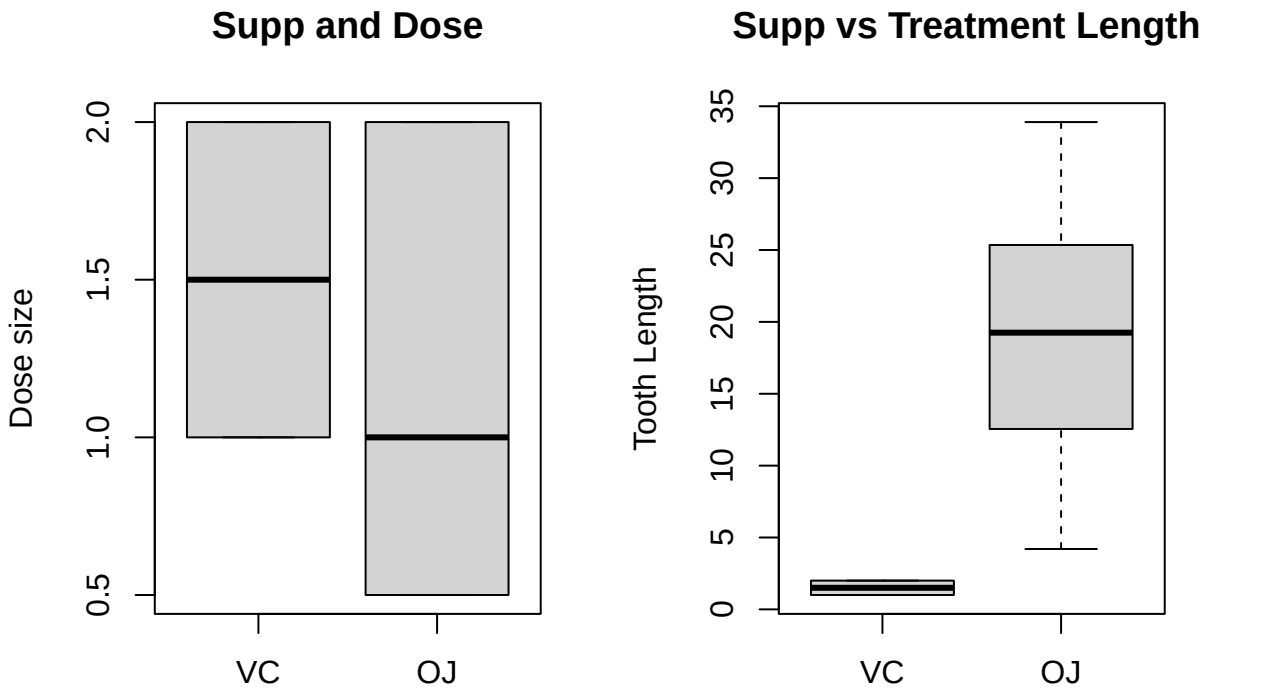
Statistical_inference_project

spemurphy

2024-10-16

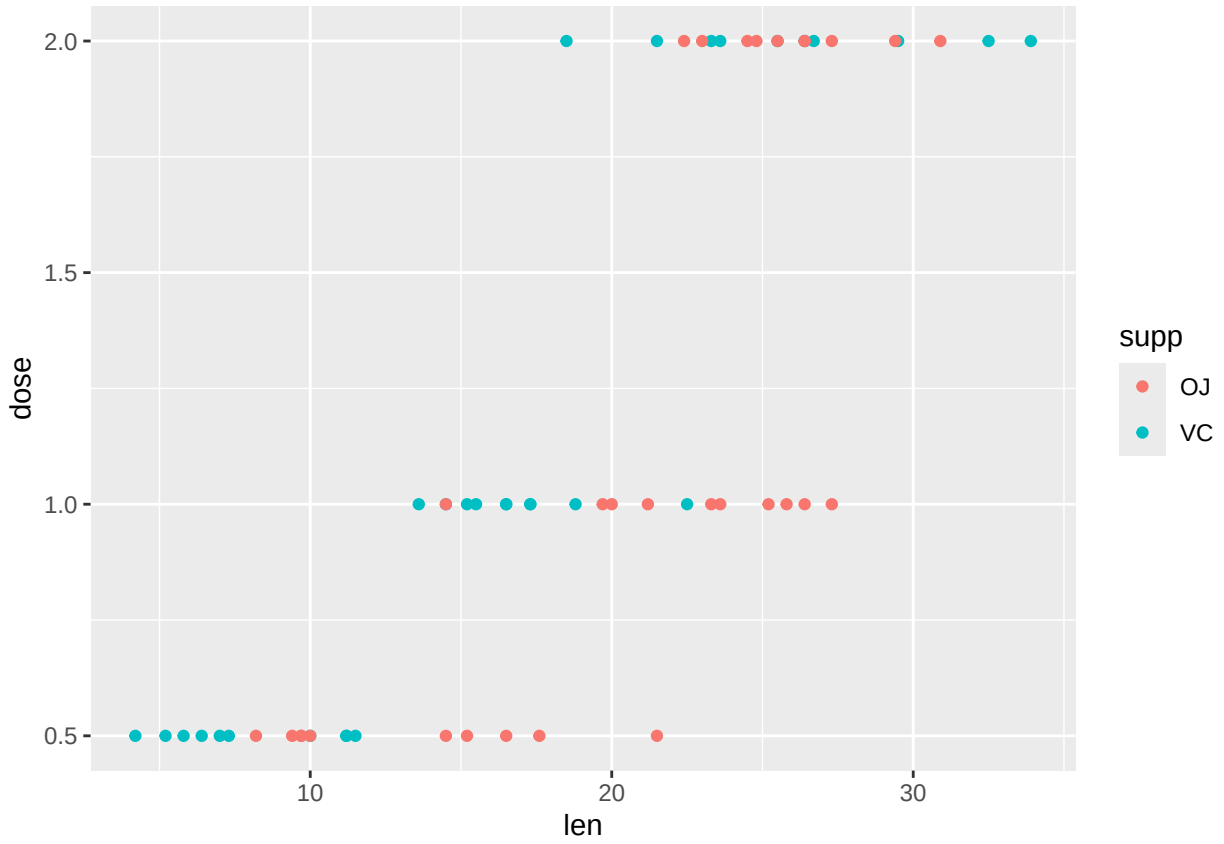
```
par(mfrow=c(1,2))
boxplot(tg$supp, tg$dose, main="Supp and Dose",
        ylab="Dose size", names=c("VC","OJ"))
boxplot(tg$supp, tg$len, main="Supp vs Treatment Length ",
        ylab="Tooth Length", names=c("VC","OJ"))
```



Doses are higher on average for VC compared to OJ, The average tooth length measurement also appears to be shorter for VC.

