Lecture Notes

6. Analysis of Variance

1. Introduction

- *Analysis of variance (ANOVA)* describes the partition of the response-variable sum of squares in a linear model into 'explained' and 'unexplained' components.
- The term also refers to procedures for fitting and testing linear models in which the explanatory variables are categorical.
  - A single categorical explanatory variable (*factor* or *classification*) corresponds to *one-way* analysis of variance;
  - two factors to *two-way* analysis of variance;
  - three factors to *three-way* analysis of variance;
  - and so on.

2. Goals:

- To introduce statistical models for one- and two-way analysis of variance.
- To show how the models can be fit to data by placing restrictions on their parameters and appropriately coding regressors.
- To explain how interaction is reflected in two-way analysis of variance.
- To show how the incremental-sum-of-squares approach can be adapted to testing main and interaction effects in two-way analysis of variance.

3. One-Way ANOVA

- Dummy regressors can be employed to code a one-way ANOVA model.
- For example, for a three-category classification:

$$Y_i = \alpha + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \varepsilon_i$$

with

Group	D_1	D_2
1	1	0
2	0	1
3	0	0

- The response variable expectation (population mean) in group j is μ_j.

- Because the error ε has a mean of 0 under the usual linear-model assumptions, taking the expectation of both sides of the model produces the following relationships between group means and model parameters:

$$\begin{array}{ll} \text{Group 1:} & \mu_1 = \alpha + \gamma_1 \times 1 + \gamma_2 \times 0 = \alpha + \gamma_1 \\ \text{Group 2:} & \mu_2 = \alpha + \gamma_1 \times 0 + \gamma_2 \times 1 = \alpha + \gamma_2 \\ \text{Group 3:} & \mu_3 = \alpha + \gamma_1 \times 0 + \gamma_2 \times 0 = \alpha \end{array}$$

  - There are three parameters (α, γ_1, and γ_2) and three group means, so we can solve uniquely for the parameters in terms of the group means:

$$\begin{aligned} \alpha &= \mu_3 \\ \gamma_1 &= \mu_1 - \mu_3 \\ \gamma_2 &= \mu_2 - \mu_3 \end{aligned}$$

  - Thus α represents the mean of the baseline category (group 3), and γ_1 and γ_2 capture differences between the other group means and the mean of the baseline category.

- One-way analysis of variance focuses on testing for differences among group means.
  - The omnibus F-statistic for the model tests H_0: $\gamma_1 = \gamma_2 = 0$, which corresponds to H_0: $\mu_1 = \mu_2 = \mu_3$, the null hypothesis of no differences among the population group means.
- Our consideration of one-way analysis of variance might well end here, but for a desire to develop methods that generalize easily to higher-way ANOVA.

3.1 The One-Way ANOVA Model

- New notation:
  - Y_{ij} denotes the ith observation within the jth of m groups.
  - n_j is the number of observations in the jth group.
  - $n = \sum_{j=1}^{m} n_j$ is the total number of observations.
  - $\mu_j \equiv E(Y_{ij})$ represents the population mean in group j (as before).
- The one-way ANOVA model:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

where:
  - μ should represent the general level of the response variable in the population.
  - α_j should represent the effect on the response variable of membership in the jth group.
  - ε_{ij} is an error variable that follows the usual linear-model assumptions.

- Upon taking expectations: $\mu_j = \mu + \alpha_j$.
  - The parameters of the model are, therefore, under-determined, for there are $m + 1$ parameters (including μ) but only m population group means.
  - For example, for $m = 3$:

$$\begin{aligned} \mu_1 &= \mu + \alpha_1 \\ \mu_2 &= \mu + \alpha_2 \\ \mu_3 &= \mu + \alpha_3 \end{aligned}$$

    – Even if we knew the three population group means, we could not solve uniquely for the four parameters.
  - Because the parameters of the model are under-determined, they cannot be uniquely estimated.
    – To estimate the model, we would need to code one dummy regressor for each group-effect parameter α_j, and the resulting dummy regressors would be perfectly collinear.

- One solution is to place a linear restriction on the parameters of the model:

$$w_0\mu + w_1\alpha_1 + \cdots + w_m\alpha_m = 0$$

where the w's are pre-specified constants, not all equal to 0.
  - All linear restrictions yield the same F-test for the null hypothesis of no differences in population group means.
    - For example, if we employ the restriction $\alpha_m = 0$, we are in effect deleting the parameter for the last category, making it a baseline category. The result is the dummy-coding scheme.
    - Alternatively, we could use the restriction $\mu = 0$, which is equivalent to deleting the constant term from the linear model, in which case the 'effect' parameters and group means are identical: $\alpha_j = \mu_j$.

