

# STAT 705 Chapter 19: Two-way ANOVA

Adapted from Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

This material is covered in Sections 19.2–19.4, but a bit differently.

We have *two* factors, A and B.

- Levels of A are indexed by  $i = 1, \dots, a$ .
- Levels of B are indexed by  $j = 1, \dots, b$ .
- The number of experimental units sampled when  $A = i$  and  $B = j$  is  $n_{ij}$ . If data are *balanced*, then  $n_{ij} \equiv n$  for all  $i, j$ .
- $n_T = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$ . For balanced data,  $n_T = nab$ .
- $Y_{ijk}$  is the  $k^{\text{th}}$  replicate of factor  $A = i$  &  $B = j$ .
- Let  $\mu_{ij} = E\{Y_{ijk}\}$ .

# Two-way ANOVA

We have  $ab$  different means.

Factor A	Factor B			
	1	2	...	$b$
1	$\mu_{11}$	$\mu_{12}$	...	$\mu_{1b}$
2	$\mu_{21}$	$\mu_{22}$	...	$\mu_{2b}$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
$a$	$\mu_{a1}$	$\mu_{a2}$	...	$\mu_{ab}$

The model is written

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2).$$

The most general, least restrictive case is when each combination of levels  $(i, j)$  has its own distinct mean. We'll look at special cases that impose *structure* on  $\{\mu_{ij}\}$ .

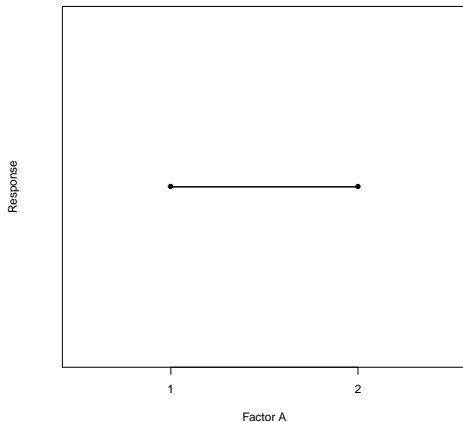
- I  $Y_{ijk} = \mu_{..} + \epsilon_{ijk}$
- II  $Y_{ijk} = \mu_{..} + \alpha_i + \epsilon_{ijk}$
- III  $Y_{ijk} = \mu_{..} + \beta_j + \epsilon_{ijk}$
- IV  $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk}$
- V  $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

# I. $\mu_{ij} = \mu_{..}$ ; neither A nor B are important

When  $a = b = 2$ , the means are

Factor A	Factor B	
	1	2
1	$\mu_{..}$	$\mu_{..}$
2	$\mu_{..}$	$\mu_{..}$

Intercept only model



# I. $\mu_{ij} = \mu_{..}$ ; neither A nor B are important

If this model fits, you're done! There is nothing further to look at.

Overall  $\mu_{..} = \sum_i \sum_j \mu_{ij} / ab$  is estimated by  $\hat{\mu}_{..} = \frac{1}{ab} \sum_i \sum_j \bar{Y}_{ij\cdot}$ , which will not equal  $\bar{Y}_{..}$  in the unbalanced case.

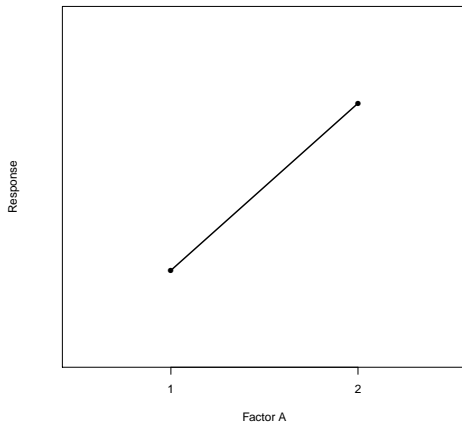
Fit in SAS `proc glm` as `model response = ;`

## II. $\mu_{ij} = \mu + \alpha_i$ , only A important

When  $a = b = 2$ , the means are

Factor A	Factor B	
	1	2
1	$\mu_{..} + \alpha_1$	$\mu_{..} + \alpha_1$
2	$\mu_{..} + \alpha_2$	$\mu_{..} + \alpha_2$

Factor A model



## II. $\mu_{ij} = \mu + \alpha_j$ , only A important

If this fits, we have a oneway model in A. We are interested in

$$L = \sum_{i=1}^a c_i \alpha_i.$$

SAS sets  $\hat{\mu}_{..} = \frac{1}{b} \sum_j \bar{Y}_{aj}$ . (not  $\bar{Y}_{a..}$ ) and  $\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{a..}$ .