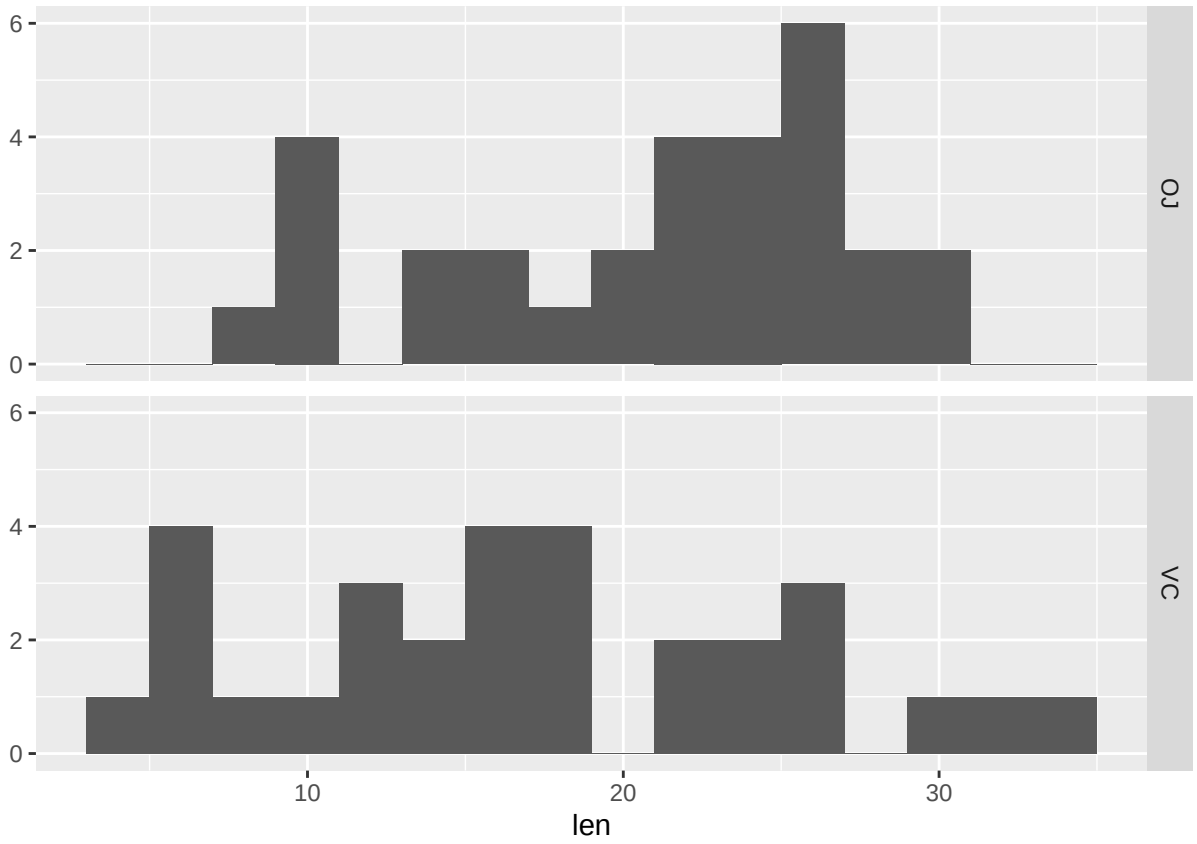
```
qplot(len, dose, data=tg, color=supp)
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



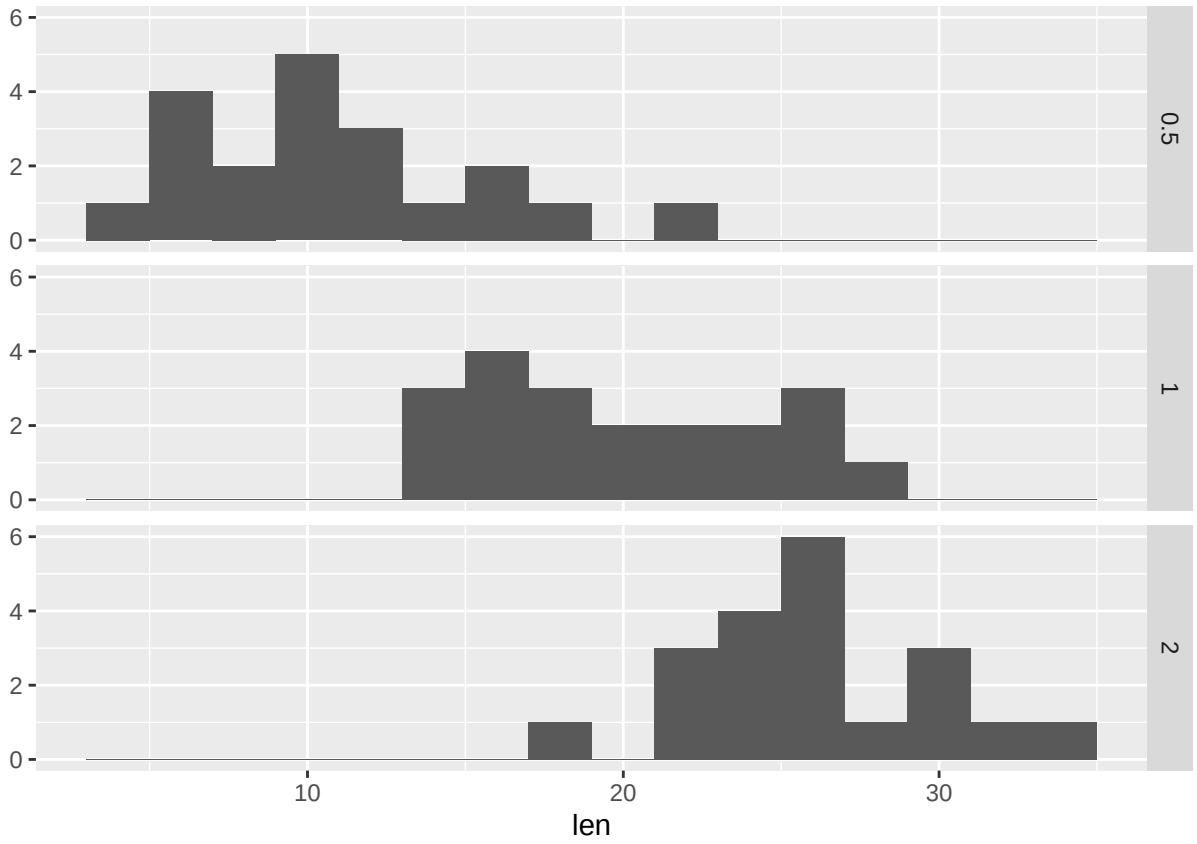
VC has shorter tooth length measurements on average compared to OJ for dose sizes of 0.5 & 1.0. However, in the larger dose case of 2.0, the tooth length measurement appears to be have greater variance than OJ: ranging from around 19 : 33 compared to between 21 and 31 respectively.

