Part III
Analysis of Variance (ANOVA)

Contents

1 One-way design . . . 2
1.1 The fixed-effects model . . . 3
1.2 Estimation . . . 4
1.3 Analysis of variance . . . 5
1.4 Linear contrasts . . . 8
1.5 The random-effects model . . . 10
2 Two-way design . . . 13
2.1 Complete randomised block design . . . 13
2.2 The additive fixed-effects model . . . 14
2.3 Analysis of variance . . . 16
2.4 Linear contrasts . . . 17
2.5 The interacting fixed-effects model . . . 19
2.6 Analysis of variance in the presence of an interaction . . . 21
3 Nested two-way design . . . 24
3.1 The nested fixed-effects model . . . 24
3.2 Analysis of variance . . . 25

Part III
Analysis of Variance (ANOVA)

Analysis of Variance goes back to early work by Ronald Fisher in 1918 on mathematical genetics and was further developed by him at Rothamsted Experimental Station in Harpenden, Hertfordshire in the 1920s. The convenient acronym ANOVA was coined much later, by the American statistician John W. Tukey, the pioneer of exploratory data analysis (EDA) in statistics, and coiner of the terms hardware, software and bit from computer science.

Lecture 9
Suggested Reading:

- ***W. J. Krzanowski***, *An Introduction to Statistical Modelling, Wiley, 2008* Section 4
- ***M. H. Kutner et al.***, *Applied Linear Statistical Models, 5th Edition, McGraw-Hill Irwin, 2005* Chapters 15-18
- ***D. C. Montgomery***, *Design and Analysis of Experiments, 9th Edition, Wiley, 2017* Chapter 3

Fisher's motivation, which arose directly from the agricultural field trials carried out at Rothamsted, was to compare yields of several varieties of crop, or of one crop under several fertiliser treatments. He realised that if there was more variability between groups (of yields with different treatments) than within groups (of yields with the same treatment) than one would expect if the treatments were the same, then this would be evidence against believing that they were the same. In other words, Fisher set out to compare means by analysing variability.

Variance – the term is due to Fisher – is simply a short form of ***variability***

1 *One-way design*

We begin with the simplest situation, namely where the investigator wishes merely to compare the responses observed in several ***a priori*** groups (factor levels). If there are just two such groups, then we can use the simple two-sample t-test familiar from introductory statistics courses, but what can we do if there are more than two groups?

One approach would be to take the groups two at a time, and to compare means by conducting all possible such two–sample tests. While this is clearly a feasible procedure, it is unsatisfactory and inefficient because:

- only the data from the pair of samples under consideration are used in each test, whereas often the complete set of data has been collected as a single experiment and every observation (response) has a contribution to make towards calculation of such quantities as standard errors;
- even with a moderate number of groups there will be many pair-wise comparisons, and the more comparisons that are made the higher is the probability of obtaining false significant differences for some of the tests.

These arguments imply that we should use all the data and conduct just a single analysis of differences between the groups, so the way forward is to formulate an appropriate model for the whole situation.

1.1 The fixed-effects model

Let us assume that

- there are a a priori populations (treatment groups) from which random samples of sizes $n_1, n_2, \ldots, n_a$ respectively have been taken, and
- that y_{ij} denotes the observation (trial) on the j-th unit in the sample from the i-th population (treatment group); here $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$,
- the value of y_{ij} is determined by the unknown ***within group mean***, denoted μ_i, plus a random error, ε_{ij}, usually assumed to be distributed normally, $\varepsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2)$.

Then, following the previous chapters, a regression model for the data is

$$y_{ij} = \mu_i + \varepsilon_{ij}.$$

In other words, for each unit j within treatment group i, the y_{ij} can be thought of being sampled from a normal distribution with mean μ_i and constant variance σ^2.

As mentioned above, ***the main aim of the ANOVA analysis is to determine if there is a systematic difference in the effect of the response variable between each of the treatment groups***. In particular, when we talk about a difference between groups, we are actually interested in a difference between the corresponding within group means, μ_i, of the underlying population and therefore, it is convenient for us to slightly alter the structure of the above model to make this more apparent.

Write $N = n_i + \ldots + n_a$ for the total number of trials. Denote by μ the ***weighted overall mean*** of the N individuals,

$$\mu = \frac{1}{N} \sum_{i=1}^{a} n_i \mu_i.$$

Then we may write

$$\mu_i = \mu + (\mu_i - \mu) \equiv \mu + \alpha_i,$$

where $\alpha_i = \mu_i - \mu$ is called the ***treatment effect*** of the i-th population.

Definition 1.1:

The fixed-effects one-way ANOVA model is

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2) \tag{1}$$

where $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$.

Note that

$$\sum_{i=1}^{a} n_i \alpha_i = \sum_{i=1}^{a} n_i(\mu_i - \mu) = N\mu - N\mu = 0.$$

The effects α_i are called ***fixed-effects*** since the treatments groups were chosen ***a priori***.

We will require to sum and average quantities repeatedly throughout this section, so some shorthand notation will be convenient to avoid both overuse of the summation sign and possible confusion with overbars.

