

Analysis of variance

Analysis of variance (ANOVA) refers to a specific set of methods for data analysis and to a way of summarizing multilevel models:

- As a tool for data analysis, ANOVA is typically used to learn the relative importance of different sources of variation in a dataset. For example, Figure 13.8 displays success rates of pilots at a flight simulator under five different treatments at eight different airports. How much of the variation in the data is explained by treatments, how much by airports, and how much remains after these factors have been included in a linear model?
- If a multilevel model has already been fit, it can be summarized by the variation in each of its batches of coefficients. For example, in the radon modeling in Chapter 12, how much variation in radon levels is explained by floor of measurement and how much by geographical variation? Or, in the analysis of public opinion by state in Section 14.1, how much of the variation is explained by demographic factors (sex, age, ethnicity, education), and how much by states and regions?

These “analysis of variance” questions can be of interest even for models that are primarily intended for prediction, or for estimating particular regression coefficients.

The sections of this chapter address the different roles of ANOVA in multilevel data analysis. We begin in Section 22.1 with a brief review of the goals and methods of classical analysis of variance, outlining how they fit into our general multilevel modeling approach. Sections 22.2 and 22.3 explain how ANOVA (or, more precisely, a set of computations that are inspired by classical ANOVA) can be used to summarize inferences from multilevel models.

Having showed how to compute the ANOVA corresponding to a given multilevel model, we discuss the converse in Section 22.4: if an ANOVA decomposition is desired, how to set up and interpret the corresponding model. We start with simple one-way and two-way structures and move to more complicated designs such as latin squares and split plots.

In Section 22.5, we discuss two statistical methods associated with ANOVA—the analysis of covariance and contrast analysis—and interpret them as multilevel models with individual-level and group-level predictors, respectively. We cannot cover all the topics of the analysis of variance in this chapter, but we hope to show the connections with multilevel modeling to make it clear how to construct appropriate models for the highly structured data that arise in many application areas, especially those with designed experiments.

22.1 Classical analysis of variance

In classical statistics, ANOVA refers either to a family of additive data decompositions, or to a method of testing the statistical significance of added predictors in a linear model. We shall discuss each of these interpretations in turn and then explain

```
> summary (aov (y ~ factor (treatment) + factor(airport)))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(treatment)	4	0.0783	0.0196	0.3867	0.8163
factor(airport)	7	3.9437	0.5634	11.1299	1.187e-06 ***
Residuals	28	1.4173	0.0506		

Figure 22.1 *Classical analysis of variance (as computed in R) for the flight simulator data. The usual focus of this sort of analysis is on the p-values, which indicate that the variation among treatments is not statistically significant (that is, it could be explained by chance alone), but the airport variation cannot be plausibly attributed to chance. Compare to the multilevel ANOVA display in Figure 22.5 on page 495.*

what parts of classical ANOVA we will keep and what parts we will discard when moving to multilevel analysis.

Classical ANOVA as additive data decomposition

In many examples with multilevel structure, it is helpful to perform a simple “data decomposition.” For the flight simulator data indexed by treatments i and airports j , we can write

$$y_i = \mu + \gamma_{j[i]} + \delta_{k[i]} + \epsilon_i,$$

or, equivalently,

$$y_{jk} = \mu + \gamma_j + \delta_k + \epsilon_{jk},$$

in either case decomposing the data y into treatment effects, airport effects, and residuals. In this case, with one observation i per cell (j, k) , the residuals are equivalent to treatment \times airport interactions.

In general, additive decompositions are equivalent to regressions on index variables and their interactions, and the classical analysis of variance can be viewed as a summary of an additive decomposition. In classical ANOVA, the model is estimated using least squares, with the estimate of each batch of coefficients (except for the mean level μ) constrained to sum to 0. However, the focus of interest is typically not the coefficient estimates but rather their variances, as we discuss next, in the context of the ANOVA table.

Sources of variation and degrees of freedom. Figure 22.1 illustrates for the flight simulator data. Each row of the table represents a set of index variables: the treatments j , airports k , and residuals i , with *degrees of freedom* (Df) defined as the number of coefficients in that group, minus the number of constraints required for the coefficients to be identifiable in a classical regression. Thus,

- 5 treatment effects minus 1 constraint = 4 degrees of freedom
- 8 airport effects minus 1 constraint = 7 degrees of freedom
- 40 residuals minus 12 constraints (1 mean, 4 treatment effects, and 7 airport effects) = 28 degrees of freedom.

The degrees of freedom can be more formally defined in the language of matrix algebra, but we shall not go into such details here.

Sums of squares. To continue with the description of Figure 22.1, the *sum of squares* for each row of the table is derived from the classical coefficient estimates (recall that these are constrained to sum to 0 in the least squares estimate). Thus, the sums of squares for treatments, airports, and residuals are $\sum_{i=1}^{40} \hat{\gamma}_{j[i]}^2$, $\sum_{i=1}^{40} \hat{\delta}_{k[i]}^2$, and $\sum_{i=1}^{40} \hat{\epsilon}_i^2$, respectively.

