

STAT 213

Two-Way ANOVA II

Colin Reimer Dawson

Oberlin College

May 2, 2018

1 / 21

Notes

Outline

Two-Way ANOVA: Additive Model

FIT: Estimating Parameters

ASSESS: Variance Decomposition

Pairwise Comparisons

2 / 21

Notes

Outline

Two-Way ANOVA: Additive Model

FIT: Estimating Parameters

ASSESS: Variance Decomposition

Pairwise Comparisons

3 / 21

Notes

Some students were interested in the effect of acidic environments on plant growth. They planted alfalfa seeds in fifteen cups and randomly chose five to get plain water, five to get a moderate amount of acid and five to get a stronger acid solution. The cups were arranged indoors near a window in five rows of three with one cup from each Acid level in each row (with row *a* nearest the window, and row *e* farthest away). The response variable was average Height of the alfalfa sprouts after four days.

A model:

$$\text{Height}_i = \mu + \alpha_{\text{Acid}_i} + \varepsilon_i$$

$$\text{Acid}_i \in \{\text{water, moderate, strong}\}$$

Any concerns about the ANOVA/regression conditions? **The residuals might not be independent within rows!**

Notes

Alfalfa Data

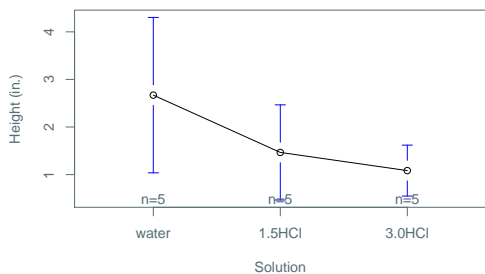
Treatment/Row	a	b	c	d	e	Trt. mean
water	1.45	2.79	1.93	2.33	4.85	2.67
moderate acid	1.00	0.70	1.37	2.80	1.46	1.47
strong acid	1.03	1.22	0.45	1.65	1.07	1.08
Row mean	1.16	1.57	1.25	2.26	2.46	1.74

Since each treatment is applied to each row, we can include row as a predictor.

Notes

Means Plots

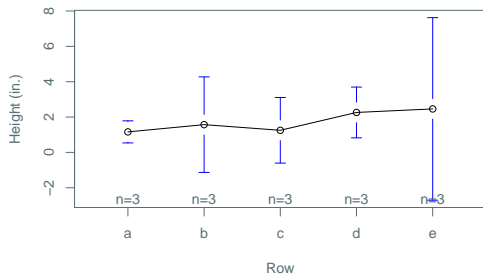
```
library("Stat2Data"); library("mosaic"); library("ggplots")
data("Alfalfa")
## Using factor() to reorder the categories
plotmeans(Ht4 ~ factor(Acid, levels = c("water", "1.5HCl", "3.0HCl")),
  data = Alfalfa, xlab = "Solution", ylab = "Height (in.)")
```



Notes

Means Plots

```
plotmeans(Ht4 ~ factor(Row), data = Alfalfa,
          xlab = "Row", ylab = "Height (in.)")
```



7 / 21

Notes

The One-way ANOVA Population Model (X categorical)

$$Y_i = f(X_i) + \varepsilon_i$$

$$Y = \mu + \alpha_{X_i} + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

One α_X for each level of X : group deviation from overall mean

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

$$Y_i = f(A_i, B_i) + \varepsilon_i$$

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

One α_A for each level of A ("row" deviation from overall mean)

One β_B for each level of B ("column" deviation from overall mean)

8 / 21

Notes

Concretely: Alfalfa Sprouts

$$\widehat{\text{Height}}_i = \begin{cases} \mu + \alpha_{\text{Water}} + \beta_a & \text{if Acid = Water and Row = a} \\ \mu + \alpha_{\text{Water}} + \beta_b & \text{if Acid = Water and Row = b} \\ \dots & \dots \\ \mu + \alpha_{\text{Water}} + \beta_e & \text{if Acid = Water and Row = e} \\ \dots & \dots \\ \mu + \alpha_{\text{HC11.5}} + \beta_a & \text{if Acid = HC11.5 and Row = a} \\ \dots & \dots \\ \mu + \alpha_{\text{HC11.5}} + \beta_e & \text{if Acid = HC11.5 and Row = e} \\ \mu + \alpha_{\text{HC13.0}} + \beta_a & \text{if Acid = HC13.0 and Row = a} \\ \dots & \dots \\ \mu + \alpha_{\text{HC13.0}} + \beta_e & \text{if Acid = HC13.0 and Row = e} \end{cases}$$

9 / 21

Notes

Outline

Two-Way ANOVA: Additive Model

FIT: Estimating Parameters

ASSESS: Variance Decomposition

Pairwise Comparisons

10 / 21

Notes

FIT: Parameter Estimation

- Population model:

$$y_{A,B,i} = \mu + \alpha_A + \beta_B + \varepsilon_{A,B,i}$$

where we let i index observations within combinations of A and B

- Estimate terms by

$$\begin{aligned} \hat{\mu} &= \bar{Y} \text{ ("grand" mean)} \\ \hat{\alpha}_A &= \bar{Y}_A - \bar{Y} \text{ ("row" deviation)} \\ \hat{\beta}_B &= \bar{Y}_B - \bar{Y} \text{ ("column" deviation)} \\ \hat{Y}_{A,B,i} &= \hat{\mu} + \hat{\alpha}_A + \hat{\beta}_B \text{ (predicted value)} \\ \hat{\varepsilon}_{A,B,i} &= Y_{A,B,i} - \hat{Y}_{A,B,i} \text{ (residual)} \end{aligned}$$