- The sample mean of the i-th treatment group is given by

$$\bar{y}_{i\bullet} = \frac{1}{n_i} y_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

- The overall sample mean (the "*grand*" mean) is given by

$$\bar{y}_{\bullet\bullet} = \frac{1}{N} y_{\bullet\bullet} = \frac{1}{N} \sum_{i=1}^{a} \sum_{j=1}^{n_i} y_{ij}.$$

Note that in the case when the number of responses within each treatment group is equal, $n := n_1 = \ldots = n_a$ and $N = an$, then

$$\mu = \frac{1}{N} \sum_{i=1}^{a} n_i \mu_i = \frac{1}{a} \sum_{i=1}^{a} \mu_i$$

becomes the ***unweighted overall mean***. A model where all n_i's are equal is known as a ***balanced design***.

1.2 *Estimation*

To estimate parameters μ and α_i in the one-way model (1) we can either use the method of maximum likelihood or the method of least-squares. Both approaches lead to the minimization of the sum of errors squared,

$$S = \sum_{i=1}^{a} \sum_{j=1}^{n_i} \varepsilon_{ij}^2 = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i)^2.$$

Differentiating S with respect to μ and α_i in turn, setting the resulting expressions to zero and solving the ensuing $a+1$ simultaneous equations, and remembering in the process that $\sum_{i=1}^{a} n_i \alpha_i = 0$, we find the following result.

The least-squares estimates of μ and α_i are

$$\hat{\mu} = \bar{y}_{\bullet\bullet}, \qquad \hat{\alpha}_i = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}.$$

The fitted value of the observation y_{ij} is the corresponding group mean,

$$\hat{y}_{ij} = \bar{y}_{i\bullet}.$$

The residuals, defined as the differences of the observed and fitted values, are

$$e_{ij} = y_{ij} - \bar{y}_{i\bullet}.$$

An important property of the residuals for the one-way model is that they sum to zero for each group i:

$$\sum_{j=1}^{n_i} e_{ij} = 0\,.$$

Moreover, it can be shown that

$$\mathbb{E}(e_{ij}) = 0, \qquad \mathrm{Cov}(e_{ij}, e_{kl}) = \delta_{ik}\left(\delta_{jl} - \frac{1}{n_i}\right)\sigma^2.$$

Here δ_{ik} and δ_{jl} are the Kronecker delta symbols:

- $\delta_{ab} = 1$ if $a = b$ and
- $\delta_{ab} = 0$ otherwise.

Hence

$$\mathrm{Var}(e_{ij}) = \mathrm{Cov}(e_{ij}, e_{ij}) = \delta_{ij} \frac{n_i - 1}{n_i} \sigma^2$$

The standardised residuals are thus given by

$$r_{ij} = \frac{e_{ij}}{\sqrt{\frac{n_i - 1}{n_i} \hat{\sigma}^2}},$$

where $\hat{\sigma}^2 = \frac{1}{N-a} \sum_{i=1}^{a} \sum_{j=1}^{n_i} e_{ij}^2$ $(\equiv MS_E)$ is the least-squares estimate of the unknown variance σ^2.

As for linear regression, analysis of normality assumptions for residuals is useful for examining the appropriateness of the ANOVA model.

1.3 Analysis of variance

To test for differences among the a groups, we must look for differences among the a population means μ_i. This is equivalent to test whether any of the treatment effects, α_i, are different from zero. Hence, we need to consider the hypothesis

$$H_0: \text{ all } \alpha_i = 0 \qquad \text{vs.} \qquad H_1: \text{ at least one } \alpha_i \neq 0. \tag{2}$$

The general procedure for testing these hypothese is the ***analysis of variance***. We have already met this procedure in the context of linear regression. However, it was originally developed within the framework of the one-way experimental design of the type we are considered in this chapter.

Recall that ***analysis of variance is a procedure for partitioning a sum of squares into components associated with recognized sources of variation.***

To derive the test for differences among groups, we first need to define the ***total sum of squares*** (total squared variation of responses)

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\bullet\bullet})^2.$$

We then define the ***between groups sum of squares***, denoted by SS_B, which is the sum of squared difference between the group sample means and the overall sample mean,

$$SS_B = \sum_{i=1}^{a} n_i \left(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}\right)^2,$$

and the ***residual sum of squares*** or within group sum of squares, denoted by SS_E, which describes the residual variation occurring via random fluctuations (not accounted for by the model),

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i\bullet}\right)^2.$$

These quantities satisfy the analysis of variance (AoV) identity,

$$SS_T = SS_B + SS_E.$$

In other words, we have partitioned the total sum of squared variation into the sum of variation accounted for by the model (potential difference in group means) and the residual variation.

