R ToothGrowth Dataset Exploratory Data Analysis and Inferential Study

This report investigates the response of the length of 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) (from ToothGrowth R help).

Exploratory Data Analysis and Data Summary

The ToothGrowth data set consists of the following variables and types:

```
## '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 ...
```

This data set has 0 incomplete cases. Additionally, each variable's summary statistics are

```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

The unique values of Vitamin C dosage are 0.5, 1, 2, so this variable may be considered a factor variable.

Hypothesis Tests Around Variation of Mean Tooth Length

Here, we will investigate several hypotheses of the effects of Vitamin C dosage and its devlivery method on mean tooth length:

1. No differnce between delivery by orange juice or absorbic acid overall.
2. No difference based on dosage level, regardless of delivery method.
3. No difference based on delivery method for each dosage.

Question 1

The summary statistics are as follows:

```
##   supp mean.len var.len num.obs
## 1   OJ    20.66   43.63      30
## 2   VC    16.96   68.33      30
```

Using a two-sided t-test with $\alpha = 0.05$, we will reject the null hypothesis of the means being equal if the computed t-value is greater than 2.0017. For these data, we get $t = 1.9153$. Comparing, we fail to reject the null hypothesis.

Question 2

The summary statistics are as follows:

```
##   dose mean.len var.len num.obs
## 1  0.5    10.61   20.25      20
## 2  1.0    19.73   19.50      20
## 3  2.0    26.10   14.24      20
```

Using the same analysis as for Question 1, but for each of the three mean differences, we get the following two-sided t-scores:

```
##   mean.diff abs.t.score t.critical
## 1       1-2       6.477      2.024
## 2       2-3       4.900      2.024
## 3       1-3      11.799      2.024
```

Snce $|t| > t_{critical}$ for all three scenarios, we reject the null hypotheses.

Question 3

The summary statistics are as follows:

```
##   supp dose mean.len var.len num.obs
## 1   OJ  0.5    13.23  19.889      10
## 2   OJ  1.0    22.70  15.296      10
## 3   OJ  2.0    26.06   7.049      10
## 4   VC  0.5     7.98   7.544      10
## 5   VC  1.0    16.77   6.327      10
## 6   VC  2.0    26.14  23.018      10
```

Again, repeating the same t-tests for the three mean differences, we get the following two-sided t-scores:

```
##   dose abs.t.score t.critical
## 1  0.5     3.16973      2.101
## 2  1.0     4.03277      2.101
## 3  2.0     0.04614      2.101
```

From these results, we fail to reject the null hypothesis only when the dosage is 2.0.

Conclusion

For the t-tests conducted above, firstly, we see that we don't have enough data to reject the claim that the mean tooth length differs based on delivery method overall. Secondly, we see that the dosage level does make a difference in mean tooth length. Finally, we see that the mean tooth length is statistically the same only when the dosage is 2.0 when disregarding the delivery method.