Fit in SAS proc glm as model response = A;

We can get pairwise differences in factor A levels via, e.g. lsmeans A / pdiff adj=tukey;

General contrasts are available in estimate or lsmestimate.

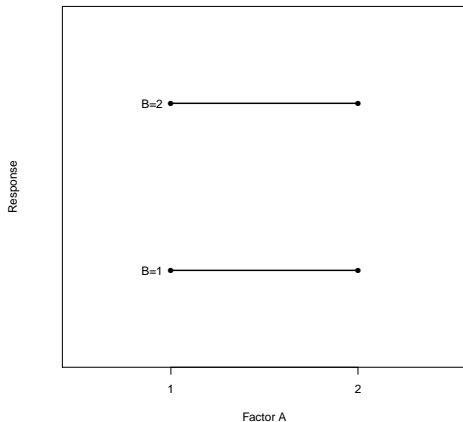


### III. $\mu_{ij} = \mu + \beta_j$ ; only B is important

When  $a = b = 2$ , the means are

Factor A	Factor B	
	1	2
1	$\mu_{..} + \beta_1$	$\mu_{..} + \beta_2$
2	$\mu_{..} + \beta_1$	$\mu_{..} + \beta_2$

Factor B model



### III. $\mu_{ij} = \mu + \beta_j$ , only B important

If this fits, we have a one-way model in B. We are interested in  $L = \sum_{j=1}^b c_j \beta_j$ .

SAS sets  $\hat{\mu}_{..} = \frac{1}{a} \sum_i \bar{Y}_{ib}$  and  $\hat{\beta}_j = \bar{Y}_{.j} - \bar{Y}_{.b.}$ .

Fit in SAS proc glm as model response = B;

We can get pairwise differences in factor B levels via, e.g. lsmeans B / pdiff adj=tukey;

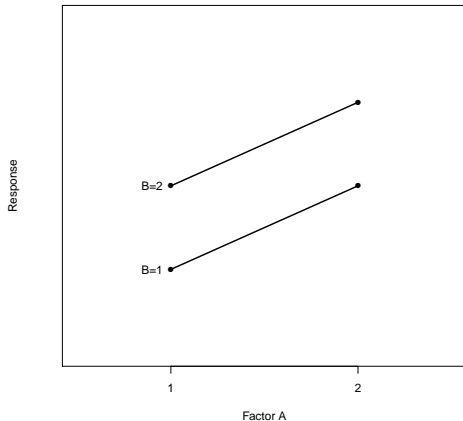
General contrasts are available in estimate or lsmestimate.

# IV. $\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$ , A and B additive

When  $a = b = 2$ , means are

Factor A	Factor B	
	1	2
1	$\mu + \alpha_1 + \beta_1$	$\mu + \alpha_1 + \beta_2$
2	$\mu + \alpha_2 + \beta_1$	$\mu + \alpha_2 + \beta_2$

Additive model



## IV. $\mu_{ij} = \mu + \alpha_i + \beta_j$ , both A and B important, but additive

Differences in factor A level means *are the same* for each level of B. Differences in factor B level means *are the same* for each level of A. For example, comparing mean differences for  $A = 1$  to  $A = 2$  we have

$$\mu_{1j} - \mu_{2j} = \mu + \alpha_1 + \beta_j - (\mu + \alpha_2 + \beta_j) = \alpha_1 - \alpha_2,$$

independent of  $j$ ! Similarly,  $\mu_{i1} - \mu_{i2} = \beta_1 - \beta_2$  indep. of  $i$ . SAS computes the LS estimates as  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , which doesn't simplify much in the unbalanced case. SAS sets  $\alpha_a = \beta_b = 0$ .

