Statistical Inference - Part 2

Graeme Smith

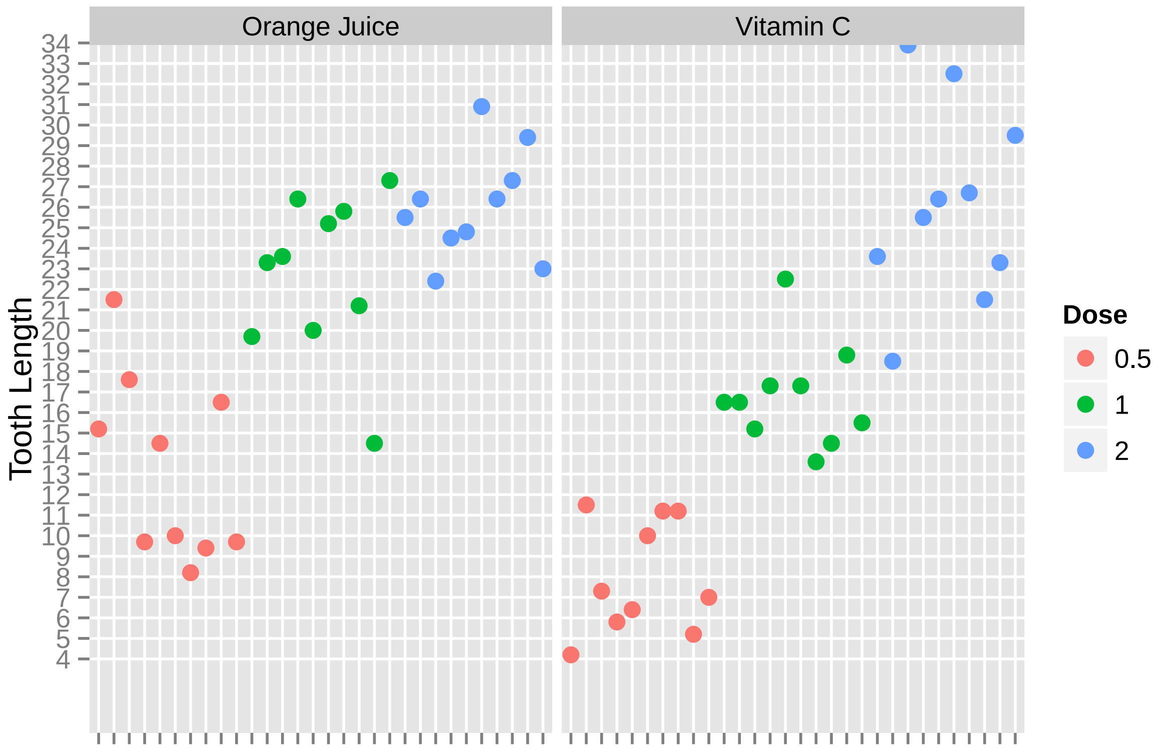
This is the second part of the assignment, it is an anlysis of the publically available ToothGrowth data set to infer whether vitamin c supplements increase the length of teeth in guinea pigs.

Load the ToothGrowth data and perform some basic exploratory data analyses

The data was loaded and summaries created for each supplement and dosage by the following code :

```
data(ToothGrowth)

ToothGrowth$dose <- as.factor(ToothGrowth$dose)
names(ToothGrowth) <- c("Len", "Supp", "Dose")
supplements <- c("Orange Juice", "Vitamin C")

byDose <- as.data.frame(t(
  sapply(unique(as.character(ToothGrowth$Dose)),
         function(Dose) data.frame(
           Mean=round(mean(ToothGrowth$Len[ToothGrowth$Dose==Dose]), 2),
           StdDev=round(sd(ToothGrowth$Len[ToothGrowth$Dose==Dose]), 2),
           N=as.numeric(sum(ToothGrowth$Dose==Dose)),
           LowerConfidenceInterval=round(mean(ToothGrowth$Len[ToothGrowth$Dose==Dose])
                                          - qt(0.975, df=sum(ToothGrowth$Dose==Dose)-1)
                                          * sd(ToothGrowth$Len[ToothGrowth$Dose==Dose])
                                          / sqrt(sum(ToothGrowth$Dose==Dose)), 2),
           UpperConfidenceInterval=round(mean(ToothGrowth$Len[ToothGrowth$Dose==Dose])
                                          + qt(0.975, df=sum(ToothGrowth$Dose==Dose)-1)
                                          * sd(ToothGrowth$Len[ToothGrowth$Dose==Dose])
                                          / sqrt(sum(ToothGrowth$Dose==Dose)), 2)
         ))
))

bySupplement <- as.data.frame(t(
  sapply(unique(as.character(ToothGrowth$Supp)),
     function(Supplement) data.frame(
       Mean=round(mean(ToothGrowth$Len[ToothGrowth$Supp==Supplement]), 2),
       StdDev=round(sd(ToothGrowth$Len[ToothGrowth$Supp==Supplement]), 2),
       N=sum(ToothGrowth$Supp==Supplement),
       LowerConfidenceInterval=round(mean(ToothGrowth$Len[ToothGrowth$Supp==Supplement])
                                     - qt(0.975, df=sum(ToothGrowth$Supp==Supplement)-1)
                                     * sd(ToothGrowth$Len[ToothGrowth$Supp==Supplement])
                                     / sqrt(sum(ToothGrowth$Supp==Supplement)), 2),
       UpperConfidenceInterval=round(mean(ToothGrowth$Len[ToothGrowth$Supp==Supplement])
                                     + qt(0.975, df=sum(ToothGrowth$Supp==Supplement)-1)
                                     * sd(ToothGrowth$Len[ToothGrowth$Supp==Supplement])
                                     / sqrt(sum(ToothGrowth$Supp==Supplement)), 2)
     ))
))

rownames(bySupplement) <- c("Vitamin C", "Orange Juice")
```

The data set can be visually inspected below :



Provide a basic summary of the data.

