

The effects model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk}, \quad \begin{array}{l} i = 1, \dots, p \text{ (treatments)} \\ j = 1, \dots, p \text{ (time periods).} \\ k = 1, \dots, n \text{ (individuals)} \end{array}$$

where we make the usual assumptions that $e_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2)$

Parameter interpretations:

- μ : is the overall mean
- α_i : is the effect of the i^{th} treatment
- β_j : is the effect of the j^{th} time period
- γ_k : is the effect of the k^{th} individual (mouse)

Constraints:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$$

In the above model we have treated subject effects as fixed. This is occasionally appropriate, but usually subject effects should be random.

Under random subject effects, we assume $\gamma_1, \dots, \gamma_n \stackrel{iid}{\sim} N(0, \sigma_\gamma^2)$ and we require constraints only on the α_i 's and the β_j 's.

- Notice that our model is purely additive one with no interactions. Therefore, as in the RCBD, treating subject effects as random versus fixed affects the interpretation of the model and standard errors and confidence intervals on treatment means, but not the ANOVA or F tests for main effects or contrasts on the treatment factor.

ANOVA Table:

Source of Variation	Sum of Squares	d.f.	F
Treatments	$n \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$	$p - 1$	$\frac{MS_{Trt}}{MS_E}$
Times	$n \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$	$p - 1$	
Subjects	$p \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2$	$n - 1$	
Error	SS_E (by subtraction)	$(n - 2)(p - 1)$	
Total	$\sum_{ijk} (y_{ijk} - \bar{y}_{...})^2$	$np - 1$	

We reject the hypothesis that the treatment means are equal (i.e., $H_0 : \mu_1 = \dots = \mu_p$) if

$$F = \frac{MS_{Trt}}{MS_E} > F_\alpha(p - 1, (n - 2)(p - 1)).$$

- Notice that although the design is balanced with respect to carry-over effects, carry-over effects aren't explicitly accounted for in the model.
- Therefore, if carry-over effects are present, their effect will be lumped into the MS_E rather than separated out as a distinct source of variability.
- For this reason, the analysis that I've presented here is recommended only for situations in which carry-over effects are assumed to be small or nonexistent.
- In crossover designs, a **washout period** between treatment applications is often used to eliminate carry-over effects.
- If carry-over effects are present, more complex analyses of crossover data are appropriate, but such methods are beyond the scope of this course.
- Note that the heifers example which we presented as a replicated Latin square design, is in fact a crossover design. The ANOVA table we presented for that design is the same as that the one given above, except that the "columns" (heifers) and "squares" (farms) sources of variability have been collapsed (combined) as "subjects" in the ANOVA at the top of the page. This has no effect on inferences for treatments and rows (time periods), so the analyses are equivalent.

Example – Mice:

Two contrasts are of interest in this example:

$$\begin{array}{rcccc} & & \text{A} & \text{B} & \text{C} \\ \psi_1 : & 1 & 1 & -2 & & \text{(control vs. drugs)} \\ \psi_2 : & 1 & -1 & 0 & & \text{(Drug A vs. Drug B)} \end{array}$$

- See `mazemice.sas` and `mazemice.pdf`. Here we will consider effects for mice as random rather than fixed, since we are not interested in the results for particular mice, but instead regard them as representative of the population of all mice that might be used to run the maze under the experimental conditions. Since mice will be considered random, we will use PROC MIXED to ensure that we produce the right standard errors for the treatment means.
- The F statistic for main effects of treatments is highly significant ($F = 99.70$, $p < 0.0001$), indicating that the average time to complete the maze differs for at least one pair of treatments.
- The contrasts are both significant indicating that there is a difference between the average time to completion under the control treatment and the average time to completion under the influence of either of the two drugs ($F = 149.54$), and that the average time to completion under drug A differs from the average time to completion under drug B ($F = 49.85$).
- From the least squares means, we can see that the drugs help the mice in the maze task, and that drug A helps the mice more than drug B.
- In PROC MIXED, the LSMEANS statement automatically produces standard errors. The `/CL` option requests $100(1 - \alpha)\%$ confidence intervals for the lsmeans. α is set with the `ALPHA=` option on the LSMEANS statement. So, here we obtain 90% confidence intervals for the treatment means. For example, a 90% confidence interval for the mean time to completion among undrugged mice is

$$\bar{y}_{3..} \pm t_{\alpha/2}((n - 2)(p - 1))\text{s.e.}(\bar{y}_{3..})$$

Since

$$\begin{aligned}
\text{var}(\bar{y}_{i..}) &= \text{var} \left(\frac{1}{np} \sum_{j=1}^p \sum_{k=1}^n y_{ijk} \right) \\
&= \text{var} \left(\frac{1}{np} \sum_{j=1}^p \sum_{k=1}^n \{ \mu + \alpha_i + \beta_j + \gamma_k + e_{ijk} \} \right) \\
&= \text{var} (\mu + \alpha_i + \bar{\beta}_{.} + \bar{\gamma}_{.} + \bar{e}_{i..}) \\
&= \text{var} (\bar{\gamma}_{.} + \bar{e}_{i..}) \\
&= \text{var} (\bar{\gamma}_{.}) + \text{var} (\bar{e}_{i..}) \\
&= \frac{\sigma_{\gamma}^2}{n} + \frac{\sigma^2}{n}
\end{aligned}$$

it follows that

$$\text{s.e.}(\bar{y}_{i..}) = \sqrt{\frac{\hat{\sigma}_{\gamma}^2}{n} + \frac{\hat{\sigma}^2}{n}}$$

so that

$$\text{s.e.}(\bar{y}_{3..}) = \sqrt{\frac{37.9096}{12} + \frac{16.1500}{12}} = 2.1225$$

which agrees with the result on the bottom of p.2 of mazemice.lst. The 90% confidence interval for μ_3 is then given by

$$\begin{aligned}
&\bar{y}_{3..} \pm t_{\alpha/2}((n-2)(p-1))\text{s.e.}(\bar{y}_{3..}) \\
&= 81.5833 \pm \underbrace{t_{.05}((12-2)(3-1))}_{=1.725}(2.1225) = 81.5833 \pm 3.661 \\
&= (77.92, 85.24)
\end{aligned}$$

which also agrees with the result given by PROC MIXED given on the bottom of p.2 of mazemice.pdf.

Incomplete Block Designs

In some situations in which we would like to perform an experiment in complete blocks, we are unable to observe all treatments in each block. Typically, we would be limited by cost, size, or other constraints at the block level. In this situation experimenters often use **incomplete block designs** in which fewer than the full number of treatments occur in each block.

Balanced Incomplete Block Designs:

When all treatment comparisons are equally important the treatments combinations observed in the blocks should be balanced. That is, each pair of treatments should occur together in the same block the same number of times as each other pair of treatments. When this is the case we have a **Balanced Incomplete Block Design**, or BIBD.

E.U. 1		E.U. 1
E.U. 2		E.U. 2
⋮		⋮
E.U. k		E.U. k
Block 1	⋯	Block b

In the BIBD, there are

- a treatments in the experiment
- $b > r$ blocks in the experiment
- $k < a$ treatments in each block
- r occurrences of each treatment in the experiment

There are a total of $N = ar = bk$ observations in the experiment.

Since this is the *Balanced Incomplete Block Design*, each pair of treatments occurs together the same number of times, λ . To obtain an expression for λ in terms of a, b, k, r we reason as follows:

Consider a particular treatment, treatment 1, say. Treatment 1 appears in r blocks and there are $k - 1$ other treatments in each of those blocks; so there are $r(k - 1)$ pairs in the design involving treatment 1. Only $(\frac{1}{a-1})^{\text{th}}$ of these involve treatment 2, $(\frac{1}{a-1})^{\text{th}}$ of these involve treatment 3, etc. So there are

$$\lambda = \frac{r(k - 1)}{a - 1}$$

occurrences of each pair of treatments in the experiment.

λ must be an integer. This restricts r, k, b , and a somewhat. That is, a BIBD does not exist for all choices of r, k, b and a .

- In particular, balance may not be possible for a given choice of a and k unless the overall size of the experiment is very large (perhaps larger than we want).

E.g., for $a = 8$ treatments and block size of $k = 5$, we need $\lambda = r(k - 1)/(a - 1) = 4r/7$ to be a whole number. This requires r to be a multiple of 7, $r = 7x$, say.

We also need $b = ra/k = (7x)8/5$ to be a whole number, so x must be a multiple of 5 \Rightarrow the smallest BIBD for this situation has $r = 7(5) = 35$ and $b = (35)8/5 = 56$, $N = ar = 8(35) = 280$. That's a pretty big experiment!

Example – Dishwashing:

The following data are taken from an experiment in which the treatments were dishwashing soaps and the response observed was the number of plates washed under standard test conditions before the foam disappeared. In the experiment, detergent solutions were made up and dirty plates were washed one at a time until they were clean. Available for the experiment were three basins and three operators (dishwashers) who washed dishes simultaneously to test three of the total of nine soaps under study. The dish soaps were tested three at a time for 12 sessions until each dish soap had been tested 4 times. Dishwashers washed at the same speed during each session, but the speed of washing varied from session to session. The data are as follows:

Session	Basin 1		Basin 2		Basin 3	
	Detergent	Plates	Detergent	Plates	Detergent	Plates
1	A	19	B	17	C	11
2	D	6	E	26	F	23
3	G	21	H	19	I	28
4	A	20	D	7	G	20
5	B	17	E	26	H	19
6	C	15	F	23	I	31
7	A	20	E	26	I	31
8	B	16	F	23	G	21
9	C	13	D	7	H	20
10	A	20	F	24	H	19
11	B	17	D	6	I	29
12	C	14	E	24	G	21

Here, there are 9 dish soaps ($a = 9$), 12 sessions that form the blocks ($b = 12$), each dish soap is used 4 times in the experiment ($r = 4$), 3 dish soaps are used in each block ($k = 3$), and each pair of dish soaps appears together in the same block 1 time in the entire experiment ($\lambda = 1$).

- In addition, there was some structure underlying the 9 treatments. Dish soaps A–D were all made from the same base detergent (detergent I) with 3, 2, 1, and 0 units of an additive, respectively; dishsoaps E–H were made from base detergent II with 3, 2, 1, and 0 units of additive; and dish soap I was a control.

The effects model:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad \begin{array}{l} i = 1, \dots, a \text{ (treatments)} \\ j = 1, \dots, b \text{ (blocks)} \end{array},$$

where y_{ij} is the observation, if there is one, on the i^{th} treatment in the j^{th} block. We make the usual assumptions that $e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$.

Parameter interpretations:

- μ : is the overall mean
- α_i : is the effect of the i^{th} treatment
- β_j : is the effect of the j^{th} block

Constraints:

$$\sum_i \alpha_i = \sum_j \beta_j = 0$$

Sums of Squares in the BIBD:

In the RCBD we had

$$SS_{Trt} = \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{i.} - \bar{y}_{..})^2 = b \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 = b \sum_{i=1}^a \hat{\alpha}_i^2,$$

where $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$ was that least-squares estimator of the i^{th} treatment effect.

In incomplete block designs, it does not make sense to estimate a treatment effect simply as the deviation of the i^{th} treatment mean from the grand mean.

Why?

Because the i^{th} treatment occurs in a subset of all of the blocks in the experiment, and this subset of the blocks may be systematically different from the collection of all of the blocks in the experiment.

I.e., we believe that there are block effects (or we wouldn't have chosen to use blocks in the first place), therefore some blocks will differ from others. In particular, the subset of blocks in which the i^{th} treatment occurs may differ from the others.

Therefore, when computing an estimate of the i^{th} treatment effect, we should adjust it for block effects (i.e., for block differences).

This reasoning explains why least squares leads to something other than $\hat{\alpha}_i = \bar{y}_i - \bar{y}..$. In particular, in the BIBD, the least-squares estimator of α_i is

$$\hat{\alpha}_i = \frac{kr}{\lambda a} \underbrace{\left(\bar{y}_i - \frac{1}{r} \sum_{j=1}^b n_{ij} \bar{y}_{.j} \right)}_{=\bar{Q}_i}, \quad \text{where } n_{ij} = \begin{cases} 1 & \text{if trt } i \text{ occurs in block } j, \\ 0 & \text{otherwise.} \end{cases}$$

Here, \bar{Q}_i is the deviation of the i^{th} treatment mean from the average of all of the observations in those blocks in which treatment i occurs.

- We call this quantity \bar{Q}_i the **adjusted treatment mean** for treatment i . $Q_i = r\bar{Q}_i = y_i - \sum_j n_{ij} \bar{y}_{.j}$ is called the **adjusted treatment total** for treatment i .

