

# Statistical Inference Course Project - part 2

Felix G Lopez

2024-08-01

## Part 2: Basic Inferential Data Analysis Instructions

This dataset involves the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods: orange juice or ascorbic acid (coded as VC). In this part, I analyze this dataset to understand the effect of dose and supplements on tooth growth. (Source: help file of the dataset)

Loading and providing a basic summary of the dataset.

```
library(ggplot2)
library(datasets)
data("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 ...
```

```
head(ToothGrowth)
```

```
##      len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
summary(ToothGrowth)
```

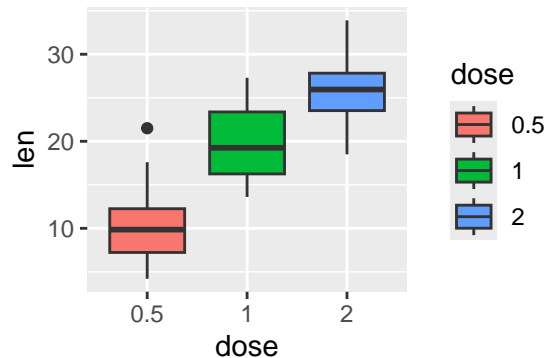
```
##      len      supp      dose
## Min.   : 4.20   OJ:30   Min.    :0.500
## 1st Qu.:13.07   VC:30   1st Qu.:0.500
## Median :19.25                Median :1.000
## Mean   :18.81                Mean    :1.167
## 3rd Qu.:25.27                3rd Qu.:2.000
## Max.   :33.90                Max.    :2.000
```

Using boxplots to check tooth growth as a function of dose.

```

ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))

```



It seems dose has a strong effect in tooth growth. Let's, then, use confidence intervals and hypothesis test to compare tooth growth by dose. To do so I will subset doses in three groups.

```

#Create three sub-groups per dose level pairs in order to check for group differences.
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 of (0.5, 1.0). Assume unequal variances between the two groups.

```

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 between group 0.5 and group 1 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 of (0.5, 2.0). Assume unequal variances between the two groups.

```

ttest1 <- t.test(len ~ dose, data = ToothGrowth.doses_0.5_2.0)
print(ttest1, options=999)

```

```

##
##  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 between group 0.5 and group 2 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 of (1.0, 2.0). Assume unequal variances between the two groups.

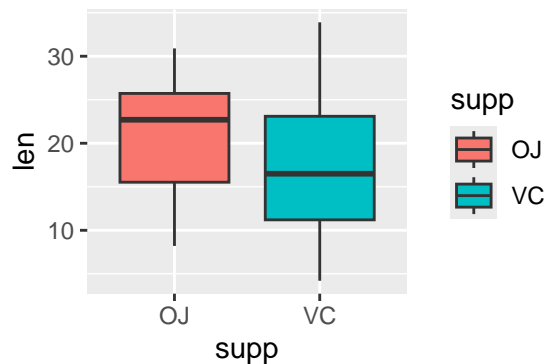
```
t.test(len ~ dose, data = ToothGrowth.doses_1.0_2.0)
```

```
##
## 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 between group 1 and group 2 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 three t-tests, the p-values are significantly less than 0.05, and the confidence intervals do not include zero. Therefore, we reject the null hypothesis and conclude that increasing the dose level results in greater tooth length.

Now lets see tooth growth as a function of supplement type.

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
```



Using confidence intervals and hypothesis test to compare the effect of supp on length. Assume unequal variances between the two groups.

```
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 between group OJ and group VC 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
```

As we see above, that the p-value is equal to 0.06 and the confidence interval contains zero. Thus, we fail to reject the null hypothesis that the different supplement types have no effect on tooth length.

## Stating my conclusions, based on my assumptions

Supplement type has no effect on tooth growth. Increasing the dose level leads to increased tooth growth. The experiment involved randomly assigning guinea pigs to different dose levels and supplement types to control for potential confounding variables that could influence the results. The 60 guinea pigs in the sample are considered representative of the entire population of guinea pigs, allowing for the generalization of the findings. For the t-tests, it is assumed that the variances of the two groups being compared are different. This assumption is less strict than assuming equal variances between the groups.

Thank you!