

STAT 705 Chapter 19: Two-way ANOVA

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

Two-way ANOVA

Material covered in Sections 19.2–19.4, but a bit differently.

Have *two* factors, A and B.

- Levels of A indexed by $i = 1, \dots, a$.
- Levels of B indexed by $j = 1, \dots, b$.
- Number sampled when $A = i$ and $B = j$ is n_{ij} . If data are *balanced* then $n_{ij} = n$ for all i, j .
- $n_T = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$. If balanced, $n_T = nab$.
- Y_{ijk} is k th replicate of factor $A = i$ & $B = j$.
- Let $\mu_{ij} = E\{Y_{ijk}\}$.
- Have ab different means.

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

Model is written

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

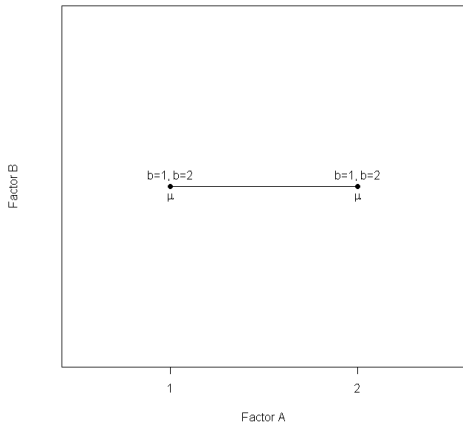
Most general, least restrictive case is when each combination of levels (i, j) has its own distinct mean. We'll look at special cases that impose *structure* on $\{\mu_{ij}\}$.

- 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 important

When $a = b = 2$, means are

Factor B	Factor A	
	1	2
1	μ	μ
2	μ	μ



I. $\mu_{ij} = \mu$, neither A nor B important

If this model fits, you're done! Nothing to look at.

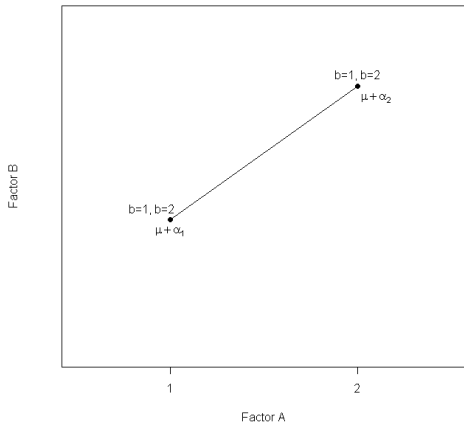
Overall μ is estimated by $\hat{\mu} = \bar{Y}_{\bullet\bullet\bullet}$.

Fit in SAS proc glm as model response = ;

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

When $a = b = 2$, means are

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



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

If this fits, have oneway model in A. Interested in $L = \sum_{j=1}^a c_j \alpha_j$.

SAS sets $\hat{\mu} = \bar{Y}_{a\bullet\bullet}$ and $\hat{\alpha}_j = \bar{Y}_{j\bullet\bullet} - \bar{Y}_{a\bullet\bullet}$.

Fit in SAS proc glm as model response = A;

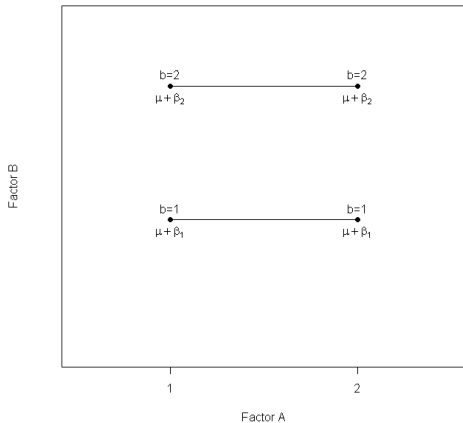
Can get pairwise differences in factor A levels via, e.g. lsmeans A / pdiff adj=tukey;

General contrasts available in estimate or lsestimate.

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

When $a = b = 2$, means are

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



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

If this fits, have oneway model in B. Interested in $L = \sum_{j=1}^b c_j \beta_j$.

SAS sets $\hat{\mu} = \bar{Y}_{\bullet b \bullet}$ and $\hat{\beta}_j = \bar{Y}_{\bullet j \bullet} - \bar{Y}_{\bullet b \bullet}$.