Indeed, we can rewrite SS_T as

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[(y_{ij} - \bar{y}_{i\bullet}) + (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})\right]^2.$$

Now notice that

$$\begin{aligned}
&\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}) \\
&\quad = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}\bar{y}_{i\bullet} - \bar{y}_{i\bullet}\bar{y}_{i\bullet} - y_{ij}\bar{y}_{\bullet\bullet} + \bar{y}_{i\bullet}\bar{y}_{\bullet\bullet}) \\
&\quad = \sum_{i=1}^{a}(n_i\,\bar{y}_{i\bullet}\bar{y}_{i\bullet} - n_i\,\bar{y}_{i\bullet}\bar{y}_{i\bullet} \\
&\qquad\quad - n_i\bar{y}_{i\bullet}\bar{y}_{\bullet\bullet} + n_i\bar{y}_{i\bullet}\bar{y}_{\bullet\bullet}) \\
&\quad = 0.
\end{aligned}$$

Thus expanding the square in terms of the quantities in the square brackets of the above SS_T, we find

$$\begin{aligned}
SS_T &= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{\bullet\bullet})^2 \\
&= \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 \\
&= SS_B + SS_E,
\end{aligned}$$

which is the wanted AoV identity.

For computation purposes it is more convenient to use these expressions:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - N\bar{y}_{\bullet\bullet}^2$$

$$SS_B = \sum_{i=1}^{a} n_i\bar{y}_{i\bullet}^2 - N\bar{y}_{\bullet\bullet}^2$$

$$SS_E = SS_T - SS_B$$

Define the ***mean squares***

$$MS_B = \frac{SS_B}{a-1} \quad \text{and} \quad MS_E = \frac{SS_E}{N-a}.$$

The quantities $\nu_A = a - 1$ and $\nu_E = N - a$ are the degrees of freedom of SS_B and SS_E, respectively. It can be shown that

$$\mathbb{E}(MS_B) = \sigma^2 + \sum_{i=1}^{a}\frac{n_i(\mu_i - \mu)^2}{a-1}, \qquad \mathbb{E}(MS_E) = \sigma^2.$$

See ***Montgomery***, Section 3.3.1

Recall that the null hypothesis H_0 is that all the group means μ_i are equal, or equivalently all the treatment effects α_i are equal to zero, cf. (2). Then, under the null hypothesis, the value of MS_B should be comparable to MS_E. Conversely, a value of MS_B much bigger than MS_E suggests (but does not prove) that at least one α_i is nonzero.

Under the null hypothesis H_0, the normality and independence of the errors ε_{ij} imply that

$$\frac{SS_B}{\sigma^2} \sim \chi^2_{a-1} \quad \text{and} \quad \frac{SS_E}{\sigma^2} \sim \chi^2_{N-a}$$

are independent random variables. Consequently, the ratio

$$F = \frac{SS_B/(\sigma^2(a-1))}{SS_E/(\sigma^2(N-a))} = \frac{MS_B}{MS_E} \sim F_{a-1,N-a}$$

is F-distributed random variable.

We can now use this property and compare the calculated F ratio F_{cal} with the corresponding critical value $F_{crit} = F_{a-1,N-a,\alpha}$ from the F tables, at a given significance level α. This will conduct a formal hypothesis test on H_0 in the same way as we did in previous chapters for the SLR and MLR.

All of these important quantities used within the analysis of variance are usually recoded in the ANOVA table below.

Recall that we reject H_0 if $F_{cal} > F_\alpha$.

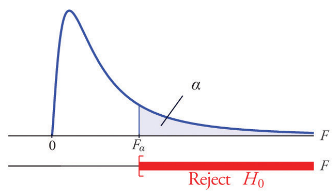


Table 1: One-way ANOVA table.

Source of variation	d.o.f.	SS	MS	F
Between groups	$a-1$	SS_B	$MS_B = \frac{SS_B}{a-1}$	$F = \frac{MS_B}{MS_E}$
Residual/Within groups	$N-a$	SS_E	$MS_E = \frac{SS_E}{N-a}$	
Total	$N-1$	SS_T		

Example 1: Tyre manufacturer

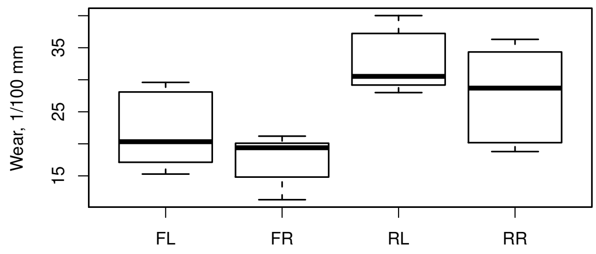
A tyre manufacturer wishes to investigate the rate of wear on their products and whether this rate differs substantially among the four possible positions that the tyre can occupy on the car:

$$FL = \text{front left}, \qquad FR = \text{front right},$$
$$RL = \text{rear left}, \qquad RR = \text{rear right}.$$

The manufacturer sets up an experiment in which the tyres are fitted to a car, the car is driven at a fixed speed for a fixed distance and the reduction in depth of tread caused by the test is measured (in hundredths of mm) for each tyre. The process is then repeated nine times with a new set of tyres each time. The table below shows the measurements for each tyre position of each of the nine experiments:

FL	20.935	19.013	20.332	17.123	15.919	15.285	29.59	28.092	28.304
FR	18.279	21.2	19.389	14.815	11.28	12.153	19.973	20.096	20.477
RL	28.535	27.998	30.073	37.227	38.853	40.017	30.529	29.177	30.795
RR	20.182	18.792	19.203	34.34	34.707	36.307	29.023	28.176	28.701

A boxplot of wear partitioned by position indicates that not all positions have the same effect on tyres:



There are $a = 4$ groups with $n = 9$ units in each, thus $N = 4 \times 9 = 36$. The sums of squares are

$$\begin{aligned} SS_T &= 24357.36 - 36 \times 24.858^2 = 2112.132, \\ SS_B &= 9 \times (467.487 + 306.880 + 1061.341 + 768.097) - 36 \times 24.858^2 = 1189.113, \\ SS_E &= 2112.133 - 1189.014 = 923.019. \end{aligned}$$

The total, between groups, and residual degrees of freedom are

$$\nu_T = N - 1 = 36 - 1 = 35, \qquad \nu_B = a - 1 = 4 - 1 = 3, \qquad \nu_E = N - a = 36 - 4 = 32.$$

We can now draw up the ANOVA table:

Source of variation	d.o.f.	SS	MS	F
between groups	3	1189.113	396.371	13.742
Residual/Within Groups	32	923.019	28.844	
Total	35	2112.132		

The calculated F-value is $F_{cal} = 13.742$. The corresponding P-value is $P_{cal} = 6.27 \times 10^{-6} \approx 0$, which indicates a strong evidence for differences among the four tyre positions.

The R code for this example is:

```
> # tyre wear data
> FL <- c(20.935,19.013,20.332,17.123,15.919,15.285,29.590,28.092,28.304)
> FR <- c(18.279,21.200,19.389,14.815,11.280,12.153,19.973,20.096,20.477)
> RL <- c(28.535,27.998,30.073,37.227,38.853,40.017,30.529,29.177,30.795)
> RR <- c(20.182,18.792,19.203,34.340,34.707,36.307,29.023,28.176,28.701)
> # combine data into a 4-column dataframe, then stack into two columns:
> tyre.data <- stack(data.frame(cbind(FL,FR,RL,RR)))
> # rename columns to "wear" and "position"
> names(tyre.data) <- c("wear", "position")
> # visual inspection
> boxplot(wear~position, data = tyre.data)
> # fit a one-way anova model
> tyre.aov <- aov(wear~position, data = tyre.data)
> summary(tyre.aov)
```

A key element in this code is the function `stack()` which stacks the 4-column and 9-row dataframe `data.frame(cbind(FL,FR,RL,RR))` into a 2-column and 36-row dataframe `tyre.data`. The first column, `wear`, contains all the measurements. The second column, `position`, indicates the tyre position. Function `aov()` is a modification of `lm()` for fitting ANOVA models.

1.4 Linear contrasts

Preparing the ANOVA table and conducting the F-test for differences between groups is just the first stage of a full analysis of the data. Often, an experiment is planned to provide answers to a number of separate questions, e.g. does the "control" treatment exert a different effect from the average of all the other treatments? Do two specified treatments have different effects? Do two sets of treatments have different effects on average? In such circumstances the overall F-test is not that helpful, as it is directed at a mixture of all these questions. We need to disentangle the answers to the individual questions from the overall result.

The overall F-test is an average of as many independent comparisons as there are degrees of freedom in the numerator mean square. In other words, while the null hypothesis that all population means are equal is specified uniquely, the alternative that at least one differs from the rest can be satisfied in many different ways.

Definition 1.2:

Any linear combination of group totals

$$z_w = l_{w1}\, y_{1\bullet} + l_{w2}\, y_{2\bullet} + \ldots + l_{wa}\, y_{a\bullet} \tag{3}$$

is called a ***linear comparison***, or ***linear contrast***, among the group totals if the coefficients l_{wi} satisfy the condition

$$n_1\, l_{w1} + n_2\, l_{w2} + \ldots + n_a\, l_{wa} = 0. \tag{4}$$

Two linear contrasts z_w and z_u are said to be orthogonal if

$$n_1\, l_{w1} l_{u1} + n_2\, l_{w2} l_{u2} + \ldots + n_a\, l_{wa} l_{ua} = 0. \tag{5}$$

Note that in the balanced design the relations (4) and (5) are equivalent to

$$\sum_{i=1}^{a} l_{wi} = 0$$

and

$$\sum_{i=1}^{a} l_{wi}\, l_{ui} = 0.$$

Linear contrasts are used to partition the between groups sum of squares, SS_B, into ***single degree of freedom components*** in the following manner:

- if z_w is any linear contrast among the $y_{i\bullet}$, and

  $$D_w = n_1 l_{w1}^2 + n_2 l_{w2}^2 + \ldots + n_a l_{wa}^2,$$

  then the quantity z_w^2/D_w is a single degree of freedom component of SS_B;
- if z_w and z_u are orthogonal, then z_u^2/D_u is similarly a single degree of freedom component of $(SS_B - z_w^2/D_w)$;
- if SS_B has $a-1$ degrees of freedom and $z_1, z_2, \ldots, z_{a-1}$ are any mutually orthogonal linear contrasts, then

  $$SS_B = \frac{z_1^2}{D_1} + \frac{z_2^2}{D_2} + \ldots + \frac{z_{a-1}^2}{D_{a-1}}.$$

