# Two Way ANOVA: Special Cases

Lecture 21

Alexandra Chronopoulou

ILLINOIS

COLLEGE OF LIBERAL ARTS & SCIENCES

Department of Statistics
101 Illini Hall, MC-374
725 S. Wright St.
Champaign, IL 61820-5710

## Unbalanced ANOVA

- When the treatment sample sizes are unequal, the analysis of variance for two-factor studies becomes more complex.
- The least-squares equations are no longer of a simple structure and the regular analysis of variance formulas are now inappropriate.
- Furthermore, the factor effect component sum of squares are no longer orthogonal; that is, they do not sum up to TSS.

- Consider the `rats` example from the previous lecture.
- Remove the first observation to make the data *unbalanced*.
- Use the `anova()` command for *each* models:
  (1) `poison*treat` - i.e. `poison` first, `treat` second
  (2) `treat*poison` - i.e. `treat` first, `poison` second
- Observe that the results change depending on the order the factors are introduced in the model.

```r
anova(lm(1/time ~ treat*poison, newrats))
```

```
## Analysis of Variance Table
##
## Response: 1/time
##               Df Sum Sq Mean Sq F value    Pr(>F)
## treat          3 20.136  6.7121 29.6807 9.986e-10 ***
## poison         2 35.102 17.5510 77.6094 1.362e-13 ***
## treat:poison   6  1.980  0.3300  1.4592    0.2207
## Residuals     35  7.915  0.2261
##
```

```r
anova(lm(1/time ~ poison*treat, newrats))
```

```
## Analysis of Variance Table
##
## Response: 1/time
##               Df Sum Sq Mean Sq F value    Pr(>F)
## poison         2 36.672 18.3358 81.0799 7.276e-14 ***
## treat          3 18.567  6.1889 27.3670 2.706e-09 ***
## poison:treat   6  1.980  0.3300  1.4592    0.2207
## Residuals     35  7.915  0.2261
##
```

# Regression Approach to 2-Factor ANOVA

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

- Due to the lack of orthogonality, the ANOVA $F$-tests are not applicable.
- We will express the ANOVA model as a regression model with indicator (dummy) variables.
- We need $a-1$ indicator variables for factor $A$ main effects and $b-1$ indicator variables for factor $B$ main effects. The interactions correspond to the cross products of the indicator variables for $A$ and $B$.

# Partial $F$-Test

$$\begin{cases} H_0 : \text{Smaller model} \\ H_\alpha : \text{Larger model} \end{cases}$$

- The partial $F$-test is:

$$F = \frac{(RSS_0 - RSS_\alpha)/(df_0 - df_\alpha)}{RSS_\alpha/df_\alpha} \sim F_{df_0 - df_\alpha, df_\alpha}$$

- Reject $H_0$, if $F$ test statistic is large, that is, the variation missed by the reduced model, when being compared with the error variance, is significantly large.

# Partial $F$-tests in R

```
ginv1 <- lm(time^-1 ~ poison*treat, data=newrats)
ginv2 <- lm(time^-1 ~ poison + treat, data=newrats)
anova(ginv2, ginv1)
```

```
## Analysis of Variance Table
##
## Model 1: time^-1 ~ poison + treat
## Model 2: time^-1 ~ poison * treat
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     41 9.8951
## 2     35 7.9151  6      1.98 1.4592 0.2207
```

```
ginv2 <- lm(time^-1 ~ poison + treat, data=newrats)
ginv3 <- lm(time^-1 ~ poison, data=newrats)
anova(ginv3, ginv2)
```

```
## Analysis of Variance Table
##
## Model 1: time^-1 ~ poison
## Model 2: time^-1 ~ poison + treat
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     44 28.4618
## 2     41  9.8951  3    18.567 25.644 1.669e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Type III Sums of Squares

- Use the Anova() command with *Type="III"* specification from the car library.
- This type tests for the presence of an effect given that both the other effects are in the model.

```
Anova(lm(1/time ~ treat*poison, data=newrats), type="III")
```

```
## Anova Table (Type III tests)
##
## Response: 1/time
##              Sum Sq Df F value    Pr(>F)
## (Intercept)  15.0605  1 66.5967 1.298e-09 ***
## treat         2.1340  3  3.1455   0.03723 *
## poison       11.7375  2 25.9514 1.225e-07 ***
## treat:poison  1.9800  6  1.4592   0.22073
## Residuals     7.9151 35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ILLINOIS