The appropriate sum of squares due to treatments in the BIBD is

$$SS_{Trt(adj)} = \frac{kr^2}{\lambda a} \sum_{i=1}^a \bar{Q}_i^2 = \frac{k}{\lambda a} \sum_i Q_i^2 \quad (*)$$

- We label this quantity $SS_{Trt(adj)}$ to emphasize that it represents treatment to treatment variability adjusted for block differences.
- Note that in the RCBD, the multiplier $\frac{kr^2}{\lambda a}$ simplifies to b , and \bar{Q}_i becomes $(\bar{y}_i - \bar{y}..)$, so we recover the formula for SS_{Trt} from that design. That is, our formula (*) for $SS_{Trt(adj)}$ in the BIBD generalizes the corresponding formula from the RCBD.
- It is possible to show that

$$SS_{Trt(adj)} = R(\alpha|\mu, \beta) = \text{Type II } SS_{Trt} = \text{Type III } SS_{Trt} = \text{Type I } SS_{Trt}$$

where the last equality holds only if we include block effects before treatment effects in our model.

ANOVA Table for the BIBD:

ANOVA Table:

Source of Variation	Sum of Squares	d.f.	F
Trts(adj)	$SS_{Trt(adj)} = \frac{k}{\lambda a} \sum_i Q_i^2$	$a - 1$	$\frac{MS_{Trt(adj)}}{MS_E}$
Blocks	$SS_{Blocks} = \sum_i \sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2$	$b - 1$	
Error	SS_E (by subtraction)	$N - a - b + 1$	
Total	$\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2$	$N - 1$	

We reject the hypothesis that the treatment means are equal (i.e., $H_0 : \mu_1 = \dots = \mu_a$) if

$$F = \frac{MS_{Trt(adj)}}{MS_E} > F_\alpha(a - 1, N - a - b + 1).$$

Example – Dishwashing:

See SAS program, `diswash.sas`, and output, `dishwash.lst`. Notice that in the SAS program we have requested Type I, II, and III SS s. Because treatments are not **orthogonal** to blocks in the BIBD (that is, all treatments do not occur in all blocks), the Type I, II, and III SS s are not identical. Also, the Type I SS s depend upon the order in which terms are added to the model. The order is determined by the positioning of the terms in the MODEL statement.

- Since dishsoap (the treatment factor) appears first, the Type I SS 's are $SS_{Trt(unadj)}$ and $SS_{Blocks(adj)}$. That is, the Type I SS here for dishsoap is wrong for this design (it would have been right if we had put block on our model statement before dishsoap).
- The Type II and III SS 's give $SS_{Trt(adj)}$ and $SS_{Blocks(adj)}$ which are the correct SS 's for inference on the treatments (and blocks, as well, if the randomization restriction didn't prevent us from doing inference on blocks).
- In the example we base our conclusions on $SS_{Trt(adj)}$ and conclude that there are differences among the dishsoaps in terms of average number of plates washed ($F = 164.85$).
- Here, I also used Dunnett's "multiple comparison with the control" procedure to compare each treatment mean with the mean of dishsoap I, the control treatment. The results indicate that dishsoap I is significantly better than any other dishsoap in the experiment.
- Alternatively, we might be interested in analyzing the treatment structure through orthogonal contrasts. That is, we might
 - a. compare the control treatment mean to the means of all others ($p < .0001$);
 - b. compare detergent I to detergent II ($p < .0001$);
 - c. determine whether the mean response changes linearly with amount of additive for detergent I (there is a trend, $p < .0001$ but it is not a linear trend ($p < .0001$);
 - d. determine whether the mean response changes linearly with amount of additive for detergent II (there is a trend, $p < .0001$ which is linear ($p = .8984$)).

Notice that the d.f. for these contrasts sum to $d.f._{Trt}$ and the SS s for these contrasts sum to SS_{Trt} , so we have indeed decomposed the treatment variability into orthogonal (non-redundant) components.

Estimating Treatment Means:

The least squares estimators of the treatment effects in the BIBD are

$$\hat{\alpha}_i = \frac{kQ_i}{\lambda a}, \quad i = 1, \dots, a.$$

The least squares estimator of the overall mean turns out to be $\hat{\mu} = \bar{y}_{..}$. Since the population treatment means are $\mu_i = \mu + \alpha_i$, the least squares estimators of the μ_i s are

$$\hat{\mu}_i = \hat{\mu} + \hat{\alpha}_i = \bar{y}_{..} + \frac{kQ_i}{\lambda a}, \quad i = 1, \dots, a.$$

To compute the standard error of $\hat{\mu}_i$, we need to find

$$\text{var}(\hat{\mu}_i) = \text{var}\left(\bar{y}_{..} + \frac{kQ_i}{\lambda a}\right) = \text{var}(\bar{y}_{..}) + \left(\frac{k}{\lambda a}\right)^2 \text{var}(Q_i)$$

(plus a covariance term that can be shown to be equal to 0).

$\bar{y}_{..}$ is just a mean of N terms, each with variance σ^2 , so $\text{var}(\bar{y}_{..}) = \sigma^2/N$. In addition, using the rules for computing variance and covariances that we discussed at the beginning of the semester, it can be shown that $Q_i = r\bar{y}_{i.} - \sum_j n_{ij}\bar{y}_{.j}$ has variance

$$\text{var}(Q_i) = \frac{r(k-1)\sigma^2}{k}$$

It follows that

$$\text{var}(\hat{\mu}_i) = \frac{\sigma^2}{N} + \left(\frac{k}{\lambda a}\right)^2 \frac{\sigma^2 r(k-1)}{k} = \frac{\sigma^2}{N} \left\{ 1 + \frac{k(a-1)^2}{a(k-1)} \right\}$$

after some simplifications.

Therefore,

$$\text{s.e.}(\hat{\mu}_i) = \sqrt{\widehat{\text{var}}(\hat{\mu}_i)} = \sqrt{\frac{MS_E}{N} \left[1 + \frac{k(a-1)^2}{a(k-1)} \right]},$$

so that, for example, a $100(1 - \alpha)\%$ confidence interval for μ_i is given by

$$\left(\bar{y}_{..} + \frac{kQ_i}{\lambda a} \right) \pm t_{\alpha/2} \underbrace{(N - a - b + 1)}_{=\text{d.f.}_E} \sqrt{\frac{MS_E}{N} \left[1 + \frac{k(a-1)^2}{a(k-1)} \right]}.$$

Another Example – Miles/Gallon with Gasoline Additives:

An engineer is studying the mileage performance characteristics of five types of gasoline additives. Suppose that additives 1,2 and 3 are increasing dosages of an additive based on chemical 1 and additives 4 and 5 are two dosages of an additive based on chemical 2. In the road test she wishes to use cars as blocks; however, because of a time constraint, she can not perform the five road tests for each car that would be necessary for a randomized complete block design. Instead data are collected as follows (a dash indicates that no data were collected under those conditions).

Additive	Car					Mean	Total
	1	2	3	4	5		
1	–	17	14	13	12	14	56
2	14	14	–	13	10	12.75	51
3	12	–	13	12	9	11.5	46
4	13	11	11	12	–	11.75	47
5	11	12	10	–	8	10.25	41
Mean	12.5	13.5	12.0	12.5	9.75	$\bar{y}_{..} = 12.05$	
Total	50	54	48	50	39		

- See handout mpg.sas.

- In mpg.sas I have requested both the lsmeans and the ordinary sample means at each level of additive. The ordinary means are not adjusted for block differences and they are not appropriate estimators of the corresponding population means, but it is instructive to compare these unadjusted means with the adjusted (least squares) means to understand the reason for adjusting for block differences.
- The lsmeans and unadjusted means appear below.

Additive	Car					Marginal Means	LS Means
	1	2	3	4	5		
1	–	17	14	13	12	14.00	14.25
2	14	14	–	13	10	12.75	12.78
3	12	–	13	12	9	11.50	11.85
4	13	11	11	12	–	11.75	11.12
5	11	12	10	–	8	10.25	10.25
Marginal Mean:	12.50	13.50	12.00	12.50	9.75		

- The last two columns compare the lsmeans with the sample marginal means for the 5 additives. Notice that additive four has the largest disparity between the two means.

The lsmean is preferable, why is the marginal mean inappropriate?

- Notice that additive 4 is the only additive that was not tested in car 5 and that car 5 was the car which got (by far) the lowest gas mileage. The unadjusted mean is based on the responses for the four cars in which additive 4 was tested making no allowances for the possibility that this subset of cars may be systematically different from the entire set of cars used in the experiment.
- In fact this is the case: cars 1,2,3, and 4 get the best gas mileage, so the mean for additive 4 is computed under an especially good set of circumstances as compared to the other additives. No adjustment for this is made with the unadjusted mean, hence its value is especially high.
- The lsmean adjusts for block-block differences, deflating additive 4's mean from 11.75 down to 11.12 to reflect the sort of deflation that one would expect would occur to the sample marginal mean if it were possible to compute it based on all 5 cars.

- We have also seen that the $SS_{Trt(adj)}$ (adjusted for block differences) is more appropriate than the unadjusted SS_{Trt} for testing for treatment effects.
- Roughly speaking, SS_{Trt} quantifies variability in the unadjusted means, and $SS_{Trt(adj)}$ quantifies variability in the adjusted (LS) means.
- Look at how the LS means have been adjusted. Relative to the unadjusted means, most of the LS means have been moved away from the grand mean. That is, the LS means are more variable than the unadjusted means.
- This is seen in the ANOVA in that $SS_{Trt} = 31.70$ and $SS_{Trt(adj)} = 35.73$.
- Based on the results in mpg.pdf, we would conclude that there are differences among the treatment means, with the mean for chemical 1 being significantly higher than the mean for chemical 2. In addition, there appears to be a significant linear trend downward in increasing doses of chemical 1.

Contrasts:

We define a contrast as usual as a linear combination of the population means, $\psi = \sum_i c_i \mu_i$ where $\sum_i c_i = 0$. The population contrast ψ is estimated by its sample analog,

$$C = \sum_i c_i \hat{\mu}_i = \underbrace{\sum_i c_i \bar{y}_{..}}_{=\bar{y}_{..} \sum_i c_i = 0} + \frac{k}{\lambda a} \sum_i c_i Q_i = \frac{k}{\lambda a} \sum_i c_i Q_i.$$

Since C is a linear combination of normally distributed random variables, C is normally distributed and we can test $H_0 : \psi = 0$ using a t test statistic of the form

$$t = \frac{|C|}{\sqrt{\hat{\text{var}}(C)}}$$

or, equivalently, an F test statistic of the form

$$F = \frac{C^2}{\hat{\text{var}}(C)}.$$

It remains to find formulas for $\hat{\text{var}}(C)$.

The variance of the estimated contrast, C , is

$$\text{var}(C) = \left(\frac{k}{\lambda a}\right)^2 \text{var}\left(\sum_i c_i Q_i\right).$$

Since Q_i , $i = 1, \dots, a$ are not independent,

$$\text{var}\left(\sum_i Q_i\right) = \sum_i \text{var}(Q_i) + \underbrace{\sum_i \sum_{i'} \text{cov}(Q_i, Q_{i'})}_{i' \neq i}.$$

Therefore,

$$\begin{aligned} \text{var}(C) &= \left(\frac{k}{\lambda a}\right)^2 \left[\sum_i \text{var}(c_i Q_i) + \underbrace{\sum_i \sum_{i'} \text{cov}(c_i Q_i, c_{i'} Q_{i'})}_{i' \neq i} \right] \\ &= \left(\frac{k}{\lambda a}\right)^2 \left[\sum_i c_i^2 \text{var}(Q_i) + \underbrace{\sum_i \sum_{i'} c_i c_{i'} \text{cov}(Q_i, Q_{i'})}_{i' \neq i} \right]. \end{aligned}$$

We have already shown (p.188) that $\text{var}(Q_i) = \frac{r(k-1)\sigma^2}{k}$. Using similar, straight-forward but tedious applications of the algebra of variances and covariances, it can also be shown that for each $i \neq i'$,

$$\text{cov}(Q_i, Q_{i'}) = -\frac{r(k-1)\sigma^2}{k(a-1)}.$$

Therefore,

$$\text{var}(C) = \left(\frac{k}{\lambda a}\right)^2 \frac{r(k-1)\sigma^2}{k} \left[\sum_i c_i^2 - \frac{1}{a-1} \underbrace{\sum_i \sum_{i'} c_i c_{i'}}_{i' \neq i} \right],$$

where the term in square brackets can be shown to equal $\frac{a}{a-1} \sum_i c_i^2$. Therefore,

$$\begin{aligned} \text{var}(C) &= \frac{kr(k-1)\sigma^2}{\lambda^2 a^2} \left[\frac{a}{a-1} \sum_i c_i^2 \right] \\ &= \frac{k\lambda(a-1)\sigma^2}{\lambda^2 a^2} \left[\frac{a}{a-1} \sum_i c_i^2 \right] \\ &= \frac{k}{\lambda a} \sigma^2 \sum_i c_i^2. \end{aligned}$$

$\text{var}(C)$ is estimated by

$$\hat{\text{var}}(C) = \frac{k}{\lambda a} MS_E \sum_i c_i^2$$

so our t statistic is

$$t = \frac{|C|}{\sqrt{\frac{k}{\lambda a} MS_E \sum_i c_i^2}},$$

and our F statistic is

$$F = t^2 = \frac{C^2}{\frac{k}{\lambda a} MS_E \sum_i c_i^2} = \frac{\left(\frac{k}{\lambda a} \sum_i c_i Q_i\right)^2}{\frac{k}{\lambda a} MS_E \sum_i c_i^2} = \frac{SS_C}{MS_E},$$

where

$$SS_C = \frac{\frac{k}{\lambda a} \left(\sum_i c_i Q_i\right)^2}{\sum_i c_i^2}.$$

We reject $H_0 : \psi = 0$ if $F > F_\alpha(1, N - a - b + 1)$.