```
qplot(len, data=tg, facets=supp~., binwidth=2)
```



OJ has a greater number of tooth length observations ranging from 20 to 30 while VC more observations in the 0-10 tooth and 30-35 tooth length measurements. So, while VC may have a smaller average tooth length measurement, it may have a larger standard deviation. Particularly, if the dose size for VC is 2.0, it will yield a longer tooth measurement.

```
qplot(len, data=tg, facets=dose~., binwidth=2)
```



To highlight these differences numerically.

```
vcSupp <- tg %>%
  filter(supp=="VC")

mean(vcSupp$len) ## The average tooth length measurement for Vitamin C is 16.96
```

```
## [1] 16.96333
```

```
sd(vcSupp$len)  ## The standard deviation of Vitamin C is 8.27
```

```
## [1] 8.266029
```

```
ojSupp <- tg %>%
  filter(supp=="OJ")

mean(ojSupp$len) ## The average tooth length measurement for Orange Juice is 20.67
```

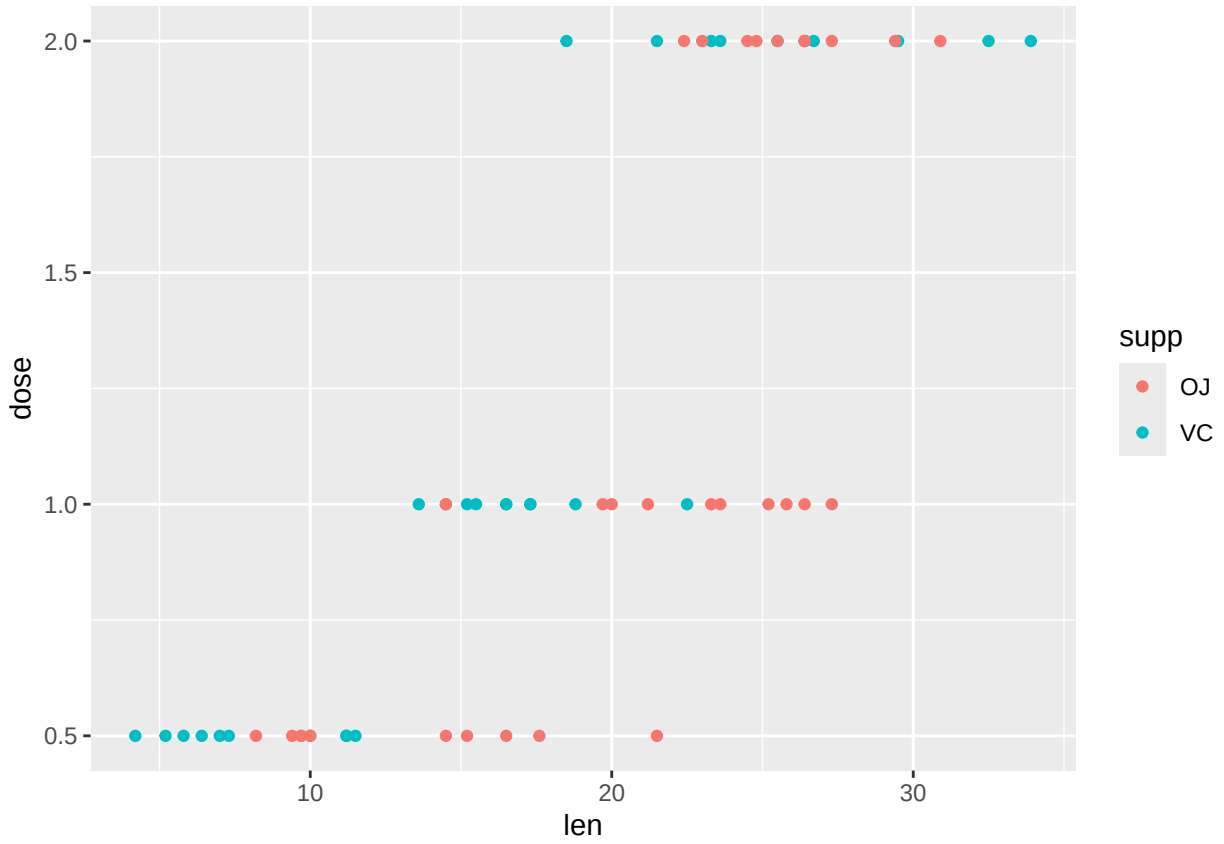
```
## [1] 20.66333
```

