BasicAnalysis

martinvanelp

Sunday, September 21, 2014

Basic inferential data analysis

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package. 1. Load the ToothGrowth data and perform some basic exploratory data analyses. 2. Provide a basic summary of the data. 3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering) 4. State your conclusions and the assumptions needed for your conclusions.

Output

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

2. Provide a basic summary of the data.

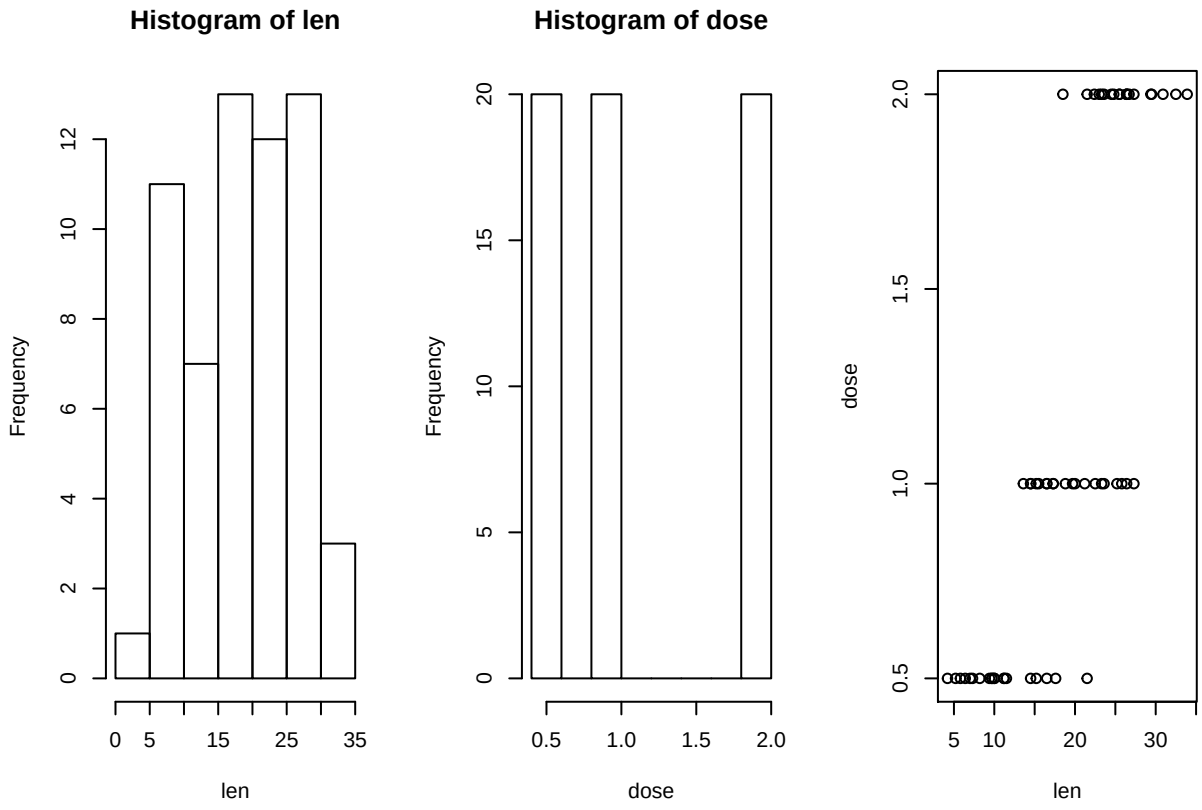
```
library(datasets)
summary(ToothGrowth)
```

```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
with(ToothGrowth, {
        par(mfrow = c(1,3))
        hist(len)
        hist(dose)
        plot(len, dose)
})
```



There are three variables in the dataset: two numeric (len & dose) and one factor (supp). Variable len seems sort of normally distributed. It seems dose and length are correlated given the last plot.

3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)

On a 95% confidence interval, the supplement is not significantly influencing tooth growth in this sample. Each increase in dose increases the tooth growth significantly on a 95% confidence interval in this sample. (see Appendix for the underlying t-tests)

4. State your conclusions and the assumptions needed for your conclusions.

For tooth growth the dose of the supplement significantly influential. The kind of supplement on the other hand does not significantly influence tooth growth, given a 95% confidence interval in both cases. The underlying assumption is that the sample is representative for the population.

Appendix

```
TGsupp <- subset(ToothGrowth, supp %in% c("OJ", "VC"))
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = TGsupp)
```

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 58, p-value = 0.06039
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.167  7.567
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

```
TG12 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = TG12)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

```
TG13 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = TG13)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15 -12.84
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

```
TG23 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, data = TG23)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994 -3.736
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```