Example – Dishwashing (Again):

We estimate the mean number of dishes washed for the i^{th} dishsoap using the estimator:

$$\hat{\mu}_i = \bar{y}_{..} + \frac{kQ_i}{\lambda a}.$$

For the dishsoaps in the experiment we have $\bar{y}_{..} = 19.417$, $k/\lambda a = 1/3$ and $Q_1 = 1, Q_2 = -6.667, Q_3 = -18.667, \dots$. The estimated means are $\hat{\mu}_1 = 19.75, \hat{\mu}_2 = 17.194, \hat{\mu}_3 = 13.194, \dots$. Dishsoap D clearly washes fewer dishes than the others ($\hat{\mu}_4 = 6.528$) and dishsoap I washes the most dishes ($\hat{\mu}_9 = 29.528$). A 95% confidence interval for dishsoap D is

$$6.528 \pm \underbrace{t_{0.05/2}(36 - 9 - 12 + 1)}_{=2.120} \sqrt{\frac{.824}{36} \left[1 + \frac{3(8)^2}{9(2)} \right]}$$
$$6.528 \pm 2.120(.517) = (5.432, 7.624).$$

To illustrate the computations that go into testing a contrast, consider testing whether the control mean is different from the non-control means, on average. That is, consider the contrast

$$\psi = \mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6 + \mu_7 + \mu_8 - 8\mu_9$$

The corresponding sample contrast is

$$C = (\hat{\mu}_1 + \hat{\mu}_2 + \hat{\mu}_3 + \hat{\mu}_4 + \hat{\mu}_5 + \hat{\mu}_6 + \hat{\mu}_7 + \hat{\mu}_8 - 8\hat{\mu}_9)$$
$$= (19.75 + 17.194 + \dots + (-8)29.528) = -91.00$$

The F statistic is, therefore,

$$F = \frac{C^2}{\frac{k}{\lambda a} MS_E \sum_i c_i^2} = \frac{(-91.0)^2}{\frac{3}{(1)(9)} .824 (1^2 + 1^2 + \dots + (-8)^2)} = 418.70$$

which we compare to the $F(1, \text{d.f.}_E) = F(1, 16)$ distribution, leading to a p -value less than .0001.

The Interblock Analysis of the BIBD:

- In our discussion of the analysis of a BIBD, we have only presented what's known as the **intra**block analysis. This is the classical analysis that is done for a BIBD.
- However, because different subsets of the treatments are observed in different blocks, there is also information concerning treatment differences that can be obtained by comparing block effects. Such a comparison is known as the **inter**block analysis.
- A complete analysis of an incomplete block design consists of doing both the intra-block analysis and the interblock analysis, and then combining the results.
- In the past, the complete analysis was seldom done, because
 - a. the interblock analysis is typically much less precise (less powerful) than the intra-block analysis, and adding interblock information to the analysis typically yields very little extra precision (or power).
 - b. In addition, it *used to be true* that the interblock analysis requires calculations that were difficult, computationally.
- However, with modern software for fitting mixed-effects models, the combined inter- and intra-block analysis can now be done routinely by treating block effects as random in the BIBD model.

Example — Dish Soaps Again

- Refer once again to `dishwash.sas` and its output `dishwash.pdf`.
- At the end of `dishwash.sas`, PROC MIXED is used to fit the model

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad \begin{array}{l} i = 1, \dots, a \text{ (treatments)} \\ j = 1, \dots, b \text{ (blocks)} \end{array},$$

where the block effects, β_j are iid $N(0, \sigma_\beta^2)$ independent of the error terms e_{ij} s, which are iid $N(0, \sigma^2)$.

- Note that the results of this analysis are (in this example) qualitatively the same as the intrablock analysis done by PROC GLM, but quantitatively different.
- The PROC GLM (intrablock) analysis is
 - Completely correct when block effects should be treated as fixed.
 - Mostly valid, but somewhat inefficient, when block effects should be regarded as random. Validity of the analysis holds in all respects except for standard errors and confidence intervals on treatment means and non-contrast functions of the treatment means, which are wrong.

Other Incomplete Block Designs:

We have only discussed the BIBD. There are many other designs that use incomplete blocks.

- The **partially balanced incomplete block design (PBIBD)** is similar to the BIBD, except that all pairs of treatments don't occur together (in the same block) the same number of times. Some pairs occur together λ_1 times, others occur together λ_2 times, etc. In such a design, not all pairwise contrasts between treatments are estimated with the same precision.
 - The advantage of this is greater flexibility. In particular, letting go of the requirement that all pairs occur together the same number of times allows a PBIBD to be done in fewer blocks than would be necessary for a BIBD.
- **Cyclic designs** are a class of incomplete block designs that are simple to construct, but don't necessarily have any nice balance features. Treatments are allocated to some multiple of a blocks in cycles.
 - E.g., Suppose we have $a = 6$ treatments (t1, t2, t3, t4, t5, t6), and we have block sizes of $k = 4$. Then a cyclic design with $b = 1(a) = 6$ blocks with initial block containing treatments t1, t2, t3, and t6 is given by

Block					
1	2	3	4	5	6
t1	t2	t3	t4	t5	t6
t2	t3	t4	t5	t6	t1
t3	t4	t5	t6	t1	t2
t6	t1	t2	t3	t4	t5

- The choice of initial block affects the efficiency of the design. See Appendix C.3 of our text for some good choices of initial blocks for various choices of a , k and r .

- A **Youden square design** is not a square design at all, it is better described as a “Latin rectangle”. That is, it is a Latin square design where one or more of the rows have been deleted, or one or more of the columns have been deleted.
 - The following design is a Youden square formed by deleting the last column of a 4×4 Latin square:

A	B	C
B	C	D
C	D	A
D	A	B

Note that we have complete blocks with respect to the columns, but incomplete blocks with respect to the rows.

- Another type of design that uses incomplete blocks is the **Split-Plot Design**, which we consider next.

Split Plot Designs

Split plot designs are characterized by three basic features:

1. a two (or higher)-way factorial treatment structure;
2. a design structure consisting of incomplete blocks; and
3. at least two types of experimental units, each with its own randomization scheme.

We will consider split plot designs in which we have a factorial arrangement of treatments involving two factors, A and B.

Two sizes of experimental unit are involved, one randomized to the levels of factor A and the other randomized to the levels of factor B.

- The main motivation for using a split-plot design is convenience. When we are concerned with the main effects and interaction between two (or more) factors, it is sometimes the case that one of the two factors is much easier to apply to large experimental units than to small ones. In such situations, a split-plot design is a good choice.
- For example, suppose we are interested in the effects of burning (with levels yes and no) and application of fertilizer (with levels 0, 100, and 200 lbs./acre) as site preparation methods before growing pine trees. Because fire is hard to control, rather than burning small plots, it is easier to burn or not burn large patches of a growing site and then subdivide those patches (whole plots) and apply different levels of fertilizer to the subdivisions (the split plots).

Example – Chocolate Cake:

An experiment was conducted to determine the effect of baking temperature on quality for three recipes of chocolate cake. Recipes I and II differed in that the chocolate was added at 40° C. and 60° C., respectively, while recipe III contained extra sugar. Six different baking temperatures were considered: 175° C., 185° C., 195° C., 205° C., 215° C., and 225° C. 45 batches of cake batter were prepared using each of the 3 recipes 15 times in a completely random order. Each batch was large enough for 6 cakes, and the six baking temperatures were randomly assigned to the 6 cakes per batch in a random manner. One of several measurements of quality made on each cake was the breaking angle of the cake. The data from this experiment are as follows:

TABLE 7.5 BREAKING ANGLES (DEGREES)

Rep.	Temperature						Unit totals	Rep. totals
	175°	185°	195°	205°	215°	225°		
Recipe I	1	42	46	47	39	53	42	269
	2	47	29	35	47	57	45	260
	3	32	32	37	43	45	45	234
	4	26	32	35	24	39	26	182
	5	28	30	31	37	41	47	214
	6	24	22	22	29	35	26	158
	7	26	23	25	27	33	35	169
	8	24	33	23	32	31	34	177
	9	24	27	28	33	34	23	169
	10	24	33	27	31	30	33	178
	11	33	39	33	28	33	30	196
	12	28	31	27	39	35	43	203
	13	29	28	31	29	37	33	187
	14	24	40	29	40	40	31	204
	15	26	28	32	25	37	33	181
Totals	437	473	462	503	580	526	2981	
Recipe II	1	39	46	51	49	55	42	282
	2	35	46	47	39	52	61	280
	3	34	30	42	35	42	35	218
	4	25	26	28	46	37	37	199
	5	31	30	29	35	40	36	201
	6	24	29	29	29	24	35	170
	7	22	25	26	26	29	36	164
	8	26	23	24	31	27	37	168
	9	27	26	32	28	32	33	178
	10	21	24	24	27	37	30	163
	11	20	27	33	31	28	33	172
	12	23	28	31	34	31	29	176
	13	32	35	30	27	35	30	189
	14	23	25	22	19	21	35	145
	15	21	21	28	26	27	20	143
Totals	403	441	476	482	517	529	2848	
Recipe III	1	46	44	45	46	48	63	292
	2	43	43	43	46	47	58	280
	3	33	24	40	37	41	38	213
	4	38	41	38	30	36	35	218
	5	21	25	31	35	33	23	168
	6	24	33	30	30	37	35	189
	7	20	21	31	24	30	33	159
	8	24	23	21	24	21	35	148
	9	24	18	21	26	28	28	145
	10	26	28	27	27	35	35	178
	11	28	25	26	25	28	28	170
	12	24	30	28	35	33	28	178
	13	28	29	43	28	33	37	198
	14	19	22	27	25	25	35	153
	15	21	28	25	25	31	25	155
Totals	419	434	476	463	516	536	2844	
Temp. totals	1259	1348	1414	1448	1613	1591	8673	

Terminology:

Whole Plot: The largest type of experimental unit (batches of batter in the example).

Whole Plot Factor: The treatment factor whose levels are randomized to the whole plots. Optimally, this is the treatment factor of secondary interest, but more commonly it is simply the factor that is easier to apply to big units. In general, we'll refer to this as factor A (recipes, in the example).

Split Plot: The smaller type of experimental unit (cakes in the example).

Split Plot Factor: The treatment factor whose levels are randomized to the split plots. Optimally, this is the treatment factor of primary interest. In general, we'll refer to this as factor B (baking temperature in the example).

- In a split plot design, the entire design has a treatment structure (two-way factorial) and a design structure (incomplete block).
- There is also a treatment structure and design structure for each type of experimental unit and an error term for each type of experimental unit.
- The model for the entire design is constructed from the structures of each type of experimental unit.

Cake Example:

- **Treatment structure:** two-way factorial because we are interested in the effects of recipes and temperatures and these factors are crossed.
- **Design structure:** incomplete block because blocks correspond to batches and only a third of all possible treatments occur in each block (there are 3×6 total treatments, and only six of them are observed in each block).
- **Whole plot treatment structure:** one-way (recipes).
- **Whole plot design structure:** completely randomized. (The 3 recipes were assigned to the 45 batches completely at random.)
- **Split plot treatment structure:** one-way (temperatures).
- **Split plot design structure:** randomized complete block. (The temperatures are assigned to cakes at random in complete blocks (batches).)

Common Types of Split Plot Designs:

1. Whole plot factor arranged in a completely randomized design. E.g., chocolate cake experiment.
 2. Whole plot factor arranged in a randomized complete block design. This is the most common type. For example, alfalfa experiment (later).
 3. Others. E.g., the whole plot factor could be arranged in a Latin Square Design, or the whole plot treatments could be the factorial combinations of two or more separate treatment factors, or the split plot treatments could be the factorial combinations of two or more separate treatment factors, etc.
- In this course we will only consider in detail split-plot designs of types 1 and 2, and we will only consider the analysis of balanced split-plot designs. That is, split-plot experimental designs where the whole plots occur in a balanced design, and split plots occur in a balanced RCBD.

