Lab 3 – Completely Randomized ANOVA (a.k.a. One-way ANOVA)

FANR 6750

Richard Chandler and Bob Cooper

Today's Topics

One-way ANOVA

Scenario

- We have independent samples from $a > 2$ groups
- We assume the residuals are normally distributed with a mean of zero and a common variance

Questions

- Do the means differ?
- By how much? (What are the effect sizes?)

Null hypothesis

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_a$
- Or:
- $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_a = 0$

Additive model

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where:

- Residuals: $\varepsilon_{ij} \sim \text{Norm}(0, \sigma^2)$
- Group means: $\mu_i = \mu + \alpha_i$

Chain Saw Data

The data as 4 vectors

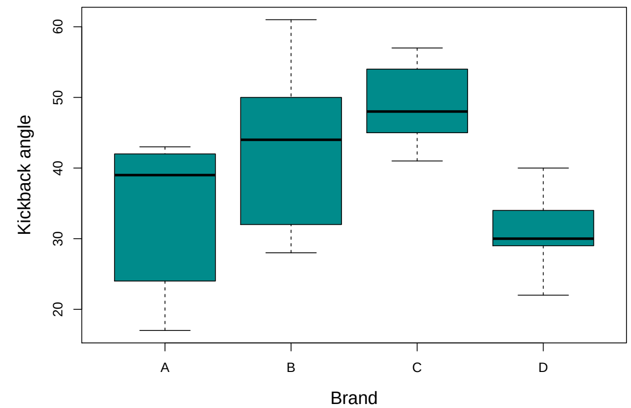
```r
kick.angle.brandA <- c(42,17,24,39,43)
kick.angle.brandB <- c(28,50,44,32,61)
kick.angle.brandC <- c(57,45,48,41,54)
kick.angle.brandD <- c(29,40,22,34,30)
```

Format as a `data.frame`

```r
n <- length(kick.angle.brandA)
a <- 4
sawData <- data.frame(
    Kick.angle=c(kick.angle.brandA, kick.angle.brandB,
        kick.angle.brandC, kick.angle.brandD),
    Brand=rep(c("A","B","C","D"), each=n))
```


Vizualize the data

```r
boxplot(Kick.angle ~ Brand, data=sawData, xlab="Brand",
        ylab="Kickback angle", cex.lab=1.3, col="darkcyan")
```



Two ANOVA functions: aov and lm

R has 2 common functions for doing ANOVA: `aov` and `lm`

We will primarily use `aov` in this class

Crude characterization

	aov	lm
Emphasis	ANOVA tables	Linear models
Typical use	Designed experiments	Regression analysis
Multiple error strata?	Yes	No

Using aov

Do the analysis

```r
aov.out1 <- aov(Kick.angle ~ Brand, data=sawData)
```

View the ANOVA table

```r
summary(aov.out1)

##             Df Sum Sq Mean Sq F value Pr(>F)
## Brand        3   1080   360.0   3.556 0.0382 *
## Residuals   16   1620   101.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estimates of means (μ's) and SE

```
model.tables(aov.out1, type="means", se=TRUE)

## Tables of means
## Grand mean
##
## 39
##
##  Brand
## Brand
##  A  B  C  D
## 33 43 49 31
##
## Standard errors for differences of means
##         Brand
##         6.364
## replic.     5
```


Estimates of effect sizes (α's) and SE

```
model.tables(aov.out1, type="effects", se=TRUE)

## Tables of effects
##
##  Brand
## Brand
##  A  B  C  D
## -6  4 10 -8
##
## Standard errors of effects
##         Brand
##           4.5
## replic.     5
```


Create ANOVA table by hand

Grand mean

```
ybar. <- mean(sawData$Kick.angle)
ybar.

## [1] 39
```

Find the group means, the hard way

```
ybar.i <- c(A=mean(kick.angle.brandA), B=mean(kick.angle.brandB),
            C=mean(kick.angle.brandC), D=mean(kick.angle.brandD))
ybar.i

##  A  B  C  D
## 33 43 49 31
```

Find the group means, the easier way

```
ybar.i <- tapply(sawData$Kick.angle, sawData$Brand, mean)
ybar.i

##  A  B  C  D
## 33 43 49 31
```


Sums of squares

Sum of squares among

```
SSa <- n*sum((ybar.i - ybar.)^2)
SSa

## [1] 1080
```

Sum of squares within

```
## Extract the response variable
y.ij <- sawData$Kick.angle
## Expand the group means and put them in the correct order
## This will only work if 'ybar.i' has names
ybar.ij <- ybar.i[as.character(sawData$Brand)]
SSw <- sum((y.ij - ybar.ij)^2)
SSw

