

Statistical Inference Peer Project 2

Mohamamd Yaqoob

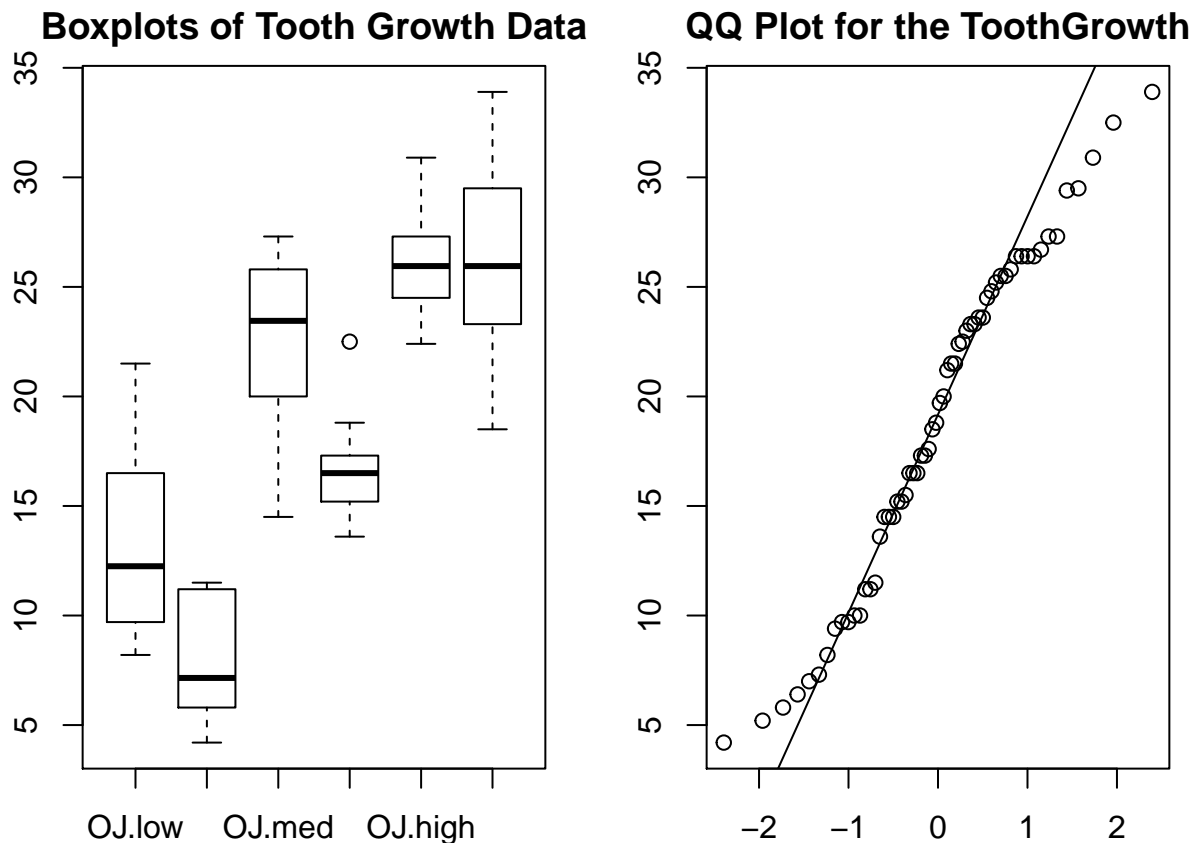
Required Analyze the ToothGrowth data in the R datasets package.

Approach First we will load/transform the data so that it is close to being normally distributed. We will then find the confidence interval for the mean and then perform a significance test to evaluate whether or not the data is away from a fixed standard. Finally, we will find the power of the test to detect a fixed difference from that standard. We will assume that a confidence level of 95% is used throughout.

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

Basic exploratory data analyses (BEDA) There are no rownames for the data; each line is numbered starting from 1. There are three columns in the dataframe: “len,” “supp,” and “dose.”

The data set contains 30 observations for each of the supplements-Orange Juice (OJ) and Vitamin C (VC). There are no Null or NA values. I replaced the dose levels with low, med, high for 0.5, 1.0, 2.0 milligrams. We subset data into two dataframes by supplement type. For each row: **len**: numeric variable giving the length of teeth. **supp**: categorical variable (factor) representing the supplement: “VC” (vitamin C) or “OJ” (orange juice). **dose** is a numeric variable representing the amount of supplement in milligrams.



From the box-plots, two things are apparent. First, the data looks relatively normal, there are no signs of severe skewness or the presence of outliers in any. Second, it looks pretty clear that increases in the dosage of vitamin C and Orange Juice given to the test subjects leads to increases in tooth length. There is one point outlier in medium dose group in VC supplement, rest of the data is arranged in distinct groups. QQ graph shows ToothGrowth is approximately normally distributed with some deviation at the especially near the extreme ends of the normal distribution quantiles.

2: Provide a basic summary of the data.

The Mean of ToothGrowth with supp OJ = 20.66 and the Mean of ToothGrowth with supp VC = 16.96. The sd of ToothGrowth with supp OJ = 6.606, and the sd of ToothGrowth with supp VC = 8.266. The length of dataframe with supp OJ = 30, and the length of dataframe with supp VC = 30. The Standard Error for OJ and VC= 1.932, and the 95% CI for the Tooth Growth is between = 16.88 20.75. The Mean for TothGrowth for OJ and VC= 18.81.

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

(Use the techniques from class even if there's other approaches worth considering)

The 95% CI for the Tooth Growth is between = 16.88 20.75. The margin of error is found based on a 95% confidence level is 1.976028. Hence the 95% confidence interval for the tooth growth is between 16.88 and 20.75.

```
##
## Welch Two Sample t-test
##
## data:  tgOJ$len and tgVC$len
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.171  7.571
## sample estimates:
## mean of x mean of y
##      20.66      16.96
```

Ho:meu_OJ = meu_VC Ha:meu_OJ != Meu_VC our hypothesis is that Orange Juice and Vitamin see have the similar effect on the totth growth. We got t-vlaue of 1.915268 and p of 0.0673165, we reject the Null hypothesis. Thus there is different effect with respect to the supplements. Next we will evaluate, lower dose levels effect is similar to higher dose levels by evlauting the Null Hypothesis: H0: meu_hd = mue_ld Ha: meu_hd != mue_ld We got t-vlaue of 3.87 and p-value of 0.9999 for higher doses and t-value 5.55 and p-value 0.9999, we reject Null Hypothesis i.e. lower doses have more effect on the tooth growth than higher doeses. From graphs we can see that effect at higher dose (2.0) is almost same for the OJ and VC.

4: State your conclusions and the assumptions needed for your conclusions.

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supp       1     205      205    15.57 0.00023 ***
## dose       2    2426     1213    92.00 < 2e-16 ***
## supp:dose   2     108       54     4.11 0.02186 *
## Residuals  54     712       13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the statistical analysis and graphs it is pretty clear that increases in the dosage of vitamin C given to the subjects leads to increases in tooth length and the same is true with orange juice. We can see in the plots and , that the tooth length increases as dosage increases, and at two low dosages, OJ works better than VC with similar effects, but at the high dosage, OJ and VC shows no difference. The effect of supplement does depend on levels of dose, therefore, there is a supp-dose interactions. Although, there is no placebo data hence we can't realy determine the real advantage of each supplement with respect no supplement, but still the graphs and statistics indicate there is significant effect especially at lowe doses with vitamin c and orange juice on tooth growth of the subjects.