Type 1 – Whole plots in CRD:

The data:

$$y_{ijk}, \quad \begin{array}{l} i = 1, \dots, a \text{ (levels of factor A, the whole plot factor)} \\ j = 1, \dots, n \text{ (replicates for A)} \\ k = 1, \dots, b \text{ (levels of factor B, the split-plot factor)} \end{array}$$

The model traditionally used for the split-plot design exemplified by the chocolate cake example is

$$y_{ijk} = \mu + \alpha_i + e_{j(i)} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}, \quad (*)$$

Parameter interpretations:

- μ = the grand mean or constant term,
- α_i = an effect for the i^{th} level of A , the whole plot factor (e.g., recipe),
- β_k = an effect for the k^{th} level of B , the split-plot factor (temperature),
- $(\alpha\beta)_{ik}$ = an interaction term for $A * B$ interactions,
- $e_{j(i)}$ = a random effect for the j^{th} whole plot nested in the i^{th} level of A ,
i.e., the whole plot error term,
- ε_{ijk} = the split plot (or overall) error term.

One way to think about the split plot model is that it is the union of the model appropriate for the whole plots and the model appropriate for the split plots.

- The whole plots occur in a one-way layout, so the one-way layout model

$$\mu + \alpha_i + e_{j(i)}$$

is appropriate for batches.

- The split plots occur in a randomized complete block design, so the RCBD model

$$\mu + \underbrace{\text{block}_{ij}}_{=\alpha_i + e_{j(i)}} + \beta_k + \varepsilon_{ijk}$$

is appropriate for cakes.

- Putting these portions of the models together and adding an interaction term $(\alpha\beta)_{ik}$ to capture interactions between the whole and split plot factors, leads to model (*).
- The random whole plot effect $e_{j(i)}$ can be thought of as the whole plot error term and ε_{ijk} as the split plot error term. Since there are two experimental units with two separate randomizations, there are two error terms in the model.

In fact, the split plot model described above is an example of a linear mixed-effects model (LMM).

- It includes fixed effects for the whole plot factor (the α_i 's for recipes), split plot factor (the β_k 's for temperatures), and their interaction (the $(\alpha\beta)_{ik}$'s).
- It also includes random effects: $e_{j(i)}$, a whole plot (or batch) effect, in addition to the overall error term ε_{ijk} , which is always present in any linear model, so is typically not categorized as a random effect, even though it is.
- The term “mixed-effects model” or sometimes simply “mixed model” refers to the fact that the *linear predictor* of the model (the right side of the model equation, excluding the overall error term) includes both fixed and random effects.

Fixed vs. random effects: The effects in the model account for variability in the response across levels of treatment and design factors. The decision as to whether fixed effects or random effects should be used depends upon what the appropriate scope of generalization is.

- If it is appropriate to think of the levels of a factor as randomly drawn from, or otherwise representative of, a population to which we'd like to generalize, then random effects are suitable.
 - Design or grouping factors are usually more appropriately modeled with random effects.
 - E.g., blocks (sections of land) in an agricultural experiment, days when an experiment is conducted over several days, lab technician when measurements are taken by several technicians, subjects in a repeated measures design, locations or sites along a river when we desire to generalize to the entire river, etc.
- If, however, the specific levels of the factor are of interest in and of themselves then fixed effects are more appropriate.
 - Treatment factors are usually more appropriately modeled with fixed effects.
 - E.g., In experiments to compare drugs, amounts of fertilizer, hybrids of corn, teaching techniques, and measurement devices, all of these factors are most appropriately modeled with fixed effects.
- A good litmus test for whether the level of some factor should be treated as fixed is to ask whether it would be of broad interest to report a mean for that level. For example, if I'm conducting an experiment in which each of four different classes of third grade students are taught with each of three methods of instruction (e.g., in a crossover design) then it will be of broad interest to report the mean response (level of learning, say) for a particular method of instruction, but not for a particular classroom of third grades.
 - Here, fixed effects are appropriate for instruction method, random effects for class.

Since the whole plot error term represents random whole plot effects (batch effects), the $e_{j(i)}$'s are random variables. Therefore, we must make some assumptions about their distribution (and the distribution of the overall error term ε_{ijk}) to complete the split-plot model.

The following assumptions are typical and fully define the Type 1 split-plot model:

$$y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + e_{j(i)} + \varepsilon_{ijk},$$

where

$$\begin{aligned} \{e_{j(i)}\} &\stackrel{iid}{\sim} N(0, \sigma_e^2) \\ \{\varepsilon_{ijk}\} &\stackrel{iid}{\sim} N(0, \sigma^2) \\ \text{cov}(\varepsilon_{ijk}, e_{j'(i')}) &= 0, \quad \text{for all } i, j, k, i', j'. \end{aligned}$$

The chocolate cake experiment is an example of a balanced split plot design with whole plots arranged in a one-way layout. More complex split-plot designs are possible. E.g., whole plots are often arranged in a RCBD, split plots could be split once again to create split-split-plots, etc.

However, the classical analysis of all of these designs is relatively straightforward *provided that the design is balanced*.

- By “balanced” here, we mean that there is an equal number of replicates for each whole plot treatment (recipe), and the same set of subplot treatments (temperatures) was observed within each whole plot (batch).

The classical analysis of the balanced split plot model with whole plots in a one-way layout is based on the model above and the following decomposition of the deviations $y_{ijk} - \bar{y}...$ of each observation from the grand sample mean:

$$y_{ijk} - \bar{y}... = (\bar{y}_{i..} - \bar{y}...) + (\bar{y}_{ij.} - \bar{y}_{i..}) + (\bar{y}_{..k} - \bar{y}...) + (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}...) + (y_{ijk} - \bar{y}_{i.k} - \bar{y}_{ij.} + \bar{y}_{i..}), (**)$$

where

$$\begin{aligned} \bar{y}... &= (anb)^{-1} \sum_{i=1}^a \sum_{j=1}^n \sum_{k=1}^b y_{ijk} & \bar{y}_{i..} &= (nb)^{-1} \sum_{j=1}^n \sum_{k=1}^b y_{ijk} \\ \bar{y}_{..k} &= an^{-1} \sum_{i=1}^a \sum_{j=1}^n y_{ijk} & \bar{y}_{i.k} &= n^{-1} \sum_{j=1}^n y_{ijk} & \bar{y}_{ij.} &= b^{-1} \sum_{k=1}^b y_{ijk} \end{aligned}$$

are sample means over all observations ($\bar{y}...$), over observations at the i^{th} level of A ($\bar{y}_{i..}$), etc.

This decomposition leads to the following analysis of variance:

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
A	SS_A	$a - 1$	$\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_{E(\text{wholeplot})}}$
B	SS_B	$b - 1$	$\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_{E(\text{splitplot})}}$
$A * B$	SS_{AB}	$(a - 1)(b - 1)$	$\frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_{E(\text{splitplot})}}$
Error (whole plot)	$SS_{E(\text{wholeplot})}$	$a(n - 1)$	$\frac{SS_{E(\text{wholeplot})}}{a(n-1)}$	
Error(split plot)	$SS_{E(\text{splitplot})}$	$a(b - 1)(n - 1)$	$\frac{SS_{E(\text{splitplot})}}{a(b-1)(n-1)}$	
Total	SS_T	$abn - 1$		

Tests:

The expected mean squares here dictate the appropriate F tests for main effects and interactions:

Source of Variation	Expected Mean Squares
A	$\sigma^2 + b\sigma_e^2 + Q(\alpha)$
B	$\sigma^2 + Q(\beta)$
$A * B$	$\sigma^2 + Q(\alpha\beta)$
Error (whole plot)	$\sigma^2 + b\sigma_e^2$
Error(split plot)	σ^2

- Here, $Q(\alpha)$ represents a sum of squares in the α_i 's which equals 0 under $H_0 : \alpha_1 = \dots = \alpha_a = 0$, and is positive otherwise. $Q(\beta)$ and $Q(\alpha\beta)$ are similarly defined.
- In addition, $\sigma_e^2 = \text{var}(e_{j(i)})$ is the whole plot error variance, and $\sigma^2 = \text{var}(\varepsilon_{ijk})$ is the split-plot error variance.

There are three hypotheses to test:

- H_0 : Mean response does not depend on the levels of A . Equivalently, $H_0 : \alpha_1 = \dots = \alpha_a = 0$, or $H_0 : \bar{\mu}_{1..} = \dots = \bar{\mu}_{a..}$.

Test: reject if

$$F = \frac{MS_A}{MS_{E(\text{whole plot})}} > F_\alpha(a - 1, a(n - 1)).$$

- H_0 : Mean response does not depend on the levels of B . Equivalently, $H_0 : \beta_1 = \dots = \beta_b = 0$, or $H_0 : \bar{\mu}_{..1} = \dots = \bar{\mu}_{..b}$.

Test: reject if

$$F = \frac{MS_B}{MS_{E(\text{split plot})}} > F_\alpha(b - 1, a(b - 1)(n - 1)).$$

c. H_0 : No $A * B$ interaction. Equivalently, $H_0 : (\alpha\beta)_{ik} = 0$ for all i, k .

Test: reject if

$$F = \frac{MS_{AB}}{MS_{E(\text{split plot})}} > F_{\alpha}((a-1)(b-1), a(b-1)(n-1)).$$

Comments:

- As in the other designs we have looked at including interactions, test (c) should be done first, before deciding if main effect tests are appropriate.
- Notice that the error degrees of freedom are smaller for test (a) than for tests (b) and (c). Test (a) has less power. This is why, ideally, if one of the two factors is of secondary interest it is better for that factor to be the whole plot factor rather than the split plot factor.

The main reason to use the split plot design is

- Convenience. E.g., randomization is easier to do at two levels, levels of A are easier to apply to larger E.U.s, etc.

A secondary consideration that may influence selection of the split plot design and choices made in the design is

- It is more powerful for detection of main effects of B and interaction effects $A * B$ than main effects of A . Therefore, it is well suited to the situation in which
 - we are more interested in B and $A * B$ than in A ; and/or
 - we expect a strong A effect (so high power not necessary) and/or a weak B effect.

Analyzing a Split-Plot Design in SAS:

- As with some of the other mixed-effect models we have studied, PROC GLM in SAS can produce a correct ANOVA table for the split plot analysis, but there are certain results that it produces that are incorrect.
- Therefore, PROC GLM is not recommended for analysis of split-plot designs. PROC MIXED or other mixed effect model software should be used.
- In particular, the type 1 split plot design can be analyzed in PROC MIXED as follows:

```
proc mixed;
  class A B replicat;
  model y=A B A*B / ddfm=satterth;
  random replicat(A);
  contrast 'contrast in marginal means for A' A a1 a2 ... aa;
  contrast 'contrast in marginal means for B' B b1 b2 ... bb;
  contrast 'any joint mean contrast' A a1 ... aa
                                     B b1 ... bb
                                     A*B c1 c2 c3 ... cab;
  estimate 'any linear combination of parameters' intercept d
          A a1 ... aa B b1 ... bb A*B c1 c2 ... cab;
  lsmeans A B A*B;
run;
```

- Here we've used the default REML method of fitting the model. For nice balanced split-plot models, METHOD=TYPE3 will give all of the same (correct) results, except that it will give a different confidence interval on σ_e^2 , the variance component for whole plots. I prefer the confidence interval for this quantity that REML gives, but REML does not give the ANOVA table and the type3 method does. So if the ANOVA table is desired, you can get it by using METHOD=TYPE3.

Chocolate Cake Example:

In this example it is natural to consider the following comparisons among the recipes:

	I	II	III	
$\psi_1 :$	1	1	-2	(normal vs extra sugar)
$\psi_2 :$	1	-1	0	(temp at which choc. added)

- Since the baking temperatures are evenly spaced we can also determine the degree of the polynomial that best describes the relationship between baking temperature and baking strength.
- Refer to `choccake.sas` and `choccake.pdf`.
- Firstly, we conclude that there is no interaction between recipe and baking temperature ($F = 1.01$, $p = .4393$) so it is appropriate to examine main effects (see profile plot as well).
- We conclude that the effects due to recipes are not significantly different from zero ($F = 0.25$, $p = 0.7809$).
- Since the overall test of recipes is not significant, the Fisher's LSD approach would say not to test our planned recipe contrasts. Even if we went ahead and tested these contrasts anyway, they are very far from significant.
- Based on the F test for main effects of baking temperatures ($F = 20.52$, $p = 0.0001$) we conclude that baking temperature has an effect on breaking strength. This effect is at least linear because the linear contrast is significant ($F = 96.07$).
- The lack of fit from linearity (i.e., nonlinearity) test statistic is

$$\begin{aligned}
 F &= \frac{MS_{LOF}}{MS_E} = \frac{(SS_{Baketemp} - SS_{linear})/(5 - 1)}{MS_E} \\
 &= \frac{(2100.3 - 1966.705)/(5 - 1)}{20.471} = 1.632.
 \end{aligned}$$

Since $F = 1.632 < F_{.05}(4, 210) = 2.4$ we do not reject the hypothesis that our model fits and conclude that the relationship between baking temperature and breaking strength is linear over the range of baking temperatures studied.