3.2 'Sigma' Constraints

- It is advantageous to select a restriction that produces easily interpretable parameters and estimates, and that generalizes usefully to more complex models:

$$\sum_{j=1}^{m} \alpha_j = \alpha_1 + \alpha_2 + \cdots + \alpha_m = 0$$

- Employing this restriction (called a *sigma constraint*) to solve for the parameters produces

$$\begin{aligned} \mu &= \frac{\sum \mu_j}{m} \equiv \mu. \\ \alpha_j &= \mu_j - \mu. \end{aligned}$$

  - The dot (in $\mu.$) indicates averaging over the range of a subscript, here over groups. The *grand* or *general mean* μ, then, is the average of the population group means, while α_j gives the difference between the mean of group j and the grand mean.

  - The hypothesis of no differences in group means

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_m$$

is equivalent to the hypothesis that all of the effect parameters are zero

$$H_0\colon \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$$

- The sigma-constrained model can be estimated by coding *deviation regressors*, an alternative to the dummy-coding scheme.
  - We require $m-1$ deviation regressors, $S_1, S_2, \ldots, S_{m-1}$, the jth of which is coded according to the following rule:

$$S_j = \begin{cases} 1 & \text{for observations in group } j \\ -1 & \text{for observations in group } m \\ 0 & \text{for observations in all other groups} \end{cases}$$

  - For example, when $m = 3$:

group	(α_1) S_1	(α_2) S_2
1	1	0
2	0	1
3	-1	-1

    - Writing out the equations for the group means in terms of the deviation regressors:

group 1: $\mu_1 = \mu + 1 \times \alpha_1 + 0 \times \alpha_2 = \mu + \alpha_1$
group 2: $\mu_2 = \mu + 0 \times \alpha_1 + 1 \times \alpha_2 = \mu + \alpha_2$
group 3: $\mu_3 = \mu - 1 \times \alpha_1 - 1 \times \alpha_2 = \mu - \alpha_1 - \alpha_2$

    - The equation for the third group incorporates the sigma constraint, since $\alpha_3 = -\alpha_1 - \alpha_2$ is equivalent to $\alpha_1 + \alpha_2 + \alpha_3 = 0$.

– The omnibus F-statistic tests the hypothesis H_0: $\alpha_1 = \alpha_2 = 0$, which, under the sigma constraint, implies that α_3 is 0 as well — and that all of the population group means are equal.

▶ Although it is often convenient to fit the one-way ANOVA model by least-squares regression, it is also possible to estimate the model and calculate sums of squares directly.

- The sample mean $\overline{Y_j}$ in group j is the least-squares estimator of the corresponding population mean μ_j. Estimates of μ and the α_j may therefore be written as follows:

$$M = \widehat{\mu} = \frac{\sum \overline{Y_j}}{m} = \overline{Y}.$$

$$A_j = \widehat{\alpha}_j = \overline{Y_j} - \overline{Y}.$$

- The fitted Y-values are the group means,

$$\widehat{Y}_{ij} = M + A_j = \overline{Y}. + (\overline{Y_j} - \overline{Y}.) = \overline{Y}_j$$

- The regression and residual sums of squares therefore take particularly simple forms in one-way analysis of variance:

$$\text{RegSS} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(\widehat{Y}_{ij} - \overline{Y}\right)^2 = \sum_{j=1}^{m} n_j \left(\overline{Y}_j - \overline{Y}\right)^2$$

$$\text{RSS} = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \left(Y_{ij} - \widehat{Y}_{ij}\right)^2 = \sum \sum \left(Y_{ij} - \overline{Y}_j\right)^2$$

- This information can be presented in an ANOVA table:

Source	SS	df	MS	F
Groups	$\sum n_j \left(\overline{Y}_j - \overline{Y}\right)^2$	$m-1$	$\dfrac{\text{RegSS}}{m-1}$	$\dfrac{\text{RegMS}}{\text{RMS}}$
Residual	$\sum\sum \left(Y_{ij} - \overline{Y}_j\right)^2$	$n-m$	$\dfrac{\text{RSS}}{n-m}$	
Total	$\sum\sum \left(Y_{ij} - \overline{Y}\right)^2$	$n-1$		

▶ I will use Duncan's occupational-prestige data to illustrate one-way analysis of variance.

- Parallel boxplots for prestige in three types of occupations appear in Figure 1 (a).
  – Prestige, recall, is a percentage, and the data push both the lower and upper boundaries of 0 and 100 percent, suggesting the logit transformation in Figure 1 (b).
  – The data are better-behaved on the logit scale, which eliminates the skew in the blue-collar and professional groups and pulls in all of the outlying observations, with the exception of store clerks in the white-collar category.

