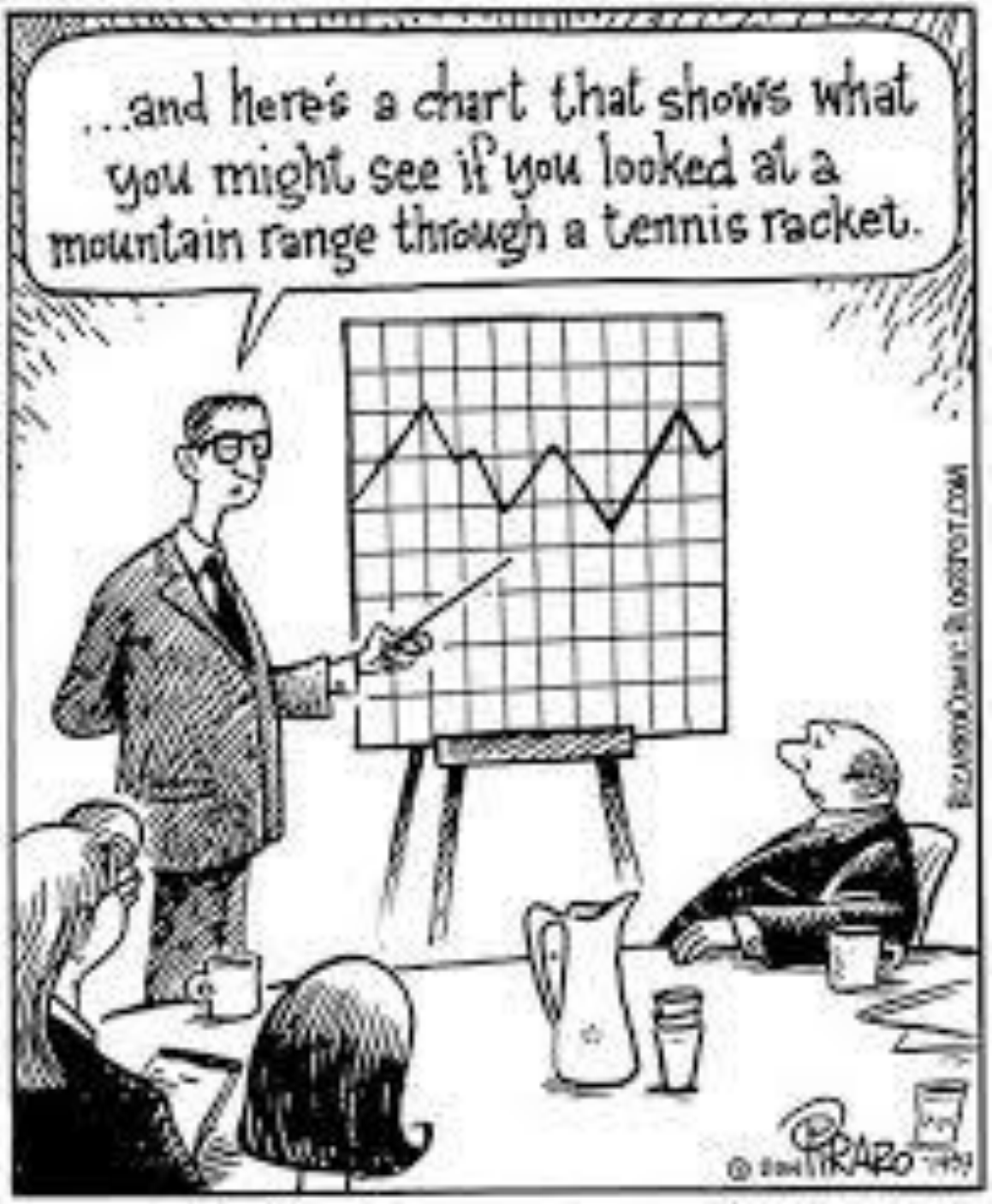




What if the Data is NOT Normally Distributed?



- **Transform the dependent variable**
- **Use a non-parametric test**



Was it useful?