Statistical Inference : Project

20th April 2015

Part 2 - Analyse ToothGrowth Data

This report analyzes the ToothGrowth data in the R datasets package.

Inspection of the data identifies the size of the dataset, names of variables, type of data and which data fields are categorical.

Inpsection of the data tells us
1. 60 observations are recorded.
+ "len" (Length of tooth growth)
+ "sup" (Type of supplement used)
+ "dose" (size of dosage)
2. there are two types of supplement used.
+ "OJ" = Orange Juice
+ "VC" = Ascorbic Acid
3. There are three dosage levels used. (0.5, 1, 2)

Plot 1 shows three levels of dosage, statistical analysis is needed to show how dosage is related to growth rate and the confidence levels of any proposed conclusions.

Plot 2 (refer Appendix) shows us the following.
* at low (0.5mg) and medium (1mg) doses more tooth growth is observed with Orange Juice than with ascorbic acid.
* at high (2mg) doses the variance in tooth grown observed is higher with ascorbic acid than with Orange Juice.
* In both cases higher doses indicate higher tooth growth.
* outliers exist for the Orange Juice @ 2mg and for Ascorbic Acid @ 1mg.
* Ascorbic Acid results in more tooth growth than Orange Juice for the same dosage (at 0.5mg & 1.0mg).

Testing hyposis using t.test at the default 95% confidence level.
H null : There is no difference in tooth growth between dosage = 0.5mg & 1.0mg

```
temp1 <- subset(ToothGrowth, dose==0.5 | dose ==1)
ts <- t.test(temp1$len ~ temp1$dose)
ts$conf.int[1:2]
```

```
## [1] -11.983781  -6.276219
```

Confidence interval does not include zero. The Alternate Hypothesis is true : The difference in means is not equal to 0.

H null : There is no difference in tooth growth between dosage = 0.5mg & 2.0mg.

```
temp1 <- subset(ToothGrowth, dose==0.5 | dose ==2)
ts <- t.test(temp1$len ~ temp1$dose)
ts$conf.int[1:2]
```

```
## [1] -18.15617 -12.83383
```

Confidence interval does not include zero.
The Alternate Hypothesis is true : The difference in means is not equal to 0.

H null : There is no difference in tooth growth between dosage = 1.0mg & 2.0mg.

```
temp1 <- subset(ToothGrowth, dose==0.5 | dose ==2)
ts <- t.test(temp1$len ~ temp1$dose)
ts$conf.int[1:2]
```

```
## [1] -18.15617 -12.83383
```

Confidence interval does not include zero. The Alternate Hypothesis is true : The difference in means is not equal to 0.
From the Hypothesis tests above we can conclude Tooth Growth increases with dosage at all dosage levels.

Now test the hypothesis that Orange Juice results in different tooth growth than Ascorbic Acid.
H0 : There is no difference in tooth growth between Orange Juice and Ascorbic Acid.

```
t.test(ToothGrowth$len ~ ToothGrowth$supp)$conf.int[1:2]
```

```
## [1] -0.1710156  7.5710156
```

Confidence Interval does include zero. Failed to reject the null hypothesis.
We conclude that across all dosage rates, there is no significant difference in tooth growth rates from Ascorbic Acid or Orange Juice.

Hnull = There is no difference in tooth growth rates at 0.5mg for different supplements.

```
temp1 <- subset(ToothGrowth, dose==0.5)
t.test(temp1$len ~ temp1$supp)$conf.int[1:2]
```

```
## [1] 1.719057 8.780943
```

Confidence interval does not include zero.
Alternate hypothesis (difference in means is non zero for different supplements @ 0.5mg) is True.

Hnull = There is no difference in tooth growth rates at 1.0mg for different supplements.

```
temp1 <- subset(ToothGrowth, dose==1.0)
t.test(temp1$len ~ temp1$supp)$conf.int[1:2]
```

```
## [1] 2.802148 9.057852
```

Confidence interval does not include zero.
Alternate hypothesis (difference in means is non zero for different supplements @ 1.0mg) is True.