	Mean	StdDev	N	LowerConfidenceInterval	UpperConfidenceInterval
0.5	10.61	4.50	20.00	8.50	12.71
1	19.73	4.42	20.00	17.67	21.80
2	26.10	3.77	20.00	24.33	27.87

Table 1: Summary by Dosage

	Mean	StdDev	N	LowerConfidenceInterval	UpperConfidenceInterval
Vitamin C	16.96	8.27	30.00	13.88	20.05
Orange Juice	20.66	6.61	30.00	18.20	23.13

Table 2: Summary by Supplement

Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

The data does not include a baseline or control group. This means that efficacy of the each supplement regime and dosage cannot be compared to a control. So instead I compare the two supplement regimes to each other, and I compare the highest dosage (2.0) to the medium dosage (1.0) and the lowest dosage (0.5) to see if increasing the dosage causes a significant improvement in efficacy.

Supplement Regime

```
VitaminC <- bySupplement["Vitamin C",]
OrangeJuice <- bySupplement["Orange Juice",]

sp <- sqrt(((VitaminC$N[[1]]-1)*VitaminC$StdDev[[1]]^2+(OrangeJuice$N[[1]]-1) *
              OrangeJuice$StdDev[[1]]^2)
           / (VitaminC$N[[1]] + OrangeJuice$N[[1]] - 2))
ci <- OrangeJuice$Mean[[1]] - VitaminC$Mean[[1]] +
      c(-1, 1) * qt(0.975, VitaminC$N[[1]] + OrangeJuice$N[[1]] - 2) * sp *
      sqrt(1 / VitaminC$N[[1]] + 1 / OrangeJuice$N[[1]])
p <- round(2*pnorm(0, mean=OrangeJuice$Mean[[1]]-VitaminC$Mean[[1]], sd=sp *
      sqrt(1 / VitaminC$N[[1]] + 1 / OrangeJuice$N[[1]])), 3)
```

	LowerBound	UpperBound
1	-0.17	7.57

Table 3: Confidence interval of the difference between Supplement Regimes

Dosage Regime

```
LowestDose <- byDose[1,]
HighestDose <- byDose[3,]

sp <- sqrt(((HighestDose$N[[1]]-1)*HighestDose$StdDev[[1]]^2+
              (LowestDose$N[[1]]-1)*LowestDose$StdDev[[1]]^2) /
             (HighestDose$N[[1]] + LowestDose$N[[1]] - 2))
ci <- HighestDose$Mean[[1]] - LowestDose$Mean[[1]] + c(-1, 1) *
      qt(0.975, HighestDose$N[[1]] + LowestDose$N[[1]] - 2) * sp *
      sqrt(1 / HighestDose$N[[1]] + 1 / LowestDose$N[[1]])
```

	LowerBound	UpperBound
1	12.83	18.15

Table 4: Confidence interval of the difference between High (2.0) and Low (0.5) Dosages

```
MediumDose <- byDose[2,]

sp <- sqrt(((HighestDose$N[[1]]-1)*HighestDose$StdDev[[1]]^2+(MediumDose$N[[1]]-1) *
            MediumDose$StdDev[[1]]^2) / (HighestDose$N[[1]] + MediumDose$N[[1]] - 2))
ci <- HighestDose$Mean[[1]] - MediumDose$Mean[[1]] + c(-1, 1) *
      qt(0.975, HighestDose$N[[1]] + MediumDose$N[[1]] - 2) *
      sp * sqrt(1 / HighestDose$N[[1]] + 1 / MediumDose$N[[1]])
```

State your conclusions and the assumptions needed for your conclusions.

In all cases the t distribution was used, the subjects were assumed to be independant, there was assumed to be no interaction between the supplementation regime and the dosage, and two tailed tests were performed in each case.

	LowerBound	UpperBound
1	3.74	9.00

Table 5: Confidence interval of the difference between High (2.0) and Medium (1.0) Dosages

The first hypothesis tested was that there was a difference in efficacy between the orange juice and vitamin c supplement regimes. In this case the null hypothesis was that there was no difference between the two regimes. In the results the confidence interval included the value of zero meaning it was not possible to rule out the null hypothesis that there was no difference between the two regimes. However it should be noted that the p value was 0.056 suggesting that the study is worth replicating with a larger number of subjects.

The next hypotheses tested were that there are differences in efficacy between the highest dosage and the lower dosages. In these cases the null hypotheses were that there was no difference in efficacy between the different dosages. For both tests the higher dosage (2.0) caused significantly larger teeth in the subjects than in the medium (1.0) and the lower(0.5) dosages.

In summary, the highest dose of vitamin c (2.0) significantly outperformed lower dosages in increasing the length of guinea pig teeth. No control group was used, so this cannot be generalised to say that vitamin c supplementation increases the length of the rodents teeth, but it does suggest that this is possible. The study wasn't able to able to demonstrate a significant difference between the supplement regimes of orange juice or vitamin c at the 0.05 threshhold, but it would have at a 0.06 threshold, suggesting this could be worth investigating in larger trials.