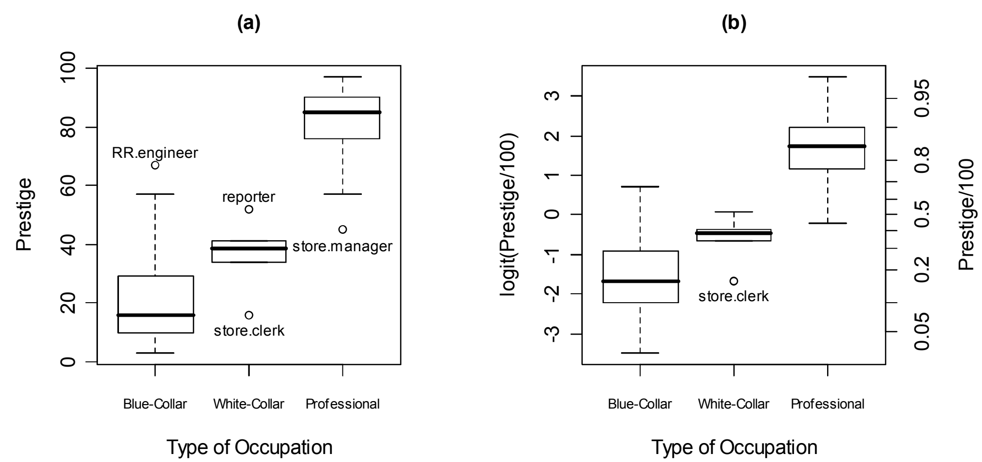

Figure 1. Parallel boxplots for (a) prestige and (b) the logit of prestige by type of occupation.

- Means, standard deviations, and frequencies for prestige within occupational types are as follows:

	Prestige		
Type of Occupation	*Mean*	*Standard Deviation*	*Frequency*
Professional and managerial	80.44	14.11	18
White collar	36.67	11.79	6
Blue collar	22.76	18.05	21

  – Professional occupations therefore have the highest average level of prestige, followed by white-collar and blue-collar occupations.

- The order of the group means is the same on the logit scale:

	logit(Prestige/100)	
Type of Occupation	*Mean*	*Standard Deviation*
Professional and managerial	1.6321	0.9089
White collar	−0.5791	0.5791
Blue collar	−1.4821	1.0696

- On both scales, the standard deviation is greatest among the blue-collar occupations and smallest among the white-collar occupations, but the differences are not very large.
- Using the logit of prestige as the response variable, the one-way ANOVA for the Duncan data is

Source	*Sum of Squares*	*df*	*Mean Square*	F	p
Groups	95.550	2	47.775	51.98	$\ll .0001$
Residuals	38.604	42	0.919		
Total	134.154	44			

  – Occupational types account for nearly three-quarters of the variation in the logit of prestige among these occupations ($R^2 = 95.550/134.154 = 0.712$).

4. Two-Way ANOVA

▶ Notation for population means in the two-way classification:

$$\begin{array}{c|cccc|c}
 & C_1 & C_2 & \cdots & C_c & \\
\hline
R_1 & \mu_{11} & \mu_{12} & \cdots & \mu_{1c} & \mu_{1\cdot} \\
R_2 & \mu_{21} & \mu_{22} & \cdots & \mu_{2c} & \mu_{2\cdot} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
R_r & \mu_{r1} & \mu_{r2} & \cdots & \mu_{rc} & \mu_{r\cdot} \\
\hline
 & \mu_{\cdot 1} & \mu_{\cdot 2} & \cdots & \mu_{\cdot c} & \mu_{\cdot\cdot}
\end{array}$$

- Within each *cell* of the design there is a population cell mean μ_{jk} for the response variable. Extending the dot notation, the *marginal mean* of the response variable in row j is

$$\mu_{j\cdot} \equiv \frac{\sum_{k=1}^{c} \mu_{jk}}{c}$$

- The marginal mean in column k is

$$\mu_{\cdot k} \equiv \frac{\sum_{j=1}^{r} \mu_{jk}}{r}$$

  and the grand mean is

$$\mu_{\cdot\cdot} \equiv \frac{\sum_j \sum_k \mu_{jk}}{r \times c} = \frac{\sum_j \mu_{j\cdot}}{r} = \frac{\sum_k \mu_{\cdot k}}{c}$$

▶ If R and C do not interact in determining the response variable, then the partial relationship between each factor and Y does not depend upon the category at which the other factor is ‘held constant.’

- This pattern is illustrated in Figure 2 (a) for the simple case where $r = c = 2$.
  – The difference in cell means across the two categories of R is the same within the two categories of C (and is therefore equal to the difference in the marginal means):

$$\mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \mu_{1\cdot} - \mu_{2\cdot}$$

  – No interaction implies parallel ‘profiles’ of cell means.

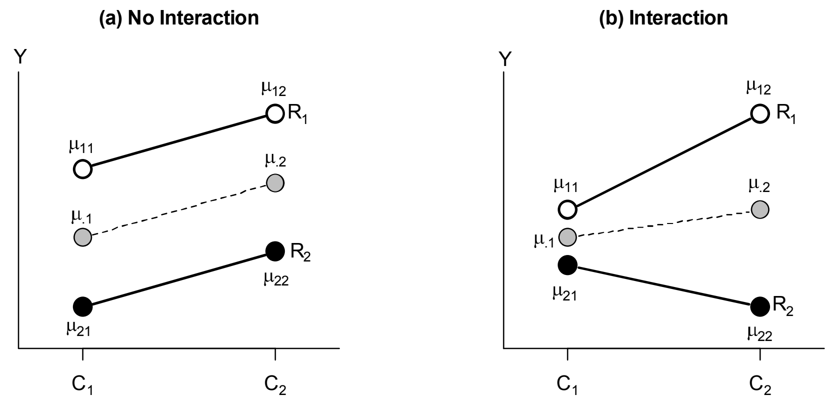

Figure 2. No interaction (a) and interaction (b) in the two-way classification.

