Tooth Growth Data Analysis

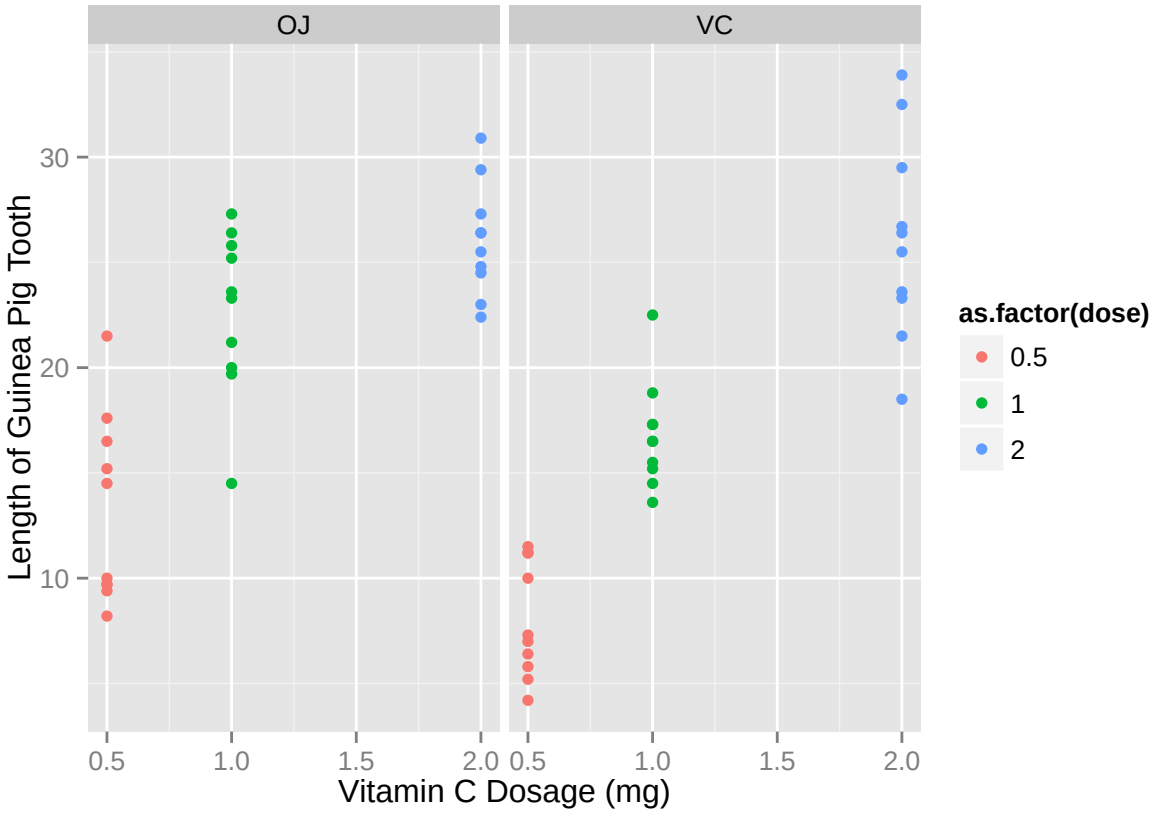
Michael McDermott

December 13, 2014

Let's consider a data set relating the Tooth length of Guinea pigs to the amount of vitamin C they were given. There were 2 ways to administer the vitamin C, one by feeding orange juice (OJ) and the other by directly giving them Ascorbic acid (VC). Within each method 3 different doses were given: 0.5, 1 and 2 mg.

A plot showing the relation of tooth growth to dosage for each method is shown below:

```
library(ggplot2)
data(ToothGrowth)
g<-ggplot(ToothGrowth,aes(dose,len))
g+geom_point(aes(color=as.factor(dose)))+facet_grid(.~supp) + labs(x="Vitamin C Dosage (mg)") + labs(y=
```



From the graph it would appear that within each group dosage plays an important role on tooth development. It would also appear that for a given dose the method of administration is important up to a point perhaps where the dosage become large and the method is no longer as important.

