Jumping into Statistics: Introduction to Study Design and Statistical Analysis for Medical Research Using JMP Pro Statistical Software

WINTER/SPRING 2021

DR. CYNDI GARVAN & DR. TERRIE VASILOPOULOS

Meet the Instructors

CYNTHIA GARVAN, MA, PHD

Research Professor in Anesthesiology

cgarvan@anest.ufl.edu

TERRIE VASILOPOULOS, PHD

Research Assistant Professor in Anesthesiology and Orthopaedics and Rehabilitation

tvasilopoulos@anest.ufl.edu

Course Objectives

- Review fundamentals of study design and research methodology
- Understand how to choose best statistical test for your research question
- Practice basic statistical analysis use JMP Pro Software

Course Topics

- Asking a Good Research Question
- Life Cycle of Research and Scientific Method
- Study Design
- Data types and Database Construction
- Descriptive Statistics
- Data Visualization
- Population and Sample, Probability, Statistical Inference
- How to Chose Correct Statistical Method and Run Some Analyses
  - T-tests, ANOVA, Non-Parametric
  - Chi-square, odds ratio, relative risk
  - Regression and Correlation
  - Survival Analysis
  - Test Diagnostics (e.g. sensitivity, specificity, etc.)
- Comparing Statistical Modeling and Machine Learning

Descriptive Statistics

Common descriptive statistics

Mean→ sum of all values/sample size (n)

- Average

Median→ middle value of all data

- Quantitative

Mode→ most frequent value of data

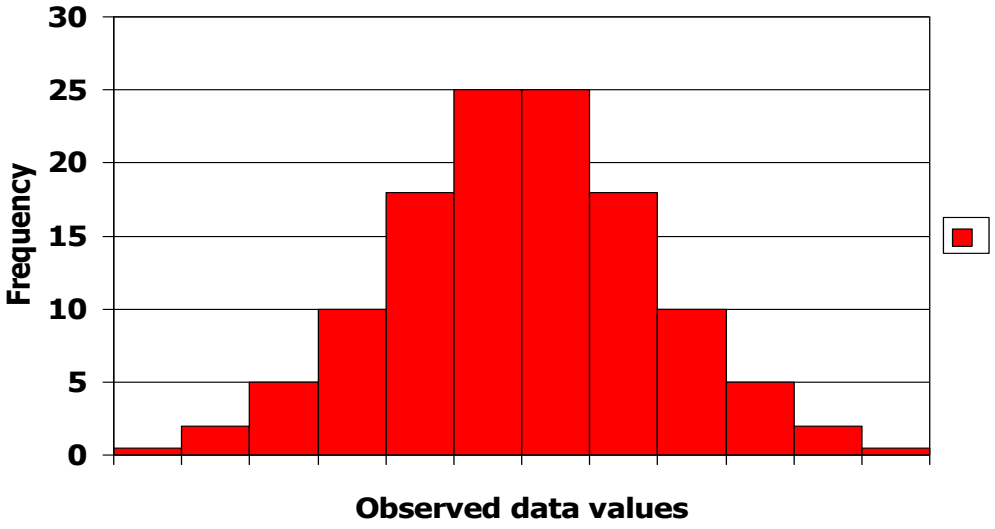
- Quantitative

Frequency→ % each value is observed in data

- Qualitative (discrete as well)

Central tendency

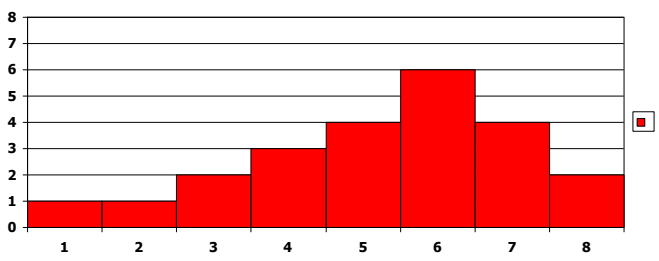
Examining Distributions of Quantitative Data