Type 2 – Whole plots in RCBD:

Example – Alfalfa:

An agronomist wants to determine the effects of variety and date of final cutting on the yield of alfalfa in the following year. Three varieties were considered: (A_1) Cossack, (A_2) Ladak, and (A_3) Ranger. The response variable was yield measured in 1944. The four cutting-time treatments (levels of factor B) all involved at least two cuttings in 1943, with the second cutting occurring on 7/27/1943 in each case. The cutting-time treatments differed in terms of the presence and timing of the third cutting time in 1943: B_1 = No cut (control), B_2 = September 1, B_3 = September 20, and B_4 = October 7. Six locations (blocks) were chosen in which to conduct the experiment. The varieties were randomized within each of the six locations so that each variety occurred in one-third of each block. The four cutting times were then randomly assigned to four plots within each one-third of a block.

- Here, it was easier to plant varieties in big units and then cut in smaller units, hence the choice of a split-plot design.

The alfalfa data and design are as follows:

Block 1			Block 2			Block 3		
A_1	A_3	A_2	A_2	A_1	A_3	A_3	A_1	A_2
B_2 1.38	B_3 1.55	B_2 1.58	B_4 2.01	B_3 1.70	B_1 1.95	B_4 1.99	B_3 1.81	B_4 1.82
B_4 2.27	B_4 1.56	B_3 2.29	B_2 1.26	B_4 1.81	B_2 1.47	B_3 1.82	B_2 1.85	B_3 1.67
B_3 1.86	B_1 1.75	B_1 2.17	B_3 1.60	B_1 2.01	B_4 1.72	B_1 2.13	B_1 1.70	B_2 1.22
B_1 2.33	B_2 1.52	B_4 2.23	B_1 1.88	B_2 1.30	B_3 1.61	B_2 1.80	B_4 2.01	B_1 1.62
Block 4			Block 5			Block 6		
A_2	A_3	A_1	A_3	A_2	A_1	A_1	A_2	A_3
B_4 2.10	B_3 1.56	B_3 1.54	B_3 1.23	B_4 1.66	B_1 1.42	B_3 0.88	B_3 1.12	B_4 1.33
B_2 1.59	B_4 1.55	B_1 1.78	B_4 1.51	B_2 1.25	B_3 1.67	B_4 1.06	B_1 1.66	B_2 1.31
B_3 1.91	B_2 1.37	B_4 1.40	B_1 1.31	B_3 1.39	B_2 1.13	B_2 1.06	B_2 0.94	B_1 1.30
B_1 2.34	B_1 1.78	B_2 1.09	B_2 1.01	B_1 1.58	B_4 1.31	B_1 1.35	B_4 1.10	B_3 1.13

- Here, the whole plots are the 18 columns in the 6 blocks, and the split plots are the plots within columns (there are 72 of them).
- The whole plot treatment factor is variety, the split plot treatment factor is cutting time.
- The whole plot treatment structure is one-way, and the whole plot design structure is randomized complete block.
- The split plot treatment structure is one-way, and the split plot design structure is randomized complete block.

The Data:

$$\begin{aligned}
 & i = 1, \dots, a \text{ (levels of factor A)} \\
 y_{ijk}, & \quad j = 1, \dots, n \text{ (whole plot blocks)} \\
 & k = 1, \dots, b \text{ (levels of factor B)}
 \end{aligned}$$

The model traditionally used for the split-plot design exemplified by the alfalfa example is

$$y_{ijk} = \mu + \alpha_i + \tau_j + e_{ij} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}, \quad (\heartsuit)$$

where we make similar assumptions about the error terms as in the type 1 case:

$$\begin{aligned}
 \{e_{ij}\} & \stackrel{iid}{\sim} N(0, \sigma_e^2) & \{\varepsilon_{ijk}\} & \stackrel{iid}{\sim} N(0, \sigma^2) \\
 \text{cov}(\varepsilon_{ijk}, \varepsilon_{i'j'}) & = 0, & & \text{for all } i, j, k, i', j'.
 \end{aligned}$$

Parameter interpretations:

- μ = the overall mean for whole plots;
- α_i = the effect of the i^{th} level of factor A (i^{th} variety);
- τ_j = the effect of the j^{th} whole plot block;
- e_{ij} = the error term for whole plots,
i.e., the effect of the $(i, j)^{\text{th}}$ whole plot;
- $(\alpha\beta)_{ik}$ = the effect of the $(i, k)^{\text{th}}$ treatment,
i.e., an effect due to interaction between A and B ;
- β_k = the effect of the k^{th} level of the split plot factor;
- ε_{ijk} = the split plot error term.

One way to think about the split plot model is that it is the union of the model appropriate for the whole plots and the model appropriate for the split plots.

- The whole plots occur in a RCBD, so the one-way layout model

$$\mu + \alpha_i + \tau_j + e_{ij}$$

is appropriate for the 18 columns or strips of field assigned to different varieties.

- The split plots also occur in a RCBD, so the model

$$\mu + \underbrace{\text{block}_{ij}}_{=\alpha_i + \tau_j + e_{ij}} + \beta_k + \varepsilon_{ijk}$$

is appropriate for the 72 individual plots in the 6 fields.

- Putting these portions of the models together and adding an interaction term $(\alpha\beta)_{ik}$ to capture interactions between the whole and split plot factors, leads to model (♡).

Model (♡) leads to the following ANOVA table:

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
A	SS_A	$a - 1$	$\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_{E(\text{wholeplot})}}$
Whole plot blocks	SS_{WPB}	$n - 1$	$\frac{SS_{WPB}}{n-1}$	
Error (whole plot)	$SS_{E(\text{wholeplot})}$	$(a - 1)(n - 1)$	$\frac{SS_{E(\text{wholeplot})}}{(a-1)(n-1)}$	
B	SS_B	$b - 1$	$\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_{E(\text{splitplot})}}$
$A * B$	SS_{AB}	$(a - 1)(b - 1)$	$\frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_{E(\text{splitplot})}}$
Error(split plot)	$SS_{E(\text{splitplot})}$	$a(b - 1)(n - 1)$	$\frac{SS_{E(\text{splitplot})}}{a(b-1)(n-1)}$	
Total	SS_T	$abn - 1$		

- Note that our model includes a interaction term $(\alpha\beta)_{ik}$ for our two treatment factors A and B, but it does not include a two-way interaction term between whole plot blocks and factor B (i.e., a $(\tau\beta)_{jk}$ term).
 - Although there are sufficient degrees of freedom to include this term, it is usually assumed that such an interaction doesn't exist and this term is omitted from the model. This is not a necessary assumption, however.
- Note that it is not possible to include three-way interactions between A, B, and whole plot blocks because of insufficient degrees of freedom in the design.

Tests:

There are again three hypotheses to test. As in case 1 (and as always), the appropriate test statistics for these hypotheses can be determined by examining expected mean squares (omitted).

The tests turn out to be the same as in the Type 1 SPD (see pp.232–233), except for test (a), for which there is a change in the error degrees of freedom.

The test for main effects of factor A changes as follows:

- H_0 : Mean response does not depend on the levels of A. Equivalently, $H_0 : \alpha_1 = \cdots = \alpha_a = 0$, or $H_0 : \bar{\mu}_{1..} = \cdots = \bar{\mu}_{a..}$

Test: reject H_0 if

$$F = \frac{MS_A}{MS_{E(wholeplot)}} > F_\alpha(a - 1, (a - 1)(n - 1)).$$

A type 2 split plot design can be analyzed in PROC MIXED as follows:

```
proc mixed;
  class A B WPB;
  model y=A WPB B A*B / ddfm=satterth;
  random A*WPB;
  contrast 'contrast in marginal means for A' A a1 a2 ... aa;
  contrast 'contrast in marginal means for B' B b1 b2 ... bb;
  contrast 'any joint mean contrast' A a1 ... aa
                                     B b1 ... bb
                                     A*B c1 c2 c3 ... cab;
  estimate 'any linear combination of parameters' intercept d
          A a1 ... aa B b1 ... bb A*B c1 c2 ... cab;
  lsmeans A B A*B;
run;
```

- Note that here we've assumed that whole plot blocks are fixed. Often, it will be more appropriate to model whole plot blocks with random effects. That is, we change the assumption on the τ_j 's to

$$\tau_1, \dots, \tau_n \stackrel{iid}{\sim} N(0, \sigma_\tau^2),$$

and we assume that the τ_j 's are independent of the other random effects in the model.

In that case, our MODEL and RANDOM statements would change to

```
model y=A B A*B / ddfm=satterth;
random WPB A*WPB;
```

Alfalfa Example:

- See alfalfa.sas and its output, alfalfa.pdf. Here, we treat whole plot blocks as random rather than fixed.
- In this example it is natural to consider a comparison between no cut and the three cutting times:

$$\begin{array}{cccccc} & \text{No Cut} & \text{Sept. 1} & \text{Sept. 20} & \text{Oct. 7} & \\ \psi : & 3 & -1 & -1 & -1 & \text{(control vs. cuts)} \end{array}$$

- Notice that main effects for cutting times, interactions between varieties and cutting times, and the above contrast in cutting times are tested using the split plot error term because they involve the split plot factor.
- In contrast, main effects for variety are tested using the whole plot error term, which is given by the mean squares for varieties*blocks.
- From the anova table we find little evidence of a variety*date interaction ($F = 1.25$, $p = .2973$) so we can proceed to examine main effects.
- We have strong, significant evidence of date effects ($F = 23.39$, $p < .0001$) and from the means we see that no cut and the late cut mean yields are greatest. There is virtually no evidence of variety effects ($F = 0.65$, $p = .5412$).
- The control vs. cuts contrast is also found to be significant ($F = 29.15$, $p < .0001$). From the means it is apparent that no cut results in a higher mean yield.

Inference on Means and Contrasts in a SPD (either type):

Marginal means for whole plot factor:

Estimate $\bar{\mu}_{i..}$, the mean at the i^{th} level of A and its standard error, using

$$\hat{\mu}_{i..} = \bar{y}_{i..}, \quad \text{s.e.}(\bar{y}_{i..}) = \sqrt{\hat{\text{var}}(\bar{y}_{i..})} = \sqrt{\frac{MS_{E(\text{wholeplot})}}{bn}}$$

CIs and contrasts for marginal means for whole plot factor:

100(1 - α)% CI for $\bar{\mu}_{i..}$:

$$\bar{y}_{i..} \pm t_{\alpha/2}(\text{d.f.}_{E(\text{wholeplot})}) \text{s.e.}(\bar{y}_{i..}) = \bar{y}_{i..} \pm t_{\alpha/2}(\text{d.f.}_{E(\text{wholeplot})}) \sqrt{\frac{MS_{E(\text{wholeplot})}}{bn}}$$

For a contrast $\psi = \sum_i c_i \bar{\mu}_{i..}$ with sample estimator $C = \sum_i c_i \bar{y}_{i..}$, we test $H_0 : \psi = 0$ in the usual way. Our test statistic is

$$t = \frac{|C|}{\text{s.e.}(C)}, \quad \text{which we compare to } t_{\alpha/2}(\text{d.f.}_{E(\text{wholeplot})})$$

for an α -level test. Equivalently, we can use an F test:

$$F = t^2, \quad \text{which we compare to } F_{\alpha}(1, \text{d.f.}_{E(\text{wholeplot})}).$$

The standard error for a contrast in the marginal means of A is given by

$$\text{s.e.}(C) = \sqrt{\hat{\text{var}}\left(\sum_i c_i \bar{y}_{i..}\right)} = \sqrt{\sum_i c_i^2 \hat{\text{var}}(\bar{y}_{i..})} = \sqrt{\frac{MS_{E(\text{wholeplot})}}{bn} \sum_i c_i^2}.$$

A 100(1 - α)% CI for ψ is given by

$$C \pm t_{\alpha/2}(\text{d.f.}_{E(\text{wholeplot})}) \text{s.e.}(C).$$

Marginal means for split plot factor and treatment (joint) means:

For marginal means for the split plot factor and for joint means in the split-plot design, the estimators that are based on the split-plot model are the nice, simple sample means that one would expect. However, the standard errors of these means are not so simple.

- As a consequence, t -statistics for means and for most contrasts are not *exactly* t distributed!

Fortunately, these quantities generally have distributions that are well approximated by a t distribution with degrees of freedom that are estimated from the data. The usual degree of freedom estimation method is due to a statistician named Satterthwaite.