Effect of delivery method

Hypothesis test and confidence interval

In order to test the effect of the delivery method we do the following:

1. Divide the data by similar dosage groups.
2. Adopt the null hypothesis H_0 on the way that vitamin C was administered, i.e. H_0: The way that vitamin C was administered has no effect on the length of tooth growth.
3. Run the appropriate tests.

```
Tooth_5<-subset(ToothGrowth,dose==0.5)
Tooth_1<-subset(ToothGrowth,dose==1)
Tooth_2<-subset(ToothGrowth,dose==2)

t.test(len ~ supp, Tooth_5,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
```

```
t.test(len ~ supp, Tooth_1,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
```

```
t.test(len ~ supp, Tooth_2,paired=FALSE,var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, 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
```

So the lower two dosages 95% confidence interval do not include 0 (for the t statistic) which means we reject the null hypothesis H_0. We are justified in claiming that delivery method does have an effect on tooth growth.

The largest dosage (2mg), however, does have a 95% confidence interval which contains 0 and thus we are not justified in rejecting the null hypothesis H_0. We can't say that at large dosages there is any effect of delivery method on tooth growth.

Effect of dosage

Hypothesis test and confidence interval

Now we investigate the effect of dosage within a given delivery method. There are 3 dosages and so 3 pairs of possible null hypothesis to investigate (for each delivery method). The null hypothesis now is:

-H_0: For a given delivery method the dosage means are identical.

```
Tooth_5_OJ<-subset(Tooth_5,supp=='OJ')
Tooth_5_VC<-subset(Tooth_5,supp=='VC')
Tooth_1_OJ<-subset(Tooth_1,supp=='OJ')
Tooth_1_VC<-subset(Tooth_1,supp=='VC')
Tooth_2_OJ<-subset(Tooth_2,supp=='OJ')
Tooth_2_VC<-subset(Tooth_2,supp=='VC')
```

All the combinations within delivery method OJ:

```
t.test(Tooth_5_OJ$len, Tooth_1_OJ$len ,paired=FALSE,var.equal=FALSE)$conf.int
```

```
## [1] -13.415634  -5.524366
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(Tooth_5_OJ$len, Tooth_2_OJ$len ,paired=FALSE,var.equal=FALSE)$conf.int
```

```
## [1] -16.335241  -9.324759
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(Tooth_1_OJ$len, Tooth_2_OJ$len ,paired=FALSE,var.equal=FALSE)$conf.int
```

```
## [1] -6.5314425 -0.1885575
## attr(,"conf.level")
## [1] 0.95
```

All the combinations within delivery method VC:

```
t.test(Tooth_5_VC$len, Tooth_1_VC$len ,paired=FALSE,var.equal=FALSE)$conf.int
```

```
## [1] -11.265712  -6.314288
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(Tooth_5_VC$len, Tooth_2_VC$len ,paired=FALSE,var.equal=FALSE)$conf.int
```

```
## [1] -21.90151 -14.41849
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(Tooth_1_VC$len, Tooth_2_VC$len ,paired=FALSE,var.equal=FALSE)$conf.int
```

```
## [1] -13.054267  -5.685733
## attr(,"conf.level")
## [1] 0.95
```

So we see that in every case we are statistically justified in rejecting the null hypothesis since the 95% confidence interval for the t-statistic does not contain 0. Within each delivery system OJ and VC, we find evidence that an increase in dosage is commensurate with and increase in average tooth growth.

Asumptions

We used the following assumptions:

- The difference in averages divided by the sample standard deviations follow a t-distribution. There are only 10 samples in each case so a normal distribution would be a bad approximation.
- The data are not paired: this is because each sample of 10 Guinea pigs is different. It would be paired if the same Guinea pigs were used in different trials (note that this probably wouldnt make sense to do for tooth growth since a guinea pig can't start from scratch after growing fangs!).
- The variance of the samples is different: this is simply because there is no good reason to expect Guinea pig populations with different dosages and delivery methods to have the same fluctuations.
- We have tacitly assumed that there are no spurious correlations in the data set. For example we have certainly assumed that the following contrived scenario is **not** the case:
  - for a given dosage of vitamin C, Guinea pigs given ascorbic acid (VC) receive it in liquid form whereas those given orange juice (OJ) get an orange which rolls around which makes the Guinea pigs run after the orange and running is correlated with larger teeth.