8.1 A Tale of Two Thieves

Methods:	Analysis of Variance Random Effects Mixed Models
Data:	Two Experiments Response: Assay of Active Ingredient 2 Sampling Instruments Postmixing Quality Analysis

8.1.1 Analysis of Variance with Mixed Effects

Prescription and over-the-counter drugs contain a mixture of both active and inactive ingredients, with the dosage determined by the amount of active ingredient in each tablet. Making sure the tablets contain the correct dosage is an important problem in the drug manufacturing industry and in this case study, we consider an experiment conducted by a pharmaceutical company to investigate sampling variability and bias associated with the manufacture of a certain type of tablet.

Outline of the Problem

Tablet Manufacture The tablets were manufactured by mixing the active and inactive ingredients in a "V-blender," so-named because it looks like a large **V**. (See Figure 8.1.) Mixing was achieved by rotating the V-blender in the vertical direction.

After the mixture was thoroughly blended, the powder was discharged from the bottom of the V-blender and compressed into tablet form.

Uniform Content The most important requirement of this manufacturing process was that the tablets have uniform content. That is, the correct amount of active ingredient must be present in each tablet.

The content uniformity of the mixture within the V-blender will need to be assessed.

Thief Sampling A "thief" instrument was used to obtain samples from different locations within the V-blender. This was essentially a long pole with a closed scoop at one end, which was plunged into the powder mixture by a mechanical device. At the appropriate depth for a given location, the scoop was opened and a sample collected.

Considerable force was needed to insert a thief into the powder mixture and it was of interest to compare two types of thieves.

- The **Unit Dose** thief collects three individual unit dose samples at each location.

- The **Intermediate Dose** collects one large sample which is itself sampled to give three unit dose samples.

Experiment Procedure

The objective of this experiment was to study bias and variability differences between the two thieves and to compare the thief-sampled results with those of the tablets. The experiment was implemented as follows.

1. Blend the mixture in the V-blender for 20 minutes.
2. Tie the thieves together and use them to obtain samples from six locations within the V-blender. A schematic of the V-blender and sampling locations is shown in Figure 8.1.
3. Discharge the powder from the V-blender and compress it to form tablets. Load tablets into 30 drums.
4. Select 10 drums and sample three tablets from each of these drums.
5. Assay all samples to determine the amount of active ingredient in each sample. The specified assay value is: 35 mg/100 mg.

In this example, applying the analysis of variance technique to the two sets of assay results (Thief and Tablet) from this experiment is more complicated since *random* and *fixed effects* are involved. When both types of factor are present, this is called a *mixed model* analysis of variance.

The Data

The datasets `c81.thief.dat` and `c81.tablet.dat` contain the results from this experiment which we have presented in a slightly more compact format in Tables 8.1 and 8.2, respectively. The variables and factors of interest for analysis purposes are summarized in the following table.

Variable	Levels	Definition
Y	(mg)	Assay Value per 100 mg
METHOD	INTM	Intermediate Dose Thief
	UNIT	Unit Dose Thief
LOC	1,2,...,6	Sampling Location
REP	1,2,3	Replicate per Location
DRUM	mod(n, 3)	Randomly Selected Drum
TABLET	1,2,3	Tablet Sample (per Drum)

The locations shown in Figure 8.1 represented the "desired" sampling positions for the thieves. In the actual experiment, these "fixed" positions were subject to a certain amount of variability. The samples collected by the thieves can be regarded as random within each location.