Fit in SAS proc glm as model response = A B;

We can get pairwise differences in factor A levels via, e.g. lsmeans A / pdiff adj=tukey;

We can get pairwise differences in factor B levels via, e.g. lsmeans B / pdiff adj=tukey;

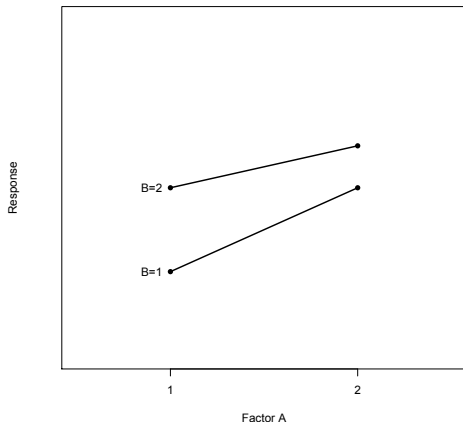
General forms  $L = \sum_{i=1}^a \sum_{j=1}^b c_{ij} \mu_{ij}$  can be computed in estimate or lsmestimate (more later).

# V. $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ , interaction model

When  $a = b = 2$ , means are

Factor A	Factor B	
	1	2
1	$\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}$	$\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}$
2	$\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}$	$\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}$

Interaction model



## V. $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ , interaction model

Now we have

$$\mu_{1j} - \mu_{2j} = \alpha_1 - \alpha_2 + \underbrace{(\alpha\beta)_{1j} - (\alpha\beta)_{2j}}_{\text{depends on } B = j} .$$

Also

$$\mu_{i1} - \mu_{i2} = \beta_1 - \beta_2 + \underbrace{(\alpha\beta)_{i1} - (\alpha\beta)_{i2}}_{\text{depends on } A = i} .$$

We no longer have parallel curves; mean differences in A *change* with levels of B and vice-versa.

SAS sets  $\alpha_a = \beta_b = 0$ ,  $(\alpha\beta)_{aj} = 0$  for  $j < b$ , and  $(\alpha\beta)_{ib} = 0$  for  $i < a$ .

Estimates can be obtained from solving  $\bar{Y}_{ij} = \hat{\mu}_{..} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij}$ .

## Comments on model V

- The interaction model gives each pairing  $(i, j)$  its own distinct mean, with no structure on  $\{\mu_{ij}\}$ . It's the same as a one-way model on  $r = ab$  groups.
- Your book focuses on the model where  $\sum_{i=1}^a \alpha_i = 0$ ,  $\sum_{j=1}^b \beta_j = 0$ ,  $\sum_{i=1}^a (\alpha\beta)_{ij} = 0$  for each  $j$ , and  $\sum_{j=1}^b (\alpha\beta)_{ij} = 0$  for each  $i$ . This model has enhanced interpretability, but is not straightforward to fit in SAS (PROC GENMOD to the rescue!).
- Interaction plots estimate the means using model V, e.g.  $\hat{\mu}_{ij} = \bar{Y}_{ij}$ . The overall shape of these plots give clues as to which is the most appropriate model from I, II, III, IV, V. More shortly.
- Fit in SAS using either model `response = A B A*B;` or `model response = A|B;`
- General forms  $L = \sum_{i=1}^a \sum_{j=1}^b c_{ij} \mu_{ij}$  can be computed in `estimate` or `lsmestimate` (more later).

# Model fitting

Say  $a = 3$ ,  $b = 2$ , and  $n_{ij} = n = 2$  for each pairing  $(i, j)$ . The parameters are  $\beta = (\mu_{..}, \alpha_1, \alpha_2, \beta_1, (\alpha\beta)_{11}, (\alpha\beta)_{21})$ . Note that  $\alpha_3 = \beta_2 = (\alpha\beta)_{12} = (\alpha\beta)_{22} = (\alpha\beta)_{32} = (\alpha\beta)_{31} = 0$ .

