Statistical Inference Course Project - 2

Author: Krupa Rajendran

The **ToothGrowth** data consists of measurements of the mean length of the odontoblast cells harvested from the incisor teeth of a population of 60 guinea pigs. These animals were divided into 6 groups of 10 and consistently fed a diet with one of 6 Vitamin C supplement regimes for a period of 42 days. The Vitamin C was administered either in the form of **Orange Juice (OJ)** or chemically pure **Vitamin C in aqueous solution (VC)**. Each animal received the same daily dosage of Vitamin C (either 0.5, 1.0 or 2.0 milligrams) consistently. Since each combination of supplement type and dosage was given to 10 randomly selected animals this required a total of 60 animals for the study. After 42 days, the animals were euthanized, their incisor teeth were harvested and subject to analysis via optical microscopy to determine the length (in microns) of the odontoblast cells.

In this report, the **ToothGrowth** data in the R datasets package is analysed and the initial analysis is as follows:

```
> # Load the dataset ToothGroth
> library(datasets)
> data(ToothGrowth)
>
> # Check the dimension and fields of dataset 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 ...
>
> # check the summary of the dataset ToothGrowth
> summary(ToothGrowth)
      len        supp         dose
 Min.   : 4.20   OJ:30   Min.   :0.500
 1st Qu.:13.07   VC:30   1st Qu.:0.500
 Median :19.25           Median :1.000
 Mean   :18.81           Mean   :1.167
 3rd Qu.:25.27           3rd Qu.:2.000
 Max.   :33.90           Max.   :2.000
>
```

ToothGrowth Dataset has 60 observations and each observation has three values

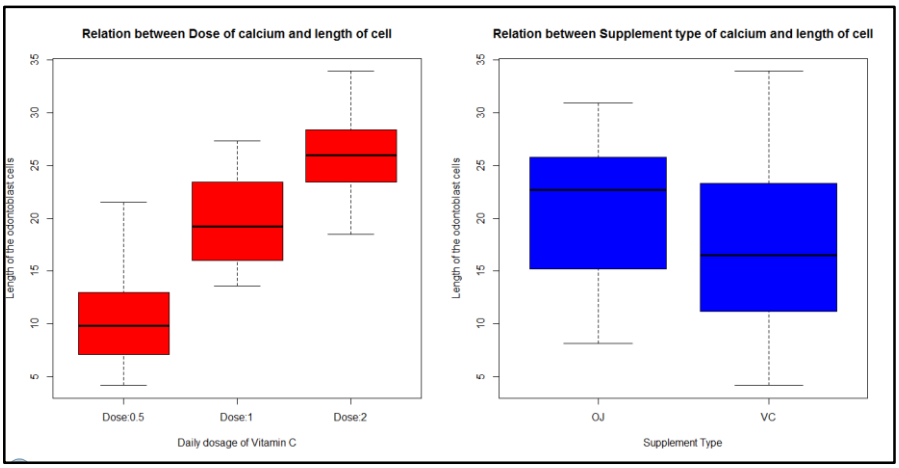
a. **len** : Numeric variable indicating mean length of odontoblasts (microns)
b. **supp** : Factor variable indicating supplement type and having two values “OJ” and “VC”
c. **dose**: Numeric variable indicating Vitamin C dosage (milligrams/day) and takes only three values 0.5, 1 and 2
d. There are 30 observation with supp = “OJ” and 30 observations with sup = “VC”
e. If we change the **dose** variable to Factor, then we see that there are 10 observations each for the three values 0.5, 1 and 2 for each type of supplement

```
> ToothGrowth$dose <-factor(ToothGrowth$dose,levels=c(0.5,1,2),labels=c("Dose:0.5","Dose:1","Dose:2"))
> table(ToothGrowth$dose,ToothGrowth$supp)

          OJ VC
  Dose:0.5 10 10
  Dose:1   10 10
  Dose:2   10 10
>
```

Plotting the Length of the odontoblast cells against Daily dosage of Vitamin C shows that **higher the dose of calcium has a positive influence on the length of the odontoblast cells**.

Similarly, Plotting the Length of the odontoblast cells against Supplement type shows that Vitamin C was administered in the form of **Orange Juice (OJ) has a slightly higher influence on the length of the odontoblast cells compared to Vitamin C in aqueous solution (VC) but this influence is not significant.**