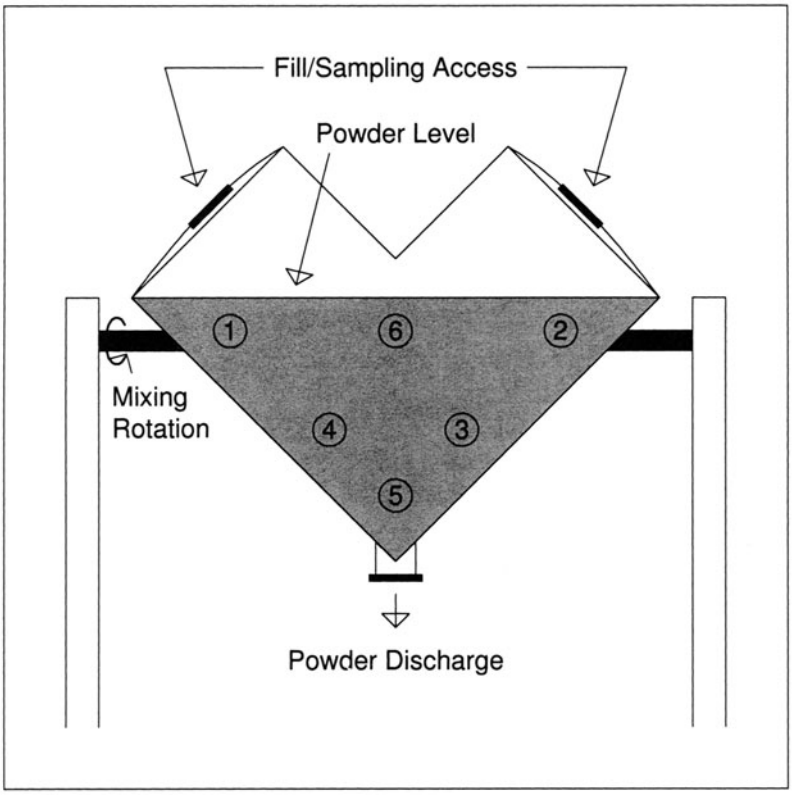

FIGURE 8.1. V-Blender Schematic and Sampling Locations

In the Tablet experiment, the order in which the drums were filled was recorded and this information was incorporated into the random selection procedure. Specifically, one drum was randomly selected from each triple sequence: $\{1,2,3\}$ $\{4,5,6\}$ ... $\{28,29,30\}$. The factor DRUM could therefore be used to test for a "time" effect in the Tablet data.

Methodology

To help motivate the discussion, recall the balanced one-way ANOVA model introduced in Chapter 3 (*Standard Methods*):

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}, \tag{8.1}$$

where the τ_i represent the levels of a factor, TRT say, and the ϵ_{ij} are assumed to be independent $N(0,\sigma^2)$ innovations. The key difference between a fixed effects model and a *random* effects model lies in the interpretation of the inference associated with these models.

TABLE 8.1. Thief Data

Location	Replicate	Assay Method	
		Intermediate	Unit Thief
1	1	34.38	33.94
	2	34.87	34.72
	3	35.71	34.10
2	1	35.31	39.11
	2	37.59	37.51
	3	38.02	37.79
3	1	36.71	37.46
	2	36.56	34.12
	3	35.92	35.94
4	1	37.80	38.05
	2	37.41	34.82
	3	38.00	35.42
5	1	36.28	36.52
	2	36.63	38.60
	3	36.62	38.16
6	1	38.89	39.16
	2	39.80	32.77
	3	37.84	36.95

Fixed Effect Inference is conditional on the "fixed" levels of the TRT factor. In this example, the factor METHOD is clearly fixed since any inference associated with the effect of METHOD applies specifically to the two types of thief involved.

Random Effect Inference is made about a "population" of TRT levels. The factor TRT is considered *random* if the τ_i were randomly selected from the TRT levels. Hence, τ_i is assumed to be a $N(0, \sigma_\tau^2)$ random variable (independent of ϵ_{ij}) for testing purposes. The model (8.1) is also called the *components of variance* model.