Marginal means for split plot factor:

Estimate $\bar{\mu}_{..k}$, the mean at the k^{th} level of B and its standard error, using

$$\hat{\mu}_{..k} = \bar{y}_{..k}, \quad \text{s.e.}(\bar{y}_{..k}) = \sqrt{\hat{\text{var}}(\bar{y}_{..k})} = \sqrt{\frac{MS_{E(\text{wholeplot})} + (b-1)MS_{E(\text{splitplot})}}{abn}}$$

- Notice that this standard error is a function of two MS_E 's. As a consequence $\frac{\bar{y}_{..k} - \mu_{..k}}{\text{s.e.}(\bar{y}_{..k})}$ is only approximately t distributed. This leads to approximate (but generally quite accurate) inference methods on the $\hat{\mu}_{..k}$'s.

An approximate $100(1 - \alpha)\%$ CI for $\bar{\mu}_{..k}$ is given by

$$\bar{y}_{..k} \pm t_{\alpha/2}(\nu) \text{s.e.}(\bar{y}_{..k}). \quad (*)$$

Usually, ν is the d.f. associated with the mean square in the formula for the s.e. of our point estimate ($\bar{y}_{..k}$ in this case). However, in this case the formula for $\text{s.e.}(\bar{y}_{..k})$ involves two mean squares.

So, what should ν be?

The answer is given by **Satterthwaite's formula**. Satterthwaite showed that a linear combination of independent mean squares of the form $MS = a_1MS_1 + \cdots + a_kMS_k$ has approximate degrees of freedom given by

$$\text{d.f.} = \frac{(MS)^2}{\frac{(a_1MS_1)^2}{\text{d.f.}_1} + \cdots + \frac{(a_kMS_k)^2}{\text{d.f.}_k}},$$

where here d.f._i is the d.f. associated with MS_i and the a_i 's are constants.

In our case, MS is $MS_{E(\text{wholeplot})} + (b-1)MS_{E(\text{splitplot})}$ so we have

$$\nu = \frac{[MS_{E(\text{wholeplot})} + (b-1)MS_{E(\text{splitplot})}]^2}{\frac{(MS_{E(\text{wholeplot})})^2}{\text{d.f.}_{E(\text{wholeplot})}} + \frac{[(b-1)MS_{E(\text{splitplot})}]^2}{\text{d.f.}_{E(\text{splitplot})}}},$$

which is intermediate between $\text{d.f.}_{E(\text{wholeplot})}$ and $\text{d.f.}_{E(\text{splitplot})}$.

Joint means:

Similarly, we estimate $\bar{\mu}_{i.k}$, the population mean at the i^{th} level of A and the k^{th} level of B and its standard error, using

$$\hat{\mu}_{i.k} = \bar{y}_{i.k}, \quad \text{s.e.}(\bar{y}_{i.k}) = \sqrt{\hat{\text{var}}(\bar{y}_{i.k})} = \sqrt{\frac{(b-1)MS_{E(\text{splitplot})} + MS_{E(\text{wholeplot})}}{nb}}$$

Again, the standard error of our point estimate ($\bar{y}_{i.k}$) involves two mean squares rather than just one, so we must use Satterthwaite's approximation to get the appropriate d.f. for our confidence interval. A $100(1 - \alpha)\%$ CI for $\bar{\mu}_{i.k}$ is given by

$$\bar{y}_{i.k} \pm t_{\alpha/2}(\nu)\text{s.e.}(\bar{y}_{i.k}),$$

where ν is as given above.

Contrasts in the joint or marginal means for factor B:

As usual, for a contrast ψ , we form test statistics in the usual way. That is, t and F statistics for $H_0 : \psi = 0$ are given by

$$t = \frac{|C|}{\text{s.e.}(C)}, \quad \text{and} \quad F = t^2, \quad \text{respectively.}$$

However, the formula for $\text{s.e.}(C)$ and the distribution of the test statistics depend on the nature of the contrast. For contrasts in the marginal means for factor B and for certain joint mean contrasts (but not others) $\text{s.e.}(C)$ involves $MS_{E(\text{splitplot})}$ only; for others, $\text{s.e.}(C)$ involves both MS_{ES} . The former case is easy and yields an exact t and F tests, the latter case is harder because we need Satterthwaite's formula again and we get approximate (but accurate) tests.

- The case where Satterthwaite's formula is needed is when our contrast ψ involves a comparison of joint means at different levels of factor A .
- Good mixed model software can determine the correct standard errors and inference methods. E.g., SAS' PROC MIXED can do so if we use the `ddfm=SATTERTH` option on the MODEL statement.

Repeated Measures

Repeated measures consist of multiple observations on each of the subjects, plots, animals, etc. under study in an experiment. Multiple observations are usually, but not always, taken on multiple occasions in time. Almost any experimental design can be altered to include repeated measures (an exception is the crossover design which is already a repeated measures design).

Strengths of Repeated Measures Studies:

- Economizes on subjects.
- In crossover studies, subjects “serve as their own controls”; i.e., comparisons between treatments can be made within a subject. More generally, comparisons of the response across time can be made within a subject.
- Between-subjects sources of variability are excluded from experimental error.
- Provide more efficient estimators (more power) of some parameters than cross-sectional designs with the same number and pattern of measurements.
- Information is more reliably quantified than in a cross-sectional study.
- Provide information concerning individual patterns of change.

Drawbacks to Repeated Measures Studies:

- Dependence of repeated measures on a particular individual must be accounted for in analysis. This makes the analysis more intricate and challenging.
- Unbalanced designs due to missing data and/or attrition more common.
- Carry-over effects (when different treatments are applied at different measurement occasions).
- If one is not interested in patterns of change over time and if a crossover design is inappropriate, then repeated measures designs are not recommended.

Methods of Analysis for (Continuous) Repeated Measures:

1. Univariate Analysis (a.k.a., Summary Statistic Approach, Response Feature Analysis).
2. Multivariate Analysis (profile analysis, growth curve analysis).
3. Repeated Measures ANOVA (use linear mixed models based on split-plot analysis).
 - We will briefly mention all three of these approaches, but concentrate on 3.

Example – Methemoglobin in Sheep:

An experiment was designed to study trends in methemoglobin (M) in sheep following treatment with 3 equally spaced levels of NO_2 (factor A). Four sheep were assigned to each level and each animal was measured at 6 sampling times (factor B), 5 of them following treatment. The response was $\log(M + 5)$.

The data from this experiment are as follows:

NO ₂	Sheep	Sampling Time					
		1	2	3	4	5	6
1	1	2.197	2.442	2.542	2.241	1.960	1.988
1	2	1.932	2.526	2.526	2.152	1.917	1.917
1	3	1.946	2.251	2.501	1.988	1.686	1.841
1	4	1.758	2.054	2.588	2.197	2.140	1.686
2	5	2.230	3.086	3.357	3.219	2.827	2.534
2	6	2.398	2.580	2.929	2.874	2.282	2.303
2	7	2.054	3.243	3.653	3.811	3.816	3.227
2	8	2.510	2.695	2.996	3.246	2.565	2.230
3	9	2.140	3.896	4.246	4.461	4.418	4.331
3	10	2.303	3.822	4.109	4.240	4.127	4.084
3	11	2.175	2.907	3.086	2.827	2.493	2.230
3	12	2.041	3.824	4.111	4.301	4.206	4.182

General form of repeated measures data:

In general we have t measurements on S units (subjects, plots, animals, etc.):

$$y_{ij}, \quad \begin{array}{l} i = 1, \dots, S \text{ (units)} \\ j = 1, \dots, t \text{ (repeated measures)} \end{array}$$

For a particular design we may have slightly different notation, where y is indexed by several indices, the last of which indexes repeated measures and the rest index the design on the S units. For example, in the sheep experiment we have

$$y_{ijk}, \quad \begin{array}{l} i = 1, \dots, a \text{ (treatments=NO}_2 \text{ levels)} \\ j = 1, \dots, n \text{ (replicates/treatment)} \\ k = 1, \dots, t \text{ (repeated measures=sampling times)} \end{array}$$

and $S = an$.

1. Univariate Methods:

This approach avoids the problems of correlated repeated measures by reducing y_{i1}, \dots, y_{it} , the collection of responses on the i^{th} unit (e.g., subject) to a single summary statistic. Once this reduction from a **multivariate** to a **univariate** response has been done, the usual (univariate) analysis appropriate to the non-repeated measures version of the experimental design can be conducted using the summary statistic as the (univariate) response variable.

In the simplest case we have two observations on each subject and instead of analyzing two responses per subject we analyze the difference between the two observations for each subject. That is, instead of analyzing y_{i1}, y_{i2} over all i , fit the appropriate univariate model to $z_i = y_{i2} - y_{i1}$. In this case there is no loss of information if interest is focused on the change over time.

- An example of this approach is the paired t -test.

More generally, we reduce the t observations on each subject to a single statistic, in the process losing some information, but simplifying the analysis. If interest is in the change over time, one possibility is to fit a simple linear regression for each unit: for each i fit

$$y_{ij} = \alpha_i + \beta_i x_j, \quad j = 1, \dots, t,$$

where x_1, \dots, x_t are the observation times. Thus, we obtain S slopes, $\hat{\beta}_1, \dots, \hat{\beta}_S$, that we can use as the values of a single response variable in a univariate analysis.

2. Multivariate Analysis:

If the number of repeated measures t is the same for each subject, and if the times at which the repeated measures were taken is the same for all subjects (as in the sheep example), then we can think of $y_{i1}, y_{i2}, \dots, y_{it}$ as the values of t separate variables measured on the i^{th} subject.

That is, we can think of having t separate response variables (the response at time 1, the response at time 2, ... the response at time t) for each subject. From this perspective, we have a multivariate response variable rather than a univariate one, and all of the methods of multivariate analysis are then available to us.

- See courses STAT 8210 (for statistics students) and STAT 8250 (for non-statistics students) for coverage of multivariate analysis. In particular, the areas of profile analysis and growth curve analysis are most relevant to the analysis of repeated measures data.
- Multivariate analyses of repeated measures data tend to have somewhat lower power/efficiency relative to repeated measures ANOVA. The reason for this is that, in multivariate analysis, no assumptions are made about the correlation/covariance structure among repeated measures. With some assumptions on this structure, we can do better.

3. Repeated Measures ANOVA:

This approach can be described as a two step method:

- i. Perform a split plot analysis treating time as the split-plot factor.
- ii. Since time is not randomized in a repeated measures study, its not really quite the same as a split plot factor. Therefore, modify or “fix” the split-plot analysis to make it valid.

Consider the sheep example. These data are laid out in a manner very similar to the data from the chocolate cake example. That is, we have $S = 12$ whole plot experimental units (sheep), which are randomly assigned to a whole plot treatment factor (NO_2 level) in a one-way layout. Within these experimental units we have another level of experimental unit. The whole plots (sheep) are “divided” into six “split plots” (the sheep at the six sampling times) on which a second treatment factor (time) is applied.

Difference: the treatment factor playing the role of the split plot factor (time) is not randomized to the “split plots” in the repeated measures design (time can’t be randomized).

Question: Is it valid to ignore this difference and do the split plot analysis anyway?

Answer: Yes, if and only if the variance-covariance matrix of $\mathbf{y}_i = (y_{i1}, \dots, y_{it})$ is **circular** (another terminology is that $\text{var}(\mathbf{y}_i)$ must satisfy the *Huynh-Feldt conditions*).

The Variance-Covariance Matrix:

Let $\sigma_{jj'} = \text{cov}(y_{ij}, y_{ij'})$ for all i . $\Rightarrow \sigma_{jj} = \text{var}(y_{ij})$. The variance-covariance matrix of $\mathbf{y}_i = (y_{i1}, \dots, y_{it})$ is denoted as Σ and is defined to be

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1t} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2t} \\ \vdots & & \ddots & \vdots \\ \sigma_{t1} & \sigma_{t2} & \cdots & \sigma_{tt} \end{pmatrix}.$$

- Σ is just the collection of all of the variances of y_{i1}, \dots, y_{it} (on the diagonal) and all of the covariances between pairs of $(y_{i1}, y_{i2}), (y_{i1}, y_{i3}), \dots$

Circularity, Compound Symmetry:

It turns out that it is valid to use the univariate ANOVA (e.g., the split plot analysis) if and only if Σ is circular.

- Σ is circular if all pairwise differences between repeated measures have equal variance; that is, if

$$\text{var}(y_{ij} - y_{ij'}) \quad \text{is constant for all time points } j, j'.$$

- Equivalently, Σ is circular if the variances of the elements $\sigma_{jj'}$ can be expressed as

$$\sigma_{jj'} = \gamma_j + \gamma_{j'} + \lambda\delta_{jj'},$$

where

$$\delta_{jj'} = \begin{cases} 1 & \text{if } j = j' \\ 0 & \text{if } j \neq j' \end{cases},$$

and $\lambda, \gamma_1, \gamma_2, \dots, \gamma_t$ are (unknown) constants.

