

CLASS NOTES  
FOR  
ELEMENTARY STATISTICS  
LETU MATH-1423

Nicholas Capo

September 6, 2012  
9:19pm

# Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Data</b>                             | <b>4</b> |
| 1.1      | Data Sets . . . . .                     | 4        |
| 1.2      | Types of Data . . . . .                 | 4        |
| <b>2</b> | <b>Sample Mean and Median</b>           | <b>4</b> |
| 2.1      | Definition . . . . .                    | 4        |
| <b>3</b> | <b>Measures of Variability</b>          | <b>5</b> |
| 3.1      | Standard Deviation . . . . .            | 5        |
| 3.1.1    | Sample Variance . . . . .               | 5        |
| 3.1.2    | Sample Standard Deviation . . . . .     | 5        |
| <b>4</b> | <b>Descriptive Statistics</b>           | <b>5</b> |
| 4.1      | Quartiles . . . . .                     | 5        |
| 4.2      | Range and Interquartile Range . . . . . | 5        |
| 4.2.1    | Range . . . . .                         | 5        |
| 4.2.2    | Interquartile Range . . . . .           | 5        |
| 4.3      | Box and Whisker Plot . . . . .          | 5        |
| 4.3.1    | Example . . . . .                       | 6        |
| <b>5</b> | <b>Stem and Leaf Plots</b>              | <b>6</b> |
| 5.1      | Key Notation . . . . .                  | 6        |
| 5.2      | Double Stem and Leaf . . . . .          | 6        |
| <b>6</b> | <b>Frequency Distribution</b>           | <b>7</b> |
| 6.1      | Midpoint of a Class . . . . .           | 7        |
| 6.2      | Relative Frequency . . . . .            | 7        |
| <b>7</b> | <b>Scatter Plots</b>                    | <b>7</b> |
| 7.1      | Example Scatter Plot . . . . .          | 7        |
| <b>8</b> | <b>Homework</b>                         | <b>8</b> |

### **Abstract**

This document comprises classroom notes from Statistics Class at [LeTourneau University](#), in the Fall of 2012.

Although the author will attempt to be complete and correct in these notes, it is the readers responsibility to learn and understand the material. The author assumes no responsibility for the completeness or accuracy of this content.

The latest version of this document and the L<sup>A</sup>T<sub>E</sub>X sourcecode is available at:

<https://bitbucket.org/nicholascapo/statisticsnotes/src/tip/StatisticsNotes.pdf>

Copyright © 2012 Nicholas Capo  
Licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported  
<http://creativecommons.org/licenses/by-sa/3.0/>

### Definition of Statistics

“Statistics is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions.”

## 1 Data

### 1.1 Data Sets

**Population** The collection of all outcomes, responses, measurements, or counts, that are of interest.

**Sample** A subset of the population.

**Parameter** A number that describes a population characteristic.

**Statistic** A number that describes a sample characteristic.

### 1.2 Types of Data

**Qualitative Data** Attributes, labels, or non-numerical entries.

**Quantitative Data** Numerical measurements or counts.

## 2 Sample Mean and Median

### 2.1 Definition

**Sample Mean** The average of the sample data points, however it may not be a data point.

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{x_1 + x_2 + x_3 \cdots x_n}{n}$$

**Sample Median** The middle value of the data.

$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}) & \text{if } n \text{ is even} \end{cases}$$

**Trimmed Mean** A trimmed mean is computed by trimming off the largest and smallest set of values. For example a 10% trimmed mean is found by eliminating the largest 10% and smallest 10% and computing the mean of the remaining values. This may be useful for data that contains possible outliers. Denoted by  $x_{tr(\text{percent})}$

## 3 Measures of Variability

### 3.1 Standard Deviation

#### 3.1.1 Sample Variance

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

#### 3.1.2 Sample Standard Deviation

$$s = +\sqrt{s^2}$$

The standard deviation is 0 when all the data points are the same.

## 4 Descriptive Statistics

### 4.1 Quartiles

Quartiles approximately divide an ordered data set into four equal parts.

**First Quartile,  $Q_1$**  About 25% of the data fall on or below  $Q_1$

**Second Quartile,  $Q_2$**  About 50% of the data fall on or below  $Q_2$

**Third Quartile,  $Q_3$**  About 75% of the data fall on or below  $Q_3$

### 4.2 Range and Interquartile Range

#### 4.2.1 Range

$$\text{range} = \text{max value} - \text{min value}$$

#### 4.2.2 Interquartile Range

$$IQR = Q_3 - Q_1$$

To help find outliers, compute  $1.5 \times IQR$ , and any values that lie outside the interval  $[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$  is a possible (and probable) outlier.