Since the τ_i are randomly selected, testing individual effects is meaningless and the hypotheses of interest are: $H_0 : \sigma_\tau^2 = 0$ versus $H_1 : \sigma_\tau^2 > 0$. Under H_0, the test statistic

$$F_o = \frac{\text{SS(TRT)}/(a-1)}{\text{SSE}/(N-a)} = \frac{\text{MS(TRT)}}{\text{MSE}} \sim \mathcal{F}_{a-1,N-a}$$

is constructed in exactly the same manner as in the fixed effects case. However, the expected mean squares associated with the random effects model are different and are needed to construct estimators of the variance components. For the model (8.1), it can be shown that:

$$E[\text{MS(TRT)}] = \sigma^2 + n\sigma_\tau^2 \qquad E[\text{MSE}] = \sigma^2 \ .$$

TABLE 8.2. Tablet Data

Drum	Tablet	Assay	Drum	Tablet	Assay
	1	35.77		1	35.43
1	2	39.44	17	2	33.80
	3	36.43		3	35.15
	1	35.71		1	34.56
5	2	37.08	19	2	35.33
	3	36.54		3	37.69
	1	35.08		1	35.82
7	2	34.25	22	2	35.67
	3	33.09		3	35.06
	1	35.21		1	35.75
11	2	34.36	25	2	37.32
	3	35.94		3	35.06
	1	35.17		1	38.58
14	2	36.54	28	2	36.63
	3	36.45		3	35.60

Equating the observed and expected mean squares provides the following estimates of the variance components

$$\hat{\sigma}_\tau^2 = \big(\text{MS}(\texttt{TRT}) - \text{MSE}\big)/n \qquad \hat{\sigma}^2 = \text{MSE}\ .$$

8.1.2 Mixed Model Analysis

The two-way ANOVA model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases} \tag{8.2}$$

is said to be a *mixed model* analysis of variance when one of the factors is fixed and the other is random. Here, we consider the situation where factor A is fixed, factor B is random, and $n > 1$ so that the interaction term can be estimated. The case where **both** A and B are fixed, or both are random, was discussed in Chapter 3.

Statistical Analysis

The interaction term AB is always assumed to be a random effect, but different assumptions can be imposed on the random components. This leads to different versions of the mixed model:

A Fixed effect: $\sum_{i=1}^{a} \alpha_i = 0$.

B Random effect: $\beta_j \sim N(0, \sigma_\beta^2)$.

AB Random effect:

standard model: $\gamma_{ij} \sim N\big(0, (1-(1/a))\sigma^2_\gamma\big)$ and $\sum_{i=1}^{a} \gamma_{ij} = 0$.

alternative model: γ_{ij} are uncorrelated with common variance σ^2_γ .

The main difference is that in the standard model (which is more commonly used), interaction effects are **not** uncorrelated at different levels of the fixed factor: $Cov(\gamma_{ij}, \gamma_{i'j}) = -(1/a)\sigma^2_\gamma$ for $i \neq i'$. Under the assumptions above, the expected mean squares provide the appropriate test statistics and variance component estimates as follows.

$H_0 : \alpha_i = 0$ versus $H_1 :$ *some* $\alpha_i \neq 0$

$F_o = \text{MS(A)}/\text{MS(AB)} \sim \mathcal{F}_{(a-1),(a-1)(b-1)}$

$E[\text{MS(A)}] = \sigma^2 + n\sigma^2_\gamma + \frac{bn}{n-1}\sum \alpha_i^2$

$\hat{\sigma}^2 = \text{MSE}$

$H_0 : \sigma^2_\beta = 0$ versus $H_1 : \sigma^2_\beta > 0$

standard model:

$F_o = \text{MS(B)}/\text{MSE} \sim \mathcal{F}_{(b-1),ab(n-1)}$

$E[\text{MS(B)}] = \sigma^2 + an\sigma^2_\beta$

$\hat{\sigma}^2_\beta = \big(\text{MS(B)} - \text{MSE}\big)/an$

alternative model:

$F_o = \text{MS(B)}/\text{MS(AB)} \sim \mathcal{F}_{(b-1),(a-1)(b-1)}$

$E[\text{MS(B)}] = \sigma^2 + n\sigma^2_\gamma + an\sigma^2_\beta$

$\hat{\sigma}^2_\beta = \big(\text{MS(B)} - \text{MS(AB)}\big)/an$

$H_0 : \sigma^2_\gamma = 0$ versus $H_1 : \sigma^2_\gamma > 0$

$F_o = \text{MS(AB)}/\text{MSE} \sim \mathcal{F}_{(b-1),ab(n-1)}$

$E[\text{MS(AB)}] = \sigma^2 + n\sigma^2_\gamma$

$\hat{\sigma}^2_\gamma = \big(\text{MS(AB)} - \text{MSE}\big)/n$.