In general, the degrees of freedom for A, B, and A\*B are the number of free parameters associated with each of these:

$$\begin{array}{l} df_A = a - 1 \qquad (\alpha_1, \alpha_2, \dots, \alpha_{a-1}) \\ df_B = b - 1 \qquad (\beta_1, \beta_2, \dots, \beta_{b-1}) \\ df_{AB} = (a - 1)(b - 1) \quad \begin{array}{cccc} (\alpha\beta)_{11} & (\alpha\beta)_{12} & \cdots & (\alpha\beta)_{1,b-1} \\ (\alpha\beta)_{21} & (\alpha\beta)_{22} & \cdots & (\alpha\beta)_{2,b-1} \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha\beta)_{a-1,1} & (\alpha\beta)_{a-1,2} & \cdots & (\alpha\beta)_{a-1,b-1} \end{array} \end{array}$$



# Matrix formulation

As usual,  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ . For our example,

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_{..} \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{21} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{311} \\ \epsilon_{312} \\ \epsilon_{321} \\ \epsilon_{322} \end{bmatrix}.$$

Your textbook has a lot on fitting, parameter estimation under balance, etc. In general, easy-to-compute closed-form estimates do not exist, especially with unbalanced data. In that case, matrix algebra saves the day. The LS estimates are easily computed as  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ . Recall  $\hat{\beta} \sim N_p(\beta, (\mathbf{X}'\mathbf{X})^{-1}\sigma^2)$  where  $p$  is the number of mean parameters in  $\beta$ . Let  $\mathbf{c} = (c_1, \dots, c_p)$ . Then

$$\hat{L} = \mathbf{c}'\hat{\beta} \sim N(\mathbf{c}'\beta, \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}\sigma^2).$$

Any linear combination of mean parameters  $L$  is easily estimated. Recall that

$$\frac{\hat{L} - L}{\hat{\sigma}(\hat{L})} \sim t(n_T - p).$$

- You can estimate any linear combination of model parameters in  $\beta$  using estimate.
- Say  $a = 3$ ,  $b = 2$ . To estimate  $L_1 = \alpha_2 - \alpha_1$  use estimate 'L1' A -1 1 0;
- To estimate  $L_2 = \beta_2 - \beta_1 + (\alpha\beta)_{22} - (\alpha\beta)_{21}$  use estimate 'L2' B -3 0 3 A\*B 0 0 -1 1 0 0/divisor=3;
- To find the order of levels for main effects and interaction, look at the table of estimated coefficients.
- $\mu_{..}$  is called the intercept.
- Confidence intervals are obtained via the `clparm` option.

- By definition, the quantities estimated by least-squares means are either the raw  $\mu_{ij}$  or simple averages of these under any of the models I, II, III, IV, V. You can get estimates of these from `lsmeans`. For two-way models there are three, `lsmeans A`, `lsmeans B`, and `lsmeans A*B`, yielding estimates of

$$\bar{\mu}_{i.} = \frac{1}{b} \sum_{j=1}^b \mu_{ij}$$

$$\bar{\mu}_{.j} = \frac{1}{a} \sum_{i=1}^a \mu_{ij}$$

$$\mu_{ij}$$

- These are defined on p. 818 in your text; your book uses, e.g.  $\mu_{i.}$  instead of  $\bar{\mu}_{i.}$ .

- `pdiff` gives all pairwise differences in LS means. If the additive model IV fits, then `lsmeans A / pdiff;` gives all  $\bar{\mu}_{i_1.} - \bar{\mu}_{i_2.} = \alpha_{i_1} - \alpha_{i_2}$  and `lsmeans B / pdiff;` gives all  $\bar{\mu}_{.j_1} - \bar{\mu}_{.j_2} = \beta_{j_1} - \beta_{j_2}$ . You can adjust these using Tukey or Bonferroni.
- If V fits, and you are looking at a single factor A or B, then `pdiff` can still be used, but pairwise differences are *averaged* over the levels of the remaining factor(s); then  $\bar{\mu}_{i_1.} - \bar{\mu}_{i_2.} = \alpha_{i_1} - \alpha_{i_2} + (\bar{\alpha\beta})_{i_1.} - (\bar{\alpha\beta})_{i_2.} \neq \alpha_{i_1} - \alpha_{i_2}$  and `lsmeans` gives the former. `c1` adds confidence intervals; the intervals are adjusted if using, e.g. `adjust=tukey`.

- We can also look at `lsmeans A*B`; it gives all estimates of each  $\{\mu_{ij}\}$ . `pdiff` is indiscriminate; it gives all possible pairwise comparisons here too.
- Alternatively, we can specify `lsmeans A*B / slice=A` or `lsmeans A*B / slice=B` to look at all pairwise differences at *each level* of the other factor. The accompanying F tests aren't so useful; `slice A*B/sliceby=A` is a much better option.

