Chapter 5: One-Way ANOVA  
by: Allison Theobold

Sums of Squares, Mean Squares, F-tests

ANOVA F-test:

Sums of Squares:

\[
\text{Sum of Squares Total} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2
\]

\[
\text{Sum of Squares Group (Variable A)} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{I} n_i (\bar{x}_i - \bar{x})^2
\]

\[
\text{Sum of Squares Error (Residual)} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^{I} (n_i - 1)s_i^2
\]

Sum of Square Decomposition

\[
SS_{Total} = SS_{Group} + SS_{Error}
\]
Idea Behind an ANOVA F-test

In order to detect differences in means of more than one group we use a one-way ANOVA. But now our test statistic needs to measure how much variability there is between all of the groups. This statistic is the $SS_{Group}$, or the sum of squares of variable A. If there truly are differences in the means of the groups, then we expect $SS_{Group}$ to be large.

There are two different approaches to hypothesis testing for a one-way ANOVA: a parametric F-test and a non-parametric test.

Parametric ANOVA F-test

Mean Squares:

\[
\text{Mean Square Error} = MS_E = \frac{SS_E}{N-I}
\]

\[
\text{Mean Square for Group (Variable A)} = MS_{Group} = \frac{SS_{Group}}{I-1}
\]

\[
\text{ANOVA F-statistic} = \frac{MS_{Group}}{MS_E} = \frac{\text{between group variability}}{\text{within group variability}}
\]
F-distribution

- Right skewed
- Defined by degrees of freedom, but now there’s two!
  - Numerator degrees of freedom, (J-1)
  - Denominator degrees of freedom, (N-J)
- Always use right tail to get p-values

\[ F(2, 111) \]

\[ F(2, 10) \]

\[ F(6, 10) \]

\[ F(6, 111) \]

ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>Sums of Squares</th>
<th>Mean Squares</th>
<th>F-Ratio</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable A</td>
<td>J-1</td>
<td>SS_A</td>
<td>MS_A = SS_A / (J - 1)</td>
<td>F = MS_A / MS_E</td>
<td>Right tail of F(J - 1, N - J)</td>
</tr>
<tr>
<td>Residuals</td>
<td>N-J</td>
<td>SS_E</td>
<td>MS_E = SS_E / (N - J)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>N-1</td>
<td>SS_Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1:

Remark: Note in Al’s notes i refers to the group and j refers to the observation. In the table above these are reversed, but the only differences are:

- degrees of freedom for Group (Variable A) are I - 1
- degrees of freedom for Residuals are N - I
library(mosaic)
library(heplots)

MockJury <- MockJury

head(MockJury)

## Attr Crime Years Serious exciting calm independent sincere warm
## 1 Beautiful Burglary 10 8 6 9 9 8 5
## 2 Beautiful Burglary 3 8 9 5 9 3 5
## 3 Beautiful Burglary 5 5 3 4 6 3 6
## 4 Beautiful Burglary 1 3 3 6 9 8 8
## 5 Beautiful Burglary 7 9 1 1 5 1 8
## 6 Beautiful Burglary 7 9 1 5 7 5 8

## phyattr sociable kind intelligent strong sophisticated happy ownPA
## 1 9 9 9 6 9 9 5 9
## 2 9 9 4 9 5 5 5 7
## 3 7 4 2 4 5 4 5 5
## 4 9 9 9 9 9 9 9 9
## 5 8 9 5 8 9 9 8 7
## 6 8 9 5 8 9 9 9 9

boxplot(Years ~ Attr, data = MockJury, las = 1,
col = c("dark blue", "gray", "forest green"))

Which groups differ the most?
```r
lm_attr <- lm(Years ~ Attr, data = MockJury)
anova(lm_attr)
```

## Analysis of Variance Table
## Response: Years
## Df  Sum Sq Mean Sq F value Pr(>F)
## Attr 2   70.94 35.469  2.77 0.067 .
## Residuals 111 1421.32 12.805
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fill out the ANOVA table for the above linear model.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>Sums of Squares</th>
<th>Mean Squares</th>
<th>F-Ratio</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals</td>
<td>111</td>
<td>1421.32</td>
<td>12.805</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td>1492.26</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2:

What distribution would you use to find the p-value above for Attr?

What hypotheses are being tested in the Attr row in the ANOVA table?
Non-Parametric Permutation Test Using $SS_{Group}$

What is our null hypothesis for a one-way ANOVA?

Think back to the permutation test you did for a difference in two means. How would you translate those steps into a one-way ANOVA setting for the MockJury data?

**Steps:**
1. Write the sentence lengths from all the women on cards, 114 total cards

2. Shuffle __________________________

3. Split into new groups of __________________________

4. Find the ________ of each new group

This is a similar idea to the permutation test you saw in the last chapter, but at the end we are not left with a difference in two means. We need a way to convert the variation in the group means into a single value, can you think of something that would work?

Below we have plotted the original group means against different shuffles (permutations).
Which plot shows the most evidence of a difference in group means?

Our process for a non-parametric permutation test is exactly the same as what we did for a non-parametric t-test, but we now have to save the sum of squares for Attr ($SS_{Group}$) of each permutation.

The process is:

1. Find the observed $SS_{Group}$ and save it as SSobs
2. Set-up how many permutations you want (1000 is usually good)
3. Set-up an empty matrix to store your shuffled $SS_{Group}$'s
4. Use a for loop:
   - For each number in the loop we shuffle the Attr of each observation (woman), keeping the original sample sizes (38, 38, 37)
   - Use the `anova` function of our linear model (`lm`) to get our ANOVA table
   - Pick-out the $SS_{Group}$ from the ANOVA table using matrix notation [1,2] (row one, column two)
SSobs <- anova(lm(Years ~ Attr, data = MockJury))[[1, 2]]
## Observed Sum of Squares for Attractiveness Pulling off the entry in the
## first row, second column of the ANOVA table

N <- length(MockJury$Attr)
## To get total sample size (114)
table(MockJury$Attr)

##
## Beautiful Average Unattractive
## 39  38  37

## To get group sample sizes

B <- 10000
## Number of permutations to do

SSstar <- matrix(NA, nrow = B)
## Empty vector to store permuted SS for Attr

MockJury$PermuteAttr <- NA
## Empty permuted Attr column in the dataset

for (b in 1:B) {
  beautiful <- sample(1:N, 39, replace = FALSE)
  ## To get a new 'Beautiful' group Sampling numbers 1-114 without replacement

  average <- sample(c(1:N)[-beautiful], 38, replace = FALSE)
  ## To get a new 'Average' group Sampling numbers 1-114 that were NOT drawn
  ## previously

  unattractive <- c(1:N)[-c(beautiful, average)]
  ## To get a new 'Unattractive' group Taking numbers 1-114 and eliminating
  ## those drawn previously

  MockJury$PermuteAttr[beautiful] <- "Beautiful"
  MockJury$PermuteAttr[average] <- "Average"
  MockJury$PermuteAttr[unattractive] <- "Unattractive"

  SSstar[b] <- anova(lm(Years ~ PermuteAttr, data = MockJury))[[1, 2]]
}

par(mfrow = c(1, 2))
## Two plots in one row

hist(SSstar, labels = TRUE)
## Histogram of Permuted Sum of Squares for Attractiveness
abline(v = SSobs, col = "red", lwd = 3)
## Observed Sum of Squares for Attractiveness

plot(density(SSstar), main = "Density Curve of SSstar")
## Density Curve of Permuted Sum of Squares for Attractiveness
How can we have a p-value of 0.068 when the difference between *Average* and *Unattractive* has a p-value of around 0.03?