

# Statistical Inference

*Dorian Kcira*

*10/24/2015*

## Tooth Growth Data

The dataset `ToothGrowth` is loaded and a simple exploratory data analysis is done. From the help (`?ToothGrowth`): “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).”

We table the dose and supplement type. We also plot the dependency of the tooth length on the dose for the two delivery methods (OJ, VC).

```
library(lattice)
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 ...
```

```
table(ToothGrowth$dose)
```

```
##
## 0.5    1    2
## 20   20   20
```

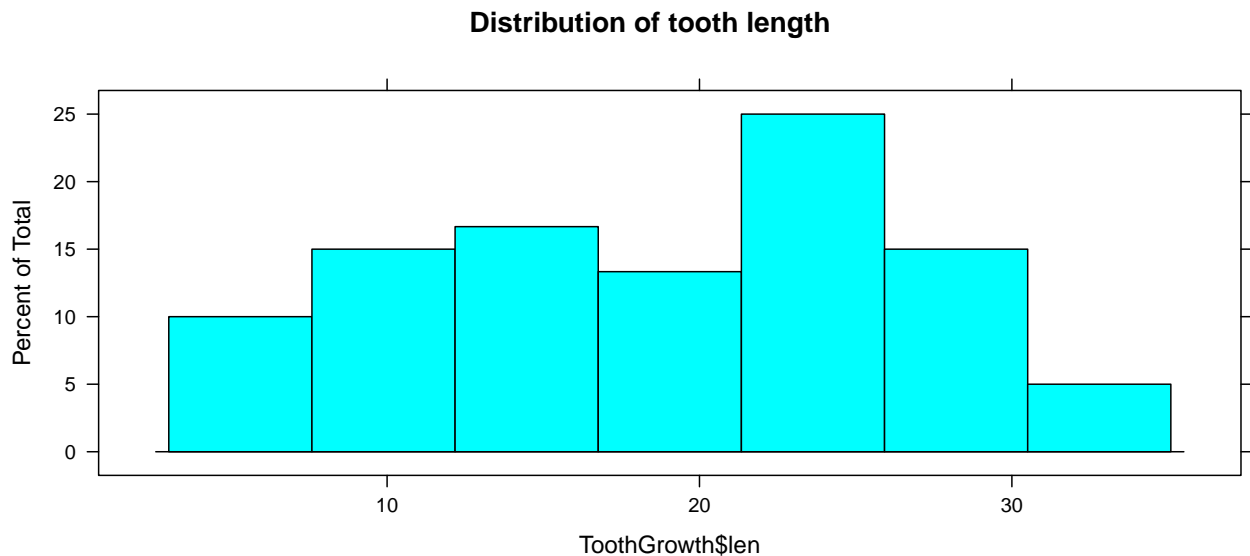
```
table(ToothGrowth$supp)
```

```
##
## OJ VC
## 30 30
```

```
with(data=ToothGrowth,
      xyplot(len~dose|supp,
              main="Tooth length vs dose for two delivery methods",
              ylab="Tooth length", xlab="dose")
)
```



```
histogram(ToothGrowth$len, main="Distribution of tooth length")
```



```
#histogram(~fastest/sex,data=m111survey,
#  type="density",
#  xlab="speed (mph)",
#  main="Fastest Speed Ever Driven,\nby Sex",
#  breaks=seq(from=60,to=190,by=10))
#with(data=ToothGrowth,
#  xyplot(len~dose/supp,
#    main="Tooth length vs dose for two delivery methods",
#    ylab="Tooth length", xlab="dose")
#)
```

We can see that there are large variations of the tooth length for each dose, but there is a clear tendency of teeth being longer for larger doses of vitamin C. Also, visually, the orange juice (OJ) method seems to do better than the ascorbic acid (VC) method. We need to confirm this analytically.

## Hypotheses

The problem formulation does not clarify whether the measurements were paired, therefore we will use the relaxed assumption that they are independent. Our first hypothesis will be that the tooth length increases with the dose.

## Varia

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package. Load the ToothGrowth data and perform some basic exploratory data analyses. Provide a basic summary of the data. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering) State your conclusions and the assumptions needed for your conclusions. Some criteria that you will be evaluated on: Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data? Did the student perform some relevant confidence intervals and/or tests? Were the results of the tests and/or intervals interpreted in the context of the problem correctly? Did the student describe the assumptions needed for their conclusions?