```
sd(ojSupp$len)  ## The standard deviation of Orange Juice is 6.61
```

```
## [1] 6.605561
```

These calculations confirm that while Vitamin C results in , on average, lower tooth measurements. However, the variability of the data is larger, as Orange Juice has a lower standard deviation.

```
# This is highlighted by the graph:
qplot(len, dose, data=tg, color=supp)
```



Where greater variability comes in when the dose of Vitamin C is increased to 2.0 mg/day

Therefore, we want to test two hypotheses:

1. Does Orange Juice result in larger tooth growth than Vitamin C?

2. Does a 2.0 mg/day dose of Vitamin C result in longer tooth length than for an equivalent dose of Orange Juice?

Question 1:

Null hypothesis (H0): There is no significant difference in the tooth length measurement of Orange Juice compared to the tooth length measurement of Orange Juice.

Alternative Hypothesis (H1): Orange Juice results in a greater tooth length measurement than Vitamin C.

We will use a 95% confidence level with a 5% alpha value

Here we put the supplements into their respective groups

```
ojGroup <- tg %>%
  filter(supp == "OJ")

vcGroup <- tg %>%
  filter(supp=="VC")
```

Then we run a Welch's t-test for our hypothesis.

```
t.test(ojGroup$len, vcGroup$len, alternative="greater", var.equal=FALSE, conf.level=0.95)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  ojGroup$len and vcGroup$len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Conclusion:

The t-test returned a p-value of 0.03032. With a alpha value of 0.05 and with 95% confidence, we lend support to the alternative hypothesis that Orange Juice results in greater tooth length measurements compared to Vitamin C. For this t-test, we assumed: Normality of the data, independence of observations in that they are independent from each other, that both groups have unequal variances, and that the sample is unbiased and represents a random sample.

Question 2:

Null hypothesis (H0): There is no significant difference in the tooth length measurement for a 2.0 mg/day dose of Vitamin C compared to an equivalent dose of Orange Juice.

Alternative Hypothesis (H1): A 2.0 mg/day dose of Vitamin C results in a longer tooth length measurement than a 2.0 mg/day dose of Orange Juice.

First, I need to sample the appropriate data from the dataset. That is those figures that have a dose of 2.0 mg/day

We will use a 95% confidence level with a 5% alpha value

This samples all observations that have a dose of 2.0.

```
tgSample <- tg %>%
  filter(dose == 2.0)
nrow(tgSample)
```

```
## [1] 20
```

```
# This function separates those observations into their respective groups.
vcSample <- tgSample %>%
  filter(supp == "VC")

ojSample <- tgSample %>%
  filter(supp == "OJ")
```

Performing a Welch's t-test as I am working with continuous data comparing two groups. However, we have unequal variance, and are performing a one-sided test on a small sample size (<30).

```
t.test(vcSample$len, ojSample$len, alternative="greater", var.equal=FALSE, conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  vcSample$len and ojSample$len
## t = 0.046136, df = 14.04, p-value = 0.4819
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  -2.9735     Inf
## sample estimates:
## mean of x mean of y
##     26.14     26.06
```

Conclusion:

Through the hypothesis test, we can conclude that the likelihood of the larger mean in the case of 2.0 mg/day dose of Vitamin C occurring due to random chance is approximately 48.20%. At a 95% confidence interval, we fail to reject the null hypothesis: There is no significant difference in the tooth length measurement for a 2.0 mg/day dose of Vitamin C compared to an equivalent dose of Orange Juice. For this t-test, we assumed: Normality of the data , independence of observations in that they are independent from each other, that both groups have unequal variances, and that the sample is unbiased and represents a random sample.

Final Conclusion

Through both hypothesis test ran during the course of analysis, we can conclude that Orange Juice results in larger average tooth length measurements compared to Vitamin C, and, that there is no significant difference in the tooth length measurement for a 2.0 mg/day dose of Vitamin C compared to an equivalent dose of Orange Juice.