– Parallel profiles also imply that the column difference for categories C_1 and C_2 is constant across rows, and is equal to the difference in column marginal means:

$$\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22} = \mu_{\cdot 1} - \mu_{\cdot 2}$$

- Interaction — where the row difference changes across columns (and the column difference changes across rows) — is illustrated in Figure 2 (b).

▶ The generalization of this is as follows:
- For any number of categories r of R and c of C, no interaction implies that all corresponding row differences are constant across columns,
$$\mu_{jk} - \mu_{j'k} = \mu_{jk'} - \mu_{j'k'} = \mu_{j\cdot} - \mu_{j'\cdot} \text{ for all } j, j' \text{ and } k, k'$$
and, equivalently, that all corresponding column differences are constant across rows,
$$\mu_{jk} - \mu_{jk'} = \mu_{j'k} - \mu_{j'k'} = \mu_{\cdot k} - \mu_{\cdot k'} \text{ for all } j, j' \text{ and } k, k'$$
- When interactions are absent, the partial effect of each factor — the factor's *main effect* — is therefore given by differences in the population marginal means.

4.1 Patterns of Means in the Two-Way Classification

▶ Several patterns of relationship in the two-way classification, all showing no interaction, are graphed in Figure 3:
- in (a) there are both row and column main effects;
- in (b) only column main effects;
- in (c) only row main effects;
- in (d) neither row nor column main effects.

▶ Figure 4 shows two different patterns of interactions:
- In (a), the interaction is dramatic: The order of row effects changes across columns and vice-versa. Interaction of this sort is sometimes called *disordinal*.
- In (b), the interaction is less dramatic.

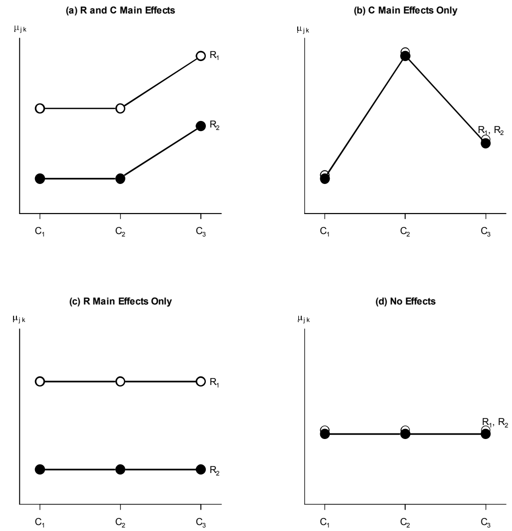

Figure 3. Patterns of association: (a) Row and Column main effects; (b) Column main effects only; (c) Row main effects only; (d) no effects.

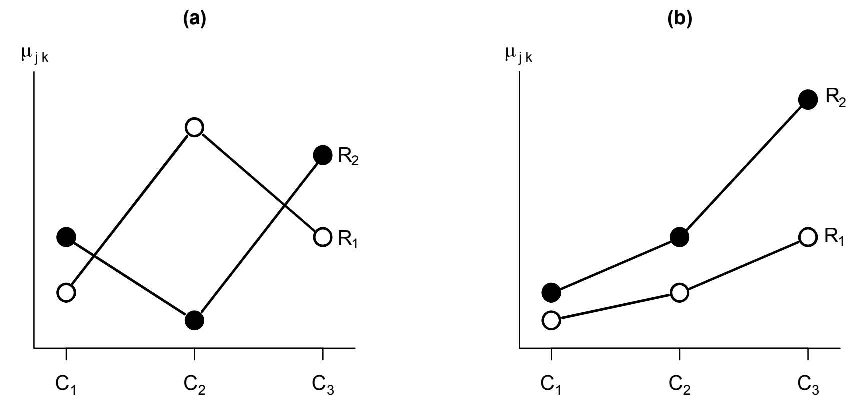

Figure 4. Two patterns of interaction in the two-way classification.

- Even when interactions are absent in the population, we cannot expect perfectly parallel profiles of *sample* means: There is sampling error in sampled data.
  - We have to determine whether departures from parallelism observed in a sample are sufficiently large to be statistically significant, and, if significant, are sufficiently large to be of interest.
  - In general, if interactions are non-negligible, then we do not interpret the main effects of the factors — consistent with the principle of marginality.
- The following table shows means ($\overline{Y}_{jk}$), standard deviations (S_{jk}), and cell frequencies (n_{jk}) for data from a social-psychological experiment, reported by Moore and Krupat (1971), designed to determine how the relationship between conformity and social status is influenced by 'authoritarianism.'

