Lab 6 – Randomized Complete Block Design

September 24 & 25, 2018
FANR 6750

Richard Chandler and Bob Cooper

Recap

Like one-way ANOVA but experimental units are organized into blocks to account for extraneous sources of variation

Blocks could be regions, time periods, individual subjects, etc...

Blocking must occur during the design phase of the study

Additive model:

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

Gypsy Moth Data

```
gypsyData <- read.csv("gypsyData.csv")
gypsyData$region <- factor(gypsyData$region)
gypsyData

##    caterpillars pesticide region
## 1            16        Bt      1
## 2             3        Bt      2
## 3            10        Bt      3
## 4            18        Bt      4
## 5            25   Control      1
## 6            10   Control      2
## 7            15   Control      3
## 8            32   Control      4
## 9            14   Dimilin      1
## 10            2   Dimilin      2
## 11           16   Dimilin      3
## 12           12   Dimilin      4
```

Note: Numeric grouping variables must be coded as factors.

Compute the means

Grand mean ($\bar{y}_.$)

```
caterpillars <- gypsyData$caterpillars
(grand.mean <- mean(caterpillars))

## [1] 14.41667
```

Treatment means ($\bar{y}_i$)

```
pesticide <- gypsyData$pesticide
(treatment.means <- tapply(caterpillars, pesticide, mean))

##      Bt Control Dimilin
##   11.75   20.50   11.00
```

Block means ($\bar{y}_j$)

```
region <- gypsyData$region
(block.means <- tapply(caterpillars, region, mean))

##        1        2        3        4
## 18.33333  5.00000 13.66667 20.66667
```

Treatment sums-of-squares

$$b \times \sum_{i=1}^{a} (\bar{y}_i - \bar{y}_.)^2$$

```
b <- 4
b <- nlevels(region)
SS.treat <- b*sum((treatment.means - grand.mean)^2)
SS.treat

## [1] 223.1667
```


Block sums-of-squares

$$a \times \sum_{j=1}^{b} (\bar{y}_j - \bar{y}_.)^2$$

```
a <- nlevels(pesticide)
SS.block <- a*sum((block.means - grand.mean)^2)
SS.block

## [1] 430.9167
```


Within groups sums-of-squares

$$\sum_{i=1}^{a} \sum_{j=1}^{b} (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_.)^2$$

```
treatment.means.long <- rep(treatment.means, each=b)
block.means.long <- rep(block.means, times=a)
SS.within <- sum((caterpillars - treatment.means.long -
                  block.means.long + grand.mean)^2)
SS.within

## [1] 114.8333
```

NOTE: For this to work, `treatment.means` and `block.means` must be in the same order as in the original data.

Create ANOVA table

```
df.treat <- a-1
df.block <- b-1
df.within <- df.treat*df.block
ANOVAtable <- data.frame(
    df = c(df.treat, df.block, df.within),
    SS = c(SS.treat, SS.block, SS.within))
rownames(ANOVAtable) <- c("Treatment", "Block", "Within")
ANOVAtable

##           df       SS
## Treatment  2 223.1667
## Block      3 430.9167
## Within     6 114.8333
```

Create ANOVA table, continued...

Mean squares

```
MSE <- ANOVAtable$SS / ANOVAtable$df
ANOVAtable$MSE <- MSE
```

F values

```
F <- c(MSE[1]/MSE[3], MSE[2]/MSE[3], NA)
ANOVAtable$F <- F
```

P-values

```
P <- c(1 - pf(F[1], 2, 6), 1 - pf(F[2], 3, 6), NA)
ANOVAtable$P <- P
round(ANOVAtable, 3)

##            df      SS     MSE     F     P
## Treatment  2 223.167 111.583 5.830 0.039
## Block      3 430.917 143.639 7.505 0.019
## Within     6 114.833  19.139    NA    NA
```


Reminder about P-values

Critical value

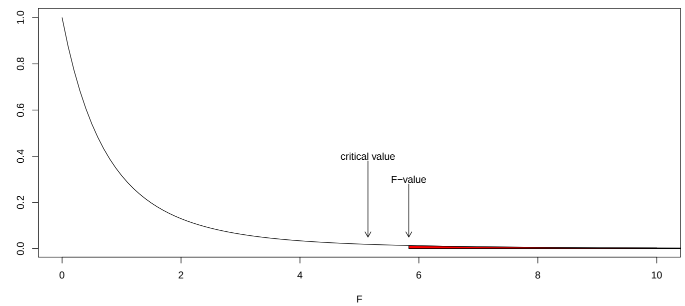
```
qf(0.95, df1=2, df2=6) # 95% of the distribution is before this value of F

## [1] 5.143253
```

P-value

```
1-pf(F[1], df1=2, df2=6) # Proportion of the distribution beyond this F value

## [1] 0.03921514
```



Using aov

```
aov1 <- aov(caterpillars ~ pesticide + region, gypsyData)
summary(aov1)

##             Df Sum Sq Mean Sq F value Pr(>F)
## pesticide    2  223.2  111.58   5.830 0.0392 *
## region       3  430.9  143.64   7.505 0.0187 *
## Residuals    6  114.8   19.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
round(ANOVAtable, 3)

##            df      SS     MSE     F     P
## Treatment  2 223.167 111.583 5.830 0.039
## Block      3 430.917 143.639 7.505 0.019
## Within     6 114.833  19.139    NA    NA
```


Using aov

Look what happens if we ignore the blocking variable

```
aov2 <- aov(caterpillars ~ pesticide, gypsyData)
summary(aov2)

##             Df Sum Sq Mean Sq F value Pr(>F)
## pesticide    2  223.2  111.58    1.84  0.214
## Residuals    9  545.8   60.64
```

Why is the effect of pesticide no longer significant?

Treating block effects as random effects

```
aov3 <- aov(caterpillars ~ pesticide + Error(region), gypsyData)
summary(aov3)

##
## Error: region
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  3  430.9   143.6
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## pesticide  2  223.2  111.58    5.83 0.0392 *
## Residuals  6  114.8   19.14
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The values of the ANOVA table are the same as before, and there is no reason to use random effects here if interest only lies in testing the null hypothesis concerning pesticides. Later, we will see cases where it is important to use random and fixed effects.

Assignment: Due before lab next week

Plantations of *Pinus caribaea* were established at four locations on Puerto Rico. Four spacings were used at each location to determine the effect of stocking density on tree height. Twenty years after the plantations were established, the following tree heights were recorded:

Location	Spacing (ft)	Height (ft)
Caracoles	5	72
	7	80
	10	85
	14	91
Utado	5	75
	7	90
	10	94
	14	112
Guzman	5	88
	7	95
	10	94
	14	91
Lares	5	79
	7	94
	10	104
	14	106

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

(1) What are the null and alternative hypotheses?.

(2) Test for effects of location and spacing on plant height using the `aov` function. Do the ANOVA again but without `aov`. Treat the block effects as fixed, not random. HINTS:
- Spacing must be treated as a factor.
- You must put the group means and block means in the correct order when computing the sums-of-squares.

(3) Perform a Tukey test to determine which spacings differ.

(4) Summarize the main results in 2-3 sentences. Upload the script to ELC the day before your next lab.