Basic Stats

Lilly Winfree & Danielle Robinson

September 6, 2016

Let's learn how to read in data and then do some basic statistical analysis.

Read in data:

```
data<-read.table("example_lengths_for_R.csv", header=TRUE, sep=',')
```

We can look at the data quickly using summary(), nrow() to see the row numbers, colnames() for column names

```
summary(data)
```

```
##  coverlip.code       treatment       values
##  K      :119   Control    :393   Min.   :  3.083
##  Q      :110   Treatment A:413   1st Qu.: 19.660
##  F      :109   Treatment B:414   Median : 39.663
##  C      :106                     Mean   : 49.700
##  E      :105                     3rd Qu.: 71.938
##  A      :102                     Max.   :228.891
##  (Other):569
```

```
nrow(data)
```

```
## [1] 1220
```

```
colnames(data)
```

```
## [1] "coverlip.code" "treatment"     "values"
```

We can subset to different columns with $ and only see select rows with [:]

```
data$coverlip.code[1:10]
```

```
##  [1] G G G G G G G G G G
## Levels: A B C D E F G I K O P Q
```

```
data$treatment[1:10]
```

```
##  [1] Treatment A Treatment A Treatment A Treatment A Treatment A
##  [6] Treatment A Treatment A Treatment A Treatment A Treatment A
## Levels: Control Treatment A Treatment B
```

```
summary(data$values)
```

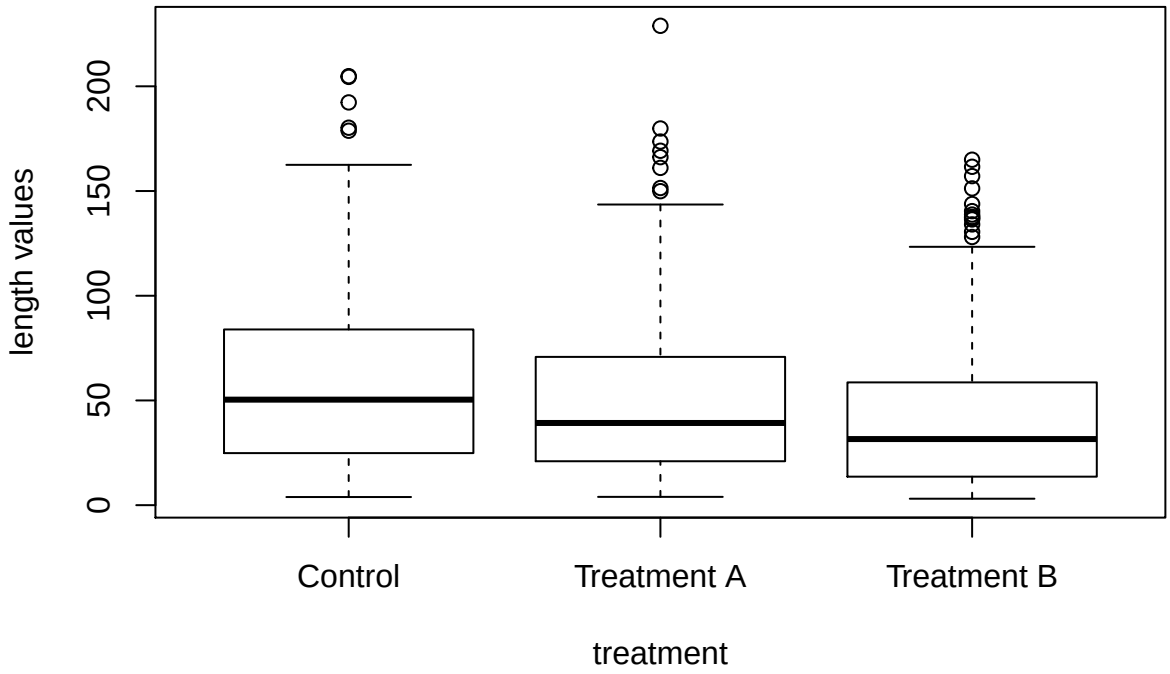
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.083  19.660  39.660  49.700  71.940 228.900
```

```
summary(data$values[data$treatment=='Control'])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   3.876  24.850  50.390  57.880  83.900 204.800
```

Let's plot our data!

```
plot(data$treatment, data$values, xlab='treatment', ylab='length values')
```



Now let's look at how to do some basic stats:

```
mean(data$values[data$treatment=='Treatment A'])
```

```
## [1] 50.00254
```

```
mean(data$values[data$treatment=='Control'])
```

```
## [1] 57.88106
```

```
mean(data$values[data$treatment=='Treatment B'])
```

```
## [1] 41.63183
```

```r
median(data$values[data$treatment=='Control'])
```

```
## [1] 50.39
```

```r
var(data$values[data$treatment=='Control'])
```

```
## [1] 1572.929
```

```r
min(data$values[data$treatment=='Control'])
```

```
## [1] 3.876
```

```r
max(data$values[data$treatment=='Control'])
```

```
## [1] 204.844
```

notice how summary() gave us the same info as the above functions OK let's do a t-test now! We can use this to compare the means between 2 of the groups

```r
t.test(data$values[data$treatment=='Control'], data$values[data$treatment=='Treatment B'])
```

```
##
##  Welch Two Sample t-test
##
## data:  data$values[data$treatment == "Control"] and data$values[data$treatment == "Treatment B"]
## t = 6.2243, df = 773.83, p-value = 7.922e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  11.12453 21.37393
## sample estimates:
## mean of x mean of y
##  57.88106  41.63183
```

What does this tell us? How do we change the code to look at the other groups?

We can also do an ANOVA with this data:

```r
results=aov(values ~ treatment, data=data)
summary(results)
```

```
##               Df  Sum Sq Mean Sq F value   Pr(>F)
## treatment      2   53291   26645   19.71 3.77e-09 ***
## Residuals   1217 1645402    1352
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

See ANOVA doc for more info on how an ANOVA works.