Statistical Inference Course Project Part 2~~1~~

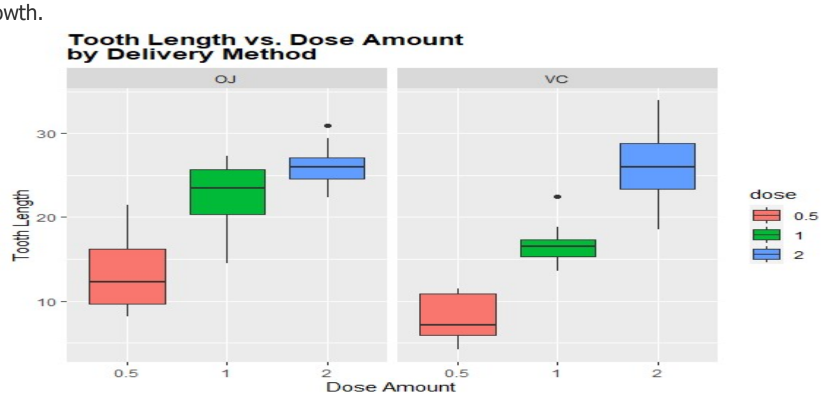
1- Introduction A few EDA methods will be carried out as well as applying two-sample t-test to analyze the relationships between factors.

2- Project Mind Flow

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

3- Conclusion & Output

1. Fig1. Dos amount
   this graph shows the relationship between tooth length and dose amount. It demonstrates a positive relationship between dose amount and tooth length, indicating that increasing dose amount within particular level can benefit tooth growth. Meanwhile, this effect differs from two types of supplement delivery, which worths further investigation.
2. Fig2. Supplement amount
   this graph shows the relationship between tooth length and supplement delivery method. When the dose amount is at 0.5 and 1.0, supplement delivery of "OJ" has a stronger positive effect of tooth length. However, this effect is minor when the dose amount is at 2.0.
3. Since p value is greater than 0.05, it can be concluded that supplement delivery methods have no impact on Tooth growth.
4. As can be seen, the p-value of each test was far below 0.05. We can assume that the average tooth length increases with an increasing dose, and therefore the null hypothesis can be rejected.
5. In reviewing our t-test analysis from above, we can conclude that supplement delivery method has no effect on tooth growth at 95% confidence interval. While increased doses benefits tooth growth.



Formatted: Indent: Before: 0.31", No bullets or numbering

Statistical Inference Course Project Part 2~~1~~

4- **Coding**

```
## EDA
data("ToothGrowth")
head(ToothGrowth)
summary(ToothGrowth$len)

unique(ToothGrowth$supp)
table(ToothGrowth$supp)

summary(ToothGrowth$dose)
unique(ToothGrowth$dose)
```

Formatted: Font: 14 pt, Bold, Underline, Font color: Accent 2, Complex Script Font: 14 pt, Bold

Formatted: Normal, Indent: Before: 0"

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Highlight

Statistical Inference Course Project Part 2~~1~~

```
nrow(ToothGrowth)
sapply(ToothGrowth, class)

ToothGrowth$dose <- as.factor(ToothGrowth$dose)

## Ploting the dose fig1
ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=dose))
+ xlab("Dose Amount") + ylab("Tooth Length") + facet_grid(~ supp) +
ggtitle("Tooth Length vs. Dose Amount \nby Delivery Method") +
  theme(plot.title = element_text(lineheight=.8, face="bold"))


## Ploting the Supply fig2
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
+ xlab("Supplement Delivery") + ylab("Tooth Length") + facet_grid(~ dose) +
ggtitle("Tooth Length vs. Delivery Method \nby Dose Amount") +
  theme(plot.title = element_text(lineheight=.8, face="bold"))

## t-test
t.test(len~supp,data=ToothGrowth)

ToothGrowth_sub1 <- subset(ToothGrowth, ToothGrowth$dose %in%
c(1.0,0.5))
t.test(len~dose,data=ToothGrowth_sub1)

ToothGrowth_sub2 <- subset(ToothGrowth, ToothGrowth$dose %in%
c(0.5,2.0))
t.test(len~dose,data=ToothGrowth_sub2)

ToothGrowth_sub3 <- subset(ToothGrowth, ToothGrowth$dose %in%
c(2.0,1.0))
```

~~4-~~ t.test(len~dose,data=ToothGrowth_sub2)

~~## Set Random data~~

~~set.seed(127)~~

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Indent: Before: 0.25", No bullets or numbering

Formatted: Font: Not Bold, No underline, Font color: Auto, Complex Script Font: Not Bold

Formatted: List Paragraph, Line spacing: single

Statistical Inference Course Project Part 2~~1~~

```
lambda <- 0.2
n <- 40
sample_size <- 1000
simulated_sample <- replicate(sample_size, rexp(n, lambda))
means_exponentials <- apply(simulated_sample, 2, mean)
## Meam Comparison
sample_mean <- mean(means_exponentials)
theo_mean <- 1 / lambda
sample_mean
theo_mean
mat.mean <- (sample_mean / theo_mean)
mat.mean
## Variance Comparison
sample_var <- var(means_exponentials)
theo_var <- (1 / lambda)^2 / (n)
sample_var
theo_var
mat.Var <- (sample_var / theo_var)
sample_sd <- sd(means_exponentials)
theo_sd <- 1/(lambda * sqrt(n))
sample_sd
theo_sd
mat.sd <- (sample_sd / theo_sd)
mat.sd
## Distribution _ Plotting
plotdata <- data.frame(means_exponentials)
m <- ggplot(plotdata, aes(x =means_exponentials))
m <- m + geom_histogram(aes(y=..density..), colour="grey",
                fill = "grey66")
m <- m + labs(title = "Distribution of means of 40 Samples", x = "Mean of 40 Samples", y =
"Density")
m <- m + geom_vline(aes(xintercept = sample_mean, colour = "sample"))
m <- m + geom_vline(aes(xintercept = theo_mean, colour = "theoretical"))
m <- m + stat_function(fun = dnorm, args = list(mean = sample_mean, sd = sample_sd), color =
"gold1", size = 1.0)
m <- m + stat_function(fun = dnorm, args = list(mean = theo_mean, sd = theo_sd), colour =
"red", size = 1.0)
m
```

Formatted: List Paragraph, Indent: Before: 0", Line spacing: single

Formatted: List Paragraph, Line spacing: single

Formatted: List Paragraph, Tab stops: Not at 0.76"

Formatted: List Paragraph