		Authoritarianism		
Partner's Status		*Low*	*Medium*	*High*
Low	$\overline{Y}_{jk}$	8.900	7.250	12.63
	s_{jk}	2.644	3.948	7.347
	n_{jk}	10	4	8
High	$\overline{Y}_{jk}$	17.40	14.27	11.86
	s_{jk}	4.506	3.952	3.934
	n_{jk}	5	11	7

- Because of the conceptual-rigidity component of authoritarianism, Moore and Krupat expected that low-authoritarian subjects would be *more* responsive than high-authoritarian subjects to the social status of their partner.

- The cell means are graphed along with the data in Figure 5, and appear to confirm the experimenters' expectations.
  - There are two outlying observations in the low-status partner, high-authoritarianism condition.

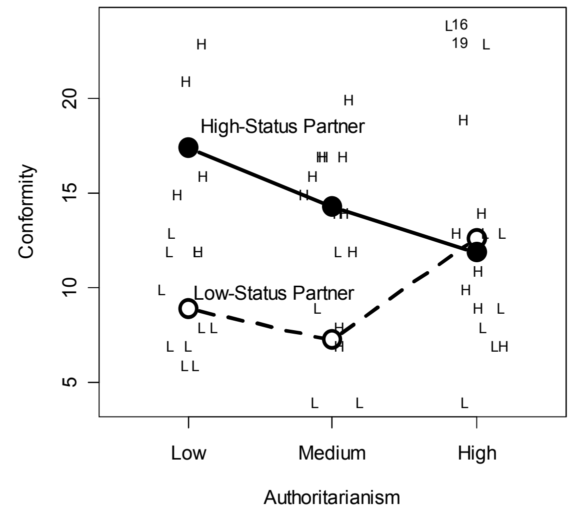

Figure 5. Mean conformity by authoritarianism and partner's status, for Moore and Krupat's data. The observations are jittered horizontally.

4.2 The Two-Way ANOVA Model

▶ Our first concern is to test the null hypothesis of no interaction.

- Based on the previous discussion, this hypothesis can be expressed in terms of the cell means:

$$H_0\colon \mu_{jk} - \mu_{j'k} = \mu_{jk'} - \mu_{j'k'} \text{ for all } j, j' \text{ and } k, k'$$

  - In words: the row effects are the same within all levels of the column factor.
- Rearranging terms,

$$H_0\colon \mu_{jk} - \mu_{jk'} = \mu_{j'k} - \mu_{j'k'} \text{ for all } j, j' \text{ and } k, k'$$

  - That is, the column effects are invariant across rows.

▶ It is convenient to express hypotheses concerning main effects in terms of the marginal means.

- Thus, for the row classification we have the null hypothesis

$$H_0\colon \mu_{1\cdot} = \mu_{2\cdot} = \cdots = \mu_{r\cdot}$$

and for the column classification

$$H_0\colon \mu_{\cdot 1} = \mu_{\cdot 2} = \cdots = \mu_{\cdot c}$$

- The main-effect hypotheses are testable whether interactions are present or absent, but these hypotheses are generally of interest only when the interactions are nil.

▶ The two-way ANOVA model provides a convenient means for testing the hypotheses about main effects and interactions. The model is

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}$$

where

- Y_{ijk} is the ith observation in row j, column k of the RC table;
- μ is the general mean of Y;

- α_j and β_k are main-effect parameters, for row-effects and column-effects, respectively;
- γ_{jk} are interaction parameters; and
- $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ and independent.

▶ Taking expectations, the model becomes

$$\mu_{jk} \equiv E(Y_{ijk}) = \mu + \alpha_j + \beta_k + \gamma_{jk}$$

- Since there are $r \times c$ population cell means and $1 + r + c + (r \times c)$ parameters, the parameters of the model are not uniquely determined by the cell means.
- As in one-way ANOVA, the indeterminacy of the model can be overcome by imposing $1 + r + c$ independent restrictions on its parameters.
  – It is convenient to select restrictions that make it simple to test the hypotheses of interest.

– With this purpose in mind, we specify the following sigma constraints on the model parameters:

$$\begin{aligned} \sum_{j=1}^{r} \alpha_j &= 0 \\ \sum_{k=1}^{c} \beta_k &= 0 \\ \sum_{j=1}^{r} \gamma_{jk} &= 0 \qquad \text{for all } k = 1, ..., c \\ \sum_{k=1}^{c} \gamma_{jk} &= 0 \qquad \text{for all } j = 1, ..., r \end{aligned}$$

– At first glance, it seems as if we have specified too many constraints, for the equations define $1 + 1 + c + r$ restrictions.

– One of the restrictions on the interactions is redundant, however.

– In short-hand form, the sigma constraints specify that each set of parameters sums to 0 over each of its coordinates.

