Statistical Inference Course Project Part 2

Melody Wolk

June 8, 2015

```
library("ggplot2")
# Load the data
data(ToothGrowth)
data <- ToothGrowth
```

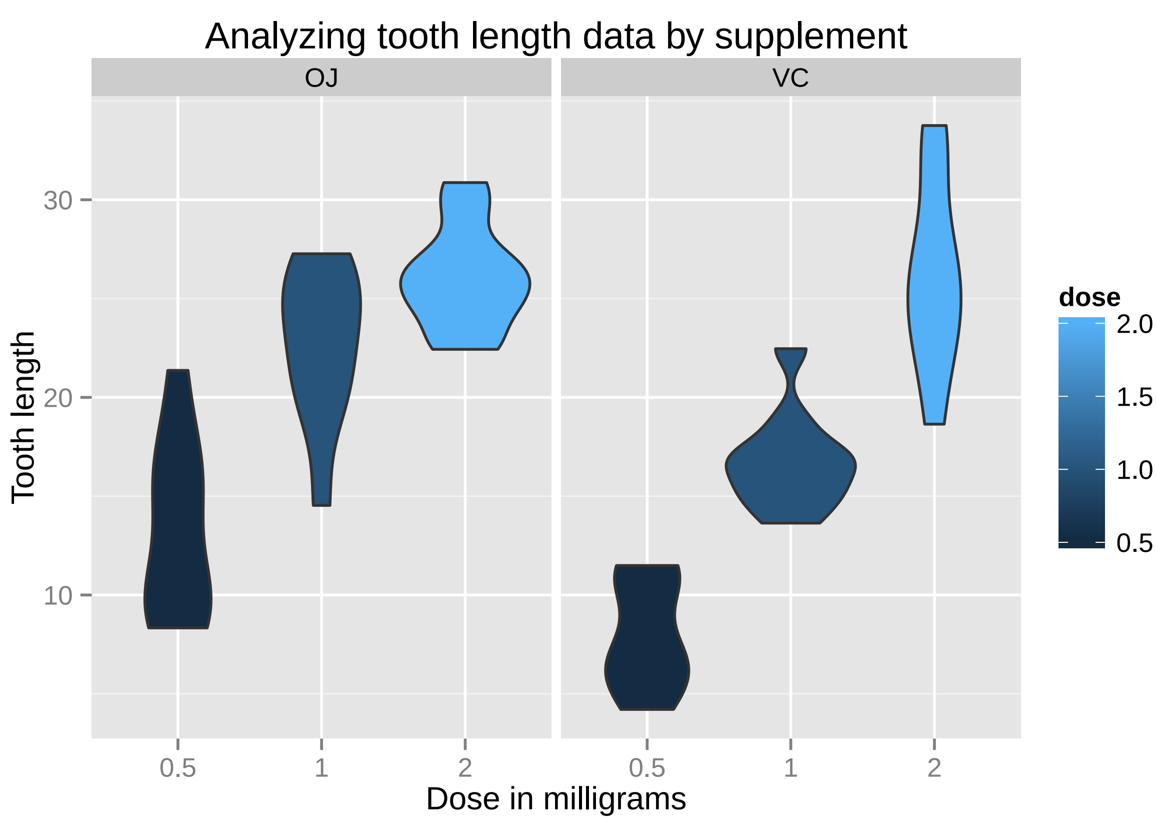
Underlying assumptions

- The experiment was done with random assignment of guinea pigs to different dose level categories and supplement type
- They are members of the same underlying population

