Statistical Inferences Course Project Part2

RAVI KUMAR YADAV

Overview:

Report includes some basic exploratory data analyses of ToothGrowth data.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

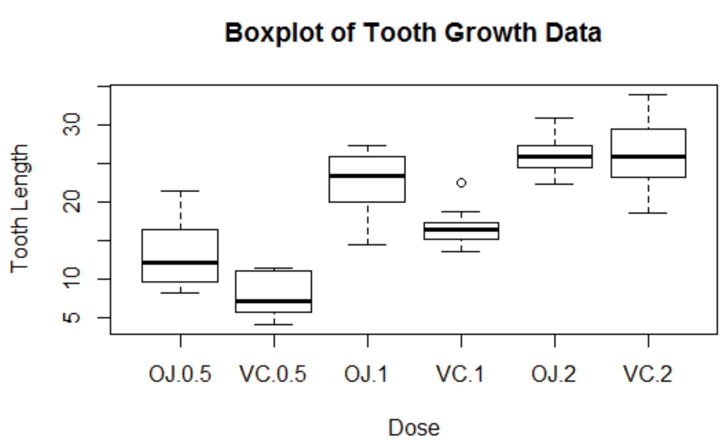
As per the help file of this dataset, the response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
> 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 ...
```

2. Basic Summary of Data

```
# Summary statics for all variables
> 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  
```

Box plot for ToothGrowth data:



Plot indicates that on an average the length of the tooth increases with the increase in dose.

3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Test for the delivery type at different dose level:

```
> ToothGrowth.doses_0.5 <- subset (ToothGrowth, dose %in% c(0.5))
> t.test(len ~ supp, data = ToothGrowth.doses_0.5)

        Welch Two Sample t-test

data:  len by supp
t = 3.1697, df = 14.969, p-value = 0.006359
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.719057 8.780943
sample estimates:
mean in group OJ mean in group VC
           13.23             7.98
```

For dose 0.5, the p-value of OJ in comparison to VC is 0.0064. Since it is less than 0.05 (strong presumption against null hypothesis), it means that there is a difference between both methods.

```
> ToothGrowth.doses_1.0 <- subset (ToothGrowth, dose %in% c(1.0))
> t.test(len ~ supp, data = ToothGrowth.doses_1.0)

        Welch Two Sample t-test

data:  len by supp
t = 4.0328, df = 15.358, p-value = 0.001038
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.802148 9.057852
sample estimates:
mean in group OJ mean in group VC
           22.70            16.77
```

For dose 1.0, the p-value of OJ in comparison to VC is 0.001. Since it is less than 0.05 (strong presumption against null hypothesis), it means that there is a difference between both methods.

```
> ToothGrowth.doses_2.0 <- subset (ToothGrowth, dose %in% c(2.0))
> t.test(len ~ supp, data = ToothGrowth.doses_2.0)

        Welch Two Sample t-test

data:  len by supp
t = -0.0461, df = 14.04, p-value = 0.9639
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.79807  3.63807
sample estimates:
mean in group OJ mean in group VC
           26.06            26.14
```

For dose 2.0, the p-value of OJ in comparison to VC is 0.963. Since it is greater than 0.05 (low presumption against null hypothesis), it means that there is a no that much of a difference between both the methods.

Test for the group differences due to different dose levels:

```
# first create three sub-groups as per dose level pairs
> ToothGrowth.doses_0.5_1.0 <- subset (ToothGrowth, dose %in% c(0.5, 1.0))
> ToothGrowth.doses_0.5_2.0 <- subset (ToothGrowth, dose %in% c(0.5, 2.0))
> ToothGrowth.doses_1.0_2.0 <- subset (ToothGrowth, dose %in% c(1.0, 2.0))
```

Check for group differences due to different dose levels (0.5, 1.0)

```
> t.test(len ~ dose, data = ToothGrowth.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 
```

Check for group differences due to different dose levels (0.5, 2.0)

```
> t.test(len ~ dose, data = ToothGrowth.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 
```

Check for group differences due to different dose levels (1.0, 2.0)

```
> t.test(len ~ dose, data = ToothGrowth.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 
```

For all the three dose level pairs, the p-value is less than 0.05, and the confidence interval does not contain zero. The mean tooth length increases on raising the dose level. This indicates that we can reject the null hypothesis, and establish that increasing the dose level leads to an increase in tooth length.

4. conclusions and the assumptions needed for the conclusions

Conclusion:

1. At lower doze level (0.5 and 1.0) there is difference between delivery types. However at dose level 2.0 delivery type has no effect on tooth growth rate.
2. Tooth growth rate increases with increase in dose level.

Assumptions:

3. Members of the sample population, i.e. the 60 guinea pigs, are representative of the entire population of guinea pigs. This assumption allows us to generalize the results.
4. For the t-tests, the variances are assumed to be different for the two groups being compared.