## Grouping like factor levels

A “lines” plot shows groups of levels that are not significantly different from each other, usually with an overall FER, say  $FER = 0.05$ .

These can be obtained automatically in `proc glimmix` by adding a `lines` subcommand to `lsmeans`.

# Fuel Cell Example

Fuel cell performance is degraded when condensation builds up in the cell. In this study, some fuel cells are operated at target voltages with no intervention; others are dried out to remove condensation. Factor A is the condensation treatment,  $i = 1, 2$  for Control, Treatment. Factor B is voltage,  $j = 1, 2, 3, 4$  is 0.50V, 0.60V, 0.70V, 0.80V. The design is balanced, with  $n_{ij} = n = 2$  cells receiving one of the  $(i, j)$  pairings.

```
data fuelcell;
  input vapor $ voltage current @@;
datalines;
Trt 0.50 32.3  Ctl 0.50 28.6
Trt 0.50 32.8  Ctl 0.50 28.5
Trt 0.60 20.3  Ctl 0.60 19.5
Trt 0.60 21.2 Ctl 0.60 19.7
Trt 0.70 10.9 Ctl 0.70 10.6
Trt 0.70 11.1 Ctl 0.70 10.6
Trt 0.80  2.53 Ctl 0.80 2.23
Trt 0.80  2.43 Ctl 0.80 2.20
;

proc glimmix data=fuelcell;
class trt voltage;
model current=trt voltage / solution;
lsmeans voltage / pdiff adjust=tukey lines;
```



$n_T = 24$  programmers were asked to predict how long a big project would take in programmer-days. After the project was over, the prediction error was calculated:  $Y_{ijk} = (\text{actual programmer-days} - \text{predicted programmer-days})$ . Programmers were classified by type of experience ( $A = 1$  small systems,  $A = 2$  small & large); and experience ( $B = 1$  is  $< 5$  years,  $B = 2$  is  $5 - 10$  years,  $B = 3$  is  $\geq 10$  years).

```
data predict;
input days exper years @@;
datalines;
  240.0 1 1  206.0 1 1  217.0 1 1  225.0 1 1  110.0 1 2  118.0 1 2  103.0 1 2  95.0 1 2
  56.0 1 3  60.0 1 3  68.0 1 3  58.0 1 3  71.0 2 1  53.0 2 1  68.0 2 1  57.0 2 1
  47.0 2 2  52.0 2 2  31.0 2 2  49.0 2 2  37.0 2 3  33.0 2 3  40.0 2 3  45.0 2 3
;
run;
```

## Example pp. 870–871

```
* gives all a*b choose 2 pairwise comparisons via Tukey;
* slice subcommand only gives F-test within each slice;
proc glm data=predict plots=all; class exper years;
model days=exper|years;
lsmeans exper*years / adjust=tukey slice=exper;

* slice command in glimmix better;
* gives PW comparisons within each slice, i.e.;
* b choose 2 comparisons within each of a slices;
proc glimmix; class exper years;
model days=exper|years;
slice exper*years / sliceby=exper adjust=tukey cl;

proc glimmix; class exper years;
model days=exper|years; * adjust=tukey not needed;
slice exper*years / sliceby=years adjust=tukey cl;
```

- You can use `lsestimate` in `proc glimmix` to obtain inference for general linear combinations  $L = \sum_{i=1}^a \sum_{j=1}^b c_{ij} \mu_{ij}$  under any of the models.
- You can obtain simultaneous inference using Bonferroni, Scheffe, or Tukey (if pairwise differences).
- The theory behind the multiple comparisons is similar to that as from the one-way model, but a bit different; pp. 848–861 Sections 19.7 - 19.9. SAS takes care of the details for us. Just make sure you know what you are estimating.
- As before, Tukey works best when looking at all pairwise comparisons.
- Bonferroni works best when looking at only a few linear combinations, Scheffe can work better when looking at a large number of combinations.

