

Statistics Class Notes

Nicholas Capo

September 3, 2012

Contents

0.1	Definition of Statistics	3
1	Data	3
1.1	Data Sets	3
1.2	Types of Data	3
2	Sample Mean and Median	3
2.1	Definition	3
3	Measures of Variability	4
3.1	Standard Deviation	4
3.1.1	Sample Variance	4
3.1.2	Sample Standard Deviation	4
4	Descriptive Statistics	4
4.1	Quartiles	4
4.2	Range and Interquartile Range	4
4.2.1	Range	4
4.2.2	Interquartile Range	4
4.3	Box and Whisker Plot	4
4.3.1	Example	5
5	Homework	5

0.1 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.

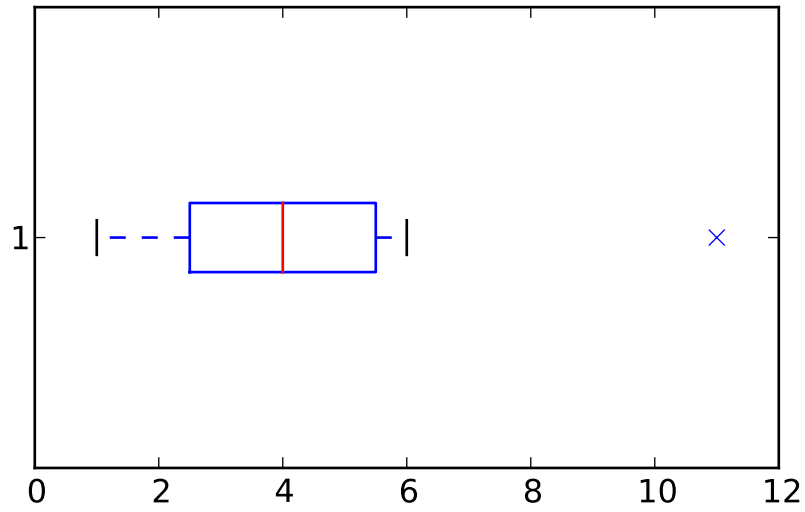
4.3 Box and Whisker Plot

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]



5 Homework

- p. 13 #'s 1.5, 1.6
- p. 17 #'s 1.11, 1.12
- p. 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>.