STAT 213
ANOVA Models With Interactions

Colin Reimer Dawson

Oberlin College

May 4-7, 2018

Notes

Outline

Notes

Outline

Notes

Additive Model

The Two-way ANOVA Additive Model (A, B categorical)

$$Y_i = f(A_i, B_i) + \varepsilon_i$$
$$Y_i = \mu + \alpha_{A_i} + \beta_{B_i} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

One α_A for each level of A

One β_B for each level of B

Assumes the "effect" of Factor A is the same at each level of Factor B (like parallel lines models in regression).

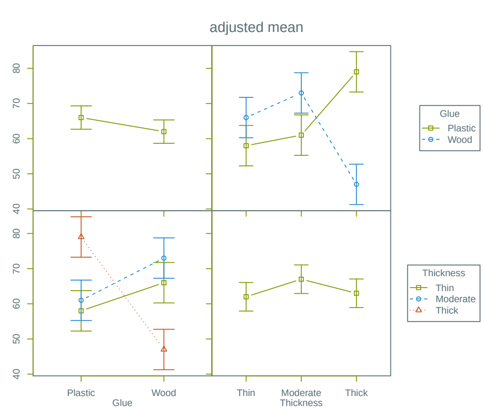
Notes

Example: Glue!

An experiment recorded the amount of force in Newtons (the response) that it takes to separate two pieces of plastic that have been glued together, for three different thicknesses of material (thin, moderate, thick), and two types of glue (wood vs. plastic). There are two cases at each combination of factors. The data is below usual one-row-per-case form.

Thickness/Glue	Plastic	Wood	**Mean**
Thin	$52, 64$	$72, 60$	62
Moderate	$67, 55$	$78, 68$	67
Thick	$86, 72$	$43, 51$	63
Mean	66	62	64

Notes



Notes

Additive Model: Parameter Estimates

```
additive <- aov(Force ~ Glue + Thickness, data = GlueData)
model.tables(additive, type = "effects")

     Tables of effects

      Glue
     Glue
     Plastic    Wood
           2      -2

      Thickness
     Thickness
         Thin Moderate    Thick
           -2        3       -1
```


Notes

Additive Model: Variance Decomposition

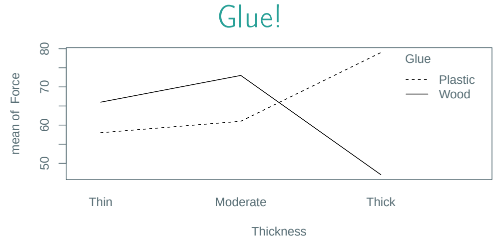
```
summary(additive)

                 Df Sum Sq Mean Sq F value Pr(>F)
     Glue         1     48    48.0   0.243  0.635
     Thickness    2     56    28.0   0.142  0.870
     Residuals    8   1580   197.5
```

From the model alone, looks like little evidence that either glue type or thickness matters

Notes

Glue!



- Effect of `Plastic` glue type is negative for thin materials; positive for thick
- Average effect close to zero
- The additive model can't capture this discrepancy: assumes the “effect” of glue type is the same at each thickness

Notes

Outline

Notes

Interaction Model

The Two-way ANOVA Interaction Model (A, B categorical)

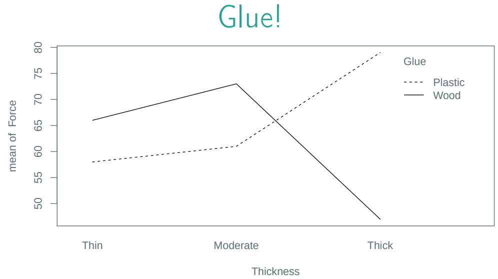
$$Y_i = f(A_i, B_i) + \varepsilon_i$$
$$Y_i = \mu + \alpha_{A_i} + \beta_{B_i} + \gamma_{A_i B_i} + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

One α_A for each level of A, and one β_B for each level of B
One γ_{AB} for each *combination* of A and B

- Predicted "effect" of level A_i of factor A, when at level B_i of factor B: $\alpha_{A_i} + \gamma_{A_i B_i}$.
- Predicted "effect" of level B_i of factor B, when at level A_i of factor A: $\beta_{B_i} + \gamma_{A_i B_i}$.
- "Effects" are modulated by the interaction term, $\gamma_{A_i B_i}$: an "adjustment to the adjustment".

Notes

Glue!



- Effect of `Plastic` glue type is negative for thin materials; positive for thick
- Average effect close to zero
- So $\alpha_{\texttt{Plastic}} \approx 0$, $\gamma_{\texttt{Plastic,Thin}} < 0$, $\gamma_{\texttt{Plastic,Thick}} > 0$

Notes

Glue!