▶ The constraints produce the following solution for model parameters in terms of population cell and marginal means:

$$\begin{aligned} \mu &= \mu_{\cdot\cdot} \\ \alpha_j &= \mu_{j\cdot} - \mu_{\cdot\cdot} \\ \beta_k &= \mu_{\cdot k} - \mu_{\cdot\cdot} \\ \gamma_{jk} &= \mu_{jk} - \mu - \alpha_j - \beta_k \\ &= \mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu_{\cdot\cdot} \end{aligned}$$

- The hypothesis of no row main effects is therefore equivalent to H_0: all $\alpha_j = 0$, for under this hypothesis

$$\mu_{1\cdot} = \mu_{2\cdot} = \cdots = \mu_{r\cdot} = \mu_{\cdot\cdot}$$

- Likewise, the hypothesis of no column main effects is equivalent to H_0: all $\beta_k = 0$, since then

$$\mu_{\cdot 1} = \mu_{\cdot 2} = \cdots = \mu_{\cdot c} = \mu_{\cdot\cdot}$$

- Finally, it is not difficult to show that the hypothesis of no interactions is equivalent to H_0: all $\gamma_{jk} = 0$.

4.3 Fitting the Two-Way ANOVA Model to Data

- Since the least-squares estimator of μ_{jk} is the sample cell mean

$$\overline{Y}_{jk} = \frac{\sum_{i=1}^{n_{jk}} Y_{ijk}}{n_{jk}}$$

least-squares estimators of the constrained model parameters follow immediately

$$\begin{aligned}
M &\equiv \widehat{\mu} = \overline{Y}_{\cdot\cdot} = \frac{\sum\sum \overline{Y}_{jk}}{r \times c} \\
A_j &\equiv \widehat{\alpha}_j = \overline{Y}_{j\cdot} - \overline{Y}_{\cdot\cdot} = \frac{\sum_k \overline{Y}_{jk}}{c} - \overline{Y}_{\cdot\cdot} \\
B_k &\equiv \widehat{\beta}_k = \overline{Y}_{\cdot k} - \overline{Y}_{\cdot\cdot} = \frac{\sum_j \overline{Y}_{jk}}{r} - \overline{Y}_{\cdot\cdot} \\
C_{jk} &\equiv \widehat{\gamma}_{jk} = \overline{Y}_{jk} - \overline{Y}_{j\cdot} - \overline{Y}_{\cdot k} + \overline{Y}_{\cdot\cdot}
\end{aligned}$$

- The residuals are just the deviations of the observations from their cell means, since the fitted values are the cell means:

$$\begin{aligned}
E_{ijk} &= Y_{ijk} - (M + A_j + B_k + C_{jk}) \\
&= Y_{ijk} - \overline{Y}_{jk}
\end{aligned}$$

- In testing hypotheses about sets of model parameters, however, we require incremental sums of squares for each set, and (unless all of the cell frequencies n_{jk} are equal) there is no way of calculating these sums of squares directly.
  - As in one-way analysis of variance, the restrictions on the two-way ANOVA model can be used to produce deviation-coded regressors.
  - Incremental sums of squares may then be calculated in the usual manner.
- To illustrate this procedure, we will examine a two-row $\times$ three-column classification:
  - In light of the restriction $\alpha_1 + \alpha_2 = 0$, α_2 can be deleted from the model, substituting $-\alpha_1$.

- Similarly, because $\beta_1 + \beta_2 + \beta_3 = 0$, β_3 can be replaced by $-\beta_1 - \beta_2$.
- The interactions in the 2×3 classification satisfy the following constraints:

$$\begin{aligned}
\gamma_{11} + \gamma_{12} + \gamma_{13} &= 0 \\
\gamma_{21} + \gamma_{22} + \gamma_{23} &= 0 \\
\gamma_{11} + \gamma_{21} &= 0 \\
\gamma_{12} + \gamma_{22} &= 0 \\
\gamma_{13} + \gamma_{23} &= 0
\end{aligned}$$

  – Although there are five such constraints, the fifth follows from the first four.)
- We can, as a consequence, delete all of the interaction parameters except γ_{11} and γ_{12}, substituting for the remaining four parameters in the following manner:

$$\begin{aligned}
\gamma_{13} &= -\gamma_{11} - \gamma_{12} \\
\gamma_{21} &= -\gamma_{11} \\
\gamma_{22} &= -\gamma_{12} \\
\gamma_{23} &= -\gamma_{13} = \gamma_{11} + \gamma_{12}
\end{aligned}$$

- These observations lead to the following coding of regressors for the 2×3 classification:

cell		(α_1)	(β_1)	(β_2)	(γ_{11})	(γ_{12})
row	column	R_1	C_1	C_2	R_1C_1	R_1C_2
1	1	1	1	0	1	0
1	2	1	0	1	0	1
1	3	1	-1	-1	-1	-1
2	1	-1	1	0	-1	0
2	2	-1	0	1	0	-1
2	3	-1	-1	-1	1	1

- Here, R_1 is the regressor for the row main effects;
- C_1 and C_2 are the regressors for the column main effects;
- R_1C_1 and R_1C_2 are the interaction regressors.
  – The notation for the interaction regressors is suggestive of multiplication, and in fact we can see that R_1C_1 is the product of R_1 and C_1, and that R_1C_2 is the product of R_1 and C_2.