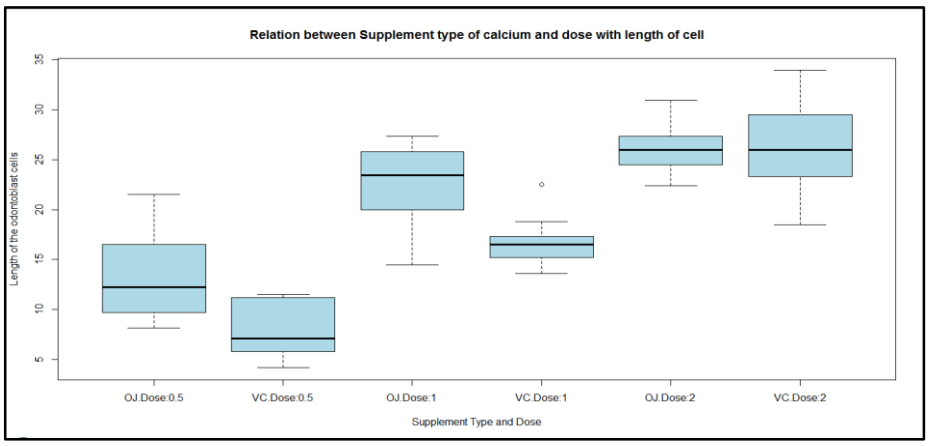


The R code for the above graph is as follows:

```
> par(mfrow=c(1,2))
>
> boxplot(len ~ dose,ToothGrowth,
+ col="red",
+         main="Relation between Dose of calcium and length of cell",
+         xlab="Daily dosage of Vitamin C",
+         ylab = "Length of the odontoblast cells")
>
> boxplot(len ~ supp,ToothGrowth,
+ col="blue",
+         main="Relation between Supplement type of calcium and length of cell",
+         xlab="Supplement Type",
+         ylab = "Length of the odontoblast cells")
>
```

Plotting the Length of the odontoblast cells against the combination of Supplement type and Daily dosage of Vitamin C shows that **Vitamin C (VC) for dosage of 2 has the highest positive influence on the length of the odontoblast cells** compared to all other combinations of supplement type and dose.

The R code of the graph is as follows:




```
> boxplot(len ~ supp*dose,ToothGrowth,
+ col="light blue",
+         main="Relation between Supplement type of calcium and dose with length of cell",
+         xlab="Supplement Type and Dose",
+         ylab = "Length of the odontoblast cells")
```

Confidence intervals and Hypothesis tests are used to compare tooth growth by supplement type and dose.

Let Null Hypothesis = There is no difference between the length of the tooth based on the supplement type

```
> # Check for length differences due to different supplement type (OJ, VC)
> t.test(len ~ supp, data = ToothGrowth)

        Welch Two Sample t-test

data:  len by supp
t = 1.9153, df = 55.309, p-value = 0.06063
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1710156  7.5710156
sample estimates:
mean in group OJ mean in group VC
        20.66333         16.96333
```

But from the above results, we see that the p value is very low (0.06) and Alternate hypothesis indicates that these is difference in the means of the length of the tooth growth due to the two supplement types. But zero is not included in the 95% confidence interval indicating that the difference is not significant.

Now let us test the impact of dose on the growth of the Tooth. Since there are three different levels of doses, we will have to create three subsets of data before performing the T tests.

```
> library(datasets)
> data(ToothGrowth)
> # Create three subgroups as per dose level pairs, to help with the analysis
> Doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
> Doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
> Doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
>
```

Case 1: Let us compare the impact of growth for daily dosage of 0.5 and 1.0 and let Null Hypothesis (H0) = There is no difference between the length of the tooth based on the dosage.

```
> # Check for lenth differences due to different dose levels (0.5, 1.0)
> t.test(len ~ dose, data = Doses_0.5_1.0)

        Welch Two Sample t-test

data:  len by dose
t = -6.4766, df = 37.986, p-value = 1.268e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.983781  -6.276219
sample estimates:
mean in group 0.5   mean in group 1
           10.605            19.735

>
```

Since the p value is very small (1.268e-07) and zero is not included in the 95% confidence interval, the Null Hypothesis (H0) is invalid and T test shows that higher dosage (1.0) positively impacts growth.

Case 2: Let us compare the impact of growth for daily dosage of 0.5 and 2.0 and let Null Hypothesis (H0) = There is no difference between the length of the tooth based on the dosage.

```
> # Check for lenth differences due to different dose levels (0.5, 2.0)
> t.test(len ~ dose, data = Doses_0.5_2.0)

        Welch Two Sample t-test

data:  len by dose
t = -11.799, df = 36.883, p-value = 4.398e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -18.15617 -12.83383
sample estimates:
mean in group 0.5   mean in group 2
           10.605            26.100

>
```

Since the p value is very small (4.398e-14) and zero is not included in the 95% confidence interval, the Null Hypothesis (H0) is invalid and T test shows that higher dosage (2.0) positively impacts growth

Case 2: Let us compare the impact of growth for daily dosage of 1.0 and 2.0 and let Null Hypothesis (H0) = There is no difference between the length of the tooth based on the dosage.

```
> # Check for lenth differences due to different dose levels (1.0, 2.0)
> t.test(len ~ dose, data = Doses_1.0_2.0)

        Welch Two Sample t-test

data:  len by dose
t = -4.9005, df = 37.101, p-value = 1.906e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.996481 -3.733519
sample estimates:
mean in group 1 mean in group 2
         19.735          26.100
```

Since the p value is very small (1.906e-05) and zero is not included in the 95% confidence interval, the Null Hypothesis (H0) is invalid and T test shows that higher dosage (2.0) positively impacts growth.

Conclusion:

1. **Supplement type (OJ, VC) does not have significant effect on tooth growth of odontoblast cells of tooth.**
2. **Increasing the dose level (0.5, 1.0 and 2.0) increases the growth of odontoblast cells of tooth.**

Assumptions

1. Members of the sample population represent of the entire population of guinea pigs.

2. While performing the T tests, the variances are assumed to be different for the two groups being compared.