Fit in SAS proc glm as model response = B;

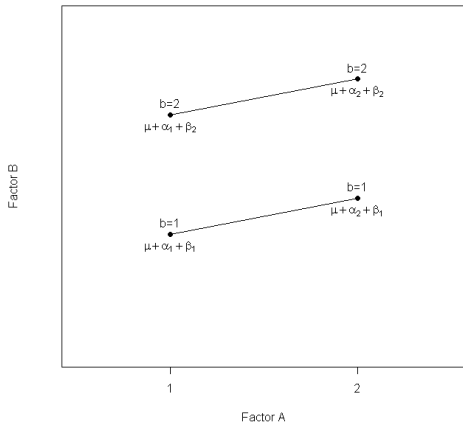
Can get pairwise differences in factor B levels via, e.g. lsmeans B / pdiff adj=tukey;

General contrasts available in estimate or lsestimate.

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

When $a = b = 2$, means are

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



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 $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, doesn't simplify much. SAS sets $\alpha_a = \beta_b = 0$.

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

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

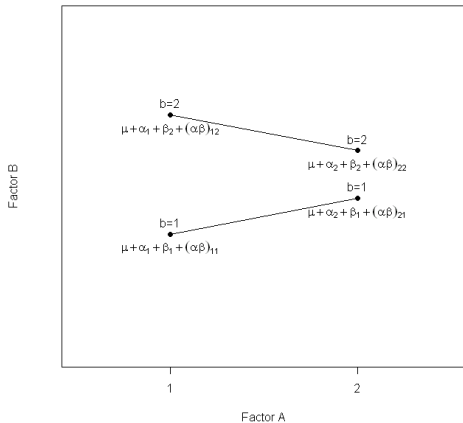
Can get pairwise differences in factor B levels via, e.g. lsmeans B / pdiff adj=tukey;

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



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

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

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

Comments on model V

- Interaction model gives each pairing (i, j) it's own distinct mean, no structure on $\{\mu_{ij}\}$. Same as oneway 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.
- Interaction plots estimate the means using model V, e.g. $\hat{\mu}_{ij} = \bar{Y}_{ij\bullet}$. 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:

$$\begin{array}{ll} dfA = a - 1 & (\alpha_1, \alpha_2, \dots, \alpha_{a-1}) \\ dfB = b - 1 & (\beta_1, \beta_2, \dots, \beta_{b-1}) \\ dfAB = (a - 1)(b - 1) & \begin{array}{cccc} (\alpha\beta)_{11} & (\alpha\beta)_{12} & \cdots & (\alpha\beta)_{1,b-1} \\ (\alpha\beta)_{21} & (\alpha\beta)_{22} & \cdots & (\alpha\beta)_{2,b-1} \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha\beta)_{a-1,1} & (\alpha\beta)_{a-1,2} & \cdots & (\alpha\beta)_{a-1,b-1} \end{array} \end{array}$$

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 β . 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

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

- You can estimate any linear combination of model parameters in β using estimate.
- Say $a = 3$, $b = 3$. To estimate $L_1 = \alpha_2 - \alpha_1$ use estimate 'L1' A -1 1 0;
- To estimate $L_2 = \beta_3 - \beta_1 + (\alpha\beta)_{23} - (\alpha\beta)_{21}$ use estimate 'L2' B -1 0 1 A*B 0 0 0 -1 0 1 0 0 0;
- To find order of levels for main effects and interaction, look at table of estimated coefficients.
- μ is called the intercept.
- Confidence intervals obtained via `clparm` option.

- By definition, the least-squares means are either the raw μ_{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

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

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

$$\mu_{ij}$$

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

- `pdiff` gives all pairwise differences in LS means. If additive model IV fits, then `lsmeans A / pdiff;` gives all $\bar{\mu}_{i_1\bullet} - \bar{\mu}_{i_2\bullet} = \alpha_{i_1} - \alpha_{i_2}$ and `lsmeans B / pdiff;` gives all $\bar{\mu}_{\bullet j_1} - \bar{\mu}_{\bullet j_2} = \beta_{j_1} - \beta_{j_2}$. You can adjust these using Tukey or Bonferroni.
- If V fits, and 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 $\bar{\mu}_{i_1\bullet} - \bar{\mu}_{i_2\bullet} \neq \alpha_{i_1} - \alpha_{i_2}$ and `lsmeans` gives the former. `cl` adds confidence intervals; the intervals are adjusted if using, e.g. `adjust=tukey`.
- Can also look at `lsmeans A*B`. Gives all estimates of each $\{\mu_{ij}\}$, `pdiff` gives all possible pairwise comparisons here too.
- Alternatively 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.