# More Fuel Cell

```
proc glimmix data=fuelcell;
  class vapor voltage;
  model current=vapor|voltage / solution;
  *Cell means for both levels of vapor at voltage=0.5 and their contrast;
  estimate "Trt 0.50V" intercept 1 vapor 0 1 voltage 1 0 0 0 vapor*voltage 0 0 0 0 1 0 0 0,
  "Ctl 0.50V" intercept 1 vapor 1 0 voltage 1 0 0 0 vapor*voltage 1 0 0 0 0 0 0 0 ,
  "Trt vs Ctl 0.50V" vapor 1 -1 vapor*voltage 1 0 0 0 -1 0 0 0 / adjust=bon alpha=0.1 cl;
run;
```

```
proc glimmix data=fuelcell;
  class vapor voltage;
  model current=vapor|voltage / solution;
  lsmestimate vapor*voltage "Trt 0.50V" 0 0 0 0 1 0 0 0,
  "Ctl 0.50V" 1 0 0 0 0 0 0 0 0
  / adjust=bon alpha=0.1 cl;
run;
```

- The ANOVA table lists rows for Treatments (depending on which model you are fitting, I–V), Error, and Total as before.
- As usual, the Pythagorean Theorem tells us

$$\|\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{1}_{n_T}\bar{Y}_{..}\|^2 + \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \|\mathbf{Y} - \mathbf{1}_{n_T}\bar{Y}_{...}\|^2.$$

- $SSTR+SSE=SSTO$ .
- The p-value in the ANOVA table tests whether *anything* is important beyond a simple intercept  $\mu$ . For example, in the interaction model V, the F-test is for  $H_0 : \alpha_i = 0, \beta_j = 0, (\alpha\beta)_{ij} = 0$ . In model IV, whether  $H_0 : \alpha_i = 0, \beta_j = 0$ , etc.

## Type 3 sums of squares and associated tests

- There are also Type 3 sums of squares and associated tests. In interaction model V, there are SSA, SSB, and SSAB; these measure variability explained by the model due to factors A, B, and the AB interaction respectively (pp. 836–841). By definition, these are the differences in sums of squared errors comparing model V to (a) a model with B and A\*B, (b) a model with A and A\*B, and (c) a model with A and B. Only (c) is a hierarchical model, so that's the only test of interest here.
- If the data are balanced,  $n_{ij} = n$  for all  $i, j$ , then  $SSTR = SSA + SSB + SSAB$ . The book notes this (pp. 837–838) and discusses associated testing at some length.

## Type 3 sums of squares and associated tests

- Tim recommends fitting  $V$  and testing  $H_0 : (\alpha\beta)_{ij} = 0$ , i.e. whether the additive model fits. If IV fits, then interpretation simplifies. Furthermore, if IV fits, you may want to test whether you can drop A or B from model IV; these two Type 3 tests are given to you automatically after fitting IV via `model response=A B`; All of these are standard “nested linear hypotheses” type F-tests.
- Your book does not recommend refitting the model when you accept  $H_0 : (\alpha\beta)_{ij} = 0$ . This goes against the book’s own advice for fitting general regression models in STAT 704. Tim recommends using the additive model if you accept the interaction is not important; discussed briefly in 19.10.

## Type III test of no interaction in simple example

Recall the data where  $a = 3$ ,  $b = 2$ , and  $n_{ij} = n = 2$ . We want to test  $H_0 : (\alpha\beta)_{ij} = 0$  in full model  $V$ .  $\mathbf{Y} = \mathbf{X}_F\boldsymbol{\beta}_F + \boldsymbol{\epsilon}_F$ ; here

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{21} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{311} \\ \epsilon_{312} \\ \epsilon_{321} \\ \epsilon_{322} \end{bmatrix}.$$

The reduced model is  $\mathbf{Y} = \mathbf{X}_R\boldsymbol{\beta}_R + \boldsymbol{\epsilon}_R$ ; here

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{311} \\ \epsilon_{312} \\ \epsilon_{321} \\ \epsilon_{322} \end{bmatrix}.$$



## Type III test, continued...

For the full model  $\hat{\beta}_F = (\mathbf{X}'_F \mathbf{X}_F)^{-1} \mathbf{X}'_F \mathbf{Y}$ ,  
 $SSE(F) = \|\mathbf{Y} - \mathbf{X}_F \hat{\beta}_F\|^2$ ,  $df_{E(F)} = 12 - 6 = 6$ , and  
 $MSE(F) = SSE(F)/df_{E(F)}$ .

