

Inference Part 2 V2

Libardo Lopez

Friday, August 22, 2014

This is the second part of the statistical inference course project, which contains an analysis of the ToothGrowth data in the R datasets package.

If you like to see the R code, please Go to Github
(https://github.com/Libardo1/Statistical_Inference_Course_Project)

It is well established that Vitamin C plays a role in tooth growth and maintenance.

In this experiment, guinea pigs were given Vitamin C through two methods: a Vitamin C supplement or orange juice. Each method was performed at three dose levels.

The tooth length of each of ten guinea pigs was measured during the six periods of Vitamin C supplementation.

The goal of this experiment is to see how Vitamin C administration affects the steady-state length of guinea pig teeth.

1. Description of the data

The ToothGrowth data explores the effect of Vitamin C on Tooth Growth in Guinea Pigs: The **length** of teeth in each of 10 guinea pigs at each of three **dose** levels of Vitamin C: 0.5, 1, and 2mg; low, med and high, respectively and with each of two delivery methods: orange juice (**OJ**) or ascorbic acid (**VC**).

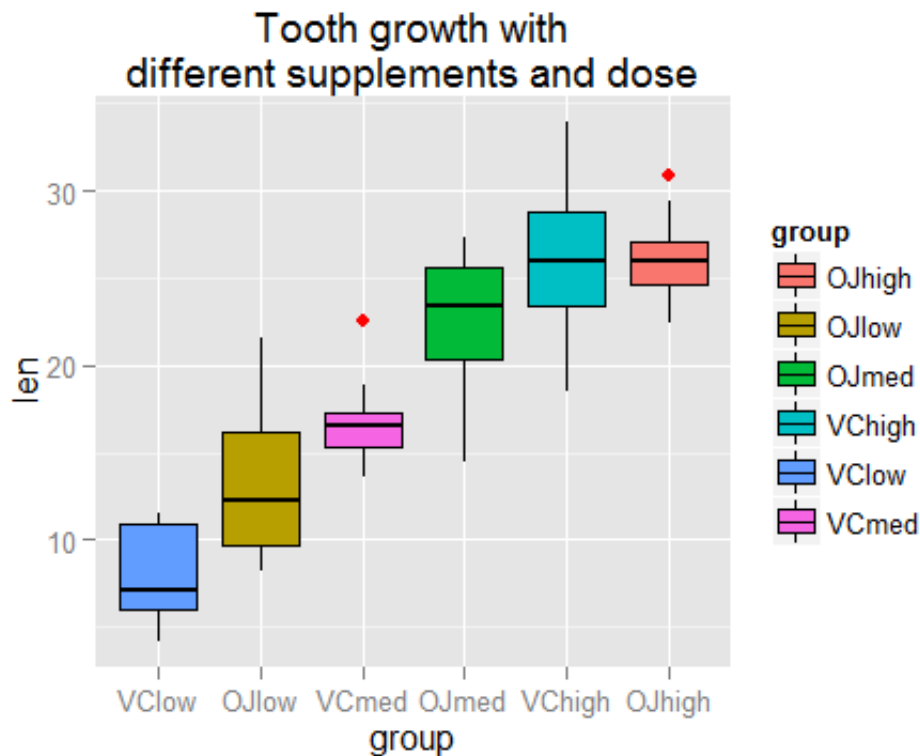
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
## '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 "low","med","high": 1 1 1 1 1 1 1 1 1 1 ...
```

2. Basic Exploratory Data Analysis

To see the effect of the dose and the delivery method in the tooth growth, grouped the info in order to have a boxplot graph.

The plot shows a clear view of the performance of each group:



There appear to be differences between the means.

Variances don't seem to be very similar.

Means don't look significantly different, but similar spread in each group.

Similar spread within each group; means of low/ medium and low/high seem to be different. But there is overlap between med and high.

There seems to be no difference between supp at high dose.

There seems to be a main effect of dose higher dose results in higher tooth length.

There doesn't seem to be much of a main effect of supp there is little difference between the 2 groups overall.

There are a potential outlier in the VC med dose.

There are a potential outlier in OJ high dose.

Since the sample groups seem to be so distinct in their grouping of tooth lengths, it can be **inferred** that there is a statistically significant effect of dosage on tooth length.

2.1 Summary of the data

Means of simple main effects i.e. each level of dose at each level of vitamin type:

```
##          low    med    high
## OJ  13.23  22.70  26.06
## VC   7.98  16.77  26.14
```

Variance:

```
##          low    med    high
## OJ  19.889  15.296   7.049
## VC   7.544   6.327  23.018
```

Summary of length by supplement, and by dose, including their 95% confidence interval.

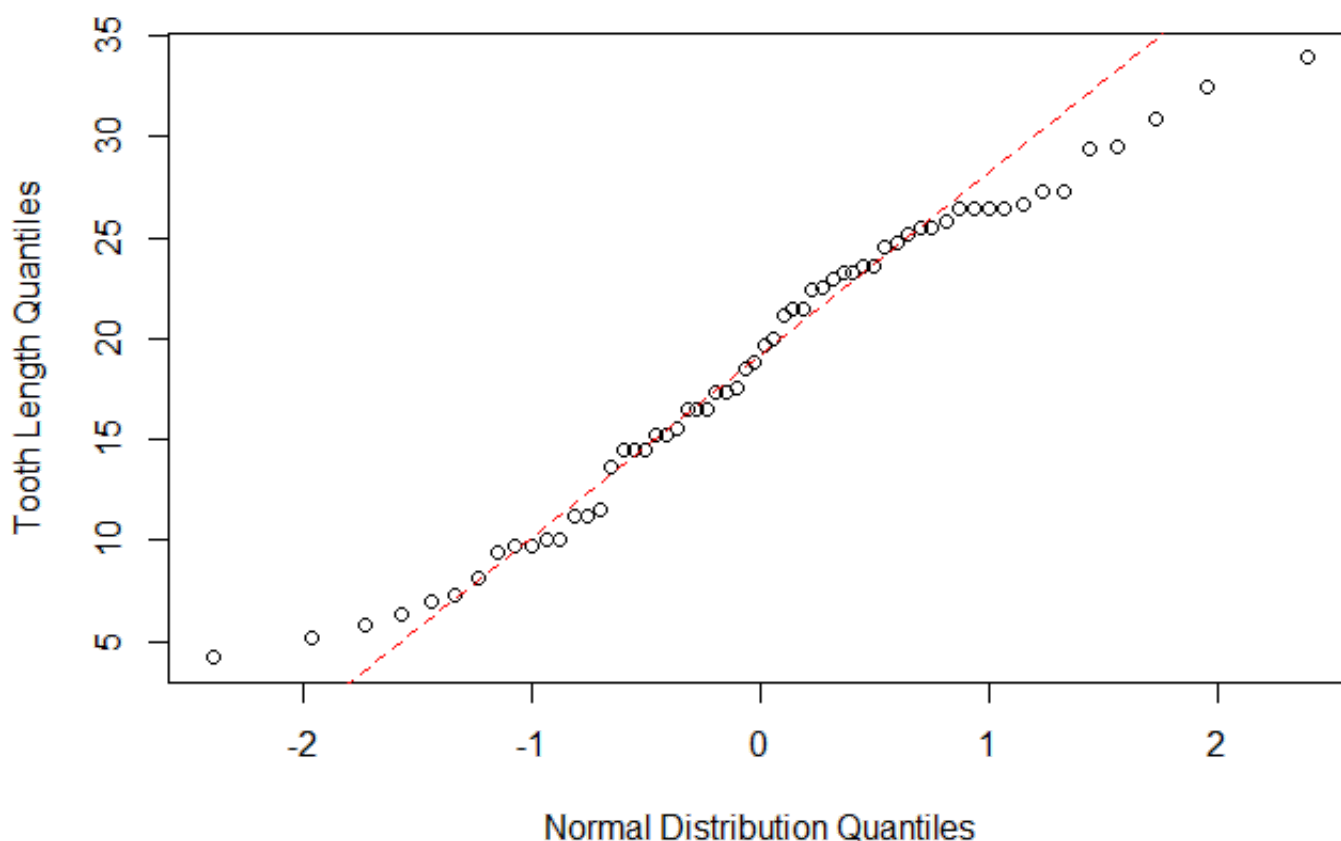
Dose (mg)	Supp.	Avg Length	Std deviation	95%lower limit	95%upper limit
low	VC	26.14	4.7977	23.1663	29.1137
low	OJ	26.06	2.6551	24.4144	27.7056
med	VC	7.98	2.7466	6.2776	9.6824
med	OJ	13.23	4.4597	10.4658	15.9942
high	VC	16.77	2.5153	15.211	18.329
high	OJ	22.7	3.911	20.276	25.124

3. Hypothesis Tests:

First at all, checked the data normalization.

When we are going to apply inferential statistics to the data, it must be normally distributed.

Q-Q Plot, Tooth Length Against Normal Distribution



The upper and lower rangers of the data are commonly where deviation from the Normal Distribution occurs. Here we see some deviation, probably what could be considered an acceptable amount.

To analyze the effect of the dose and the delivery method on the growth length of the tooth, using two-sample t-tests for len vs. supp and len vs. dose.

3.1 Test Assumptions:

No paired observations,
no equal variances accross groups,
interval confidence level is 95% and
the **null hypothesis** to be tested is that there **the differences between the means of the tested groups are 0**.

3.2 Length by Dose Testing

Let define two basic tests for len vs dose: One comparing between 0.5mg and 1mg and the other one comparing between 1mg and 2mg. The two basic tests included both delivery methods (OJ and VC). Then, these two tests were repeated for subsets of OJ only and for VC only, to neutralize the effect of the large variance added due to the difference between the two delivery methods.

The six tests have the structure:

```
t.test(len ~ dose, data = ToothGrowth, subset= (per-test-subset))
```

3.3 Test Results Summary:

Test/Subset	Statistic	DF	P-value	95% Conf. Interval	Mean diff.
low vs med, OJ+VC	-6.4766	37.9864	1.2683×10^{-7}	-11.9838, -6.2762	9.13
med vs high, OJ+VC	-4.9005	37.1011	1.9064×10^{-5}	-8.9965, -3.7335	6.365
low vs med, OJ	-5.0486	17.6983	8.7849×10^{-5}	-13.4156, -5.5244	9.47
med vs high, OJ	-2.2478	15.8424	0.0392	-6.5314, -0.1886	3.36
low vs med, VC	-7.4634	17.8624	6.811×10^{-7}	-11.2657, -6.3143	8.79
med vs high, VC	-5.4698	13.6	9.1556×10^{-5}	-13.0543, -5.6857	9.37

4 Test Conclusions:

All the test results are consistent in rejecting the null hypothesis, and concluding that **there is a very high probability that an increased dose would result with increased tooth length**.