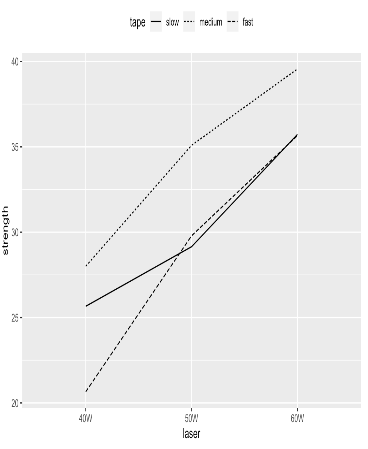
# Balanced ANOVA with $n = 1$

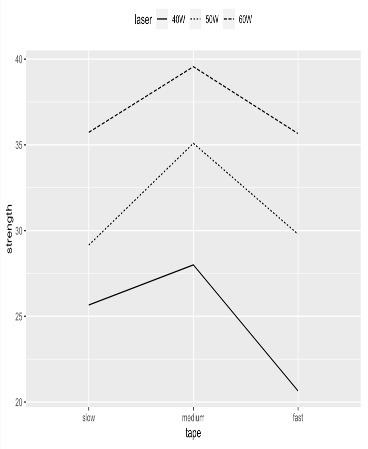
- Only one observation in each cell, so we cannot fit the interaction model.
- There are no degrees of freedom left for estimating the error.
- $RSS = 0$ when the model includes main effects and interaction term.
- All $F$-tests are valid, but the interaction model is not a candidate model.

```r
fullmodel <- lm(strength ~ laser*tape, composite)
summary(fullmodel)
```

```
##
## Call:
## lm(formula = strength ~ laser * tape, data = composite)
##
## Residuals:
## ALL 9 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)            25.66         NA      NA       NA
## laser50W                3.49         NA      NA       NA
## laser60W               10.07         NA      NA       NA
## tapemedium              2.34         NA      NA       NA
## tapefast               -5.01         NA      NA       NA
## laser50W:tapemedium     3.60         NA      NA       NA
## laser60W:tapemedium     1.49         NA      NA       NA
## laser50W:tapefast       5.65         NA      NA       NA
## laser60W:tapefast       4.94         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 8 and 0 DF,  p-value: NA
```

# Interaction Plots



```r
meffectmodel <- lm(strength ~ laser + tape, composite)
summary(meffectmodel)
```

```
##
## Call:
## lm(formula = strength ~ laser + tape, data = composite)
##
## Residuals:
##        1        2        3        4        5        6        7        8
##  1.74222 -1.34111 -0.40111  0.04556  0.56222 -0.60778 -1.78778  0.77889
##        9
##  1.00889
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   23.918      1.208  19.803 3.84e-05 ***
## laser50W       6.573      1.323   4.968 0.007661 **
## laser60W      12.213      1.323   9.231 0.000765 ***
## tapemedium     4.037      1.323   3.051 0.037991 *
## tapefast      -1.480      1.323  -1.119 0.325944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.62 on 4 degrees of freedom
## Multiple R-squared:  0.963,  Adjusted R-squared:  0.9259
## F-statistic:    26 on 4 and 4 DF,  p-value: 0.004013
```

## Tukey's Test for Additivity

- Consider the following model that includes interactions:

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta\ \alpha_i\ \beta_j + \epsilon_{ij}$$

- Here, we assume that the interactions are of *multiplicative* nature, i.e.

$$(\alpha\beta)_{ij} = \theta\alpha_i\ \beta_j$$

## SS Decomposition

- Consider the $SSA$, $SSB$ as before and :

$$SSAB^* = \frac{\left(\sum_i \sum_j (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) y_{ij}\right)^2}{\sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \ \sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}$$

- The $TSS$ is computed as usual and is decomposed as

$$TSS = SSA + SSB + SSAB^* + SSRem^*$$

where the remainder is

$$SSRem^* = TSS - SSA - SSB - SSAB^*$$

## Tukey's Test for Additivity

- We want to test the following hypothesis

$$\begin{cases} H_0: \ \theta = 0 \quad \text{(no interactions)} \\ H_\alpha: \ \theta \neq 0 \quad \text{(interactions)} \end{cases}$$

  which is essentially a test for model additivity.

- The test statistic computes as

$$F^* = \frac{SSAB^*/1}{SSRem^*/(ab - a - b)}$$

## Implementation of Tukey's Additivity Test in **R**

```
meffectmodel <- lm(strength ~ laser + tape, composite)
lasercoefs <- rep(c(0, 6.5733, 12.2133), 3)
tapecoefs <- rep(c(0, 4.0367, -1.4800), each=3)

newmod <- update(meffectmodel, .~. + I(lasercoefs*tapecoefs))
anova(newmod)
```

```
## Analysis of Variance Table
##
## Response: strength
##                             Df  Sum Sq Mean Sq F value   Pr(>F)
## laser                        2 224.184 112.092 36.8201 0.007745 **
## tape                         2  48.919  24.459  8.0344 0.062401 .
## I(lasercoefs * tapecoefs)   1   1.370   1.370  0.4501 0.550340
## Residuals                    3   9.133   3.044
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```