Huh?

A conceptually simpler *sufficient* (but not necessary) condition is **compound symmetry**. The univariate analysis is valid if (but not only if) Σ has the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}.$$

That is, if Σ has compound symmetry, the analysis is valid.

- Compound symmetry is a slightly stronger condition than circularity (Compound Symmetry implies Circularity, but Circularity does not imply Compound Symmetry). Nevertheless, as a practical matter, we can think of this as *the* condition we need for the split-plot analysis to be valid for repeated measures.
- Note that compound symmetry is not always the most intuitively reasonable structure for $\text{var}(\mathbf{y}_i)$. In some applications, we would expect that the correlation is higher for observations taken close together in time than for observations taken far apart in time. Compound symmetry says that the correlation is the same no matter how far apart in time two observations on the same subject are taken.

Mauchly has proposed a test for circularity. This test is of somewhat help-ful for determining whether circularity holds, but it has several problems:

- Low power in small samples.
- In large samples test is likely to reject circularity when non-circularity has little effect on validity of the analysis that assumes circularity.
- Sensitive to departures from normality.
- Very sensitive to outliers.

Therefore, rather than testing the null hypothesis of circularity and then adjusting the analysis only if we reject that hypothesis, the typical ap-proach is to adjust the analysis according to the degree of non-circularity present in the data.

- That is, the modification of the split-plot analysis is a function of the degree of non-circularity evidenced in the data, with no evidence resulting in no modification.

Modifications to Univariate Analysis under Non-circularity:

The degree of non-circularity in the data can be measured by

$$\theta = \frac{t^2(\bar{\sigma}_{jj} - \bar{\sigma}_{..})^2}{(t-1)(\sum_{j=1}^t \sum_{j'=1}^t \sigma_{jj'}^2 - 2t \sum_{j=1}^t \bar{\sigma}_{j.}^2 + t^2 \bar{\sigma}_{..}^2)},$$

where

$\bar{\sigma}_{..}$ = mean of all elements of Σ

$\bar{\sigma}_{jj}$ = mean of elements on main diagonal of Σ

$\bar{\sigma}_{j.}$ = mean of elements in row j of Σ

It can be shown that circularity holds when $\theta = 1$ and maximum non-circularity holds when $\theta = 1/(t-1)$.

- The “fix” of the univariate analysis is to multiply the numerator and denominator degrees of freedom by $\hat{\theta}$ for all F tests on the repeated measures factor and interactions involving the repeated measures factor (usually time).

Two estimators of θ are commonly used: Greenhouse-Geisser and Huynh-Feldt. $\hat{\theta}_{GG}$ is simply θ computed on the sample variance-covariance matrix $\hat{\Sigma}$ rather than Σ , the population (true) variance-covariance matrix. $\hat{\theta}_{HF}$ is defined as

$$\hat{\theta}_{HF} = \min \left(1, \frac{S(t-1)\hat{\theta}_{GG} - 2}{\underbrace{(t-1)(\text{d.f.}_{E(W.P.)} - (t-1)\hat{\theta}_{GG})}_{\text{value given by SAS}}} \right),$$

where S = the total number of subjects, or “whole plots”.

- Note that $\theta \leq 1$, so the value given by SAS should always be rounded down to 1, if it is > 1 .

Which $\hat{\theta}$? For true $\theta \leq 0.5$ (greater non-circularity) $\hat{\theta}_{GG}$ is better. For true $\theta \geq 0.75$ (less non-circularity) $\hat{\theta}_{HF}$ is better. In practice we don't know the true value of θ so often it is hard to say which is better.

- $\hat{\theta}_{GG}$ tends to give the larger adjustment (makes it harder to reject H_0) than $\hat{\theta}_{HF}$, so if we desire to be conservative (slow to reject) then we should use $\hat{\theta}_{GG}$.

If we have a program like SAS that can compute $\hat{\theta}_{GG}$ and $\hat{\theta}_{HF}$ and corresponding adjusted p -values easily, then we should always go ahead and do the adjustment for non-circularity.

- There is no down-side here, because if the data are circular, then we should get $\hat{\theta}_{GG} = \hat{\theta}_{HF} = 1$, and the adjustment will end up not altering the split-plot analysis at all (which is what we would want). If the data are non-circular, then an appropriate adjustment will be done.

On the other hand, if we do not have SAS or some other program to compute the adjustment factors, then we'd like to be able to proceed with the analysis without computing $\hat{\theta}_{GG}$ or $\hat{\theta}_{HF}$. In many cases, the following algorithm will allow us to do just that

Greenhouse-Geisser Algorithm for Repeated Measures ANOVA:

- Should be applied to main effects of the repeated measures factor (time), and interactions involving the repeated measures factor.

The “fix” that we use when the data do not satisfy the circularity condition requires us to compute $\hat{\theta}$. Since calculating $\hat{\theta}$ is a bit involved and there is an issue of which $\hat{\theta}$ to use, we’d like to avoid this step if possible. Greenhouse and Geisser have proposed the following algorithm which often allows us to make the appropriate conclusion without having to compute $\hat{\theta}$.

1. Conduct the unadjusted F test (from split plot analysis). If not significant, then we know that the adjusted F test will not be significant either because unadjusted $F_{\text{crit}} < \text{adjusted } F_{\text{crit}}$. Therefore, stop and “accept” H_0 . Else do step 2.
 2. Conduct most conservative test, corresponding to $\theta = 1/(t - 1)$. Compare F test statistic to $F_{\alpha}(\text{d.f.}_{\text{num}}/(t - 1), \text{d.f.}_{\text{denom}}/(t - 1))$ (the most conservative F_{crit}). If significant, then we know that the adjusted F test will be significant too because the most conservative $F_{\text{crit}} > \text{adjusted } F_{\text{crit}}$. Therefore, stop and reject H_0 . Else do step 3.
 3. Compute $\hat{\theta}$ and conduct the adjusted F test based on $F_{\alpha}(\hat{\theta}\text{d.f.}_{\text{num}}, \hat{\theta}\text{d.f.}_{\text{denom}})$. To choose estimator of θ note that $\hat{\theta}_{GG}$ is more conservative than $\hat{\theta}_{HF}$.
- Under non-circularity, F tests in repeated measures ANOVA are approximate, in the sense that the distribution of the test statistic is approximately $F(\hat{\theta}\text{d.f.}_{\text{num}}, \hat{\theta}\text{d.f.}_{\text{denom}})$ rather than exactly so.

Methemoglobin in Sheep, Continued:

- The experimental design here is a one-way layout with repeated measures. It is very similar in structure to the chocolate cake example (type 1 split plot design).

Model:

$$y_{ijk} = \mu + \alpha_i + e_{ij} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}, \quad \begin{array}{l} i = 1, \dots, a \text{ (levels of NO}_2\text{)} \\ j = 1, \dots, n \text{ (replicates)} \\ k = 1, \dots, t \text{ (sampling times)} \end{array} .$$

Here,

- α_i = the effect of the i^{th} level of NO₂,
- e_{ij} = the (whole plot) error term for NO₂,
- β_k = the effect of the k^{th} sampling time,
- $(\alpha\beta)_{ik}$ = interaction terms for NO₂ × Time interactions,
- ε_{ijk} = the (split plot) error term for Time and NO₂ × Time interactions.

$a = 3$, $n = 4$ and $t = 6$ in this design.

- See sheep1.sas and sheep1.lst for analysis of this example.
- Sheep1.sas contains two approaches to the analysis of these data. First, PROC MIXED is used to fit the split-plot model. PROC MIXED will not produce G-G or H-F estimates of θ , the non-circularity parameter, nor will it produce G-G or H-F adjusted p -values for terms involving time. However, it can be used with the G-G algorithm to generate correct inferences.
- Alternatively, PROC GLM will produce G-G and H-F estimates of θ and adjusted p -values if the REPEATED statement is used. However, the data must be input into PROC GLM in a multivariate format (i.e., one variable for each measurement occasion).
- We first consider the results from PROC MIXED, and use the G-G algorithm to make our conclusions.

- We first test for interaction between NO₂ and Time. This test must be performed using the Greenhouse-Geisser algorithm:

Step 1: Test with univariate F test: $F = 5.62$ ($p = .0001$) so we go to step 2.

Step 2: Conduct most conservative test. Compare $F = 5.62$ to $F_{.05}(10/5, 45/5) = F_{.05}(2, 9) = 4.26$. Based on this test we reject, so we can stop and conclude that there is an interaction between time and NO₂.

- We can see the interaction in the profile plot on p.5 of the output. All three NO₂ groups have a pattern over time which is clearly not linear, consisting of an initial increase followed by a decrease after the third or fourth sampling time. The interaction is seen in that the initial increase is much more dramatic for level 3 of NO₂ than for level 1. In addition, the decrease after time 4 for level 3 of NO₂ is relatively small, whereas for level 1 of NO₂ this decrease brings the mean down below its original level by time period 6.
- The interaction between time and level of NO₂ is present, however it seems to be an orderly interaction allowing for testing of main effects. We next test for a time effect, which again requires the Greenhouse-Geisser algorithm:

Step 1: Test with univariate F test: $F = 22.37$ ($p = .0001$) so we go to step 2.

Step 2: Conduct most conservative test: Compare $F = 22.37$ to $F_{.05}(5/5, 45/5) = F_{.05}(1, 9) = 5.12$. Based on this test we reject, so we can stop and conclude that there is a significant time effect (this is clear from the profile plots).

- Finally, we can test for main effects of NO₂. Since this is the “whole plot factor” we need not use the G-G algorithm. We simply compare MS_A with $MS_{E(W.P.)}$. Our test statistic is $F = 12.32$ ($p = .0026$) so we reject the null hypothesis that the mean response is the same at the three levels of NO₂ (this also is clear from the profile plot).

- Now consider the analysis using PROC GLM.
 - To use the REPEATED statement, the data must be arranged in a multivariate format, with a separate response variable for each time point (sampling time) in the study. The model is specified with all six time-specific variables on the left side of the equals sign. “REPEATED TIME;” tells SAS to regard those six response variables as repeated measures of the a repeated measures factor to be called “TIME” in the SAS output.
- The ‘/ NOUNI’ option on the MODEL statement tells SAS to omit the univariate analyses on time1, time2, ..., time6. By default SAS will fit separate univariate models for each time point as well as a multivariate model for all times simultaneously.
- The ‘/ NOM’ option on the REPEATED statement tells SAS to omit a full multivariate analysis.
- The ‘/ PRINTE’ option on the REPEATED statement is necessary to obtain Mauchly’s test of circularity. This test is identified in the SAS output as a “Sphericity Test”. The one we want is the one for “Orthogonal Components” (see p.8 of the output).
- Based on Mauchly’s test, we reject the null hypothesis of circularity (called sphericity in the SAS output) ($p < .0001$). In addition, SAS provides G-G and H-F estimates of the non-circularity parameter θ (called epsilon, in the SAS output) at the bottom of p.10. Since these estimates are much closer to the maximum non-circularity value of $1/(t - 1) = 1/(6 - 1) = .20$ than to the minimum non-circularity value of 1, they also provide clear evidence of non-circularity.
- G-G and H-F adjusted p -values for time*no2 and time are given on p.10. As before, we conclude that there is a significant main effect of time and a significant interaction here. Notice that the G-G and H-F adjustments adjust the p -value upward to account for non-circularity.
- Finally, notice that the test of no2 is on p.9. This test needs no adjustment since it is on the “whole plot factor”, and the result from PROC GLM agrees exactly with that from PROC MIXED.

- Further analysis of these data is probably appropriate. For example, we could examine the trend over time separately at each of the three NO₂ levels. It would be of interest to determine if these trends were linear; or, if not, if they were adequately described by some low-order polynomial. This would involve contrasts in the joint means at no2 by time combinations.
- If the group \times time interaction had not been significant, that would suggest testing orthogonal polynomial contrasts in time (averaging across groups). This would involve contrasts in the marginal means for time.
- For testing such contrasts it makes sense to adjust for non-circularity with a G-G or H-F adjustment. However, the numerator d.f. for any single contrast is already 1, so it is not clear how to make the adjustment (the adjusted numerator d.f. would be $\hat{\theta}$ d.f._{num}, which would be between 0 and 1, so what numerator d.f. would we use?).
- Repeated measures anova is not well-suited to inference on contrasts in the marginal means for time or for contrasts in the joint means for time by treatment combinations. Such contrasts can be investigated, but more sophisticated methods are necessary (e.g., profile analysis, polynomial growth curve analysis, or analysis based upon mixed models with non-compound symmetry variance-covariance assumptions). See STAT 8250, STAT 8210, or STAT 8630 for details.

