Statistical Inference Project Part 2

Kyle Day

Overview

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

Step 1: Loading the data

```
data(ToothGrowth)
toothgrowth <- ToothGrowth
toothgrowth$dose <- as.factor(toothgrowth$dose) ## will help things later on.
```

Step 2: Summary

Description: The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
#?ToothGrowth
#dim(toothgrowth)
#head(toothgrowth)
str(toothgrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(toothgrowth)
```

```
##       len        supp     dose   
##  Min.   : 4.20   OJ:30   0.5:20  
##  1st Qu.:13.07   VC:30   1  :20  
##  Median :19.25           2  :20  
##  Mean   :18.81                   
##  3rd Qu.:25.27                   
##  Max.   :33.90                   
```

```
meth1 <- split(toothgrowth, toothgrowth$supp) ## split by supp
meth2 <- split(toothgrowth, toothgrowth$dose) ## split by dose
par(mfrow=c(1,2))
#plot(meth1$VC[,3], meth1$VC[,1], main="len by dose for VC", col="purple")
#plot(meth1$OJ[,3], meth1$OJ[,1], main="len by dose for OJ", col="orange")
```

When we look at the relationship of len and dose for both supp's we see that a positive and possibly a slightly diffrent relationship exist for both.

```
names(meth2) <- c("half", "one", "two")
## it's difficult to use the numbers that were originally assigned.
col <- c("orange", "purple")
par(mfrow=c(1,3))
#plot(meth2$half[,2], meth2$half[,1], main="len by supp for dose 0.5", col = col)
#plot(meth2$one[,2], meth2$one[,1], main="len by dose for dose 1", col = col)
#plot(meth2$two[,2], meth2$two[,1], main="len by dose for dose 2", col = col)
```

Charts are not outputted to save space.

From these charts it seems like the main relationship of len is with dose and is varied by supp. This supports the basic ideas of the experiment.

Step 3: Compare tooth growth by supp and dose.

We know from the information supplied to us about the experiment that the Guinea Pigs are kept on a Vitamin C free diet and that the 0.5 dose is essential for them not to get scurvy. Because of this we can treat 0.5 dose as a base level or control.
There are really two questions posed by this experiment.

Question 1: Does Vitamin C effect tooth growth? This requires a two group interval T test comparing the mean of dose at 0.5 to the mean at dose 1 and dose2.

Therefore question 1 part 1 would be: Ho: mu0.5 = mu1 - length of 0.5 does equal length of 1 Ha: mu0.5 != mu1 - length of 0.5 does not equal length of 1

and part two would be: Ho: mu0.5 = mu2 - length of 0.5 does equal length of 2 Ha: mu0.5 != mu2 - length of 0.5 does not equal length of 2

Note: because we know there is a positive relationship we could test Ha: mu0.5 < mu1. For now lets stick with the more general question, which means we are using a two sided test.

```
#var.test(meth2$half[,1], meth2$one[,1])
#var.test(meth2$half[,1], meth2$two[,1])
```

Variance tests not shown to save space.

To perform a t.test we need to know if the variance is equal between the data sets. We know that they do have similar variance because the p-value is greater than 0.05.

```
t.test(meth2$half[,1], meth2$one[,1], var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  meth2$half[, 1] and meth2$one[, 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Because our p-value is lower than our alpha we can safely **reject** Ho in favor of Ha.

```r
t.test(meth2$half[,1], meth2$two[,1], var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  meth2$half[, 1] and meth2$two[, 1]
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15352 -12.83648
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

Because our p-value is lower than our alpha we can safely **reject** Ho in favor of Ha.

Question 2: Does the supp(liment), OJ or VC effect the level of growth.
Ho: muVC = muOJ - length of VC equals OJ
Ha: muVC != muOJ - length of VC does not equals OJ

```r
#var.test(meth1$VC[,1], meth1$OJ[,1]) ##our var is equal
t.test(meth1$VC[,1], meth1$OJ[,1], var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  meth1$VC[, 1] and meth1$OJ[, 1]
## t = -1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5670064  0.1670064
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Because our p-value is higher than our alpha we **fail to reject** Ho.

Step 4: Conclusions and Assumptions

In summary we have rejected the hypothesis that length of teeth will be the same regardless of Vitamin C level. We also know, because the confidence intervals were negative, the length at dose 0.5 was less than at both dose one and two. Suggesting an increased length for greater amounts of Vitamin C. We also have failed to reject the hypothesis that the supplements of orange juice and ascorbic acid have different effects on teeth length for measures across all doses. **These conclusions are based off the assumptions that this group of Guinea Pigs represents the whole population, was picked at random, dosed at random, different Guinea Pigs were used for each dose (therefore not paired),the dose 0.5 is a good baseline level of Vitamin C, and the data is normally distributed.** Lastly, in the future we might also want to test if orange juice and ascorbic acid have different effects on teeth length for specific doses.