Project Stat Inference Part 2

R. Quinn

Tuesday, October 20, 2015

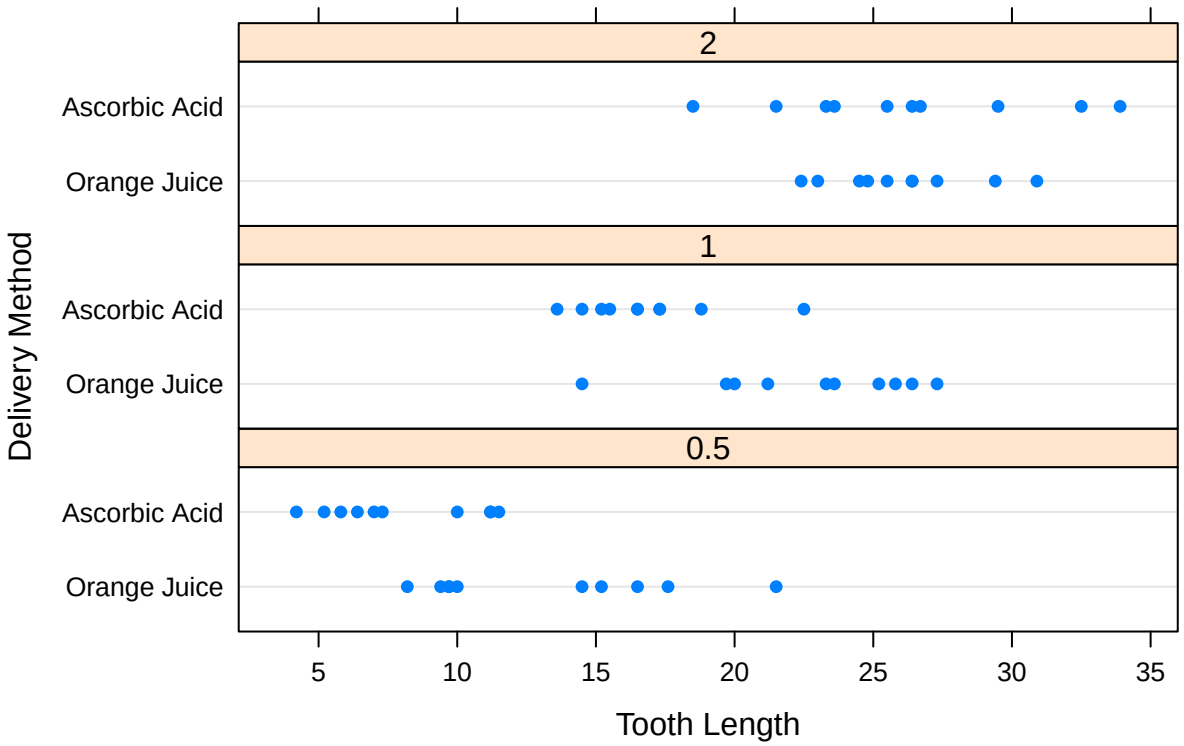
1. Load the ToothGrowth data and perform some basic exploratory analysis

```
##Ttake a look at its structure
data (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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

We can see that our data is a data frame of 60 observations of 3 variables. More information is available at the ToothGrowth dataset page This states that the data contains measurements of tooth growth in 60 guinea pigs who received one of three daily doses of vitamin C (0.5,1,2 mg) by one of two methods (orange juice or ascorbic acid).

```
library (lattice)
names(ToothGrowth) <- c("length", "method", "dose")
levels (ToothGrowth$method) <- c("Orange Juice", "Ascorbic Acid")
ToothGrowth$dose <- factor (ToothGrowth$dose)
dotplot (ToothGrowth$method ~ ToothGrowth$length|ToothGrowth$dose, xlab= "Tooth Length", ylab="Delivery
```



The plot above shows that across both methods of administration (orange juice and ascorbic acid) tooth length increased as dosage increased.

The plot above also shows that for the dosage levels of 0.5mg/day and 1mg/day administration of Vitamin C by Orange Juice resulted in greater variability of tooth length and greater overall tooth growth than administration by ascorbic acid did, however at the level of 2mg/day there looks to be little difference between methods of administration.

2. Provide a basic summary of the data.

```
##summary of complete data and summary of data broken down by group
summary (ToothGrowth)
```

```
##      length                method    dose
##  Min.   : 4.20   Orange Juice :30   0.5:20
##  1st Qu.:13.07   Ascorbic Acid:30   1  :20
##  Median :19.25                      2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
by(ToothGrowth$length, INDICES = list(ToothGrowth$dose, ToothGrowth$method), summary)
```