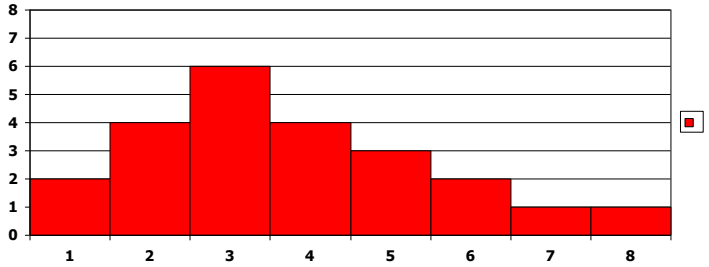
In a perfectly symmetrical normal distribution, the mean, median, and mode are the same value

Skewness

Many common statistical tests assume your data are normal distributed, but sometimes it is not (skewed)



Negative skew



Positive skew

- Mean is more affected by skewness than median
- Can transform data (e.g. take log or square root of values)
- Or use alternative tests

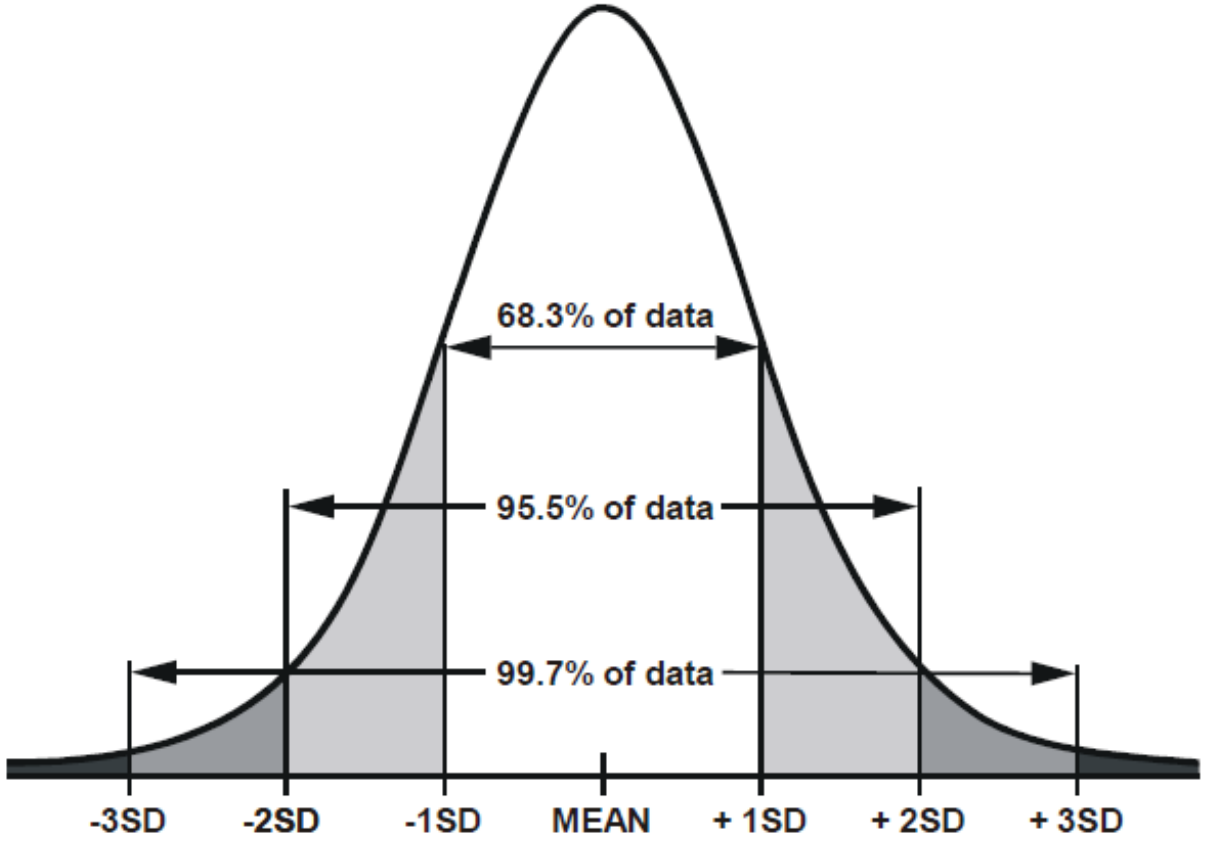
Measures of variability (spread)

