Introductory Statistics Lectures

Case Study

Descriptive Statistics

Anthony Tanbakuchi
Department of Mathematics
Pima Community College

© 2009

Contents

1 Case Study

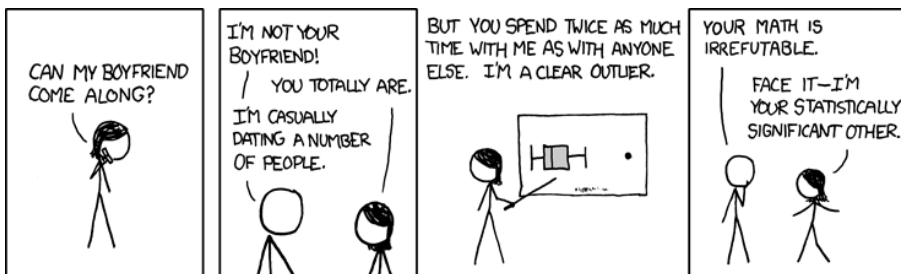## 1.1 Review

Statistically Significant Other

Figure 1: Statistically significant boyfriend. (Credit: XKCD Web Comics http://xkcd.com/539/)

Review: resistance to outliers

Question 1. Make a table listing the measures of center (mean, median, mode, midrange) and variation (range, IQR, standard deviation) we have discussed in terms of their resistance to outliers.

Demo

Validate the Empirical Rule using percentiles and 100 randomly generated data points for a normal distribution. Show how the empirical rule fails with a uniform distribution. (Time permitting.)

1.2   Case Study

Research Question

Are male and female students at this institution the same height on average?

Question 2. What is the population of interest?

Question 3. If a census is not possible, what method of sampling would be appropriate.

To answer this question we will sample the heights of students in this class.

Question 4. What type of sample is this?

Collected Data

```
R: load("ClassData.RData")
R: height.male = class.data$height[class.data$gender ==
+     "MALE"]
R: height.female = class.data$height[class.data$gender ==
+     "FEMALE"]
R: height.male
[1]  71  77  70  67  73  68  70
R: height.female
 [1]  65  68  66  68  65  62  68  62  69  65  63
```

Descriptive Statistics: Visual

```
R: boxplot(class.data$height ~ class.data$gender,
+      xlab = "Gender", ylab = "Height (in inches)")
```
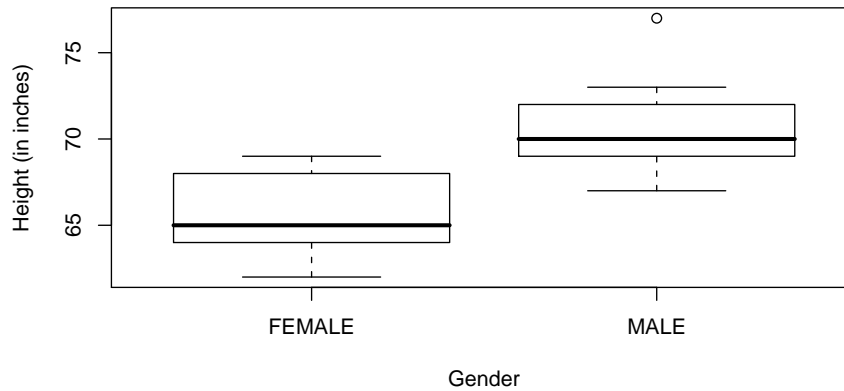

Question 5. What is the shape of the distribution for the males and females?

Question 6. Are there any outliers?

Question 7. How do the heights compare versus gender?

Descriptive Statistics: Numerical

Male Data

```
R: summary(height.male)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   67.0    69.0    70.0    70.9    72.0    77.0
R: sd(height.male)
[1] 3.3381
```

Typical Values $\bar{x} \pm 2s = (67.5, 74.2)$

Female Data

```
R: summary(height.female)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   62.0    64.0    65.0    65.5    68.0    69.0
R: sd(height.female)
[1] 2.5045
```

Typical Values $\bar{x} \pm 2s = (63, 68)$

Thinking about our research question.

Question 8. Does the sample data indicate that the mean height of males and females is different at this institution?

Question 9. What two possible explanations are there for the observed difference in the mean heights from the sample?

We will be able to answer the research question in a quantitative manner using **inferential statistics**.