▶ I have constructed these regressors to reflect the constraints on the model, but they can also be coded mechanically by applying these rules:

1. There are $r-1$ regressors (and hence degrees of freedom) for the row main effects; the jth such regressor, R_j, is coded according to the following scheme:
$$R_{ij} = \begin{cases} 1 & \text{if obs. } i \text{ is in row } j \\ -1 & \text{if obs. } i \text{ is in row } r \\ 0 & \text{if obs. } i \text{ is in any other row} \end{cases}$$

2. There are $c-1$ regressors (and *df*) for the column main effects; the kth such regressor, C_k, is coded according to the following scheme:
$$C_{ik} = \begin{cases} 1 & \text{if obs. } i \text{ is in column } k \\ -1 & \text{if obs. } i \text{ is in column } c \\ 0 & \text{if obs. } i \text{ is in any other column} \end{cases}$$
3. There are $(r-1)(c-1)$ regressors (and *df*) for the RC interactions. These interaction regressors consist of all pairwise products of the $r-1$ main-effect regressors for rows and $c-1$ main-effect regressors for columns.

4.4 Testing Hypotheses in Two-Way ANOVA

▶ I have specified constraints on the two-way ANOVA model so that testing hypotheses about the parameters of the constrained model is equivalent to testing hypotheses about interactions and main effects of the two factors.

▶ Tests for interactions and main effects can be constructed by the incremental sum of squares approach.
- Let SS(α, β, γ) denote the regression sum of squares for the full model, which includes both sets of main effects and the interactions.
- The regression sums of squares for other models are similarly represented.
  – For example, for the no-interaction model, we have SS(α, β);
  – and for the model that omits the column main-effect regressors, we have SS(α, γ).

– This last model violates the principle of marginality, but it plays a role in constructing the incremental sum of squares for testing the column main effects.

- As usual, incremental sums of squares are given by differences between the regression sums of squares for alternative models:
$$\begin{aligned} \text{SS}(\gamma|\alpha,\beta) &= \text{SS}(\alpha,\beta,\gamma) - \text{SS}(\alpha,\beta) \\ \text{SS}(\alpha|\beta,\gamma) &= \text{SS}(\alpha,\beta,\gamma) - \text{SS}(\beta,\gamma) \\ \text{SS}(\beta|\alpha,\gamma) &= \text{SS}(\alpha,\beta,\gamma) - \text{SS}(\alpha,\gamma) \\ \text{SS}(\alpha|\beta) &= \text{SS}(\alpha,\beta) - \text{SS}(\beta) \\ \text{SS}(\beta|\alpha) &= \text{SS}(\alpha,\beta) - \text{SS}(\alpha) \end{aligned}$$
  – We read SS$(\gamma|\alpha, \beta)$, for example, as 'the sum of squares for interaction *after* the main effects,' and SS$(\alpha|\beta)$ as 'the sum of squares for the row main effects *after* the column main effects and *ignoring* the interactions.'

– The residual sum of squares is

$$\begin{aligned} \text{RSS} &= \sum\sum\sum E_i^2 \\ &= \sum\sum\sum (Y_{ijk} - \overline{Y}_{jk})^2 \\ &= \text{TSS} - \text{SS}(\alpha, \beta, \gamma) \end{aligned}$$

- The incremental sum of squares for interaction, SS$(\gamma|\alpha, \beta)$, is appropriate for testing the null hypothesis of no interaction, H_0: all $\gamma_{jk} = 0$.
- In the presence of interactions, we can use SS$(\alpha|\beta, \gamma)$ and SS$(\beta|\alpha, \gamma)$ to test hypotheses concerning main effects (i.e., differences among row and column marginal means), but these hypotheses are usually not of interest when the interactions are important.

- In the *absence* of interactions, SS$(\alpha|\beta)$ and SS$(\beta|\alpha)$ can be used to test for main effects, but the use of SS$(\alpha|\beta, \gamma)$ and SS$(\beta|\alpha, \gamma)$ is also appropriate.
  - If, however, interactions are *present*, then F-tests based on SS$(\alpha|\beta)$ and SS$(\beta|\alpha)$ *do not* test the main-effect null hypotheses H_0: all $\alpha_j = 0$ and H_0: all $\beta_k = 0$; instead, the interaction parameters become implicated in these tests.