Balance. If the data are *balanced*, then the sums of squares in the table add up to the “total sum of squares” of the data, $\sum_{i=1}^{40} (y_i - \hat{\mu})^2$. Roughly speaking, in a balanced design there are the same number of observations in each row and each column of the data. The flight simulator example is trivially balanced because the data are complete, with 8 observations for each treatment, 5 for each airport, and 1 for each treatment \times airport interaction. The data would be unbalanced if, for example, we were missing the data from the YN condition for Nagoya (recall Figure 13.8 on page 290), or if we had replications for some cells and not for others. Balance is much less important in multilevel modeling than in classical ANOVA, and we do not discuss it further here.

Mean squares, F ratios, and p-values. For each source of variation in the ANOVA table, the *mean square* is defined as the sum of squares divided by the degrees of freedom. The mean square is important in classical ANOVA because it can be used for hypothesis testing. If a given row in the table actually has zero variation, then its mean square will, on average, equal the mean square of the residuals in the model. The ratios of mean squares are called the *F statistics*, and the usual goal of classical ANOVA is to find F-ratios that are significantly greater than 1.

In the flight simulator example, the treatment mean square is less than the residual mean square, indicating no evidence for variation between treatments. Airport effects, however, seem to be present—their mean square is 11.1 times greater than the residual mean square. The discrepancy of a mean square from the “null hypothesis” value of 1 is tested using the F_{ν_1, ν_2} , where ν_1 and ν_2 are the degrees of freedom of the numerator and denominator, respectively. The *p*-values in the table indicate the statistical significance of the F-tests. As usual, the *p*-values are only considered significant if they are close to 0 or 1 (typically, if less than 0.05 or greater than 0.95). Thus, that the treatment mean square is *not* statistically significantly less than what would be expected under the null hypothesis, but the airport mean square is statistically significant, indicating that we can reject the hypothesis that the “airport” factor has no effect on the outcome. Equivalently, we can say that the between-airport variation is greater than what could be expected by chance, given the variation in the data.

Classical ANOVA for model comparison

We have just illustrated how ANOVA is used to summarize data, and how this ANOVA corresponds to a “default” linear model with indicator variables. We now discuss the other classical role of the analysis of variance, which is to summarize hypothesis tests within an existing family of models. The basic idea is that an analyst has fit a linear regression model and is considering fitting a larger model formed by adding predictors to the first model. (The two models are said to be *nested*, with this term having a different meaning from its use in multilevel models as in Chapters 12 and 13.)

When comparing nested models, ANOVA is related to the classical test of the hypothesis that the smaller model is true, which is equivalent to the hypothesis that the additional predictors all have coefficients of zero when included in the larger model. Suppose the smaller model has k_1 predictors, the larger model has $k_1 + k_2$ predictors, and each model is fit to n data points. If the predictions from the smaller and larger regressions are $X_1\hat{\beta}_1$ and $X_2\hat{\beta}_2$, respectively, then the classical ANOVA for testing the model expansion can be written as

	Df	SS	MS
Model expansion	k_2	$\sum_i (y_i - X_1 \hat{\beta}_1)^2 - \sum_i (y_i - X_2 \hat{\beta}_2)^2$	SS/k_2
Residuals	$n - k_1 - k_2$	$\sum_i (y_i - X_2 \hat{\beta}_2)^2$	$SS/(n - k_1 - k_2)$

If the ratio of these sums of squares is statistically significantly greater than 1 (as compared to the $F_{k_2, n-k_1-k_2}$ distribution), then the improvement in fit from the model expansion cannot be reasonably explained by chance.

22.2 ANOVA and multilevel linear and generalized linear models

When moving to multilevel modeling, the key idea we want to take from the analysis of variance is the estimation of the importance of different batches of predictors (“components of variation” in ANOVA terminology). As usual, we focus on estimation rather than testing: instead of testing the null hypothesis that a variance component is zero, we estimate the standard deviation of the corresponding batch of coefficients. If this standard deviation is estimated to be small, then the source of variation is minor—we do not worry about whether it is exactly zero. In the social science and public health examples that we focus on, it can be a useful research goal to identify important sources of variation, but it is rare that anything is truly zero.

