Statistical Inference - Tooth Growth Data Analysis

Ivana Peretin

July, 2015

The purpose of this report is to explore Tooth Growth data using simple exploratory and inferential anlysis. The data measures the length of teeth (**len**) for 10 guinea pigs at each of three dose levels (**dose**) of Vitamin C (0.5, 1.0, 2.0 mg) with each of two delivery methods (**supp**: orange juice - OJ, or ascorbic acid - VC). Our goal is to study the effects of Vitamin C dosage and delivery method on teeth growth in guinea pigs.

Exploratory Analysis

Our first step will be to get a general sense of data through basic summary tables and plots.

```
library(pander)
library(dplyr)
library(ggplot2)
data(ToothGrowth)
```

Table 1: ToothGrowth data summary

len	supp	dose
Min. : 4.20	OJ:30	Min. :0.500
1st Qu.:13.07	VC:30	1st Qu.:0.500
Median :19.25	NA	Median :1.000
Mean :18.81	NA	Mean :1.167
3rd Qu.:25.27	NA	3rd Qu.:2.000
Max. :33.90	NA	Max. :2.000

Table 2: Summary of mean tooth length per delivery method

supp	N	mean	median	sd
OJ	30	20.66	22.7	6.606
VC	30	16.96	16.5	8.266

Table 3: Summary of mean tooth length per dose

dose	N	mean	median	sd
0.5	20	10.61	9.85	4.5
1	20	19.73	19.25	4.415
2	20	26.1	25.95	3.774

Table 4: Summary of mean tooth length per delivery method and dose

supp	dose	N	mean	median	sd
OJ	0.5	10	13.23	12.25	4.46
OJ	1	10	22.7	23.45	3.911
OJ	2	10	26.06	25.95	2.655
VC	0.5	10	7.98	7.15	2.747
VC	1	10	16.77	16.5	2.515
VC	2	10	26.14	25.95	4.798

From the tables above we can see that mean teeth length is generally larger for delivery through orange juice. If we take a look at the scatterplot below (*Figure 1*), we can see that teeth length values are more evenly distributed between low and high values for ascorbic acid (VC), while there is more dots for higher teeth length values in cases where delivery method was orange juice. Also, for both delivery methods the mean teeth length increases with the dosage. From the boxplot below (*Figure 2*) the trend can be easily seen.

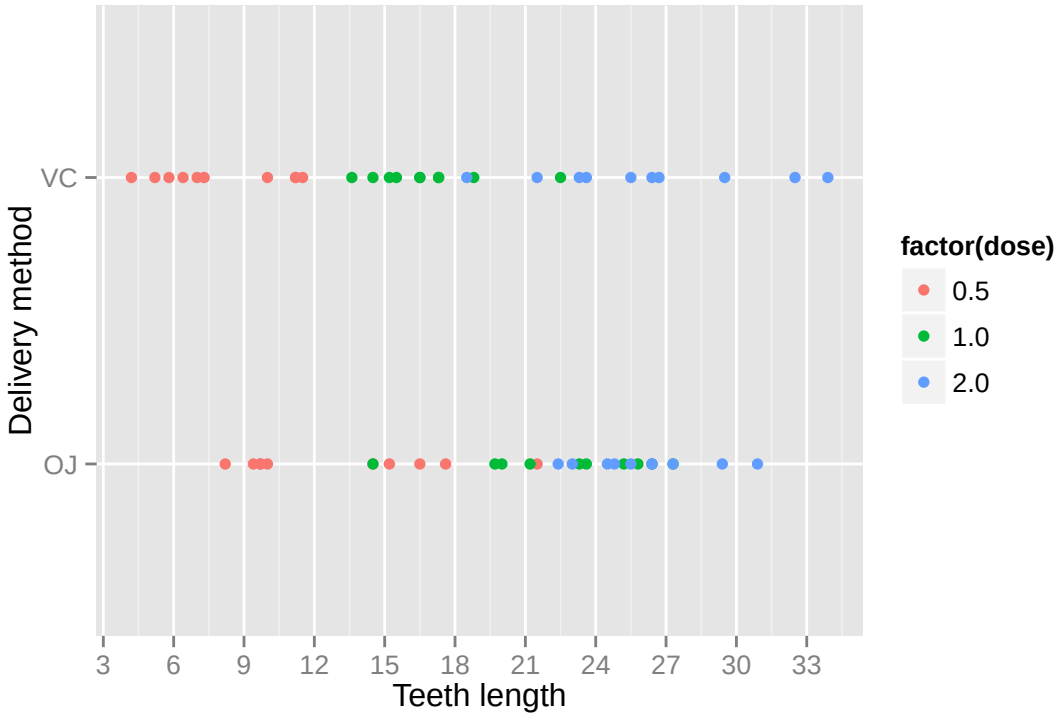

Figure 1: Teeth length per delivery method

Inferential analysis

Next step will be to confirm our observations from the previous section by performing several hypothesis tests (two-sample t-test).

Growth by delivery method

According to the t-test results p-value is 0.061 which is larger than our confidence level 0.05. Also, the confidence interval contains 0 and we don't have enough evidence to reject null hypothesis that there is no significant difference in average teeth length between OJ and VC delivery methods.

