Untitled

Nicholas Nagle

March 7, 2015

1. Two-sample T-test> ACT scores Consider the hypothesis that college-bound males and females have the same average ACT scores.

```
act <- read.table('ACT_scores.txt', header=TRUE)
```

   a. Write out the hypothesis that is tested by a two-sample t-test in this case.
   b. Using boxplots and summary statistics, report on the suitability of these data for a two-sample t-test to evaluate the hypothesis that seeding influences rainfall.
   c. Conduct a t-test to evaluate the hypothesis that seeding influences rainfall and report the result.

2. Two-sample t-tests. Cloud Seeding.

These data were collected in Southern Florida between 1968 and 1972 to test a hypothesis that seeding silver iodide into clouds can promote rainfall. They appear in a number of intriductory texts, including a favorite intermediate text of mine: Ramsay and Schafer's The Statistical Sleuth. The data contain 52 weather events that were deemed suitable for seeding. On each day an airplane flew the clouds and a payload was released. It was randomly determined whether the payload would contain silver iodide or not, but the pilots were unaware of this, i.e. the pilots were blind to the treatment. Following the flight, radar was used to measure the precipitation falling from the cloud (in units of acre-feet).

```
library(Sleuth3)
data(case0301)
head(case0301)
```

```
##   Rainfall Treatment
## 1   1202.6  Unseeded
## 2    830.1  Unseeded
## 3    372.4  Unseeded
## 4    345.5  Unseeded
## 5    321.2  Unseeded
## 6    244.3  Unseeded
```

```
# Here are two ways to calculate a boxplot
# boxplot(Rainfall~Treatment, data=case0301)
# library(ggplot2)
# ggplot(data=case0301) + geom_boxplot(aes(x=Treatment, y=Rainfall))

# Here are two ways to calculate some summary statistics
# tapply(X=case0301$Rainfall, INDEX=case0301$Treatment, FUN=mean) # Calculate mean, repeat for sd
# library(dplyr)
# case0301 %>% group_by(Treatment) %>% summarize(m=mean(Rainfall), sd=sd(Rainfall), n=n() )
```

   a. Using boxplots and summary statistics, report on the suitability of these data for a two-sample t-test to evaluate the hypothesis that seeding influences rainfall.
   b. Conduct a t-test to evaluate the hypothesis that seeding influences rainfall and report the result.

c. Repeat a. and b. using a log-transform of the rainfall measurement. Which analysis do you prefer and why?
d. Why is it important that the pilots were unaware of whether they were seeding or not?
e. Unfortunately, we do not have the date of the weather event. Why would it be helpful to have this information?

2. ANOVA

Here are some simple data describing average daily temperatures for different cities grouped by season.

```
# Create a dataset
temp = c(60,65,62,64,63,80,85,82,84,83,90,95,101,99,100)
season = rep( c('Fall', 'Spring', 'Summer'), times=c(5,5,5))

data <- data.frame(temp=temp, season=season)
```

For the air temperature data, conduct an ANOVA and use the F-statistic to test the hypothesis that air temperature varies by season.

3. Spock Conspiracy Trial.

This is another classic dataset. Dr. Benjamin Spock was a pediatrician who wrote a very popular book in 1948. In 1968, Dr Spock was accused of conspiring to violate the Selective Services Act by helping young men to medically avoid the military draft during the Vietnam War. His defence challeneged the case on the grounds that the jury did not contain a single woman, and therefore could not be random.

These data used here contain jury compositions for Spock's judge and 6 other judges in the same court. For each jury, we have the fraction of the jury that was female.

```
data(case0502)
```

a. Qualitatively describe these data using boxplots and summary statistics.
b. Using an ANOVA, test the hypothesis that all of the judges have the same jury compositions on average and report your results.
c. Grouping together Judges A-F, perform a two-sample t-test to test the hypothesis that Spock's judge consistently chooses juries with lower than expected women.
d. Use another ANOVA to evaluate whether or not it was suitable to group together the other 6 judges.
e. How should your interpretation of the results in part (c) change if you rejected the hypothesis the other six judges were not all statistically the same?