For the reduced model  $\hat{\beta}_R = (\mathbf{X}'_R \mathbf{X}_R)^{-1} \mathbf{X}'_R \mathbf{Y}$ ,  
 $SSE(R) = \|\mathbf{Y} - \mathbf{X}_R \hat{\beta}_R\|^2$ , and  $df_{E(R)} = 12 - 4 = 8$ .

Define

$$F^* = \frac{\{SSE(R) - SSE(F)\} / \{df_{E(R)} - df_{E(F)}\}}{MSE(F)}.$$

Then if  $H_0 : (\alpha\beta)_{11} = (\alpha\beta)_{21} = 0$  is true,

$$F^* \sim F(df_{E(R)} - df_{E(F)}, df_{E(F)}).$$

Use the Type III test for A\*B in SAS.

# Diagnostics and remedial measures

- Interaction plots are given from `plots=all` fitting model V. These will tell you which of models I–V are good candidates for the data.
- Residuals are defined as usual. For example, under model IV,  $e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - [\hat{\mu}_{..} + \hat{\alpha}_i + \hat{\beta}_j]$ . Look at least at  $e_{ijk}$  vs.  $\hat{Y}_{ijk}$ . You can plot the  $\{e_{ijk}\}$  vs. the indices  $i$  and  $j$  for two additional plots. Look at normal probability plot.
- If data have nonconstant variance but are normal, you can use `repeated / group=A*B`; in `proc mixed`. If variance only changes with levels of A or B, we can instead use `repeated / group=A`; or `repeated / group=B`;
- If data are nonnormal and have nonconstant variance, try a Box-Cox transformation of the  $Y_{ijk}$  in `proc transreg`. Sometimes a Box-Cox transformation of the  $Y_{ijk}$  can also get rid of a significant interaction (pp. 826–827).

- Check whether  $A$  and  $B$  interact with the Type 3 test. If not, base inference on the additive model or else model with  $A$  or  $B$  only. Typically one looks at pairwise mean differences from `lsmeans`.
- If  $A$  and  $B$  significantly interact, then you can examine pairwise differences of averaged effects, e.g.  $\bar{\mu}_{\cdot j_1} - \bar{\mu}_{\cdot j_2}$ , or else pairwise difference “slices”  $\mu_{ij_1} - \mu_{ij_2}$  for  $i = 1, \dots, a$ . These are interpreted differently. For example, the slices may be significant whereas the averaged differences may not.
- Check the appropriateness of the model with standard diagnostic plots. If have both non-constant variance and non-normal data, consider a Box-Cox transformation of the response. Often a Box-Cox transformation will also eliminate significant interactions.

# Outline of book's approach

- Everything in Chapter 19 requires balance  $n_{ij} = n$ . Balance is nice if you are computing ANOVA's by hand, but more often than not data are unbalanced and the nice formulae do not apply. Chapter 23 is an entire (short) chapter devoted to unbalanced analyses.
- 19.1 Three examples of designs leading to two-way ANOVA.
- 19.2 Interpretation of model, overall means, additive and interaction models, important and non-important interactions, transforming the data to get rid of an interaction.
- 19.3 Cell-means model with two factors.
- 19.4 Interaction model: fitting via least squares, partitioning sums of squares & degrees of freedom, an augmented ANOVA table with  $SSTR = SSA + SSB + SSAB$ .
- 19.5 Residual analysis.

# Outline of book's approach

- 19.6 F tests: three of them from fitting model V using augmented table. Kimball inequality:  
 $\alpha \leq 1 - (1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3)$  (has to do with controlling family-wise error rate on whether factors A, B and AB are important).
- 19.7 Strategy for analysis & flowchart.
- 19.8 Analyzing factor effects without an interaction (within the context of V!)
- 19.9 Analyzing factor effects with an interaction.
- 19.10 Pooling sums of squares, i.e. using IV instead of V when interaction not important.