8.1.3 Preliminary Analysis

The objective of this experiment was to study bias and variability differences between the two thieves and to compare the thief-sampled results with those of the tablets. The three main components of our preliminary analysis are as follows.

- Preliminary data screening
- Comparing the two thieves
- Thief versus Tablet samples.

Data Screening

The preliminary data screening involves checking for outliers and assessing the distributional properties of the assay values. This is important since our intention is to apply the analysis of variance technique which is sensitive to the presence of outliers and departures from normality. Of course, with a sample size of only 18 per thief, we will need to exercise a certain amount of caution in interpreting distributional features.

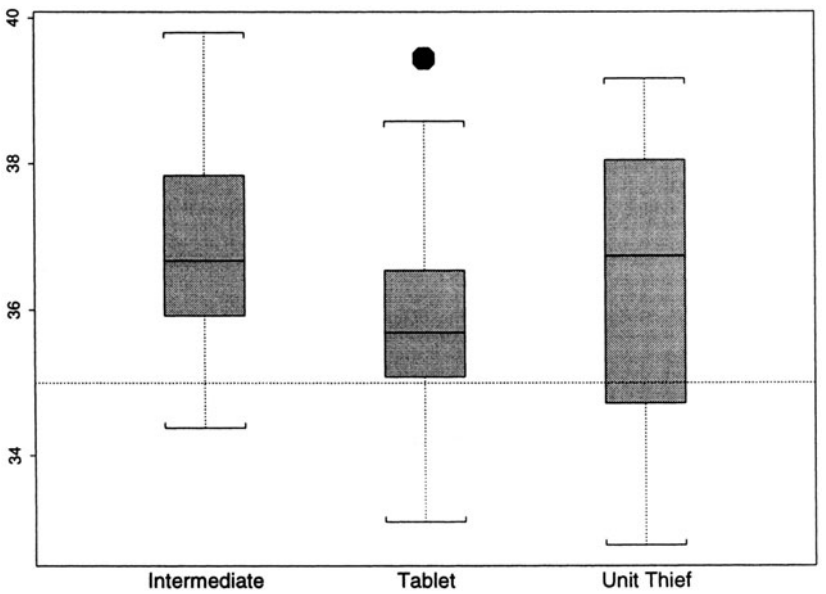

FIGURE 8.2. Parallel Boxplots of the Assay Values

From the parallel boxplot display in Figure 8.2, it can be readily seen that the methods were all positively biased relative to the specified assay value of 35 mg. However, there are some differences in the variability and distribution of the assay values among these methods. In particular, the "average" assay value of the tablet sample is noticeably lower. Where did the active ingredient go? Summary statistics for the three methods are presented in the table below.

	Thief		Tablet
Method	UNIT	INTM	DRUM
N	18	18	30
Mean	36.40	36.91	35.79
Std Dev	1.98	1.40	1.36
Max	39.16	39.80	39.44
Q_3	38.08	37.88	36.54
Median	36.74	36.67	35.69
Q_1	34.57	35.87	35.08
Min	32.77	34.38	33.09

Although the Shapiro–Wilk test and Q–Q plots with Lilliefor's confidence bounds (Conover 1980) did not indicate any significant departure from normality for any of the methods, we should regard these results as inconclusive due to the small samples involved. Similarly, a histogram of the Tablet data shows that the "outlier" indicated in Figure 8.2 is consistent with a heavy right-tail trend. This observation is also less than $Q_3 + 2 * \text{IQR}$ which is commonly used as a more conservative (upper) criterion for determining potential outliers.