Notation

As always, any multilevel model can be expressed in several ways. For ANOVA, we write the models to emphasize the grouping of regression coefficients into “sources of variation,” with each batch corresponding to one row of the ANOVA table. We use the notation $m = 1, \dots, M$ for the rows of the table. Each row m represents a batch of J_m regression coefficients $\beta_j^{(m)}$, $j = 1, \dots, J_m$. We denote the m^{th} subvector of coefficients as $\beta^{(m)} = (\beta_1^{(m)}, \dots, \beta_{J_m}^{(m)})$ and the corresponding classical least squares estimate as $\hat{\beta}^{(m)}$. These estimates are subject to c_m linear constraints, yielding $(df)_m = J_m - c_m$ degrees of freedom. We label the constraint matrix as $C^{(m)}$, so that $C^{(m)}\hat{\beta}^{(m)} = 0$ for all m . For notational convenience, we label the grand mean as $\beta_1^{(0)}$, corresponding to the (invisible) zeroth row of the ANOVA table and estimated with no linear constraints.

In classical ANOVA, a linear model is fit to the data points y_i , $i = 1, \dots, n$, and can be written as

$$y_i = \sum_{m=0}^M \beta_{j_i^m}^{(m)} \tag{22.1}$$

where j_i^m indexes the appropriate coefficient j in batch m corresponding to data point i . Thus, each data point pulls one coefficient from each row in the ANOVA table. Equation (22.1) could also be expressed as a linear regression model with a design matrix composed entirely of 0’s and 1’s. The coefficients β_j^M of the last row of the table correspond to the residuals or error term of the model.

ANOVA can also be applied more generally to regression models, in which case we can have any design matrix X , and (22.1) would be generalized to

$$y_i = \sum_{m=0}^M \sum_{j=1}^{J_m} X_{ij}^{(m)} \beta_j^{(m)}. \tag{22.2}$$

The essence of analysis of variance is in the structuring of the coefficients into

batches—hence the notation $\beta_j^{(m)}$ —going beyond the usual linear model formulation that has a single indexing of coefficients β_j . We assume that the structure (22.1), or the more general regression parameterization (22.2), has already been constructed using knowledge of the data structure. To use ANOVA terminology, we assume the sources of variation have already been set, and our goal is to perform inference for each variance component.

We shall follow our usual practice and model each batch of regression coefficients as a sample from a normal distribution with mean 0 and superpopulation standard deviation σ_m :

$$\beta_j^{(m)} \sim N(0, \sigma_m^2), \quad \text{for } j = 1, \dots, J_m, \quad \text{for each batch } m = 1, \dots, M. \quad (22.3)$$

With factors that have only a finite number of levels (for example, 50 states), the superpopulation is difficult to interpret in itself except as a tool that allows better inferences for the individual coefficients.

We model the underlying coefficients β as unconstrained (unlike the least squares estimates) but in many cases will summarize them by subtracting off their averages, as in Section 19.4. The mean of 0 for each batch in (22.3) comes naturally from the ANOVA decomposition structure (pulling out the grand mean, main effects, interactions, and so forth), and the standard deviations represent the magnitudes of the variance components corresponding to each row of the table.

The finite-population standard deviation

One measure of the importance of each row or “source” in the ANOVA table is the standard deviation of its constrained regression coefficients, the finite-population standard deviation

$$s_m = \sqrt{\frac{1}{J_m - 1} \sum_{j=1}^{J_m} \left(\beta_j^{(m)} - \bar{\beta}^{(m)} \right)^2}, \quad (22.4)$$

where $\bar{\beta}^{(m)} = \sum_{j=1}^{J_m} \beta_j^{(m)} / J_m$. As discussed in the context of definitions (21.1) on page 460, s_m captures the variation in the existing J_m levels of factor m in the data, in comparison to σ_m , which reflects the potential uncertainty in the superpopulation.

Variance estimation is often presented in terms of the superpopulation standard deviations σ_m , but in our ANOVA summaries, we focus on the finite-population quantities s_m . However, for computational reasons, the parameters σ_m are sometimes useful intermediate quantities to estimate.

Generalized linear models

The multilevel ANOVA framework does not require the normal distribution or, for that matter, even linearity. All that is needed is that the parameters β can be grouped into reasonable batches, with the magnitude of each batch summarized by a standard deviation. We illustrate this process in the next section with a logistic regression. For generalized linear models, coefficients and variance parameters on the logarithmic or logit scales can be interpreted as discussed in Chapters 5–6 and 14–15.

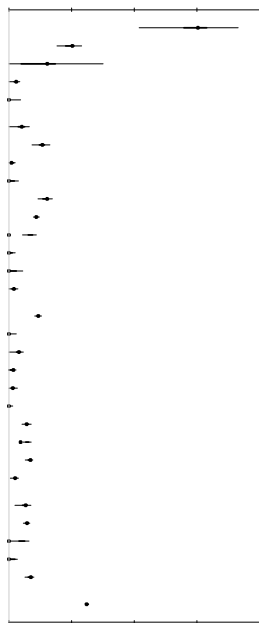


Figure 22.2 ANOVA display for the World Wide Web data. The bars indicate 50% and 95% intervals for the finite-population standard deviations s_m . The display makes apparent the magnitudes and uncertainties of the different components of variation. Since the data are on the logarithmic scale, the standard-deviation parameters can be interpreted directly. For example, $s_m = 0.20$ corresponds to a coefficient of variation of $\exp(0.2) - 1 \approx 0.2$ on the original scale, and so the exponentiated coefficients $\exp(\beta_j^{(m)})$ in this batch correspond to multiplicative increases or decreases in the range of 20%. (The dots on the bars show simple classical estimates of the variance components that were used as starting points in the multilevel analysis.)