Range→ highest value - lowest value

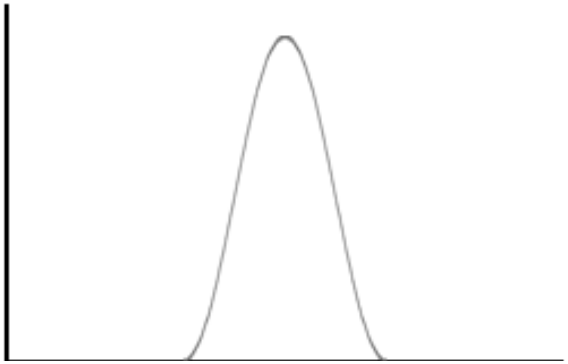
Variance is mean of squared deviations (differences) from sample mean

Standard deviation (*s*)→ square root of variance

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

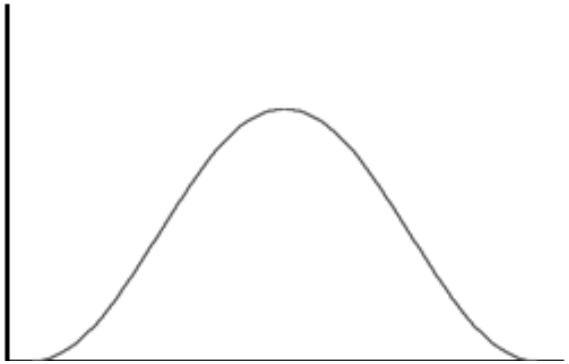

68.3% of data
95.5% of data
99.7% of data
-3SD
-2SD
-1SD
MEAN
+ 1SD
+ 2SD
+ 3SD

Low Standard Deviation



Mean height= 60 in, SD = 2 in (60 ± 2)

High Standard Deviation



Mean height= 60 in, SD = 5 in (60 ± 5)