Comparing Thieves

Since the preliminary analysis results did not indicate any obvious violation of the assumptions underlying the analysis of variance procedure, we proceed with using the mixed model analysis of variance model to compare the two thieves. Several SAS procedures can be used to perform a mixed analysis which we have presented in the program code below.

SAS Program Code

```
*-------------------------------------------------------------;
  filename in1 'c81.thief.dat' ;  * input the two datasets  ;
  filename in2 'c81.tablet.dat' ;
  data a ;
    infile in1 ;
    input method $ loc rep ya ;

  data b ;
    infile in2 ;
    input method $ drum tablet yb ;
*-------------------------------------------------------------;
                                  * repeated measures ANOVA  ;
proc glm data=a ;                 * with "loc" as random     ;
  class method loc rep ;
  model ya = method loc(method) rep rep*method ;
  random loc(method) / test  ;
```

```
  test h=methda e=loc(method) ;
*-------------------------------------------------------------;
                                  * get variance components  ;
proc varcomp data=a method=ml ;   * (subsumed by proc mixed) ;
  class method loc rep ;
  model ya = method rep rep*method loc(method) / fixed=3 ;
*-------------------------------------------------------------;
                                  * set up the data for the  ;
proc sort data=a ;                * multivariate version of  ;
     by method loc ;              * repeated measures ANOVA  ;

proc transpose out=a2(rename=(_1=y1 _2=y2 _3=y3)) ;
     by method loc ;
     id rep ;

proc glm data=a2 ;
  class methda loc ;
  model y1 y2 y3 = method ;
  repeated rep / summary ;
*-------------------------------------------------------------;
* Output from the "mixed" procedure is presented below.      ;
* ...........................                                ;
                                 * use the "mixed" procedure ;
proc mixed method=ml data=a ;    * which allows more general ;
  class method loc ;             * covariance structures to  ;
  model ya = method ;            * be specified. (SAS 6.08+) ;
  random loc ;
  repeated / subject=loc group=method ;
*-------------------------------------------------------------;
                                 * use the "mixed" procedure ;
proc mixed method=ml data=b ;    * to test for a time effect ;
  class drum ;                   * with respect to DRUM.     ;
  model yb = ;                   * ar(1) = autoregressive    ;
  random drum / s type=ar(1) ;   * covariance structure      ;
```

The `proc mixed` procedure is available in the SAS release version 6.08 and subsumes the `proc varcomp` procedure. It generalizes the `proc glm` procedure and allows a wider class of mixed models to be fitted. It also provides a more flexible "repeated" measures statement than the univariate or multivariate versions employed in `proc glm`.

In this example, we have declared `LOC` to be a random effect and treated the three `REP` levels as repeated measures. The output from `proc mixed` is presented below and shows no significant `METHOD` effect. This implies that either thief could be employed for sampling during the mixing process and for practical purposes, the Intermediate Dose Thief was found to be easier

to handle. However, there were significant differences among sampling locations which suggests the powder mixture may not have uniform content.

SAS Output from Proc Mixed

The Mixed Procedure
Covariance Parameter Estimates

Cov Parm	Subject	Group	Estimate
loc			0.9977
Residual	loc	method intm	0.7068
Residual	loc	method unit	3.4001

Fit Statistics

Log Likelihood	-64.4
Akaike's Information Criterion	-67.4
Schwarz's Bayesian Criterion	-67.1
-2 Log Likelihood	128.8

Solution for Random Effects

Effect	loc	Estimate	Std Err Pred	DF	t Value	Pr > |t|
loc	1	-1.6389	0.5498	29	-2.98	0.0058
loc	2	0.2959	0.5498	29	0.54	0.5946
loc	3	-0.4341	0.5498	29	-0.79	0.4362
loc	4	0.5308	0.5498	29	0.97	0.3422
loc	5	-0.0792	0.5498	29	-0.14	0.8864
loc	6	1.3255	0.5498	29	2.41	0.0225

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
method	1	29	1.14	0.2934

To examine the issue of bias, we can use the *bootstrap* procedure (Efron and Tibshirani 1993). The idea of the bootstrap is to resample the observed data *with replacement* and use the resampled estimates, $\hat{\theta}_i$ say, to assess the variability of $\hat{\theta}$ about some unknown true θ. Hence the *bias* of $\hat{\theta}$ may be estimated by the taking the mean of the differences: $\hat{\theta}_i - \hat{\theta}$.