22.3 Summarizing multilevel models using ANOVA

It can be helpful to graph the estimates of variance components, especially for complex data structures with many levels of variation. In basic multilevel models (that is, the models covered in this book), each variance parameter corresponds to a set of coefficients—for example, $y \sim N(X\beta, \sigma_y^2)$, or $\alpha \sim N(\mu_\alpha, \sigma_\alpha^2)$. As discussed in Section 21.2, the standard deviation of a set of coefficients gives a sense of their predictive importance in the model. An analysis-of-variance plot, which shows the relative scale of different variance components, can be a useful tool in understanding a model.

A five-way factorial analysis: internet connect times

We illustrate the analysis of variance with an example of a linear model fitted for exploratory purposes to a highly structured dataset.



Figure 22.3 Analysis of variance (ANOVA) display for two logistic regression models of the probability that a survey respondent prefers the Republican candidate for the 1988 U.S. presidential election, based on data from seven CBS News polls. Point estimates and error bars show median estimates, 50% intervals, and 95% intervals of the finite-population standard deviations s_m . The demographic factors are those used by CBS to perform non-response adjustments, and states and regions are included because we were interested in estimating average opinions by state. The large coefficients for ethnicity, region, and state suggest that it might make sense to include interactions, hence the inclusion of the ethnicity \times region and ethnicity \times state coefficients in the second model.

Data were collected by an internet infrastructure provider on connect times for messages processed by two different companies. Messages were sent every hour for 25 consecutive hours, from each of 45 locations to 4 different destinations, and the study was repeated one week later. It was desired to quickly summarize these data to learn about the importance of different sources of variation in connect times.

Figure 22.2 shows the Bayesian ANOVA display for an analysis of logarithms of connect times on the five factors: destination (“to”), source (“from”), service provider (“company”), time of day (“hour”), and week. The data have a full factorial structure with no replication, so the full five-way interaction at the bottom represents the “error” or lowest-level variability.

Each row of the plot shows the estimated finite-population standard deviation of the corresponding group of parameters, along with 50% and 95% uncertainty intervals. We can immediately see that the lowest-level variation is more important in variance than any of the factors except for the main effect of the destination. **Company** has a large effect on its own and, perhaps more interestingly, in interaction with **to**, **from**, and in the three-way interaction. (By comparison, a classical analysis of variance reveals that all the main effects and almost all the interactions are statistically significant, but it does not give a good sense of the relative magnitudes of the different variance components.)

Figure 22.2 would not normally represent the final statistical analysis for this sort of problem. The ANOVA plot represents a default model and is a tool for data exploration—for learning about which factors are important in predicting the variation in the data—and can be used to construct more focused models or design future data collection.

A multilevel logistic regression: vote preference broken down by state and demographics

We illustrate the use of ANOVA for understanding an existing model with the vote preference study described in Section 14.1. There, we focused on hierarchical modeling as a tool for estimating state opinions; here, we examine the fitted models

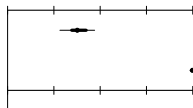


Figure 22.4 *One-way analysis of variance for the radon data. The dots, thick lines, and thin lines represent medians, 50% intervals, and 95% intervals for the finite-population standard deviation s , for each source of variation.*

themselves to see the relative importance of different inputs in predicting vote preferences. The left plot of Figure 22.3 displays the analysis of variance from the basic model, which shows that ethnicity is by far the most important demographic factor, with state also explaining quite a bit of variation.

The natural next step is to consider interactions among the most important factors, as shown in the plot on the right side of Figure 22.3. The ethnicity \times state \times region interactions are surprisingly large: the differences between African Americans and others vary dramatically by state.

22.4 Doing ANOVA using multilevel models

Our general solution to the ANOVA problem is simple: we treat *every* row in the table as a batch of “varying coefficients”—that is, a set of regression coefficients drawn from a distribution with mean 0 and some standard deviation to be estimated from the data. We illustrate in the rest of this chapter with several examples of basic data structures to which the analysis of variance is often applied.