- given a linear contrast z_w the corresponding effect, i.e. the null hypothesis

  $$H_0 \;:\; l_{w1}\mu_1 + l_{w2}\mu_2 + \ldots + l_{wa}\mu_a = 0, \tag{6}$$

  can be tested using the ratio

  $$F_w = \frac{z_w^2/D_w}{MS_E} \sim F_{1,N-a}.$$

Starting with a specified contrast z_1, we can always find appropriate $z_2, \ldots, z_{a-1}$ to construct a completely orthogonal set. The greater the value of a, the greater is the number of distinct possibilities. It is the analyst's task to choose the most relevant set of z_w's for interpretation of the results.

If non-orthogonal contrasts are selected by mistake, then the sum $\sum_{i=1}^{a} z_i^2/D_i$ will not equal SS_B.

In the balanced design the null hypothesis (6) becomes

$$H_0 \;:\; l_{w1}\alpha_1 + l_{w2}\alpha_2 + \ldots + l_{wa}\alpha_a = 0.$$

Each contrast z_w has single degree of freedom. Thus the quantity z_w^2/D_w is also a mean square.

Example 2: Orthogonal contrasts

Consider an experiment with three groups and equal sample sizes, i.e. $a = 3$ and $n_1 = n_2 = n_3 \equiv n$. Assume that group 1 is a "control" group while groups 2 and 3 correspond to the two experimental regimes, "A" and "B". Choose $l_1 = (+2, -1, -1)$ and $l_2 = (0, +1, -1)$ so that

$$\sum_{i=1}^{a} n_i\, l_{1i} = n\,(l_{11} + l_{12} + l_{13}) = n(2 - 1 - 1) = 0,$$
$$\sum_{i=1}^{a} n_i\, l_{2i} = n\,(l_{21} + l_{22} + l_{23}) = n(0 + 1 - 1) = 0$$

and

$$\sum_{i=1}^{n} n_i\, l_{1i}\, l_{2i} = n\,(l_{11}\, l_{21} + l_{12}\, l_{22} + l_{13}\, l_{23}) = n\,(0 - 1 + 1) = 0,$$

i.e. l_1 and l_2 can be used to define orthogonal linear contrasts. Then

$$z_1 = \underbrace{2\,y_{1\bullet}}_{\text{control}} \underbrace{- y_{2\bullet} - y_{3\bullet}}_{\text{A and B}}, \qquad z_2 = \underbrace{0}_{\text{control}} + \underbrace{y_{2\bullet}}_{\text{A}} - \underbrace{y_{3\bullet}}_{\text{B}}$$

are a contrast between the control group and the two experimental regimes, and a contrast between the two experimental regimes, A and B. The corresponding null hypotheses are

$$H_0^{(1)} : 2\alpha_1 = \alpha_2 + \alpha_3, \qquad H_0^{(2)} : \alpha_2 = \alpha_3.$$

Example 3: Tyre manufacturer (continued)

Suppose the manufacturer is interested in the average effects on wear of front compared with rear tyres, and of right compared with left positions. We can partition the between-group sum of squares in such a way that these effects are obtained as orthogonal contrasts with a single degree of freedom each. Recall that the treatment groups are *front left*, *front right*, *real left* and *rear right*. Choose

$$l_1 = (\underbrace{+1, +1}_{\text{front}}, \underbrace{-1, -1}_{\text{rear}}), \qquad l_2 = (\underbrace{+1, -1, +1, -1}_{\text{left vs. right}})$$

so that

$$z_1 = y_{1\bullet} + y_{2\bullet} - y_{3\bullet} - y_{4\bullet} = -190.38,$$
$$z_2 = y_{1\bullet} - y_{2\bullet} + y_{3\bullet} - y_{4\bullet} = +80.704.$$

The z_1 contrasts *front and rear* tyres, z_2 contrasts *left and right* tyres, and z_1 and z_2 are orthogonal. The corresponding quantities D_1 and D_2 are

$$D_1 = 9 \times 1^2 + 9 \times 1^2 + 9 \times (-1)^2 + 9 \times (-1)^2 = 36,$$
$$D_2 = 9 \times 1^2 + 9 \times (-1)^2 + 9 \times 1^2 + 9 \times (-1)^2 = 36.$$

The components of SS_B due to the "front–rear" and "left–right" effects are

$$SS_{FR} = \frac{(-190.38)^2}{36} = 1006.79, \qquad SS_{LR} = \frac{80.704^2}{36} = 180.92.$$

These two components have a single degree of freedom each, so are also the respective mean squares, giving the following F-ratios:

$$F_{FR} = \frac{SS_{FR}}{MS_E} = \frac{1006.79}{28.847} = 34.9, \qquad F_{LR} = \frac{SS_{LR}}{MS_E} = \frac{180.92}{28.847} = 6.27.$$