We can write down a two-way ANOVA model with an interaction as follows:

$$Y_i = \mu + \alpha_{\texttt{Glue}_i} + \beta_{\texttt{Thickness}_i} + \gamma_{\texttt{Glue}_i,\texttt{Thickness}_i} + \varepsilon_i$$

How do we interpret each coefficient?

Notes

Concretely: Six Prediction Equations

$$\widehat{\texttt{Force}} = \begin{cases} \mu + \alpha_{\texttt{Wood}} + \beta_{\texttt{Thin}} + \gamma_{\texttt{Wood,Thin}} \\ \mu + \alpha_{\texttt{Wood}} + \beta_{\texttt{Moderate}} + \gamma_{\texttt{Wood,Moderate}} \\ \mu + \alpha_{\texttt{Wood}} + \beta_{\texttt{Thick}} + \gamma_{\texttt{Wood,Thick}} \\ \mu + \alpha_{\texttt{Plastic}} + \beta_{\texttt{Thin}} + \gamma_{\texttt{Plastic,Thin}} \\ \mu + \alpha_{\texttt{Plastic}} + \beta_{\texttt{Moderate}} + \gamma_{\texttt{Plastic,Moderate}} \\ \mu + \alpha_{\texttt{Plastic}} + \beta_{\texttt{Thick}} + \gamma_{\texttt{Plastic,Thick}} \end{cases}$$

Notes

Fitting the Model

```
interaction.model <- aov(Force ~ Thickness * Glue, data = GlueData)
model.tables(interaction.model, type = "means")

    Tables of means
    Grand mean

    64

     Thickness
    Thickness
        Thin Moderate    Thick
          62       67       63

     Glue
    Glue
    Plastic    Wood
         66      62

     Thickness:Glue
              Glue
    Thickness  Plastic Wood
      Thin     58      66
      Moderate 61      73
      Thick    79      47
```


Notes

Finding γs

Glue	Thickness	i	Force	$\hat{\mu}_{A,B}$	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\varepsilon}$
Wood	Thin	1	72	66	64	-2	-2	?	6
Wood	Thin	2	60	66	64	-2	-2	?	-6
Wood	Mod	1	78	73	64	-2	3	?	5
Wood	Mod	2	68	73	64	-2	3	?	-5
Wood	Thick	1	43	47	64	-2	-1	?	-4
Wood	Thick	2	51	47	64	-2	-1	?	4
Plastic	Thin	1	52	58	64	2	-2	?	-6
Plastic	Thin	2	64	58	64	2	-2	?	6
Plastic	Mod	1	67	61	64	2	3	?	6
Plastic	Mod	2	55	61	64	2	3	?	-6
Plastic	Thick	1	86	79	64	2	-1	?	7
Plastic	Thick	2	72	79	64	2	-1	?	-7

Since we are modeling each *combination* separately, we should have

$$\hat{\mu}_{A,B} = \hat{f}(A,B) = \hat{\mu} + \hat{\alpha}_A + \hat{\beta}_B + \hat{\gamma}_{A,B}$$

Notes

Parameter Estimates

```
model.tables(interaction.model, type = "effects")

    Tables of effects

     Thickness
    Thickness
        Thin Moderate    Thick
          -2        3       -1

     Glue
    Glue
    Plastic    Wood
          2      -2

     Thickness:Glue
              Glue
    Thickness  Plastic Wood
      Thin      -6        6
      Moderate  -8        8
      Thick     14      -14
```


Notes

Sums of Squares

Glue	Thickness	i	Force	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\varepsilon}$
Wood	Thin	1	72	64	-2	-2	6	6
Wood	Thin	2	60	64	-2	-2	6	-6
Wood	Mod	1	78	64	-2	3	8	5
Wood	Mod	2	68	64	-2	3	8	-5
Wood	Thick	1	43	64	-2	-1	-14	-4
Wood	Thick	2	51	64	-2	-1	-14	4
Plastic	Thin	1	52	64	2	-2	-6	-6
Plastic	Thin	2	64	64	2	-2	-6	6
Plastic	Mod	1	67	64	2	3	-8	6
Plastic	Mod	2	55	64	2	3	-8	-6
Plastic	Thick	1	86	64	2	-1	14	7
Plastic	Thick	2	72	64	2	-1	14	-7
					$SS_A = \sum \hat{\alpha}^2$	$SS_B = \sum \hat{\beta}^2$	$SS_{AB} = \sum \hat{\gamma}^2$	$SS_E = \sum \hat{\varepsilon}^2$