Exploratory data analysis

Each sample is independent so we are going to use paired=FALSE

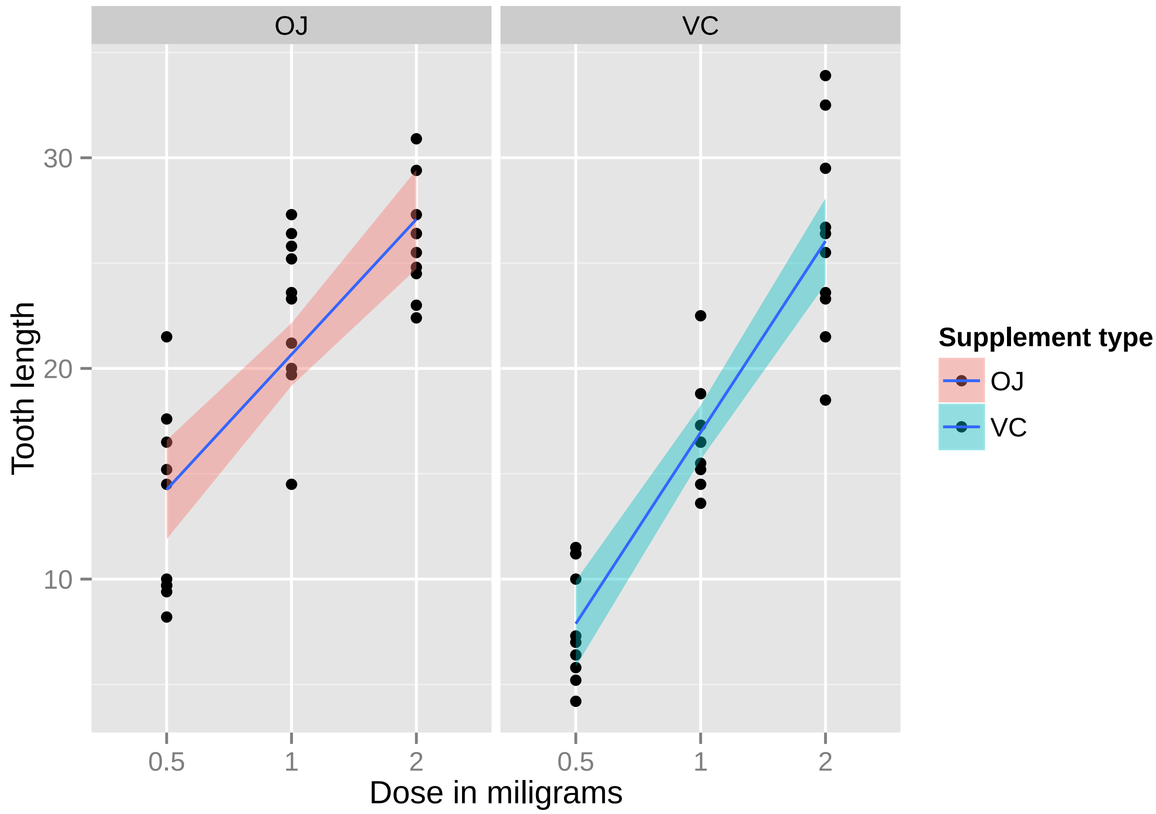
```
 ggplot(data, aes(x=as.factor(dose), y=len, fill=dose))+geom_violin()+facet_grid(.~supp)+
ggtitle("Analyzing tooth length data by supplement")+xlab("Dose in milligrams")+ylab("Tooth length")
```



We see that higher doses of supplements seem to give longer teeth. We also see that the variances amongst the samples are very different so we are going to use var.equal = FALSE.

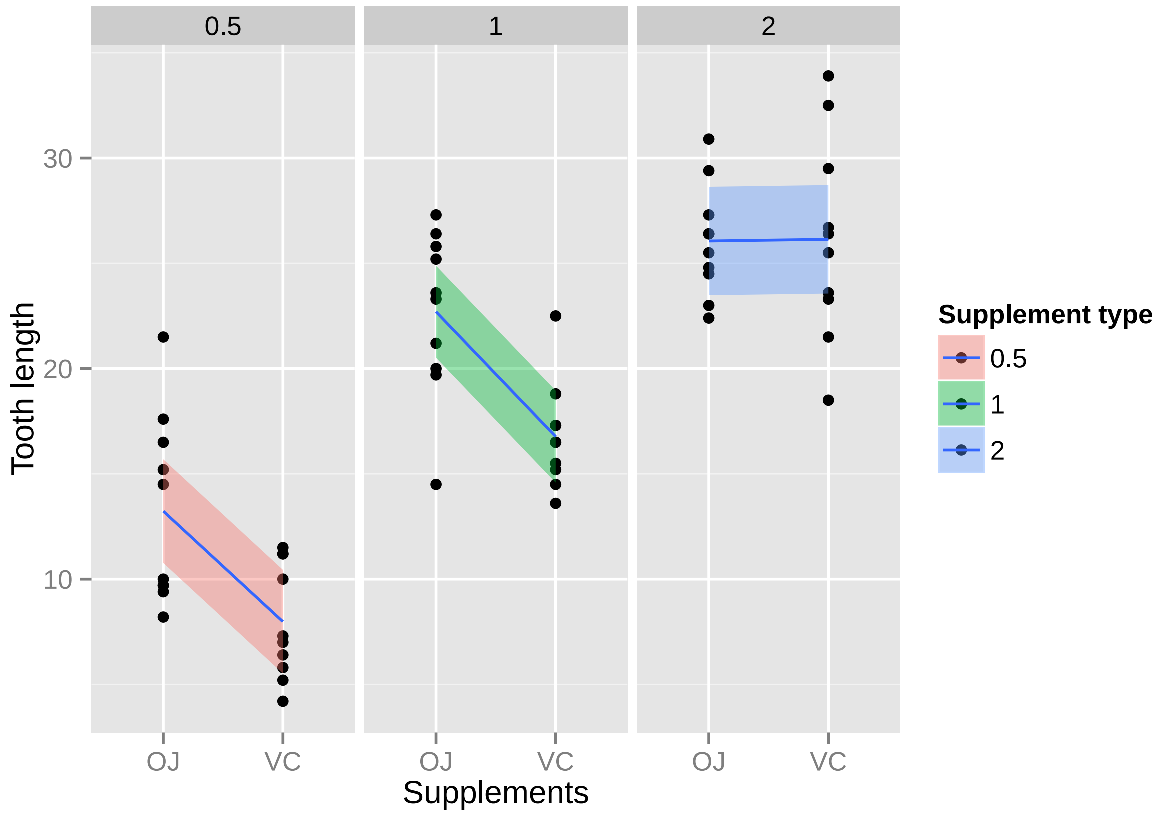
We can look a the scatter plot of the tooth length as a function of the dose:

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) + geom_point(stat="identity",) +
facet_grid(. ~ supp)  + xlab("Dose in miligrams") + ylab("Tooth length") +
guides(fill=guide_legend(title="Supplement type"))+stat_smooth(method="lm",aes(group=1))
```



which seems to confirm that an incresing dose leads to an increasing tooth length. Let's split by dose level and look at the effect of the supplement:

```
ggplot(data=ToothGrowth, aes(x=as.factor(supp), y=len, fill=factor(dose))) +
geom_point(stat="identity",) + facet_grid(. ~ dose) + xlab("Supplements") + ylab("Tooth length") +
guides(fill=guide_legend(title="Supplement type"))+stat_smooth(method="lm",aes(group=1))
```



The trend is less clear here. We need to perform some tests!

Comparing tooth growth by supplement and dose

Effect of the dose

First let's use confidence intervals to see if the trend that we have detected is meaningful. We split the data according to the different supplements:

```
dataOJ <- data[data$supp=="OJ",]
fitOJ <- lm(dataOJ$len ~ dataOJ$dose)
summary(fitOJ)
```

```
##
## Call:
## lm(formula = dataOJ$len ~ dataOJ$dose)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2557 -3.7979 -0.0643  3.3521  7.9386
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.550      1.722   6.708 2.79e-07 ***
## dataOJ$dose    7.811      1.302   6.001 1.82e-06 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.446 on 28 degrees of freedom
## Multiple R-squared:  0.5626, Adjusted R-squared:  0.547
## F-statistic: 36.01 on 1 and 28 DF,  p-value: 1.825e-06
```

