Statistical Inference: Course Project

rtaph

November 23, 2014

Part II: Tooth Growth

This paper will explore the `ToothGrowth` data in the R data sets package. In particular, the objective is to use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

Discussion of the Data and Experimental Design

The `?ToothGrowth` page gives insight into the data set, which originates from a study by C. I. Bliss in *The Statistics of Bioassay*, (1952). Three variables describe the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid):

- **len:** Tooth length
- **supp:** Supplement type (VC or OJ)
- **dose:** Dose in milligrams

To begin the analysis, we load the data and run summaritive printouts:

```
library(datasets); data(ToothGrowth);
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

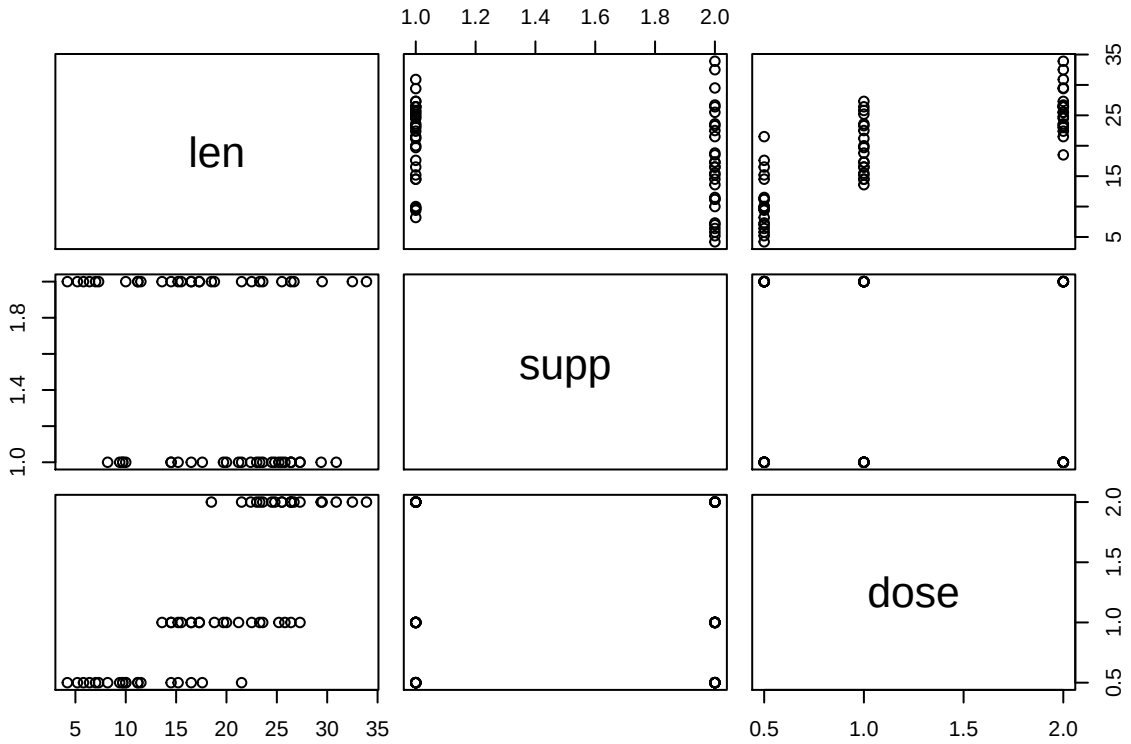
```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

The data conforms to the expectations from reading the help file. In particular, we note that the sample sizes are equally distributed across the variables `supp` and `dose`. Although the help file mentions that there were only 10 guinea pigs, and that the data fits nicely into 10-sample subsets, there is no identifier variable for each guinea pig. As a result, I will assume throughout this paper that the data is not structured in a paired order, and accordingly conduct independent-group hypothesis tests. It is also assumed that data is independent and identically distributed (IID) *within groups*, and that the variance between these groups unequal (the more conservative assumption).

To begin the analysis, I plot the pairs of variables against one another:

```
pairs(ToothGrowth)
```



At first glance, it appears that there is a strong positive relationship between tooth length and dose. A potential relationship between the length and supplement type is less apparent.

Setting Up Our Tests

The first step is to state our null and alternate hypotheses, set a decision rule, and chose the appropriate tests. According to the assumptions made in the introductory paragraphs, my choice of test is a paired, two-sided t-test. Although the plot in the exploratory phase suggests that tooth growth increases directly proportionate to dosage, it would have been equally meaniful/plausible if the inverse was true (decreasing dosage $\implies$ increased length). In order not two bias the hypothesis test with a lower tolerance resulting from having already explored the data visually, I justify the two-sided alternative as the appropriate option.

The null hypothesis is therefore that $H_0 : \mu_1 = \mu_2$, and our alternate hypothesis is $H_a : \mu_1 \neq \mu_2$. I would normally set the family-wise tolerance at $\alpha = 0.05$, and accordingly make bonferroni corrections. However, given these adjustments were not covered in class, and that 95% confidence intervals are requested, I will simply ignore the problem of multiplicity for simplicity.

```
#t.test(, alternative = "greater", paired = T)
```

Based on the results of the test, I conclude that...

Tooth Growth & Delivery Method

In this part, I will test whether tooth growth is correlated to the delivery method of a supplement (orange juice or ascorbic acid).

Other info

State your conclusions and the assumptions needed for your conclusions. Some criteria that you will be evaluated on Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data? Did the student perform some relevant confidence intervals and/or tests? Were the results of the tests and/or intervals interpreted in the context of the problem correctly? Did the student describe the assumptions needed for their conclusions?