Statistical Inference Assignment - Part 2

Phillip Telfer

Overview

This report conducts a basic inferential data analysis on the ToothGrowth data in the R datasets package.

The analysis includes generation of basic exploratory analysis data plus the conduct of confidence interval t-tests to determine whether supplement types or dosage levels effect the tooth length in the test subjects.

Data Selection

This section of the report provides a basic inferential analysis of the ToothGrowth data from within the r datasets package.

```
library(datasets)
```

'ToothGrowth' contains data from an experiment investigating 'The Effect of Vitamin C on Tooth Growth Guinea Pigs'. The data contains three columns:

- len - the length of teeth at the end of the trial (in microns)
- supp - the supplement used to provide vitamin c to the trial guinea pigs (either OJ 'Orange Juice') or VC 'Absorbic Acid')
- dose - the dose applied to the subject guinea pig (in milligrams). Was either 0.5, 1 or 2.

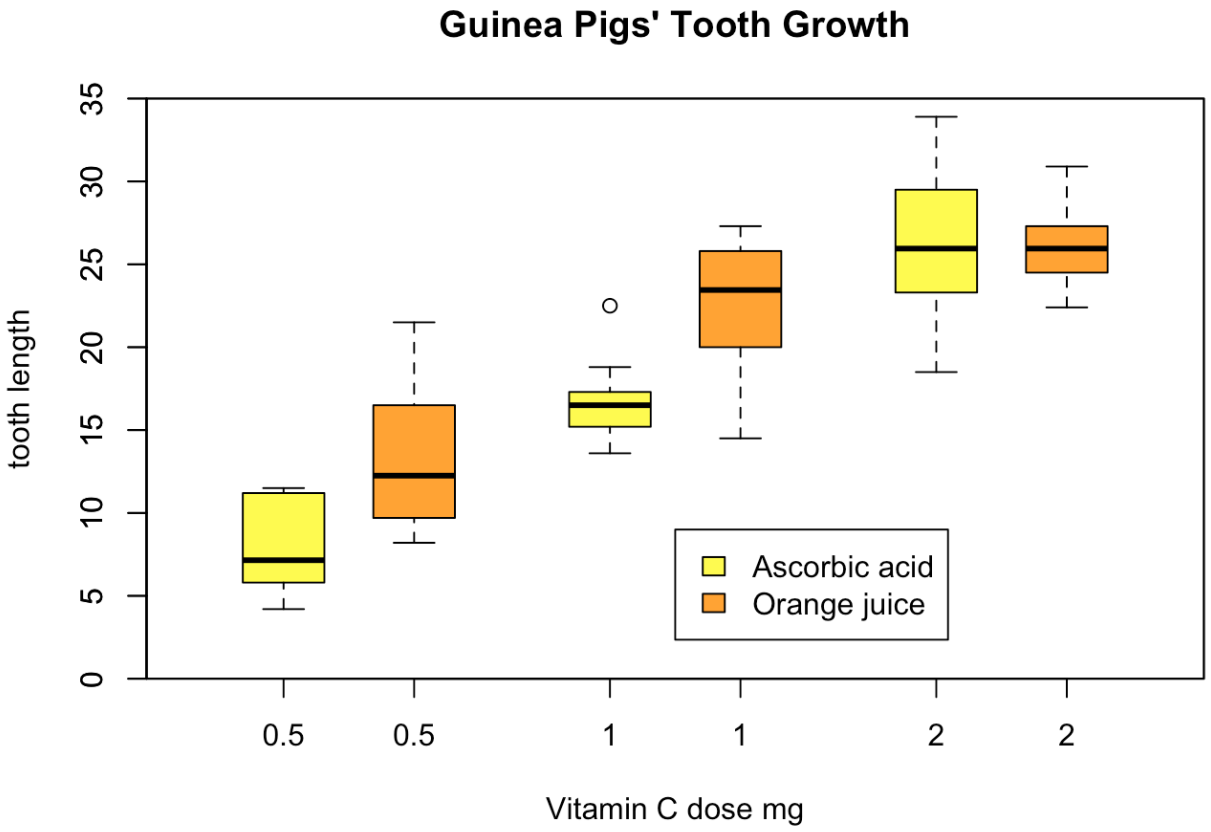
60 guinea pigs were involved in the trial, with each subject receiving a specific supplement and dose level.