11 / 21

Notes

Practice: Alfalfa Data

Treatment/Row	a	b	c	d	e	Trt. mean
water	1.45	2.79	1.93	2.33	4.85	2.67
moderate acid	1.00	0.70	1.37	2.80	1.46	1.47
strong acid	1.03	1.22	0.45	1.65	1.07	1.08
Row mean	1.16	1.57	1.25	2.26	2.46	1.74

Find: $\hat{\mu}, \hat{\alpha}_{\text{water}}, \hat{\alpha}_{\text{moderate}}, \hat{\alpha}_{\text{strong}}, \hat{\beta}_a, \dots, \hat{\beta}_e$

12 / 21

Notes

Outline

Two-Way ANOVA: Additive Model

FIT: Estimating Parameters

ASSESS: Variance Decomposition

Pairwise Comparisons

Notes

Sums of Squares

$$Y_{A,B,i} = \hat{\mu} + \hat{\alpha}_A + \hat{\beta}_B + \varepsilon_{A,B,i}$$

$$(Y_{A,B,i} - \hat{\mu})^2 = (\hat{\alpha}_A + \hat{\beta}_B + \varepsilon_{A,B,i})^2$$

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

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

$$SS_{Error} = \sum_A \sum_B \sum_{i=1}^{n_{A,B}} \varepsilon_{A,B,i}^2 \text{ (doesn't simplify)}$$

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

Notes

Alfalfa: Sums of Squares

Treatment	Row	<i>i</i>	Height	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	ε
water	a	1	1.45	1.74	0.93	-0.58	
water	b	1	2.79	1.74	0.93	-0.17	
...	
water	e	1	4.85	1.74	0.93	0.72	
moderate	a	1	1.00	1.74	-0.27	-0.58	
...	
moderate	e	1	1.46	1.74	-0.27	0.72	
strong	a	1	1.03	1.74	-0.67	-0.58	
...	
strong	e	1	1.07	1.74	-0.67	0.72	

$$SS_A = \sum \hat{\alpha}^2 \quad SS_B = \sum \hat{\beta}^2 \quad SS_E = \sum \varepsilon^2$$

Notes

The Two-Way ANOVA Table

Source	df	SS	MS	F	P
Factor A	$J - 1$				
Factor B	$K - 1$				
Residuals	$N - J - K + 1$			--	--
Total	$N - 1$			--	--

Pairs: Factor A has $J = 3$ levels, factor B has $K = 5$ levels, with one observation per cell. How many degrees of freedom in each row of the table above?

16 / 21

Notes

Two-Way ANOVA Table

```
library(mosaic); library(Stat2Data); data(Alfalfa)
alfalfa.model <- aov(Ht4 ~ Acid + Row, data = Alfalfa)
summary(alfalfa.model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Acid	2	6.852	3.426	4.513	0.0487 *
Row	4	4.183	1.046	1.378	0.3235
Residuals	8	6.072	0.759		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Caution: The F tests here amount to sequential *nested* F -tests, so order matters if there is any collinearity (here there is none, since the design is perfectly balanced)

17 / 21

Notes

Getting Means

```
## Note: this only works if you used aov(), not lm()
model.tables(alfalfa.model, type = "means")
```

```
Tables of means
Grand mean
1.74

Acid
Acid
1.5HCl 3.0HCl water
1.466 1.084 2.670

Row
Row
a b c d e
1.16 1.57 1.25 2.26 2.46
```

18 / 21

Notes

Getting "Effects" (α s and β s)

```
## Note: this only works if you used aov(), not lm()
model.tables(alfalfa.model, type = "effects")

Tables of effects

Acid
Acid
1.5HCl 3.0HCl water
-0.274 -0.656 0.930

Row
Row
a b c d e
-0.58 -0.17 -0.49 0.52 0.72

## Notice that the alphas and betas each sum to zero
## This will happen when the data is perfectly balanced
## since overall average is unweighted average of group means
## (Otherwise the weighted sum is zero)
```

Notes

Outline

Two-Way ANOVA: Additive Model

FIT: Estimating Parameters

ASSESS: Variance Decomposition

Pairwise Comparisons

Notes

Post-Hoc Pairwise Comparisons

```
library(DescTools)
comparisons <- PostHocTest(alfalfa.model, method = "hpd", ordered = TRUE)
comparisons$Acid %>% round(3)

      diff lwr.ci upr.ci pval
1.5HCl-3.0HCl 0.392 -1.193 1.957 0.774
water-3.0HCl 1.586 0.011 3.161 0.048
water-1.5HCl 1.204 -0.371 2.779 0.134

comparisons$Row %>% round(3)

      diff lwr.ci upr.ci pval
c-a 0.09 -2.368 2.549 1.000
b-a 0.41 -2.048 2.868 0.975
d-a 1.10 -1.358 3.558 0.564
e-a 1.30 -1.158 3.758 0.421
b-c 0.32 -2.138 2.778 0.990
d-c 1.01 -1.448 3.468 0.633
e-c 1.21 -1.248 3.668 0.483
d-b 0.69 -1.768 3.148 0.861
e-b 0.89 -1.568 3.348 0.725
e-d 0.20 -2.258 2.658 0.998
```

Notes