### 4.3 Box and Whisker Plot

Exploratory Data Analysis Tool

- Requires
  - Min
  - $Q_1$
  - Median
  - $Q_3$
  - Max

### 4.3.1 Example

Example Data [1, 2, 3, 4, 5, 6, 11]  
 Min 1  
 Median 4.0  
 Max 6  
 Outlier 11

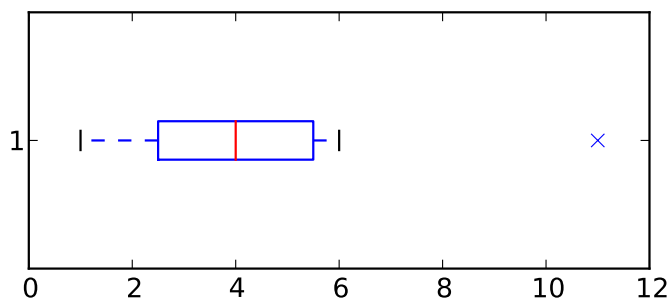


Figure 1: Example Box And Whisker Plot

## 5 Stem and Leaf Plots

These look like a sideways histogram

Data: [31, 21, 32, 33, 41, 42, 58, 25, 21]

| Stem | Leaf  | Key: $a b = ab$ |
|------|-------|-----------------|
| 2    | 1,1,5 |                 |
| 3    | 1,2,3 |                 |
| 4    | 1,2   |                 |
| 5    | 8     |                 |

### 5.1 Key Notation

Key: 4—5 = 45 Key: 4—5 = 4.5

### 5.2 Double Stem and Leaf

Separate the leaves into two groups, (0-4, and 5-9)

Data: [31, 21, 32, 33, 41, 42, 58, 25, 21]

| Stem | Leaf  | Key: $a b = ab$ |
|------|-------|-----------------|
| 2    | 1,1   |                 |
| 2    | 5     |                 |
| 3    | 1,2,3 |                 |
| 4    | 1.2   |                 |
| 4    |       |                 |
| 5    |       |                 |
| 5    | 8     |                 |

## 6 Frequency Distribution

A table that shows classes or intervals of data with a count of the number of entries in each class.

### 6.1 Midpoint of a Class

Average of the class limits.

$$\frac{(\text{lower class limit}) + (\text{upper class limit})}{2}$$

### 6.2 Relative Frequency

$$\frac{\text{class frequency}}{\text{sample size}} = \frac{f}{n}$$

## 7 Scatter Plots

Each entry in one data set corresponds to one entry in a second set, one-to-one mapping.

### 7.1 Example Scatter Plot

Data:

X: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]

Y: [10, 11, 10, 3, 1, 12, 10, 10, 8, 9, 4, 11]

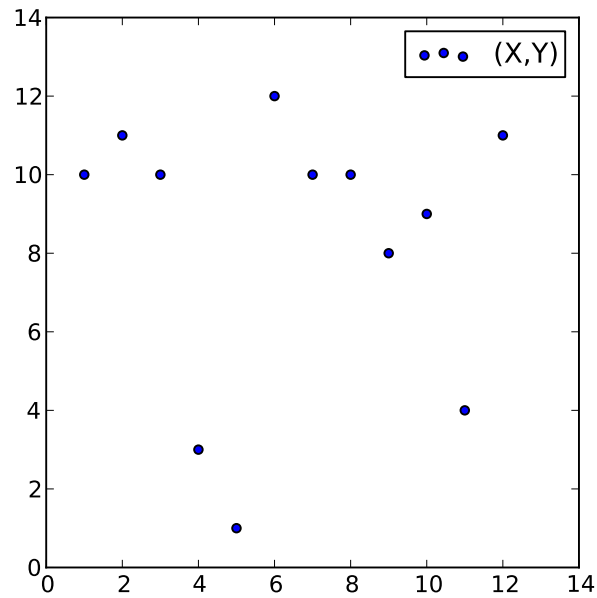


Figure 2: Example Scatter Plot

## 8 Homework

- Page 13 #'s 1.5, 1.6
- Page 17 #'s 1.11, 1.12
- Page 31 #'s 1.18, 1.19, 1.20, 1.29, 1.30



## References

- [1] R.E. Walpole et al. *Probability and Statistics for Engineers and Scientists*. Pearson Education, 2010. ISBN: 9780321629111. URL: <http://books.google.com/books?id=tzZxRQAACAAJ>.