Hnull = There is no difference in tooth growth rates at 2.0mg for different supplements.

```
temp1 <- subset(ToothGrowth, dose==2.0)
t.test(temp1$len ~ temp1$supp)$conf.int[1:2]
```

```
## [1] -3.79807  3.63807
```

Confidence interval *does* include zero.
Null hypothesis (difference in means is zero for different supplements @ 1.0mg) is True.
*****

Conclusions

Statistical analysis of the data has proven the following with 95% confidence.
* Tooth growth increases with dosage regardless of supplement type.
* When considered across the range of dosage rates, there is no difference in tooth grown rates between Ascorbic Acid or Orange Juice.
* At 0.5mg & 1.0mg dosage, there is a difference in tooth growth rates for Ascorbic Acid and Orange Juice.
* At 2.0mg dosage there is no difference in the tooth growth rates for Ascorbic Acid and Orange Juice.

Appendix A

Plots

Plot 1

```
## plot using panels to produce a plot for each supplement type.
g <- ggplot(ToothGrowth, aes(dose, len))
g + geom_point() + facet_grid(. ~ supp)
```
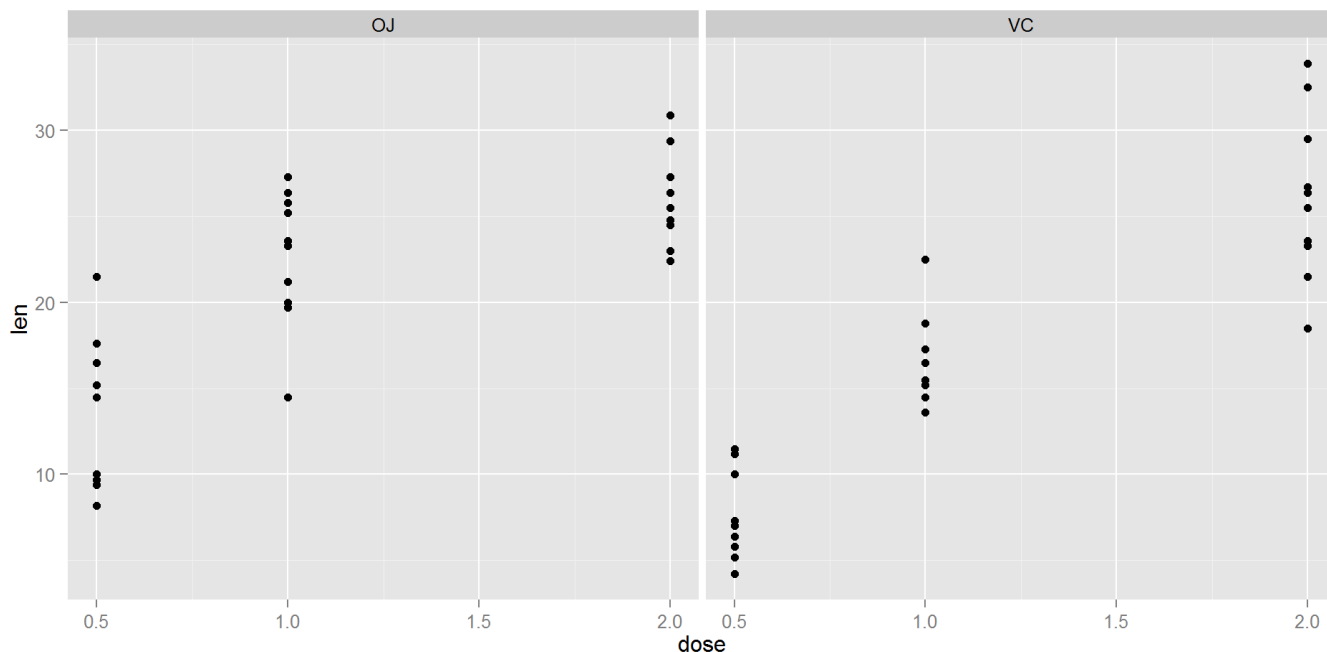
Plot 2

```
## Plotting using boxplot
## function to label the plot facets for readability.
facet_names <- list('OJ'="Orange Juice", 'VC'="Ascorbic Acid")
facet_labeller <- function(variable,value){ return(facet_names[value]) }
## NB : factor of dose required because ggplot requires factor (dose is numeric)
g <- ggplot(ToothGrowth, aes(factor(dose), len, fill = factor(dose))) +
scale_fill_discrete(name="Dosage (mg)") +
geom_boxplot(outlier.colour = "red",outlier.shape = 23, outlier.size = 5) +
geom_jitter(position=position_jitter(w=0.2)) +
facet_grid(.~supp, labeller=facet_labeller) +
ylab("Tooth Growth") +
xlab("Dose (mg)")+
ggtitle("Tooth Growth vs Supplement Doses")
## fill = factor(dose) : fills in boxplot with colour coding, easy to read graph
## geom_jitter - jitters points so we can see overlapping points. w=0.2 sets width of j
itter.
## outlier - set properties so we can see points outside the values shown by the boxplo
t
## facet_grid - plot separate panels for each value of supp. ("OJ" &" "VC")
## scale_fill_discrete : sets legend title.
g
```