```
confint(fitOJ)
```

```
##                2.5 %   97.5 %
## (Intercept) 8.022743 15.07726
## dataOJ$dose 5.145073 10.47778
```

This is consistent with a positive slope so an increase in the length with an increasing in the dose. What about the other supplement?

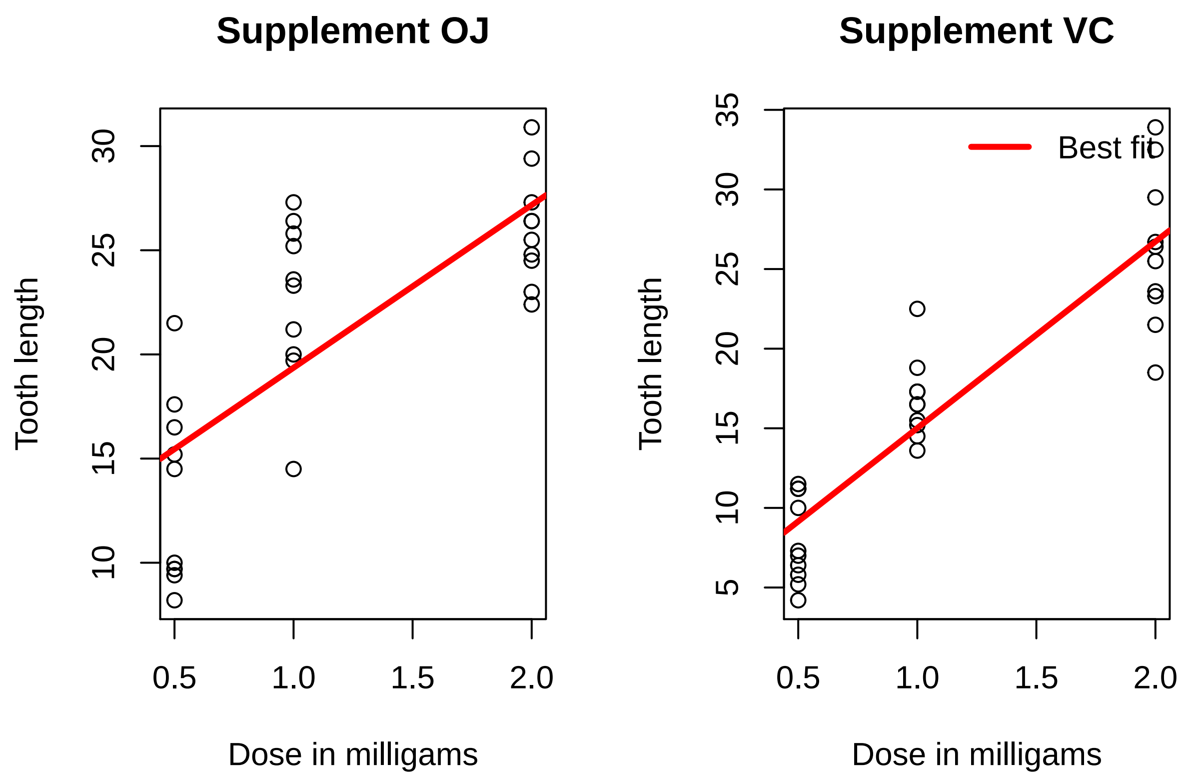
```
dataVC <- data[data$supp=="VC",]
fitVC <- lm(dataVC$len ~ dataVC$dose)

summary(fitVC)
```

```
##
## Call:
## lm(formula = dataVC$len ~ dataVC$dose)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2264 -2.6029  0.0814  2.2288  7.4893
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.295      1.427   2.309   0.0285 *
## dataVC$dose   11.716      1.079  10.860 1.51e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.685 on 28 degrees of freedom
## Multiple R-squared:  0.8082, Adjusted R-squared:  0.8013
## F-statistic: 117.9 on 1 and 28 DF,  p-value: 1.509e-11
```

```
confint(fitVC)
```

```
##                 2.5 %   97.5 %
## (Intercept) 0.3717998  6.21820
## dataVC$dose 9.5059827 13.92545
```

```
par(mfrow=c(1,2))
plot(dataOJ$dose, dataOJ$len, xlab = "Dose in milligams", ylab="Tooth length", main="Supplement OJ")
abline(fitOJ, lwd=3, col="red")
plot(dataVC$dose, dataVC$len, xlab = "Dose in milligams", ylab="Tooth length", main="Supplement VC")
abline(fitVC, lwd=3, col="red")
legend('topright',c("Best fit"), col=c("red"), lty=c(1), bty="n", lwd=c(3))
```



This is also consistent with a positive slope. To confirm this we can also perform hypothesis testing. In order to do so we are going to split the data set for each dose. Our null hypothesis H0 is "the dose has no effect on the tooth lenght".

```
t.test(data$len[data$dose==2], data$len[data$dose==1], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$len[data$dose == 2] and data$len[data$dose == 1]
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
t.test(data$len[data$dose==2], data$len[data$dose==0.5], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$len[data$dose == 2] and data$len[data$dose == 0.5]
## t = 11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83383 18.15617
```

```
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

```r
t.test(data$len[data$dose==1], data$len[data$dose==0.5], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  data$len[data$dose == 1] and data$len[data$dose == 0.5]
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

For each case, the p-value of the t-test is <0.05. Thus we can reject the null hypothesis, which subsequently means that we can assume that the dose change creates a positive effect on tooth length i.e the tooth length increases.

Effect of the supplement

We can now look at the effect of the supplements. First looking at the confidence intervals for the slopes in the previous analysis, we see that they are consistent with each other within the confidence intervals. It means that there is no obvious effect of the supplement type. To confirm this, we can perform an hypothesis testing with H0 "the supplement has no effect on the tooth lenght".

```r
t.test(len~supp, data=data, paired=FALSE, var.equal = FALSE)
```

```
##
##  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 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
```

We can see that the p-value of this test is greater than a significance level of 5%, we have not enough evidence to reject the null hypothesis i.e we cannot assume that the supplement type has en effect on tooth length.

Conclusions

We found that the supplement type has no significant effect on the tooth growth. However our results suggest that increasing the dose level leads to an increase in the tooth growth.