## [1] 1620
```

Means squares and F statistic

Mean squares among

```
df1 <- a-1
MSa <- SSa / df1
MSa

## [1] 360
```

Mean squares within

```
df2 <- a*(n-1)
MSw <- SSw / df2
MSw

## [1] 101.25
```

F statistic

```
F.stat <- MSa / MSw
F.stat

## [1] 3.555556
```


Critical values and p-values

Critical value

```
F.crit <- qf(0.95, df1, df2)
F.crit

## [1] 3.238872
```

p-value

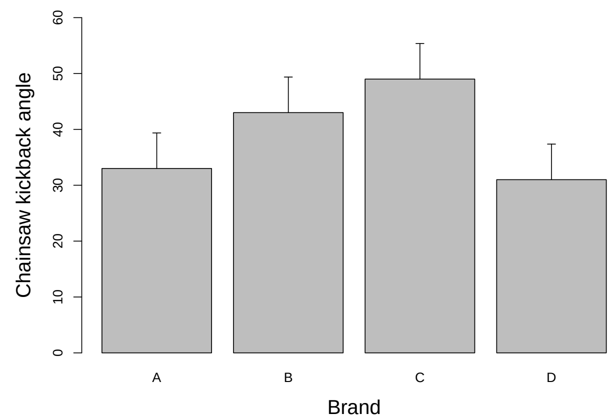
```
p.value <- 1 - pf(F.stat, df1, df2)
p.value

## [1] 0.03823275
```

Conclusion: Reject the null

Today's Topics

Group means +1 SE

```
xx <- barplot(ybar.i, ylim=c(0, 60), xlab="Brand", cex.lab=1.5,
              ylab="Chainsaw kickback angle")
mean.SE <- 6.364 # from model.tables(). See slide 9.
arrows(xx, ybar.i, xx, ybar.i+mean.SE, angle=90, length=0.05)
```

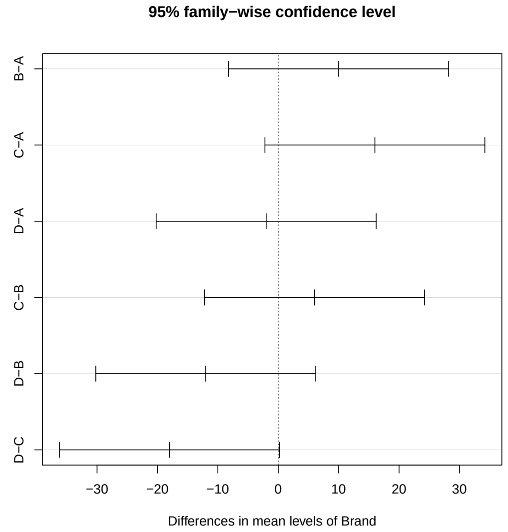


Tukey's Honestly Significant Difference test

```
TukeyHSD(aov.out1)

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Kick.angle ~ Brand, data = sawData)
##
## $Brand
##     diff        lwr        upr     p adj
## B-A   10  -8.207419 28.2074187 0.4213711
## C-A   16  -2.207419 34.2074187 0.0955690
## D-A   -2 -20.207419 16.2074187 0.9888365
## C-B    6 -12.207419 24.2074187 0.7826478
## D-B  -12 -30.207419  6.2074187 0.2726522
## D-C  -18 -36.207419  0.2074187 0.0532168
```


Plot Tukey's confidence intervals

```
plot(TukeyHSD(aov.out1))
```



Assignment

A biologist wants to compare the growth of four different tree species she is considering for use in reforestation efforts. All 32 seedlings of the four species are planted at the same time in a large plot. Heights in meters are recorded after several years. The data are in the file `treeHt.csv`:

Create an **R** script to do the following:

(1) Create an ANOVA table using the `aov` and `summary` functions.

(2) Create an ANOVA table (degrees of freedom, sums-of-squares, mean-squared error, and F-value) without using `aov`. Compute either the critical value of F or the p-value.

(3) Add a comment to the script indicating what the null and alternative hypotheses are, and whether the null can be rejected at the $\alpha = 0.05$ level.

(4) Use Tukey's HSD test to determine which pairs of means differ at the $\alpha = 0.05$ level. Add a comment, indicating which pairs are different.

(5) Create a barplot showing the means and SEs.