The significance of the "front–rear" and "left–right" contrasts is tested by comparing the calculated F-ratios agains the the critical F-ratio, $F_{crit} = F_{1,32,\alpha} = 4.15$. Since both F_{FR} and F_{LR} are greater that F_{crit}, both contrasts show a significant effect at this level. Moreover, there is still one degree of freedom left after these effects have been removed from the SS_B. The remaining contrast is the "cross" effect and must have a sum of squares equal to

$$SS_B - SS_{FR} - SS_{LR} = 1189.014 - 1006.79 - 180.92 = 1.304,$$

and thus is not significant, $F_{cross} = 1.304/28.847 = 0.045 < F_{crit}$. The explanation for the observed difference between the four groups thus resides exclusively in the former two contrasts.

The R code for this example is (continuing from Example 1):

```
> l1 <- c(+1,+1,-1,-1)  # front vs rear contrast
> l2 <- c(+1,-1,+1,-1)  # left vs right contrast
> # assign contrasts to the data
> contrasts(tyre.data$position) <- cbind(l1,l2)
> # create a list of names of contrasts
> clist <- list("front vs rear"=1, "left vs right" = 2)
> # re-fit the anova model
> tyre.aov <- aov(wear~position, data=tyre.data)
> # summary with single degree of freedom contrasts
> summary.aov(tyre.aov, split=list(position=clist))
                          Df Sum Sq Mean Sq F value   Pr(>F)
position                   3 1189.0   396.3  13.739 6.28e-06 ***
  position: front vs rear  1 1006.8  1006.8  34.901 1.42e-06 ***
  position: left vs right  1  180.9   180.9   6.272   0.0176 *
Residuals                 32  923.1    28.8
```

A crucial part of this code is the assignment of contrasts `lmat` to the tyres `weardata$tyre` and then splitting the analysis of variance table with respect to these contrasts.

1.5 The random-effects model

Sometimes the factor of interest has a large number of possible levels (possible choices of treatment groups). In such cases an analyst wants to draw conclusions about the entire population of factor levels (all possible treatment groups). If the analyst randomly selects a of these levels from the population of factor levels, we say that the factor is a ***random factor***. Because the levels of the factor actually used in the experiment are chosen randomly, the conclusions reached are valid for the entire population of factor levels.

This is a very different situation than the one we encountered in the ***fixed-factors*** case (the model we studied so far) in which the conclusions apply only for the factor levels (treatment groups) used in the experiment.

We will assume that the population of factor levels is large enough so that treatment effects α_i could be assumed to be normally distributed random variables with mean zero and unknown variance σ_α^2. They are conveniently called the ***random-effects***.

Definition 1.3:

The random-effects one-way ANOVA model is

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad \alpha_i \overset{ind}{\sim} N(0, \sigma_\alpha^2), \qquad \varepsilon_{ij} \overset{ind}{\sim} N(0, \sigma^2).$$

The σ_α^2 and σ^2 are called ***variance components***.

This model is identical in structure to the fixed-effects model, but the parameters have a different interpretation. The response variable y_{ij} is now distributed as

$$y_{ij} \overset{ind}{\sim} N(\mu, \sigma_\alpha^2 + \sigma^2).$$

However, unlike in the fixed-effect model, the observations y_{ij} are only independent if they come from different factor levels,

$$\mathrm{Cov}(y_{ij}, y_{kl}) = \delta_{ik}\,(\sigma_\alpha^2 + \delta_{jl}\sigma^2).$$

Here δ_{ik} and δ_{jl} are the Kronecker delta symbols:
- $\delta_{ab} = 1$ if $a = b$ and
- $\delta_{ab} = 0$ otherwise.

Since $\alpha_i \overset{ind}{\sim} N(0, \sigma_\alpha^2)$, a meaningful null hypothesis and the alternative hypothesis for the random-effects model are

$$H_0 : \sigma_\alpha^2 = 0 \quad \text{and} \quad H_1 : \sigma_\alpha^2 > 0.$$

If $\sigma_\alpha^2 = 0$, all treatments are identical; but if $\sigma_\alpha^2 \neq 0$, variability exists between treatments. The analysis of variance table and the test statistic remain the same as in the fixed-effects model (except the expectation values of SS_B and SS_E are now different). It can be shown that, in case of a balanced design,

See ***Montgomery***, Section 3.9.2

$$\mathbb{E}(MS_B) = \sigma^2 + n\sigma_\alpha^2 \qquad \mathbb{E}(MS_E) = \sigma^2$$

and so the estimates of σ_α^2 and σ^2 are

$$\hat{\sigma}_\alpha^2 = \frac{MS_B - MS_E}{n}, \qquad \hat{\sigma}^2 = MS_E.$$

In case of an unbalanced design, n should be replaced with

$$\frac{1}{1-a}\left(\sum_{i=1}^{a} n_i - \sum_{i=1}^{a} n_i^2 \Big/ \sum_{i=1}^{a} n_i\right).$$