One-way ANOVA: radon measurements within counties

Some of the essential elements of multilevel analysis of variance can be seen in the simplest case of a one-way structure. We illustrate with the radon example from Chapter 12, simplified to measurements within counties, ignoring all individual-level and group-level predictors. The model is then

$$\begin{aligned} y_i &\sim N(\alpha_{j[i]}, \sigma_y^2), \text{ for } i = 1, \dots, n \\ \alpha_j &\sim N(\mu_\alpha, \sigma_\alpha^2), \text{ for } j = 1, \dots, J. \end{aligned}$$

As discussed in Section 19.4, we use redundant parameters to speed the computation. We also add lines in the Bugs model to define the finite-population standard deviations, s_e and s_α :

```
Bugs code      for (i in 1:n){
                  e.y[i] ~ y[i] - y.hat[i]
                }
                s.y <- sd(e[])
                s.a <- sd(a[])
```

Figure 22.4 shows the result of fitting this model to the Minnesota radon data. The variation among houses within counties is much larger than the variation of county mean radon levels. The between-county variation is estimated to be about 0.3; the analysis is on the log scale, so this corresponds to a multiplicative variation

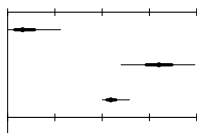


Figure 22.5 *Two-way analysis of variance for the flight simulator data. The dots, thick lines, and thin lines represent medians, 50% intervals, and 95% intervals for the finite-population standard deviation s , for each source of variation.*

of $\exp(0.3) \approx 1.3$. That is, average radon levels vary by about 30% among Minnesota counties. Although this variation is small, it is clearly “statistically significant”: its 95% confidence interval in Figure 22.4 shows that we are fairly certain that the between-county standard deviation falls between 0.2 and 0.4. (The within-county standard deviation is estimated much more accurately, which makes sense given its larger degrees of freedom.)

Two-way ANOVA: flight simulator data

We illustrate two-way analysis of variance with the flight simulator experiment described in Section 13.5. The data, displayed in Figure 13.8, are the success rates of pilots in flight simulators, under five different experimental treatments at eight different airports. Figure 22.1 on page 488 displays the classical two-way ANOVA for these data. The corresponding multilevel model is (13.9) on page 289, a model we can also write as

$$\begin{aligned} y_i &= \mu + \gamma_{j[i]} + \delta_{k[i]} + \epsilon_i, \text{ for } i = 1, \dots, n \\ \gamma_j &\sim N(0, \sigma_\gamma^2), \text{ for } j = 1, \dots, J \\ \delta_k &\sim N(0, \sigma_\delta^2), \text{ for } k = 1, \dots, K \\ \epsilon_i &\sim N(0, \sigma_\epsilon^2), \text{ for } i = 1, \dots, n. \end{aligned} \tag{22.5}$$

In the terminology of the analysis of variance, the data have $n - 1 = 39$ degrees of freedom, which can be decomposed into:

- $J - 1 = 4$ degrees of freedom for γ
- $K - 1 = 7$ degrees of freedom for δ
- The remaining 28 degrees of freedom for ϵ .

The multilevel ANOVA is performed by fitting model (22.5) and summarizing by the estimated variance components. We could use the superpopulation variance parameters $\sigma_y, \sigma_\gamma, \sigma_\delta$, but, for reasons described in Section 21.2, we prefer to work with the finite-population variance parameters s_y, s_γ, s_δ defined in (21.1) on page 460. Each of these standard deviations is defined in terms of the model parameters and thus varies across the simulations produced by the multilevel inference in Bugs.

Figure 22.5 summarizes the inference for the variance components for the flight simulator data. Treatment effects are small (with coefficients estimated to be less than 0.05 in absolute value), airport effects are on the order of 0.3 (which is large, considering that the outcomes y are proportions and thus fall between 0 and 1), and the scale of the errors is moderately large, at about 0.2.

Two-way ANOVA with replication

When replications are present, it is possible to estimate the two-way interactions separately from the measurement error. The additive model can be expanded to

$$y_i = \mu + \gamma_{j[i]} + \delta_{k[i]} + \eta_{j[i],k[i]} + \epsilon_i,$$

with separate variance components for the γ_j 's, δ_k 's, $\eta_{j,k}$'s, and ϵ_i 's. For example, if the flight simulator data had four replications per cell (thus, $5 \times 8 \times 4 = 160$ observations), the ANOVA would look like

Source	df
treatment	4
airport	7
treatment \times airport	28
error	120

The 120 degrees of freedom for error correspond to the 160 data points, minus the 40 cell means. Each row of the table would correspond to a single variance component.

Unbalanced designs

Multilevel analysis of variance works the same for unbalanced as for balanced designs. For example, in the flight simulator analysis, if we had 160 observations in the 40 cells, but not necessarily evenly distributed at 4 per cell, then the ANOVA table would have the same structure.