Interquartile Range

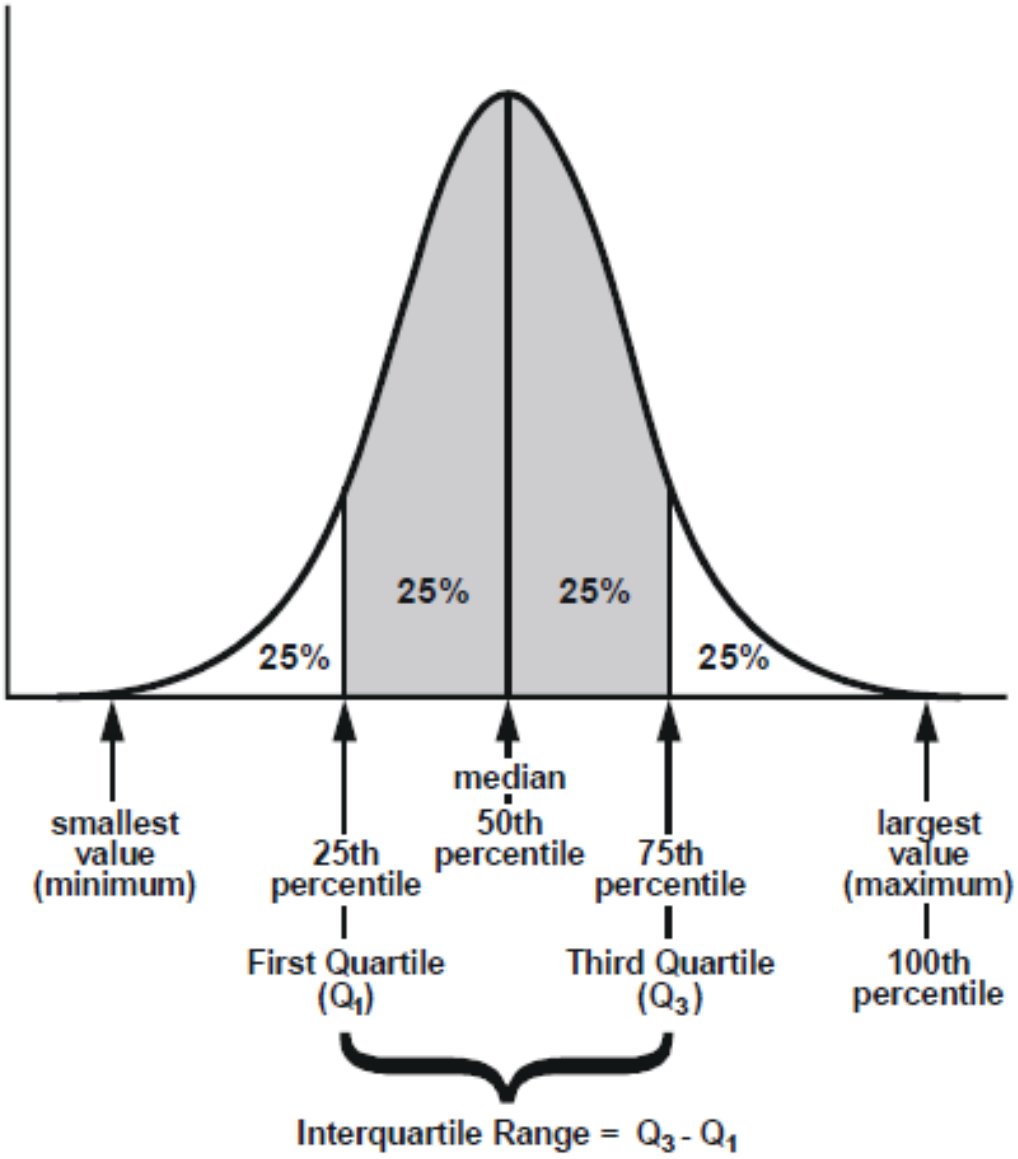
25^{th} percentile → 1^{st} quartile (Q1)

50^{th} percentile → 2^{nd} quartile (Q2) Median

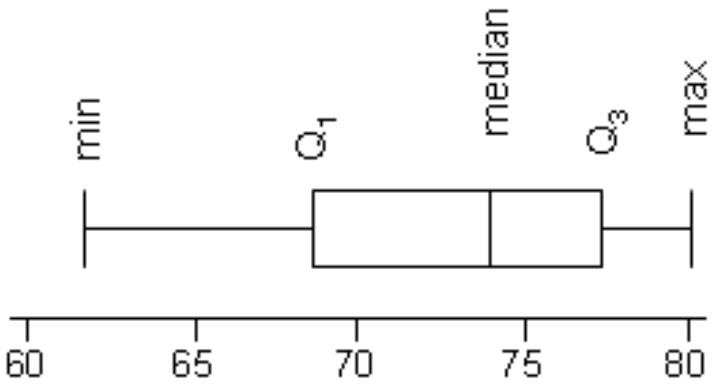
75^{th} percentile → 3^{rd} quartile (Q3)

The **inter-quartile range (IQR)** is the difference between the first and third quartiles, i.e.

$IQR = Q_3 - Q_1$


25%
25%
25%
25%
smallest value (minimum)
25th percentile
median 50th percentile
75th percentile
largest value (maximum)
First Quartile (Q_1)
Third Quartile (Q_3)
100th percentile
Interquartile Range = $Q_3 - Q_1$

Use Box Plot to Display Median and IQR



JMP Demo