Exploratory Data Analysis

To gain an intutive understanding of the spread of the data, a boxplot of the data, grouped by supplement type and dose is created.

```
library(dplyr)
```

```
boxplot(len ~ dose, data = ToothGrowth, boxwex = 0.25, at = 1:3 - 0.2, subset =
 supp == "VC", col = "yellow", main = "Guinea Pigs' Tooth Growth", xlab = "Vita
min C dose mg", ylab = "tooth length", xlim = c(0.5, 3.5), ylim = c(0, 35), yax
s = "i")
boxplot(len ~ dose, data = ToothGrowth, add = TRUE, boxwex = 0.25, at = 1:3 + 0
.2, subset = supp == "OJ", col = "orange")
legend(2, 9, c("Ascorbic acid", "Orange juice"), fill = c("yellow", "orange"))
```



The graph suggests an increase in tooth length associated with a higher dose, and at the lower doses (0.5 and 1.0) the orange juice supplement also appears to be associated with a higher tooth length than the absorbic acid.

The mean and standard deviation data corresponding to the supplement and dose levels is displayed below.

```
m <-aggregate(len ~ supp + dose, ToothGrowth, mean)
s <-aggregate(len ~ supp + dose, ToothGrowth, sd)
sumd <- cbind(m, s$len)
names(sumd) <- c("supp","dose", "mean", "sd")
arrange(sumd, supp, dose)
```

```
##   supp dose  mean       sd
## 1   OJ  0.5 13.23 4.459709
## 2   OJ  1.0 22.70 3.910953
## 3   OJ  2.0 26.06 2.655058
## 4   VC  0.5  7.98 2.746634
## 5   VC  1.0 16.77 2.515309
## 6   VC  2.0 26.14 4.797731
```

Confidence Interval Testing

Two confidence intervals will be tested:

1. Mean of tooth length as effected by supplement type, and
2. Mean of tooth length as effected by dose level.

Supplement type

For this test, an equal variance is assumed from review of the boxplot and statistics generated in the exploratory data analysis. A t-test is conducted between the two data sets defined by supplement type.

```
suppt <- t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = ToothGrowt
h)
suppt
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The confidence interval for the t.test was [-0.1670064, 7.5670064], which includes 0. Therefore the supplement type was not shown to have a statistically significant effect on tooth length.

Dose levels

For this test, an unequal variance is assumed from review of the boxplot and statistics generated in the exploratory data analysis. A t-test is conducted between the data sets defined by dosage levels. Three tests were conducted:

1. Dose of 0.5mg versus 1.0mg
2. Dose of 0.5mg versus 2.0mg
3. Dose of 1.0mg versus 2.0mg

```
d.5to1 <- filter(ToothGrowth, dose == 0.5 | dose == 1.0)
d.5to2 <- filter(ToothGrowth, dose == 0.5 | dose == 2.0)
d1to2 <- filter(ToothGrowth, dose == 1.0 | dose == 2.0)

dose_t1 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d.5to1)
dose_t2 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d.5to2)
dose_t3 <- t.test(len ~ dose, paired = FALSE, var.equal = FALSE, data = d1to2)
dose_t1
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
dose_t2
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
dose_t3
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The intervals for all three tests, [-11.9837813, -6.2762187], [-18.1561665, -12.8338335] and [-8.9964805, -3.7335195], remain below zero. Therefore an increase in dosage from 0.5 to 1mg, or 1 to 2.0mg, is statisically proven to increase tooth length.

Conclusion

The confidence interval output indicated that:

1. The supplement type was not shown to have a statistically significant effect on tooth length. This output assumed equal variance between the two groups.

2. The dose level was shown to have a statistically significant effect on tooth length. This output assumed nonequal variance between the three groups.