# Focus on model V (as in textbook)

This is Minitab's model. LS parameter estimates minimize

$$Q(\mu_{..}, \alpha, \beta, \alpha\beta) = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - (\mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij}))^2,$$

subject to  $\alpha_{.} = \beta_{.} = (\alpha\beta)_{.i} = (\alpha\beta)_{.j} = 0$ . These are given by

$$\begin{aligned}\hat{\mu}_{..} &= \bar{Y}_{...} \\ \hat{\alpha}_i &= \bar{Y}_{i..} - \bar{Y}_{...} \\ \hat{\beta}_j &= \bar{Y}_{.j.} - \bar{Y}_{...} \\ \widehat{(\alpha\beta)}_{ij} &= \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}\end{aligned}$$

We have

$$\underbrace{\bar{Y}_{ij.} - \bar{Y}_{...}}_{\hat{\mu}_{ij} - \hat{\mu}_{..}} = \underbrace{\bar{Y}_{i..} - \bar{Y}_{...}}_{\hat{\alpha}_i} + \underbrace{\bar{Y}_{.j.} - \bar{Y}_{...}}_{\hat{\beta}_j} + \underbrace{\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}}_{\widehat{(\alpha\beta)}_{ij}}.$$

Fitted values are  $\hat{Y}_{ijk} = \bar{Y}_{ij.}$ . Estimates only for balanced data!

# Matrix formulation

For example considered earlier,

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{311} \\ Y_{312} \\ Y_{321} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_{..} \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ (\alpha\beta)_{11} \\ (\alpha\beta)_{21} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \\ \epsilon_{311} \\ \epsilon_{312} \\ \epsilon_{321} \\ \epsilon_{322} \end{bmatrix}.$$

Uses  $\alpha_3 = -\alpha_1 - \alpha_2$ ,  $\beta_2 = -\beta_1$ ,  $(\alpha\beta)_{12} = -(\alpha\beta)_{11}$ ,  
 $(\alpha\beta)_{22} = -(\alpha\beta)_{21}$ ,  $(\alpha\beta)_{31} = -(\alpha\beta)_{11} - (\alpha\beta)_{21}$ , and  
 $(\alpha\beta)_{32} = -(\alpha\beta)_{12} - (\alpha\beta)_{22} = (\alpha\beta)_{11} + (\alpha\beta)_{21}$ .

Recall model V is simply a one-way model with cell-means  $\{\mu_{ij}\}$ .

$$SSTO = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{..})^2$$

$$SSTR = \sum_{i=1}^a \sum_{j=1}^b n_{ij} (\bar{Y}_{ij} - \bar{Y}_{...})^2$$

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij})^2$$

Note that each deviation can be broken up as

$$\underbrace{Y_{ijk} - \bar{Y}_{...}}_{\text{ijk}^{\text{th}} \text{ deviation}} = \underbrace{\bar{Y}_{ij} - \bar{Y}_{...}}_{\text{explained by model}} + \underbrace{Y_{ijk} - \bar{Y}_{ij}}_{\text{left over}}$$



# Model V, balanced case $n_{ij} = n$ (p. 841)

We can show that  $SSTR = SSA + SSB + SSAB$  where

$$SSA = nb \sum_{i=1}^a (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SSB = na \sum_{j=1}^b (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$SSAB = n \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

SSA, SSB, and SSAB measure the portion of variability explained by the model due to Factors A, B, and interaction, respectively. This leads to a “refined ANOVA table”

Source	SS	df	MS	F	p-value
A	SSA	$a - 1$	$\frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$P\{F(a-1, (n-1)ab) > MSA/MSE\}$
B	SSB	$b - 1$	$\frac{SSB}{b-1}$	$\frac{MSB}{MSE}$	$P\{F(b-1, (n-1)ab) > MSB/MSE\}$
AB	SSAB	$(a-1)(b-1)$	$\frac{SSAB}{(a-1)(b-1)}$	$\frac{MSAB}{MSE}$	$P\{F((a-1)(b-1), (n-1)ab) > MSAB/MSE\}$
Error	SSE	$(n-1)ab$	$\frac{SSE}{(n-1)ab}$		
Total	SSTO	$nab - 1$			

We have three tests for  $H_0 : \alpha_i = 0$ ,  $H_0 : \beta_j = 0$ , and  $H_0 : (\alpha\beta)_{ij} = 0$  respectively. Only the last one, the test for the interaction, yields a hierarchical model if accepted.