Nested designs

Consider an experiment on 4 treatments for an industrial process applied to 20 machines (randomly divided into 4 groups of 5), with each treatment applied 6 times independently on each of its 5 machines. For simplicity, we assume no systematic time effects, so that the 6 measurements on each machine are replications. The ANOVA table is then

Source	df
treatment	3
machine	16
error	100

Because the design is nested, it does not make sense to consider treatment \times machine interactions. If expressed in terms of measurements i , machines j , and treatments k , the nested model can be written identically as the two-way non-nested model (22.5). The multilevel analysis automatically accounts for the nesting.

22.5 Adding predictors: analysis of covariance and contrast analysis

Individual-level predictors and analysis of covariance

Analysis of covariance is a decomposition of the sources of variation of a dataset (as in ANOVA), after adjusting for a predictor or set of predictors. In the multilevel modeling context, analysis of covariance corresponds to an ANOVA-like decomposition of a model that includes individual-level predictors.

For example, Figure 22.4 on page 494 displays the one-way analysis of variance for the Minnesota radon data, showing the estimated county-level and house-level



Figure 22.6 ANOVA displays for a 5×5 latin square experiment (an example of a crossed three-way structure) for the data shown on page 292: (a) with no group-level predictors, (b) contrast analysis including linear trends for rows, columns, and treatments. See also the plots of coefficient estimates and trends on page 293.

variation. Suppose we are interested in estimating these sources of variation, after controlling for the floor of measurement (recall from Chapter 12 that radon levels tends to be higher in houses with basements). We then fit the model

$$\begin{aligned} y_i &\sim N(\alpha_{j[i]} + x_i\beta, \sigma_y^2), \text{ for } i = 1, \dots, n \\ \alpha_j &\sim N(\mu_\alpha, \sigma_\alpha^2), \text{ for } j = 1, \dots, J, \end{aligned}$$

and display the estimated variance components σ_β, σ_y (or their finite-sample counterparts, s_β, s_y). In this particular example, the estimated variance parameters are not much changed from the simple ANOVA, and so we do not display the analysis of covariance table here.

Group-level predictors and contrast analysis

Adding predictors at the group level in a multilevel model corresponds to the classical method of *contrasts* in the analysis of variance. We illustrate with the latin square data displayed in Figure 13.11 on page 292. First we perform the analysis of variance for the three-level model with no additional predictors; then we add group-level predictors and show the contrast analysis.

Multilevel ANOVA with no contrasts. In Section 13.5, we fit a model including row effects, column effects, treatment effects, and linear trends for each of these factors. We shall first fit the model without the linear trends and simply estimate the magnitudes of the row, column, and treatment effects. The model is the following stripped-down version of (13.10):

$$\begin{aligned} y_i &\sim N(\mu + \beta_{j[i]}^{\text{row}} + \beta_{k[i]}^{\text{column}} + \beta_{l[i]}^{\text{treat}}, \sigma_y^2), \text{ for } i = 1, \dots, 25 \\ \beta_j^{\text{row}} &\sim N(0, \sigma_{\beta, \text{row}}^2), \text{ for } j = 1, \dots, 5 \\ \beta_k^{\text{column}} &\sim N(0, \sigma_{\beta, \text{column}}^2), \text{ for } k = 1, \dots, 5 \\ \beta_l^{\text{treat}} &\sim N(0, \sigma_{\beta, \text{treat}}^2), \text{ for } l = 1, \dots, 5, \end{aligned} \quad (22.6)$$

and we summarize by the finite-sample standard deviations. Figure 22.6a shows the results: none of the effects are large compared to residual variation, and the sample size is small enough that it is difficult to distinguish column and treatment effects from zero.

Multilevel ANOVA with contrasts. We next add *linear contrasts* for the rows, columns, and treatments, expanding the model by replacing the group-level models

in (22.6) with

$$\begin{aligned}\beta_j^{\text{row}} &\sim N(\gamma^{\text{row}} \cdot (j - 3), \sigma_{\beta^{\text{row}}}^2) \\ \beta_k^{\text{column}} &\sim N(\gamma^{\text{column}} \cdot (k - 3), \sigma_{\beta^{\text{column}}}^2) \\ \beta_l^{\text{treat}} &\sim N(\gamma^{\text{treat}} \cdot (l - 3), \sigma_{\beta^{\text{treat}}}^2).\end{aligned}$$

Figure 22.6b shows the new ANOVA, with each factor decomposed into a linear contrast and residuals from the contrast. The column and treatment effects are mostly captured by the linear contrasts, whereas the variation in the row effects does not follow a linear trend.

These ANOVA displays give a reasonable quick summary, but in this particular application it is probably more useful to simply display each group of parameter estimates, along with the estimated linear contrasts, directly, as in Figure 13.12 on page 293. We have shown the ANOVA here to connect to classical contrast analysis.

22.6 Modeling the variance parameters: a split-plot latin square

Multilevel data structures can be characterized by the number of grouping factors (that is, rows of the ANOVA table) and the number of groups in each. For example, the radon example has one grouping factor—counties—which takes on 85 values. The latin square example of the previous section had three factors—row, column, and treatment—each of which took on five levels. Designed experiments sometimes go even further in this direction: for example, a so-called 2^6 design has six factors, each of which can take on two different values (thus a total of 64 data points in a complete design, or more if there is replication).