```
## : 0.5
## : Orange Juice
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
## ------------------------------------------------------------
## : 1
## : Orange Juice
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## ------------------------------------------------------------
## : 2
## : Orange Juice
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.58   25.95   26.06   27.08   30.90
## ------------------------------------------------------------
## : 0.5
## : Ascorbic Acid
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
## ------------------------------------------------------------
## : 1
## : Ascorbic Acid
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.27   16.50   16.77   17.30   22.50
## ------------------------------------------------------------
## : 2
## : Ascorbic Acid
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Firstly I looked at tooth growth by method of administration of vitamin C (supp). I used an independent samples t test to test whether there is a significant difference between tooth length in guinea pigs given vitamin C as orange juice compared with those given ascorbic acid.

```
asc <- ToothGrowth$length[1:30]
oj <- ToothGrowth$length[31:60]
t.test (asc, oj, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  asc and oj
## t = -1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5710156  0.1710156
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

The null hypothesis was that there is no difference in tooth length between guinea pigs given orange juice and guinea pigs given ascorbic acid. The results of the t test give a p value of 0.06039 which means that our result just fails to reach significance at the 0.05 level. We therefore fail to reject the null hypothesis. Guinea

pigs given daily doses of orange juice did not show significantly different tooth growth than those given daily doses of asorbic acid.

Although, as discussed already, our data appears to indicate that Orange Juice produces greater tooth growth than Ascorbic Acid, I used a two tailed test. This is because we should select the tests we wish to use prior to collecting our data and with no apriori reason to specify a direction of difference I used a two tailed test. On the two tailed test our p value approached significance so further investigation of this relationship with a larger sample may be warranted.

```
##assign low, medium and high to each of the dosage groups and run t tests on each pair of ##groups
low_dose <- ToothGrowth$length[ToothGrowth$dose==0.5]
medium_dose <- ToothGrowth$length[ToothGrowth$dose==1]
high_dose <- ToothGrowth$length[ToothGrowth$dose==2]
med_low <-t.test (medium_dose, low_dose, paired=FALSE, alternative="greater")
format (med_low, scientific=FALSE)
```

```
##                   statistic                   parameter
##                  "6.476648"                  "37.98641"
##                     p.value                    conf.int
##          "0.00000006341504"             "6.753323, Inf"
##                    estimate                  null.value
##            "19.735, 10.605"                         "0"
##                 alternative                      method
##                   "greater"  "Welch Two Sample t-test"
##                   data.name
## "medium_dose and low_dose"
```

Above are the results of the comparison of the medium dose (1 mg/day) group with the low dose (0.5 mg/day) group. In this case it seems reasonable to propose an alternate hypotheses (H1) that greater doses of vitamin C lead to greater tooth growth so I applied a one tailed test. The results of the t test indicate we should reject the null hypothesis.

```
high_low <- t.test (high_dose, low_dose, paired=FALSE, alternative="greater")
format (med_low, scientific=FALSE)
```

```
##                   statistic                   parameter
##                  "6.476648"                  "37.98641"
##                     p.value                    conf.int
##          "0.00000006341504"             "6.753323, Inf"
##                    estimate                  null.value
##            "19.735, 10.605"                         "0"
##                 alternative                      method
##                   "greater"  "Welch Two Sample t-test"
##                   data.name
## "medium_dose and low_dose"
```

The results of the comparison of the high dose (2mg/day) group with the low dose (0.5mg/day) group are once again highly significant and we should reject the null hypothesis that there is no difference in tooth growth between these groups.

```
high_med <- t.test (high_dose, medium_dose, paired=FALSE, alternative="greater")
format (high_med, scientific=FALSE)
```

```
##                   statistic                   parameter
##                  "4.900484"                  "37.10109"
##                     p.value                    conf.int
##            "0.000009532148"              "4.17387, Inf"
##                    estimate                  null.value
##            "26.100, 19.735"                         "0"
##                 alternative                      method
##                   "greater"   "Welch Two Sample t-test"
##                   data.name
## "high_dose and medium_dose"
```

The results of the comparison of high dose and medium dose groups were also significant and the null hypothesis of no difference in tooth growth between these groups should be rejected.

4. State your conclusions and the assumptions needed for your conclusions.

Conclusions
Higher dosages of Vitamin C (irrespective of whether these are administered by Ascorbic Acid or Orange Juice) lead to more tooth growth in guinea pigs than lower doses. There was a significant difference between groups administered 0.5mg per day and 1 mg per day, between groups administered 1mg per day and 2 mg per day and between groups administered 0.5mg day and 2mg day. In each case the higher dosage led to higher rates of tooth growth.

With regard to method of administration of Vitamin C the difference between Ascorbic Acid and Orange Juice did not reach a statistically significant level. From our graph it appears that there is a difference between Vitamin C and Orange Juice at 0.5mg/day and 1 mg/day dosage levels but not at 2mg/day. It might be worth investigating further the nature of this relationship, given more data.

Assumptions
We are assuming that the samples are independent. This is a safe assumption to make as each tooth length measurement represents a different guinea pig.

We are also assuming that tooth length is normally distributed in guinea pigs.

We are assuming that a random sample of guinea pigs were used in the study.

We assume that tooth length is measured on at least an interval scale. Length is a ratio measurement so it satisfies this assumption.

For the one tailed tests we are assuming the data is not significantly skewed.