Tooth Growth vs Supplement Doses
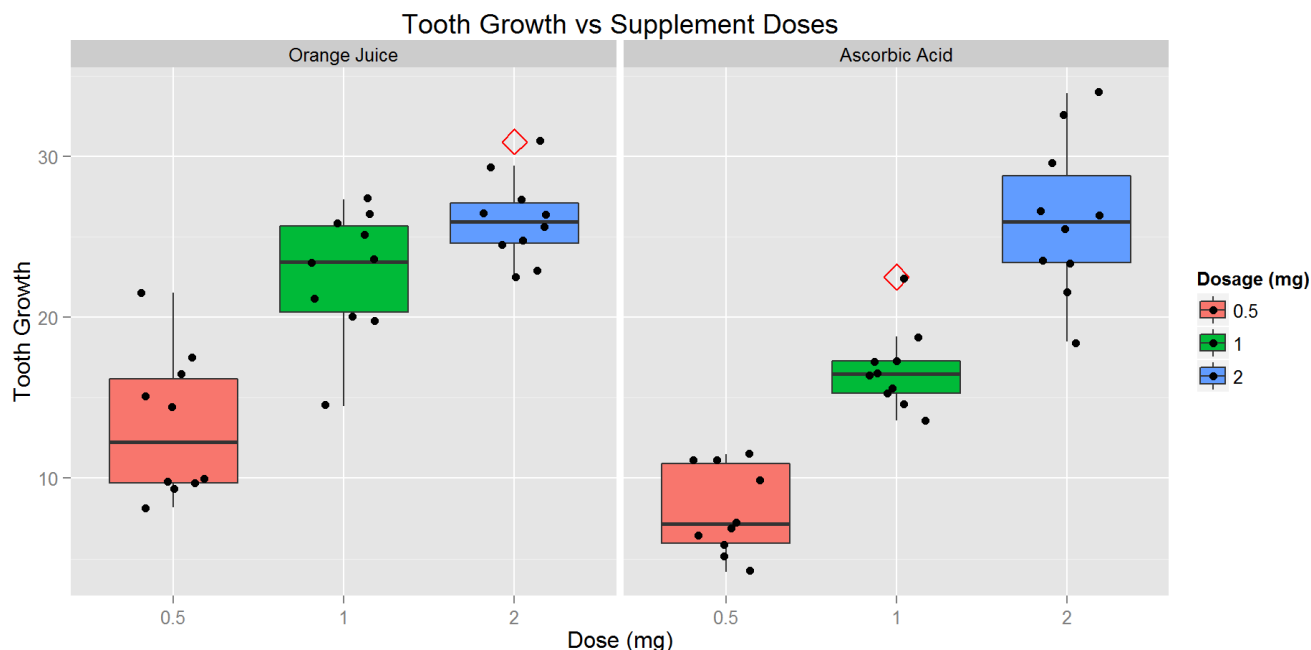
Assumptions

The Analysis are based on the following assumptions.
* The variances of each test group are different. (ie: var.equal = FALSE)
* The test subjects are not related, ie: each test subject was tested with either Orange Juice or Ascorbic Acid.
* The test subjects were randomly selected. There are no properties of the test subjects causing bias in results.
* The test subjects are independent and identically distributed.
https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html)

Description : The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Ascorbic Acid (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Git url : https://github.com/aspiringguru/Statistical-Inference-Project.git
(https://github.com/aspiringguru/Statistical-Inference-Project.git)