Different models are appropriate for differently shaped data structures. With few grouping factors and many levels per factor, the models discussed so far in this book—modeling the coefficients but leaving the variance components unmodeled (in Bayesian terms, to be estimated using noninformative prior distributions)—are appropriate. At the other extreme, when the number of levels per grouping factor is small, it can be helpful to model the variance parameters, and when there are many factors, it is possible to use partial pooling to do this estimation. We illustrate with a hierarchical data structure with two treatment factors and nine grouping factors (including interactions).

Crossed and nested ANOVA: a split-plot design

In a *split-plot* design, units are clustered, and there are two treatment factors, with one factor applied to groups and the other to individual units. For example, in an educational experiment with students within classrooms, several teaching methods might be applied at the classroom level, with individual interventions applied to individual students. In this sort of design, the individual-level treatments are typically estimated with higher precision than the group-level treatments, and part of the goal of ANOVA is to assess the importance of both factors. (The term “split-plot” refers to agricultural experiments, where the groups are large plots that are split into subplots, which play the role of individual units in our analysis.)

For example, here are the variance components for a $5 \times 5 \times 2$ split-plot latin square:

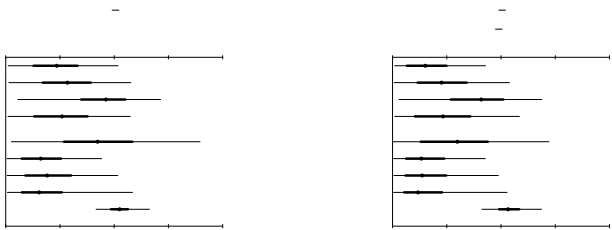


Figure 22.7 ANOVA display for a split-plot latin square experiment: posterior medians, 50%, and 95% intervals for finite-population standard deviations σ_k . (a) The left plot shows inferences given uniform prior distributions on the σ_k 's; (b) the right plot shows inferences given a hierarchical half-Cauchy model with scale fit to the data. The half-Cauchy model gives sharper estimates, indicating the power of hierarchical modeling for these highly uncertain quantities.

Source	df
row	4
column	4
(A,B,C,D,E)	4
plot	12
<hr/>	
(1,2)	1
row \times (1,2)	4
column \times (1,2)	4
(A,B,C,D,E) \times (1,2)	4
plot \times (1,2)	12

In this example, there are 25 plots with five full-plot treatments (labeled A, B, C, D, E), and each plot is divided into two subplots with subplot varieties (labeled 1 and 2). The horizontal line in the table separates the main-plot from the subplot effects. In a classical analysis, it is easy enough to decompose the 49 degrees of freedom to the rows in the ANOVA table; the tricky part of the analysis is to know which residuals are to be used for which comparisons.

The analysis is straightforward using multilevel models. We first present the results (using data from an agricultural experiment) with a simple multilevel model that leaves the variance parameters unmodeled, then with an expanded model that includes a hierarchical model for the variance parameters themselves.

Multilevel model with noninformative prior distributions for the variance parameters

To perform ANOVA using multilevel analysis, we simply set up a linear model with a batch of coefficients corresponding to each source of variation; thus,

$$y_i = \alpha^0 + \alpha_{j[i]}^{\text{row}} + \alpha_{k[i]}^{\text{col}} + \alpha_{l[i]}^{\text{ABCDE}} + \alpha_{m[i]}^{\text{plot}} + \alpha_{n[i]}^{12} + \alpha_{j[i],n[i]}^{\text{col} \times 12} + \alpha_{k[i],n[i]}^{\text{ABCDE} \times 12} + \alpha_{l[i],n[i]}^{\text{plot} \times 12}.$$

Here the data i run from 1 to 50, and we have used the variables j, k, l, m, n to index rows, columns, group-level treatments, plots, and individual-level treatments. There is no replication in the study (that is, there is only one measurement per subplot), and so the last term in the additive model, $\alpha_{l,n}^{\text{plot} \times 12}$, corresponds to data-level errors.

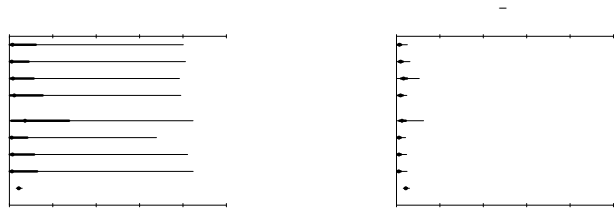


Figure 22.8 Posterior medians, 50%, and 95% intervals for standard-deviation parameters σ_k estimated from a split-plot latin square experiment. (a) The left plot shows inferences given uniform prior distributions on the σ_k 's; and (b) the right plot shows inferences given a hierarchical half-Cauchy model with scale fit to the data. The half-Cauchy model gives much sharper inferences, using the partial pooling that comes with fitting a hierarchical model. Compare to Figure 22.7 (which is on a different scale).