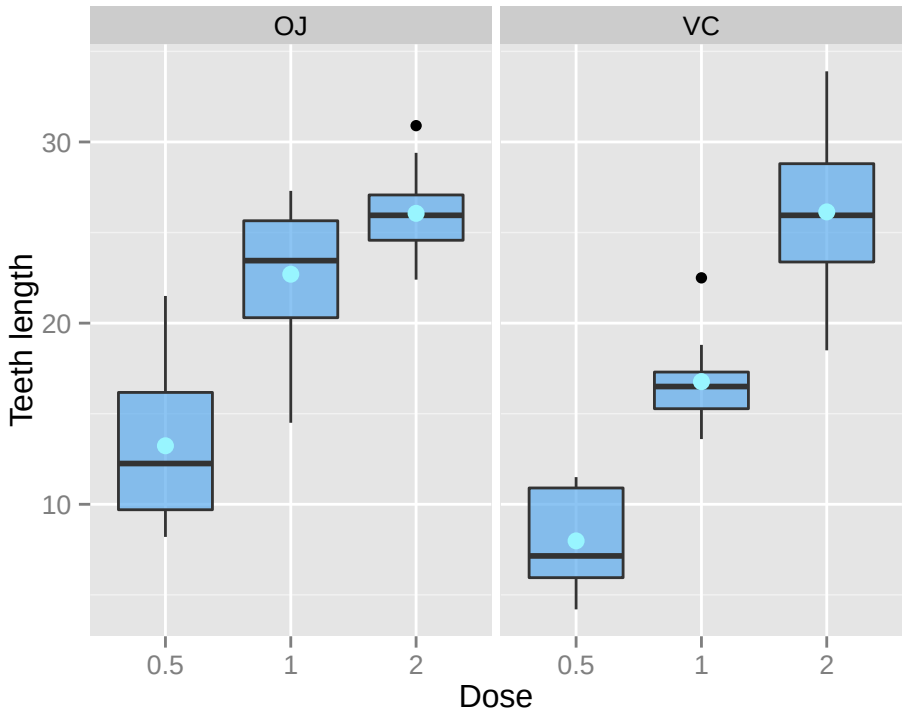

Figure 2: Teeth length per dose and delivery method

```
t1 <- t.test(len ~ supp, var.equal=T, data=ToothGrowth)
```

Table 5: Two-sample t-test for delivery method

	T	df	p.value	Lower.CI	Upper.CI	mean_OJ	mean_VC
OJ - VC	1.915268	58	0.0603934	-0.1670064	7.567006	20.66333	16.96333

Growth by dose

For each of the three t-tests we got a very small p-value. None of the confidence intervals contains 0. Therefore, for the confidence level 0.05 we will reject null hypothesis in favour of the alternative that true difference in mean teeth length between dosages is less than 0. In every test we compared smaller dose to larger, and tests showed that mean teeth length increases significantly as the dosage increases.

```
t1 <- t.test(len ~ dose, alternative="less", var.equal=T, data=ToothGrowth[ToothGrowth$dose!=2.0,])
t2 <- t.test(len ~ dose, alternative="less", var.equal=T, data=ToothGrowth[ToothGrowth$dose!=1.0,])
t3 <- t.test(len ~ dose, alternative="less", var.equal=T, data=ToothGrowth[ToothGrowth$dose!=0.5,])
```

Table 6: Two-sample t-test by dosage

	T	df	p.value	Lower.CI	Upper.CI	mean_d1	mean_d2
0.5 - 1.0	-6.476648	38	1.0e-07	-Inf	-6.753344	10.605	19.735
0.5 - 2.0	-11.799046	38	0.0e+00	-Inf	-13.280934	10.605	26.100
1.0 - 2.0	-4.900484	38	9.1e-06	-Inf	-4.175196	19.735	26.100

Growth by delivery method and dosage

In case of 0.5 and 1.0 dosage, p-value was below significance level and confidence intervals did not contain 0. We rejected null hypothesis in favour of alternative that true difference in means is greater than 0. There was enough evidence that, for 0.5 and 1.0 dosage, OJ delivery method yielded higher mean teeth length than VC delivery method. For 2.0 dosage p-value was 0.518 and confidence interval contained 0, so there was no sufficient evidence that there is a significant difference in mean teeth length between delivery methods.

```
t1 <- t.test(len ~ supp, alternative="greater", var.equal=T, data=ToothGrowth[ToothGrowth$dose==0.5,])
t2 <- t.test(len ~ supp, alternative="greater", var.equal=T, data=ToothGrowth[ToothGrowth$dose==1.0,])
t3 <- t.test(len ~ supp, alternative="greater", var.equal=T, data=ToothGrowth[ToothGrowth$dose==2.0,])
```

Table 7: Two-sample t-test by delivery method and dosage

	T	df	p.value	Lower.CI	Upper.CI	mean_d1	mean_d2
OJ - VC, 0.5	3.1697328	18	0.0026518	2.377886	Inf	13.23	7.98
OJ - VC, 1.0	4.0327696	18	0.0003904	3.380140	Inf	22.70	16.77
OJ - VC, 2.0	-0.0461361	18	0.5181451	-3.086866	Inf	26.06	26.14

Conclusion

From the performed analysis, we can conclude that higher dosages of Vitamin C generally improve teeth growth in guinea pigs. For dosages 0.5 and 1.0 there is evidence that delivery through orange juice yielded better results in teeth growth, while for 2.0 dosage delivery method had no effect. Delivery method itself, had no overall effect on teeth growth in guinea pigs.

In this analysis, our significance level was set at 0.05 and we assumed equal variances across the groups. Also, we assume that guinea pigs were randomly assigned certain dosage and delivery method so that groups we compare are independent. Finally, we assumed normality of dependent variable - teeth length.