Basic_inferential_data_analysis

Stephen Yang

August 7, 2019

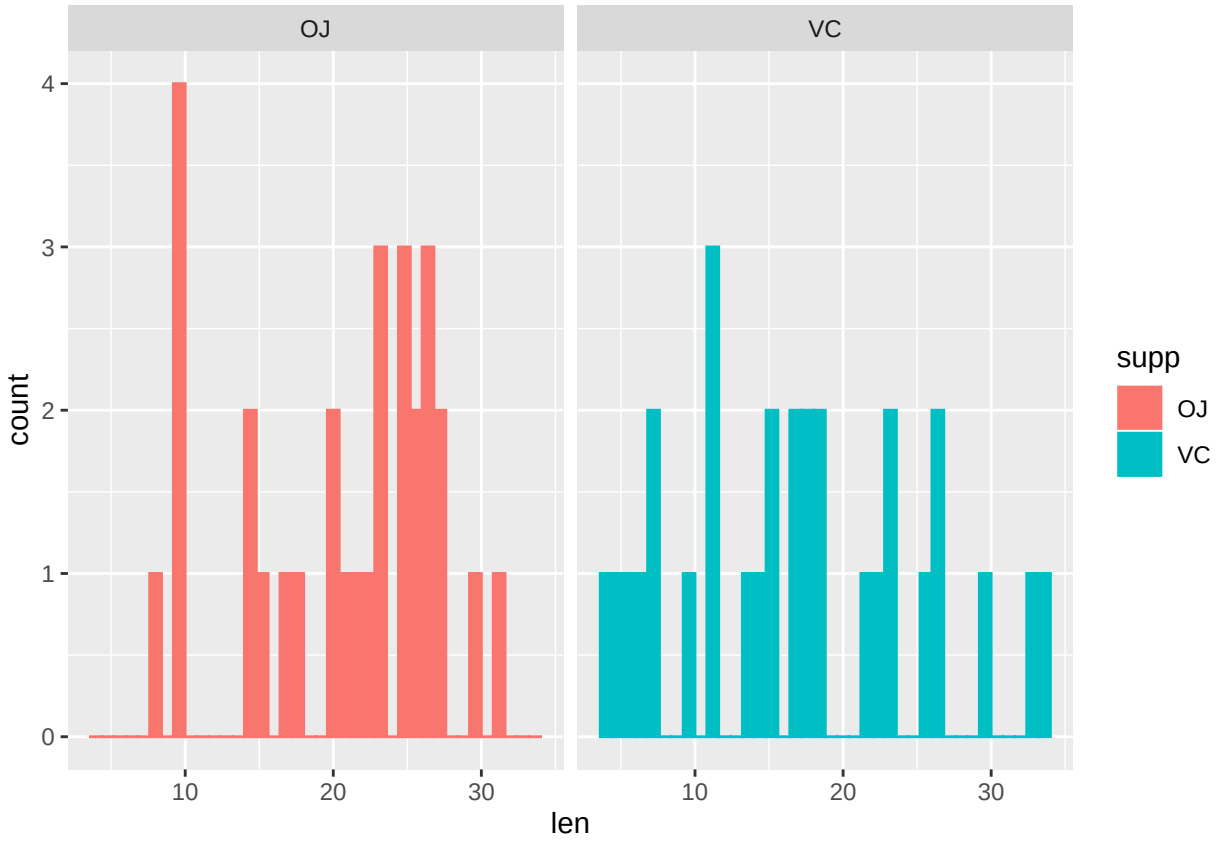
Summary

The Effect of Vitamin C on Tooth Growth in 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.

Basic exploratory analysis

The dataset is used to perform a basic inferential analysis. Let's look at the structure of the dataset.

```
library(ggplot2)
data("ToothGrowth")
ggplot(ToothGrowth, aes(x=len, fill=supp, colour = supp)) + facet_wrap(~supp) + geom_histogram(binwidth
```



We can see that the dataset with 60 observations and 3 variables.

Compare difference by group in the given dosage

First thing is to look for any difference between the supplements at the specific dosage.

```r
low_supp <-  ToothGrowth[ToothGrowth$dose == 0.5,]
t.test(len ~ supp, data = low_supp, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

The 95% confidence interval is between (1.71, 8.78) which shows that we are confident to reject the null hypothesis. There's the difference between the two supplements given the dosage is 0.5.

```r
mid_supp <- ToothGrowth[ToothGrowth$dose == 1.0,]
t.test(len ~ supp, data = mid_supp, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

The 95% confidence interval is between (2.80, 9.05) which shows that we are confident to reject the null hypothesis. There's the difference between the two supplements given the dosage is 1.0.

```r
high_supp <- ToothGrowth[ToothGrowth$dose == 2.0,]
t.test(len ~ supp, data = high_supp, paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

The 95% confidence interval is between (-3.79, 3.63) which shows the possibility of no difference in two supplements. We can't reject the null hypothesis at the dosage of 2.0. So we are sure that only at the dosage of 2.0 shows some significant value that the mean of two groups might be equal.

State your conclusions and the assumptions needed for your conclusions.

Assumptions

The confidence interval of the dosage 0.5 is between 1.71 to 8.78, which doesn't contain 0. Thus, we are certain that the difference between the two delivery ways exists. So does the difference of the dosage 1.0 exist, which show the interval between 2.80 to 9.05. But the confidence interval of the dosage 2.0 contains 0, meaning the null hypothesis could be true.

Conclusions

In dosage 0.5 and 1.0, we can see the effect on two different delivery ways are significantly different. Both of which are helpful for the teeth growth in Guinea Pigs. Especially the orange juice shows the better effect in this dosage range. The more dosage we use the better length of tooth growth would be. But when up to 2.0, the effect of the two supplements would be the same.