- These remarks are summarized in the following table:

Source	df	SS	H_0
R	$r-1$	SS$(\alpha\|\beta, \gamma)$	all $\alpha_j = 0$ $(\mu_{j\cdot} = \mu_{j'\cdot})$
		SS$(\alpha\|\beta)$	all $\alpha_j = 0\|$all $\gamma_{jk} = 0$ $(\mu_{j\cdot} = \mu_{j'\cdot}\|$no int.$)$
C	$c-1$	SS$(\beta\|\alpha, \gamma)$	all $\beta_k = 0$ $(\mu_{\cdot k} = \mu_{\cdot k'})$
		SS$(\beta\|\alpha)$	all $\beta_k = 0\|$all $\gamma_{jk} = 0$ $(\mu_{\cdot k} = \mu_{\cdot k'}\|$no int.$)$
RC	$(r-1)(c-1)$	SS$(\gamma\|\alpha, \beta)$	all $\gamma_{jk} = 0$ $(\mu_{jk} - \mu_{j'k} = \mu_{jk'} - \mu_{j'k'})$
Residual	$n - rc$	TSS − SS$(\alpha, \beta, \gamma,)$	
Total	$n-1$	TSS	

- Certain authors prefer main-effects tests based upon SS$(\alpha|\beta)$ and SS$(\beta|\alpha)$ (sometimes called 'Type-II sums of squares') because, if interactions are absent, tests based upon these sums of squares are more powerful than those based upon SS$(\alpha|\beta, \gamma)$ and SS$(\beta|\alpha, \gamma)$.

- Other authors prefer SS$(\alpha|\beta, \gamma)$ and SS$(\beta|\alpha, \gamma)$ (sometimes called 'Type-III' sums of squares) because, in the presence of interactions, tests based upon these sums of squares have a straight-forward (if usually uninteresting) interpretation.
- I believe that either approach is reasonable. It is important to understand, however, that while SS(α) and SS(β) are useful as building blocks of SS$(\alpha|\beta)$ and SS$(\beta|\alpha)$, it is in general *inappropriate* to use SS(α) and SS(β) to test hypotheses about the R and C main effects: Each of these sums of squares depends upon the other set of main effects (and the interactions, if they are present).
  - Consequently, the sequential ("Type-I") sums of squares SS(α),SS$(\beta|\alpha)$, and SS$(\gamma|\alpha, \beta)$ do not provide an appropriate test for the R main effects.

4.5 An Example: Moore and Krupat's Conformity Experiment

- For the Moore and Krupat conformity data, factor R is partner's status and factor C is authoritarianism.
- Sums of squares for various models fit to the data are as follows:

$$\begin{aligned} \text{SS}(\alpha,\beta,\gamma) &= 391.44 \\ \text{SS}(\alpha,\beta) &= 215.95 \\ \text{SS}(\alpha,\gamma) &= 355.42 \\ \text{SS}(\beta,\gamma) &= 151.87 \\ \text{SS}(\alpha) &= 204.33 \\ \text{SS}(\beta) &= 3.7333 \\ \text{TSS} &= 1209.2 \end{aligned}$$

- The ANOVA for the experiment is shown in the following table:

Source	SS	df	MS	F	p
Partner's Status		1			
$\alpha\vert\beta,\gamma$	239.57		239.57	11.43	.002
$\alpha\vert\beta$	212.22		212.22	10.12	.003
Authoritarianism		2			
$\beta\vert\alpha,\gamma$	36.02		18.01	0.86	.43
$\beta\vert\alpha$	11.62		5.81	0.28	.76
Status× Authoritarianism	175.49	2	87.74	4.18	.02
Residual	817.76	39	20.97		
Total	1209.2	44			

- A researcher would not normally report *both* sets of main-effect sums of squares.

5. Summary

- One-way analysis of variance examines the relationship between a quantitative response variable and a categorical explanatory variable (or factor).
- The one-way ANOVA model

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

  is under-determined because it uses $m+1$ parameters to model m group means.
  - The model can be solved, however, by placing a restriction on its parameters.
  - Setting one of the α_j's to 0 leads to dummy-regressor coding.
  - Constraining the α_j's to sum to 0 leads to deviation-regressor coding.
  - The two coding schemes are equivalent in that they provide the same fit to the data, producing the same regression and residual sums of squares.

- The two-way analysis of variance model

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}$$

  incorporates main effects and interactions of two factors.
  - The factors interact when the profiles of population cell means are not parallel.
- The two-way ANOVA model is over-parameterized, but it may be fit to data by placing suitable restrictions on its parameters.
  - A convenient set of restrictions is provided by sigma constraints, specifying that each set of parameters (α_j, β_k, and γ_{jk}) sums to 0 over each of its coordinates.
  - Testing hypotheses about the sigma-constrained parameters is equivalent to testing interaction-effect and main-effect hypotheses about cell and marginal means.

- There are two reasonable procedures for testing main-effect hypotheses in two-way ANOVA:
  - Tests based on SS$(\alpha|\beta,\gamma)$ and SS$(\beta|\alpha,\gamma)$ (Type-III sums of squares) employ models that violate the principle of marginality, but are valid whether or not interactions are present.
  - Tests based on SS$(\alpha|\beta)$ and SS$(\beta|\alpha)$ (Type-II sums of squares) conform to the principle of marginality, but are valid only if interactions are absent.