Example 4: Textile company

A textile company weaves a fabric on a large number of looms. It would like the looms to be homogeneous so that it obtains a fabric of uniform strength. The process engineer suspects that, in addition to the usual variation in strength within samples of fabric from the same loom, there may also be significant variations in strength between looms. To investigate this, she selects four looms at random and makes four strength determinations on the fabric manufactured on each loom. This experiment is run in random order, and the data obtained are shown in the table below:

Loom 1	98	97	99	96
Loom 2	91	90	93	92
Loom 3	96	95	97	95
Loom 4	95	96	99	98

The ANOVA table:

Source of variation	d.o.f.	*SS*	*MS*	*F*
Between groups	3	89.19	29.73	15.68
Residual/Within Groups	12	22.75	1.90	
Total	15	111.94		

The variance components σ^2 and σ^2_α are estimated by $\hat{\sigma}^2 = 1.90$ and

$$\hat{\sigma}^2_\alpha = \frac{29.73 - 1.90}{4} = 6.96.$$

Therefore, the variance of any observation on strength is estimated by

$$\hat{\sigma}^2_y := \mathrm{Var}(y_{ij}) = \hat{\sigma}^2_\alpha + \hat{\sigma}^2 = 6.96 + 1.90 = 8.86.$$

Most of this variability is attributable to differences between looms.

The R code (with a modified output) for this example is:

```
> # experimental data
> y1 <- c(98,97,99,96)
> y2 <- c(91,90,93,92)
> y3 <- c(96,95,97,95)
> y4 <- c(95,96,99,98)
> # prepare the data for anova
> textile <- stack(data.frame(cbind(y1,y2,y3,y4)))
> names(textile) <- c("strength", "loom")
> # fit a random-effects anova model
> library(lme4)
> textile.rand <- lmer(strength ~ (1 | loom), data = textile)
> summary(textile.rand)
Random effects:
 Groups   Name        Variance Std.Dev.
 loom     (Intercept) 6.958    2.638
 Residual             1.896    1.377
Number of obs: 16, groups:  loom, 4

Fixed effects:
            Estimate Std. Error t value
(Intercept)   95.438      1.363   70.01
```

The function `lmer()` from the package `lme4` is used to fit a linear random–effects model. The random effects are declared via the notation `(1 | loom)`. The fixed effect is the overall mean μ. Its estimated value is $\bar{y}_{\bullet\bullet} = 95.438$.

Please let me know if you spot any typos or mistakes.

2 Two-way design

Lecture 10

Suggested Reading:

- ***W. J. Krzanowski***, *An Introduction to Statistical Modelling, Wiley, 2008* Section 4
- ***M. H. Kutner et al.***, *Applied Linear Statistical Models, 5th Edition, McGraw-Hill Irwin, 2005* Chapter 19
- ***D. C. Montgomery***, *Design and Analysis of Experiments, 9th Edition, Wiley, 2017* Chapters 3 and 4

The methods of the previous lecture are appropriate when *no sources of variation other than treatment effects are anticipated*. In particular, it is assumed that the different individuals in each group are subject to the same systematic treatment influence but otherwise any differences between them are purely random effects.

> In many situations it is known beforehand that certain individuals are likely to behave more similarly than others. For example:
>
> - adjacent plots in agricultural field experiments will usually be more alike in response than ones far apart;
> - animals from the same litter will generally show more similarities than ones from different litters;
> - observations made on the same day, or using a particular piece of equipment, may resemble each other more than those made on different days or using different pieces of equipment.

There are two implications for such situations where the behaviour of individual units may be anticipated in part and the units classified accordingly. One concerns the ***design of the study***: it is important to arrange the experiment in such a way as to make optimal use of this partial information, and to prevent the anticipated systematic effects of the units from interfering with the objectives of the study. This aspect is a major consideration in the ***planning of statistical investigations***, and is outside the scope of the present course. We will concentrate instead on the second implication, concerning the ***analysis of the results***: we need to incorporate these anticipated effects within the systematic part of the model so that they are excluded from the estimated experimental error and the analysis is thereby made more precise.

2.1 Complete randomised block design

Consider again Example 1 (Tyre manufacturer), and suppose that we are told by the investigator that the nine trials of the experiment had actually been conducted in three batches of three, and that each batch of three repeats used a different car for the trials.

Given this information we might wish to reconsider our analysis of the trials, because we would (in general) expect a systematic difference between the three cars as regards the amount of tyre wear that they induce (due to different car weights, roadholding, balance, and so on). Each value in the table can now be categorized in two ways:

- according to which tyre position it corresponds to, and
- according to which car it came from.

Both of these categorizations need to be taken into account in the analysis. An alternative way of viewing the data is to say that the responses have been collected into three ***blocks*** (i.e. the cars), and within these blocks each of the former groups (i.e. the tyre positions) is represented by independent observations. This is an example of the ***complete randomised block*** design. We now outline the analysis of such a design, but stress that there is one vital condition that has to be satisfied in this analysis: ***the number of independent trials of each group must be the same in each block***. If this condition is not satisfied then more general methods need to be used.