Bootstrap Estimates of P-Values

	UNIT vs. INTM	UNIT vs. DRUM	INTM vs. DRUM
Observed P-Value	0.3770	0.2135	0.0090
Bootstrap Mean	0.3932	0.3204	0.0587
Bias	0.0162	0.1069	0.0498
Standard Error	0.2958	0.3092	0.1244

In this example, we have employed the bootstrap to assess the reliability of the t-test for comparing the individual differences among the three methods. Note that for simultaneous comparisons, the Bonferroni inequality provides a conservative estimate of the critical level to use for each test: $0.0167 = 0.05/3$. The results presented in the table above are based on 1000 resamples and suggest that the significant difference originally observed between the Intermediate Dose Thief and Tablet assay values may not necessarily be valid.

8.1.4 Summary

Practical considerations and time constraints often limit the size and extent to which diagnostic experiments can be performed in the manufacturing sector. Careful planning and efficient sampling procedures are needed to ensure the experiment will satisfy the objectives of the study. We should point out that having *more* than the specified level of active ingredient in the tablet would not present a problem for the consumer. From the drug manufacturer's perspective, however, achieving the specified value is clearly desirable. The results from this experiment can be analyzed further and in the questions below, we have indicated some of the areas and approaches that the interested reader may wish to pursue.

Questions

1. Are the assay values generally well behaved? Note that when we treat the Thief samples as repeated measures, the issue of correlation needs to be incorporated in the criteria for determining outliers. Find out what procedures or tests are available and apply them to these data.

2. Is there any evidence of a location effect? Our preliminary analysis suggests there is, which would be of concern to the production management. What recommendations should we make in our report to management regarding this issue?

3. Do the tablet data show any drum or time effect? We employed an autoregressive AR(1) covariance structure in the last `proc mixed` procedure for this purpose. Is this the same as using the Durbin–Watson test?

4. Are the thief-sampled values comparable to the tablet values? One approach would be to consider the concordance correlation coefficient proposed by Lin (1989): quantify the agreement between two readings from the same sample by measuring the variation from the 45° line through the origin.

8.2 Plastic Explosives Detection

Source:	Authors
Methods:	Discriminant Analysis
Data:	21 Variables Last Variable is 0/1 Response

8.2.1 Pattern Recognition

The importance of plastic explosives detection was made clear by the tragedy of Pan Am Flight 103 which exploded over Lockerbie, Scotland on December 21, 1988. One of the most effective devices for detecting plastic explosives is a type of X-ray scanner that produces a profile of the chemical composition of a small area inside the suitcase. If the profile shows a pattern similar to one of the known plastic explosives then the suitcase is classified as a bomb.

The emphasis in this case study is on the reliability of plastic explosive detection based on an early X-ray machine prototype. That is, how well can plastic explosive be detected? What type of classification method should be used? Before continuing, we must first point out that:

> The analysis presented here does NOT in any way have implications concerning the reliability of the actual inspection machines used at security points such as in airports.

The Data

In order to obtain a classification rule we performed an experiment where 2500 profiles were obtained, of which 1250 corresponded to explosive substances and the remaining 1250 were from typical substances found in suitcases. Each profile is a vector of 20 numbers $x_1, \ldots, x_{20}$ which are a summary of the signal absorbed by the material. The response y is 1 if the suitcase has a bomb and 0 otherwise. The profiles and response variables are provided in the dataset `c82.dat`.