Example – Asparagus:

An agronomist wished to compare the effects of various cutting times on subsequent yields of asparagus. Asparagus may be harvested at multiple times during a growing season. However, it was suspected that later harvest times would reduce the vigor of the plant, thereby reducing the yield in subsequent years. Four cutting times in the year 1929 were considered: (A_1) June 1, (A_2) June 15, (A_3) July 1, and (A_4) July 15. Cutting times were arranged in a randomized complete block design. The yields were observed over four successive years: 1930 (T_1), 1931 (T_2), 1932 (T_3), and 1933 (T_4).

Experimental Design: randomized complete block design with repeated measurements.

The data and design are as follows:

Block 1				Block 2			
A_1	A_2	A_4	A_3	A_2	A_4	A_3	A_1
T_1 230	T_1 212	T_1 148	T_1 183	T_1 190	T_1 126	T_1 186	T_1 216
T_2 324	T_2 415	T_2 246	T_2 320	T_2 296	T_2 201	T_2 295	T_2 317
T_3 512	T_3 584	T_3 304	T_3 456	T_3 471	T_3 289	T_3 387	T_3 448
T_4 399	T_4 386	T_4 144	T_4 255	T_4 280	T_4 83	T_4 187	T_4 361

Block 3				Block 4			
A_1	A_3	A_2	A_4	A_3	A_1	A_2	A_4
T_1 219	T_1 177	T_1 151	T_1 107	T_1 209	T_1 200	T_1 150	T_1 168
T_2 357	T_2 298	T_2 278	T_2 192	T_2 328	T_2 362	T_2 336	T_2 226
T_3 496	T_3 427	T_3 399	T_3 271	T_3 462	T_3 540	T_3 485	T_3 312
T_4 344	T_4 239	T_4 254	T_4 90	T_4 244	T_4 381	T_4 279	T_4 168

Model:

$$y_{ijk} = \mu + \alpha_i + \tau_j + e_{ij} + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}, \quad \begin{array}{l} i = 1, \dots, a \text{ (cutting times)} \\ j = 1, \dots, b \text{ (blocks)} \\ k = 1, \dots, t \text{ (years)} \end{array} .$$

Here,

- α_i = the effect of the i^{th} cutting time,
- τ_j = the effect of the j^{th} block (assumed fixed here),
- e_{ij} = the (whole plot) error term for cutting times,
- β_k = the effect of the k^{th} year,
- $(\alpha\beta)_{ik}$ = interaction terms for cutting time*year interactions,
- ε_{ijk} = the (split plot) error term for years and cutting time*year interactions.

- See SAS program `asparagus.sas` and its output. As in `sheep1.sas`, `asparagus.sas` contains a call to PROC MIXED and one to PROC GLM.
- In the call to PROC MIXED, we do the standard split-plot analysis, without any automatic adjustments. Adjustments or the G-G algorithm can be done “by hand” from this output.
- In the first call to PROC MIXED, the split-plot analysis is based on the model at the top of the page, and is essentially the same as used for a split-plot design with whole plots in a RCBD (for example, the alfalfa data).
- On pp.1–2 of `asparagus.lst`, the appropriate F statistics for the “whole plot factors” cut and block are those that use the “whole plot error term”, $MS_{cut*block}$: $F = 33.12$ and $F = 4.14$, respectively. These do not have to be adjusted for non-circularity, and are correct as given, although since blocks represent a restriction on randomization, we would not formally test for block effects here.
- However, we really should test interaction between cut and year before examining main effects of cut. The interaction between cut and year and the main effect of year both involve the repeated measures factor, year, so need to be approached with the G-G algorithm (or by using G-G or H-F adjustments, but see the call to PROC GLM for this).

- For cut*year: Step 1: Do split-plot F test. This is significant ($F = 13.22$, $p < .0001$) so we go to step 2. Step 2: The most conservative θ value is $\theta = 1/(4 - 1) = 1/3$. Compare $F = 13.22$ to $F_{.05}(9/3, 36/3) = F_{.05}(3, 12) = 3.49$. Since this is significant, stop and reject H_0 . Conclusion: there is a significant interaction that we need to examine with a profile plot.
- An examination of the profile plot (p.4) suggests that there is only a small exception (at year 0) to the general trend that treatments 1, 2, 3, and 4 are ordered from best to worst. The exception at year 0, is small, in the sense that treatments 2 and 3 are reversed, and they are close to one-another. The safest thing here is to compare cutting times within years and compare years within cutting times, but in my opinion, it is appropriate to compare marginal means of cutting time and year separately, because the pattern across years and across cutting times is almost completely consistent.
- To compare years, we again must use the G-G algorithm. Step 1: significant ($F = 401.94$, $p < .0001$) so go to step 2. Step 2: compare $F = 401.94$ to $F_{.05}(3/3, 36/3) = 4.75$. Therefore, reject H_0 and conclude that there are significant differences across years.
- Main effects of cutting time are also significant ($F = 33.12$, $p < .0001$). An examination of the lsmeans and the profile plots reveals the general pattern: for all four treatments, yield is reduced considerably in the year immediately following treatment. Yield then increases until peaking three years after treatment and then then drops again at four years. Later cutting times generally reduce the yield in subsequent years more than earlier cutting times, as one would expect.
- In the call to PROC GLM, we use the REPEATED statement to obtain the full repeated measures anova, complete with estimates $\hat{\theta}_{GG}$ and $\hat{\theta}_{HF}$, G-G and H-F-adjusted p -values, and Mauchley test of circularity.

- However, **note** that the repeated measures ANOVA model fit using PROC GLM and the REPEATED statement differs slightly from the one at the top of p.234 (the one fit with PROC MIXED in asparagus.sas)!
- By using the REPEATED statement, PROC GLM automatically generates interaction terms for year and all “whole plot” factors. That is, the model includes an interaction for cut*year $((\alpha\beta)_{ik})$ and for cut*block $((\tau\beta)_{jk})$. Given the design of the experiment, the cut*block term is probably unnecessary, so it has been omitted from the model fit in PROC MIXED. If it had been included, then the results from PROCs GLM and MIXED would agree perfectly.
- Notice in the output (p.8) that the F test statistics for cut and block given by PROC GLM agree exactly with those obtained in the split plot analysis. If we had included a block*year interaction in the split-plot model, then the F tests and unadjusted p -values from p.9 would have agreed perfectly as well (try it).
- Based on Mauchly’s test ($p = .4512$ on p.7), we fail to reject the null hypothesis of circularity. In addition, the G-G and (especially) the H-F estimates of the non-circularity parameter (bottom of p.9) are close to 1. Therefore, there is little evidence of non-circularity in these data, and the adjusted p -values for year, year*block and year*cut are only slightly adjusted by the G-G adjustment, and not adjusted at all by the H-F adjustment.
- Finally, note that in this analysis of these data, we have treated the block effects (the τ_j ’s) as fixed rather than random. If we feel that a random effects assumption on these quantities is more appropriate (which it probably is in this example), then PROC GLM cannot be used to get the correct analysis. In that case, PROC MIXED should be used to do the analysis, with the following MODEL and RANDOM statements:

```
model yield= cut year cut*year / ddfm=satterth;
random block block*cut;
```

or, if a block*year interaction term is desired:

```
model yield= cut year cut*year / ddfm=satterth;
random block block*cut block*year;
```

LMM Approach to Repeated Measures Analysis

- The RM-ANOVA is based upon using the split-plot model to analyze repeated measures. Such a model makes the assumption that the variance-covariance structure of the data satisfies compound symmetry.
- Because the compound symmetry assumption is often too simplistic for repeated measures data, RM-ANOVA attempts to “fix” the analysis from the split-plot model (the wrong model) by doing a somewhat *ad hoc* adjustment to F statistics and p -values.
- Although this approach often works reasonably well, it is a bit crude, and it has limitations (e.g., it says nothing about how to adjust contrasts across levels of time).
- Recall the split-plot model is a linear mixed model (LMM). Not all LMMs necessarily assume compound symmetry! This class of models is quite broad, and allows the variance-covariance structure of the data to be modeled in a wide variety of ways.
- So, rather than assuming a split-plot model (or, equivalently, instead of assuming compound symmetry in our LMM) and fixing the analysis according to how wrong that assumption is, it makes more sense to just fit a more appropriate variance-covariance structure model for our data to begin with.
 - We do this by changing our model from one that assumes compound symmetry, to one that assumes something more appropriate to the data at hand.

Compound symmetry assumption inherent in the split-plot model:

Recall our case 1 split plot model (e.g., the chocolate cake example):

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + e_{j(i)} + \varepsilon_{ijk} \\ &= \bar{\mu}_{i \cdot k} + e_{j(i)} + \varepsilon_{ijk} \end{aligned}$$

where the $e_{j(i)}$ s are $N(0, \sigma_e^2)$ and the ε_{ijk} s are $N(0, \sigma^2)$.

According to this model,

$$\begin{aligned} \text{var}(y_{ijk}) &= \text{var}(\bar{\mu}_{ik} + e_{j(i)} + \varepsilon_{ijk}) = \text{var}(e_{j(i)} + \varepsilon_{ijk}) \\ &= \text{var}(e_{j(i)}) + \text{var}(\varepsilon_{ijk}) + \underbrace{2 \text{cov}(e_{j(i)}, \varepsilon_{ijk})}_{=0} = \sigma_e^2 + \sigma^2. \end{aligned}$$

- Because the variance of y_{ijk} is the sum of two components, σ_e^2 and σ^2 , these quantities are often called **variance components**.

In addition,

$$\begin{aligned} \text{cov}(y_{ijk}, y_{ijl}) &= \text{cov}(\bar{\mu}_{ik} + e_{j(i)} + \varepsilon_{ijk}, \bar{\mu}_{il} + e_{j(i)} + \varepsilon_{ijl}) = \text{cov}(e_{j(i)} + \varepsilon_{ijk}, e_{j(i)} + \varepsilon_{ijl}) \\ &= \text{cov}(e_{j(i)}, e_{j(i)}) + \underbrace{\text{cov}(e_{j(i)}, \varepsilon_{ijl})}_{=0} + \underbrace{\text{cov}(\varepsilon_{ijk}, e_{j(i)})}_{=0} + \underbrace{\text{cov}(\varepsilon_{ijk}, \varepsilon_{ijl})}_{=0} \\ &= \text{var}(e_{j(i)}) = \sigma_e^2 \end{aligned}$$

for $k \neq l$, and

$$\text{cov}(y_{ijk}, y_{i'j'k'}) = 0,$$

for $i \neq i'$ or $j \neq j'$ (i.e., the covariance is 0 between subjects).

From these results we see that the correlation is zero between observations on different subjects, but

$$\text{corr}(y_{ijk}, y_{ijl}) = \frac{\sigma_e^2}{\sigma^2 + \sigma_e^2} \equiv \rho, \quad j \neq k.$$

- The within-subject correlation ρ here is called the **intra-class correlation**.

The within subject covariance structure we have just described can be represented succinctly as

$$\begin{aligned} \Sigma_{ij} \equiv \text{var}(\mathbf{y}_{ij}) &= \begin{pmatrix} \sigma^2 + \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \cdots & \sigma_e^2 \\ \sigma_e^2 & \sigma^2 + \sigma_e^2 & \sigma_e^2 & \cdots & \sigma_e^2 \\ \sigma_e^2 & \sigma_e^2 & \sigma^2 + \sigma_e^2 & \cdots & \sigma_e^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_e^2 & \sigma_e^2 & \sigma_e^2 & \cdots & \sigma^2 + \sigma_e^2 \end{pmatrix} \\ &= \tilde{\sigma}^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & \vdots & \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} \end{aligned}$$

where $\tilde{\sigma}^2 = \sigma_e^2 + \sigma^2$.

- This variance covariance structure is called **compound symmetry**.
- To summarize, the split-plot model in which we assume a random whole plot effect, which becomes a random subject effect in the repeated measures context, implies that Σ_{ij} , the variance-covariance matrix for the data from the i, j whole-plot/subject has a compound symmetric structure.
- Alternatively, instead of including a random whole-plot/subject effect, a compound symmetric structure for Σ_{ij} can be assumed directly as part of the model. That is, we can omit the whole-plot/subject effect and instead specify our model as

$$y_{ijk} = \mu + \alpha_i + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}, \quad \text{and} \quad \Sigma_{ij} = \tilde{\sigma}^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & \vdots & \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}$$

- This can be done in PROC MIXED on the REPEATED statement, and it results in the same split plot model as we fit previously using a random whole-plot/subject effect.
- More importantly, the variance-covariance structure specified on the repeated statement does not have to be compound symmetry, there are a wide variety of other choices.

- Typically, several different choices for this variance-covariance specification are considered, and the one with that produces the smallest value of the model selection criterion AIC is chosen for use in the final model. Then, results from the final model are used directly. I.e., unlike in the RM-ANOVA, no modifications or “fixes” to the analysis are necessary because we’ve fit the right variance-covariance model in the first place.

Sheep Example Revisited

- See `sheep2.sas` and its output `sheep2.pdf`.