Notes

Simplifed Sums of Squares

$$SS_A = \sum_A \sum_B \sum_{i=1}^{n_{A,B}} \hat{\alpha}_A^2 = \sum_A n_{A\cdot} \hat{\alpha}_A^2$$

$$SS_B = \sum_A \sum_B \sum_{i=1}^{n_{A,B}} \hat{\beta}_B^2 = \sum_B n_{\cdot B} \hat{\beta}_B^2$$

$$SS_{AB} = \sum_A \sum_B \sum_{i=1}^{n_{A,B}} \hat{\gamma}_{A,B}^2 = \sum_A \sum_B n_{A,B} \hat{\gamma}_{A,B}^2$$

$$SS_E = \sum_A \sum_B \sum_{i=1}^{n_{A,B}} \hat{\varepsilon}_{A,B,i}^2 \text{ (doesn't simplify)}$$

$$SS_T = \sum_A \sum_B \sum_{i=1}^{n_{A,B}} (Y_{A,B,i} - \hat{\mu})^2 = SS_A + SS_B + SS_{AB} + SS_E$$

Notes

Degrees of Freedom

```
model.tables(interaction.model, type = "effects")

    Tables of effects

     Thickness
    Thickness
        Thin Moderate    Thick
          -2        3       -1

     Glue
    Glue
    Plastic    Wood
          2      -2

     Thickness:Glue
              Glue
    Thickness  Plastic Wood
      Thin      -6       6
      Moderate  -8       8
      Thick     14     -14
```


How many of each type of parameter are "free"?

Notes

Degrees of Freedom

With J levels of factor A and K levels of factor B:

- J different αs, but $J-1$ degrees of freedom (they must sum to zero[1])
- K different βs, but $K-1$ degrees of freedom (they must sum to zero[1])
- JK different γs, but only $(J-1)(K-1)$ df (must sum to zero at *each* J and *each* K[1])

[1] assuming balanced data

Notes

Error DF

Glue	Thickness	i	Force	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\varepsilon}$
Wood	Thin	1	72	64	-2	-2	6	6
Wood	Thin	2	60	64	-2	-2	6	-6
Wood	Mod	1	78	64	-2	3	8	5
Wood	Mod	2	68	64	-2	3	8	-5
Wood	Thick	1	43	64	-2	-1	-14	-4
Wood	Thick	2	51	64	-2	-1	-14	4
Plastic	Thin	1	52	64	2	-2	-6	-6
Plastic	Thin	2	64	64	2	-2	-6	6
Plastic	Mod	1	67	64	2	3	-8	6
Plastic	Mod	2	55	64	2	3	-8	-6
Plastic	Thick	1	86	64	2	-1	14	7
Plastic	Thick	2	72	64	2	-1	14	-7
					$SS_A = \sum \hat{\alpha}^2$	$SS_B = \sum \hat{\beta}^2$	$SS_{AB} = \sum \hat{\gamma}^2$	$SS_E = \sum \hat{\varepsilon}^2$

How many εs are "free"?

Notes

Degrees of Freedom

With J levels of factor A and K levels of factor B:

- J different αs, but $J-1$ degrees of freedom (they must sum to zero[2])
- K different βs, but $K-1$ degrees of freedom (they must sum to zero[2])
- JK different γs, but only $(J-1)(K-1)$ df (must sum to zero at *each* J and *each* K)
- With balanced data, $JK(n_{A,B}-1)$ error df (in general, $df_{Error} = df_{Total} - df_A - df_B - df_{AB}$, where $df_{Total} = N - 1$)

[2]assuming balanced data

Notes

ANOVA Table

```
summary(interaction.model)

                 Df Sum Sq Mean Sq F value Pr(>F)
Thickness         2     56      28   0.424 0.6725
Glue              1     48      48   0.727 0.4265
Thickness:Glue    2   1184     592   8.970 0.0157 *
Residuals         6    396      66
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Significant evidence that differences betwen glue strengths are modulated by material type
- Typically if an interaction is present, the "main effects" are not so useful
- Thus should interpret differences between glue types separately at each thickness

Notes

Following Up: "Simple" Effects

```
library(phia)
testInteractions(
    interaction.model,
    fixed="Thickness", across="Glue", adjustment = "bonferroni")

     F Test:
     P-value adjustment method: bonferroni
               Value Df Sum of Sq       F Pr(>F)
         Thin     -8  1        64  0.9697 1.0000
     Moderate    -12  1       144  2.1818 0.5703
        Thick     32  1      1024 15.5152 0.0229 *
     Residuals        6       396
     ---
     Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appears that we only have significant evidence for a difference between glue types with thick materials

Notes

Notes

Notes