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`.

For example,

```
proc glimmix data=bakery;  
  class height width;  
  model sales=height width / solution;  
  lsmeans height / pdiff adjust=tukey lines;
```

SAS lsestimate command

- 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 19.8 & 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.

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

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

- $SSTR+SSE=SSTO$.
- The p-value in the ANOVA table tests whether *anything* is important beyond a simple intercept μ . 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
 $H_0 : \alpha_i = 0, \beta_j = 0$, 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 model due to factors A, B, and 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. 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 data where $a = 3$, $b = 2$, and $n_{ij} = n = 2$. Want to test $H_0 : (\alpha\beta)_{ij} = 0$ in full model $\mathbf{Y} = \mathbf{X}_F\boldsymbol{\beta}_F + \boldsymbol{\epsilon}_F$; here

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

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

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 $dfE(R) = 12 - 4 = 8$.

Define

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

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

$$F^* \sim F(dfE(R) - dfE(F), dfE(F)).$$

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

Castle Bakery (p. 833)

Castle Bakery supplies Italian bread to stores in a large city. They want to sell more bread so they designed an experiment. Factor A is shelf height, $i = 1, 2, 3$ for bottom, middle, or top. Factor B is width of the display, $j = 1, 2$ is regular, wide. The design is balanced, with $n_{ij} = n = 2$ stores receiving one of the (i, j) pairings.

```
data bakery;
  input sales height width store;
datalines;
  47 1 1 1
  43 1 1 2
  46 1 2 1
  40 1 2 2
  62 2 1 1
  68 2 1 2
  67 2 2 1
  71 2 2 2
  41 3 1 1
  39 3 1 2
  42 3 2 1
  46 3 2 2
;

* initial fit of model V;
proc glm data=bakery plots=all;
  class height width;
  model sales = height width height*width / solution;
run;
```

This duplicates your textbook's analyses:

```
* Example 1, pp. 853-855;
proc glm data=bakery;
  class height width;
  model sales=height width height*width / solution;
  lsmeans height / pdiff adjust=tukey alpha=0.05 cl;
run;

* Example 2, p. 855;
proc glimmix data=bakery;
  class height width;
  model sales=height width height*width / solution;
  lsmestimate height*width "reg width middle" 0 0 1 0 0 0,
    "reg width high " 0 0 0 0 1 0
    / adjust=bon alpha=0.1 cl;
run;
```

Alternatively, you could use `model sales = height width` in these commands instead of `sales=height width height*width`, following Tim's advice.

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} . 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, can use `repeated / group=A*B`; in `proc mixed`. If variance only changes with levels of A or B, 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).

Strategy for analysis

- Check whether A and B interact with Type 3 test. If not, base inference on 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. $\bar{\mu}_{\bullet j_1} - \bar{\mu}_{\bullet j_2}$, 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 Box-Cox transformation of the response. Often a Box-Cox transformation will also eliminate significant interactions.

Example pp. 870–871

$n_T = 24$ programmers asked to predict how long a big project would take in programmer-days. After project over Y_{ijk} is actual minus predicted programmer-days (prediction errors).
Programmers 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 ≥ 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
;
* 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;
```


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 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 only fitting model V using augmented table. Kimball inequality $\alpha \leq 1 - (1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3)$ (has to do with doing sequential tests on whether factors 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 $\alpha_{\bullet} = \beta_{\bullet} = (\alpha\beta)_{i\bullet} = (\alpha\beta)_{\bullet j} = 0$. These are given by

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

We have

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

Fitted values are $\hat{Y}_{ijk} = \bar{Y}_{ij\bullet}$. 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 $\alpha_3 = -\alpha_1 - \alpha_2$, $\beta_2 = -\beta_1$, $(\alpha\beta)_{12} = -(\alpha\beta)_{11}$,

$(\alpha\beta)_{22} = -(\alpha\beta)_{12}$, and

$(\alpha\beta)_{32} = -(\alpha\beta)_{12} - (\alpha\beta)_{22} = (\alpha\beta)_{11} + (\alpha\beta)_{12}$.

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

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

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

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

Note that each deviation can be broken up as

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

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

Can show that $SSTR = SSA + SSB + SSAB$ where

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

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

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

SSA, SSB, and SSAB measure portion of variability explained by model due to Factors A, B, and interaction, respectively. 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	$nab - 1$			

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