Each batch of coefficients α is then assigned its own normal distribution with mean 0 and standard deviation estimated from the data: $\alpha_j^{\text{row}} \sim N(0, (\sigma^{\text{row}})^2)$, and so forth. We can then fit the model and summarize using finite-population standard deviations. Figure 22.7a illustrates.

Multilevel model with noninformative prior distributions for the variance parameters

Each row of the ANOVA table corresponds to a different variance component, and the split-plot ANOVA can be understood as a linear model with nine variance components, $\sigma_1, \dots, \sigma_9$ —one for each row of the table. The default performed earlier assigns a uniform prior distribution to each of the parameters σ_k .

More generally, we can set up a hierarchical model, where the variance parameters have a common distribution with hyperparameters estimated from the data. We consider a half-Cauchy prior distribution with peak 0 and scale A , and with a uniform prior distribution on A . The hierarchical half-Cauchy model allows most of the variance parameters to be small but with the occasionally large σ_α , which seems reasonable in the typical settings of analysis of variance, in which most sources of variation are small but some are large. See Section 19.6 for further discussion of the half-Cauchy model for multilevel variance parameters.

Figure 22.7b shows inferences for the finite-population standard-deviation parameters s_α for each row of the latin square split-plot ANOVA under this new model. The inferences from the half-Cauchy prior distribution are slightly more precise than with the uniform, with the most pooling occurring for the variance component that has just one degree of freedom. The Cauchy scale parameter A was estimated at 1.8, with a 95% interval of $[0.5, 5.1]$.

Superpopulation and finite-population standard deviations

As discussed in Section 21.2, finite-population inferences can be much more precise than superpopulation inferences when the number of groups is small. We illustrate here by displaying, in Figure 22.8, the inferences for the superpopulation standard deviations in the split-plot latin square example, again separately for the uniform

and hierarchical half-Cauchy prior distributions for the standard-deviation parameters σ_k .

As the left plot shows, the uniform prior distribution does not rule out the potential for some extremely high values of the variance components—the degrees of freedom are low, and the interlocking of the linear parameters in the latin square model results in difficulty in estimating any single variance parameter. In contrast, the hierarchical half-Cauchy model performs a great deal of shrinkage, especially of the high ranges of the intervals. (For most of the variance parameters, the posterior medians are similar under the two models, but the 75th and 97.5th percentiles are shrunk by the hierarchical model.) This is an ideal setting for hierarchical modeling of variance parameters in that it combines separately imprecise estimates of each of the individual σ_k 's.

22.7 Bibliographic note

Searle, Casella, and McCulloch (1992) review classical analysis of variance and variance-component models. Kirk (1995) provides an introductory treatment from the perspective of experimental psychology.

The multilevel ANOVA approach described in this chapter comes from Gelman (2005), and is based on earlier work of Green and Tukey (1960), Nelder (1965a, b), Yates (1967), and Lane and Nelder (1982). The hierarchical model for variance parameters appears in Gelman (2006). See also many of the references in Section 21.9 for related ideas.

McCullagh (2005) points out that ANOVA can be applied more generally to non-exchangeable models such as arise in genetics or, more generally, in models with structured interactions.

22.8 Exercises

1. Take a varying-intercept model from one of the exercises in Part 2 of this book and construct the corresponding ANOVA plot as in Section 22.3.
2. Analysis of variance for meta-analysis: consider the magnetic-fields experiments from Section 21.8 (data in the folder `chicks`) as an analysis-of-variance problem. Identify the sources of variation and construct the ANOVA plot, making whatever assumptions are necessary to do this using the available data.
3. Three-way designs: perform the analysis of variance as described in Section 22.4 for the data in the folder `olympics`, which are figure-skating ratings classified by judges, skaters, and measurement criterion (see Exercise 11.3). You will need to identify the sources of variation, then fit the model and display the estimated standard-deviation parameters and their uncertainties.
4. Hierarchical modeling of variance parameters: consider the model for height and earnings shown in Figure 13.10 on page 291.
 - (a) Start with a varying-intercept model (with intercepts varying by ethnicity, age, and ethnicity \times age), with a hierarchical model for the variance parameters as in Section 22.6. Make plots similar to Figures 22.7 and 22.8 to compare the inferences under flat and hierarchical prior distributions.
 - (b) Repeat (a) for a varying-intercept, varying slope model (with both intercepts and slopes varying by ethnicity, age, and their interaction).
5. Multivariate analysis of variance for multilevel models: consider how the models

and displays of this chapter could be generalized to varying-intercept, varying-slope models.