See ***Krzanowski***, Section 4.5

2.2 *The additive fixed-effects model*

Let us assume that:

- there are a a priori treatment groups forming the primary focus of the investigation, from which random samples of the same size have been taken, and
- each sample is divided into b blocks with n observations, so that the total number of trials is $N = abn$,
- y_{ijk} denotes the observation on the k-th individual in the j-th block from the i-th group, i.e. the (i, j)-th cell.

The observational data can thus be put into the following table:

Group	Block 1			Blocks 2, 3, …				Block b		
1	y_{111}	$\cdots$	y_{11n}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	y_{1b1}	$\cdots$	y_{1bn}
2	y_{211}	$\cdots$	y_{21n}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	y_{2b1}	$\cdots$	y_{2bn}
3	y_{311}	$\cdots$	y_{31n}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	y_{3b1}	$\cdots$	y_{3bn}
$\vdots$	$\vdots$		$\vdots$					$\vdots$		$\vdots$
a	y_{a11}	$\cdots$	y_{a1n}	$\cdots$	$\cdots$	$\cdots$	$\cdots$	y_{ab1}	$\cdots$	y_{abn}

Since we anticipate systematic differences between the blocks, let us assume that:

- each block has a ***block effect*** β_j, to go with
- the ***group effects*** α_i already postulated in the one-way analysis, and
- each block has the ***same effect*** on a trial regardless to which group it belongs to.

These assumptions mean that the block and group effects are ***additive***.

Definition 2.1:

The additive fixed-effects two-way ANOVA model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \overset{ind}{\sim} N(0, \sigma^2) \qquad (7)$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n$.

The standard shorthand notation for sums and averages is as follows.

- The mean of the bn observations in the i-th group is

$$\bar{y}_{i\bullet\bullet} = \frac{1}{bn} y_{i\bullet\bullet} = \frac{1}{bn} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}\,.$$

- The mean of the an observations in the j-th block is

$$\bar{y}_{\bullet j\bullet} = \frac{1}{an} y_{\bullet j\bullet} = \frac{1}{an} \sum_{i=1}^{a} \sum_{k=1}^{n} y_{ijk}\,.$$

- The mean of the n observations in the j-th block from the i-th group is

$$\bar{y}_{ij\bullet} = \frac{1}{n} y_{ij\bullet} = \frac{1}{n} \sum_{k=1}^{n} y_{ijk}\,.$$

- The overall sample mean (the "grand" mean) of all $N = abn$ observations is

$$\bar{y}_{\bullet\bullet\bullet} = \frac{1}{abn} y_{\bullet\bullet\bullet} = \frac{1}{abn} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}\,.$$

Note that

$$\sum_{i=1}^{a} \alpha_i = 0, \qquad \sum_{j=1}^{b} \beta_j = 0.$$

These constraints on the effects follow from their definition,

$$\alpha_i = \bar{\mu}_{i\bullet} - \mu, \qquad \beta_j = \bar{\mu}_{\bullet j} - \mu,$$

where

- $\bar{\mu}_{i\bullet} = \frac{1}{b} \sum_{j=1}^{b} \mu_{ij}$ is the unknown mean of the i-th group,
- $\bar{\mu}_{\bullet j} = \frac{1}{a} \sum_{i=1}^{a} \mu_{ij}$ is the unknown mean of the j-th block,
- μ_{ij} is the unknown mean of the (i, j)-th cell: the j-th block in the i-th group,
- $\mu = \frac{1}{ab} \sum_i^a \sum_j^b \mu_{ij}$ is the unknown overall mean.

Therefore

$$\begin{aligned} \alpha_\bullet &= \sum_{i=1}^{a} \alpha_i = \sum_{i=1}^{a} (\bar{\mu}_{i\bullet} - \mu) \\ &= \frac{a}{a} \cdot \frac{1}{b} \sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij} - a\mu \\ &= a\mu - a\mu = 0, \\ \beta_\bullet &= \sum_{j=1}^{b} \beta_j = \sum_{j=1}^{b} (\bar{\mu}_{\bullet j} - \mu) \\ &= \frac{b}{b} \cdot \frac{1}{a} \sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij} - b\mu \\ &= b\mu - b\mu = 0, \end{aligned}$$

as required.

To estimate parameters μ, α_i, and β_j we can either use the method of maximum likelihood or the method of least-squares. Both methods lead to minimization of the sum or errors squared,

$$S = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \varepsilon_{ijk}^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \mu - \alpha_i - \beta_j)^2.$$

The least-squares estimates of μ, α_i, and β_j are

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}, \qquad \hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \qquad \hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}.$$

The fitted value of the observation y_{ijk} is

$$\hat{y}_{ijk} = \bar{y}_{i\bullet\bullet} + \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}.$$

The residuals are

$$e_{ijk} = y_{ijk} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}.$$

The least-squares estimation goes as follows. First, we compute the partial derivative

$$\begin{aligned} \frac{\partial S}{\partial \mu} &= -2 \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \mu - \alpha_i - \beta_j) \\ &= -2\,(y_{\bullet\bullet\bullet} - N\mu) \end{aligned}$$

since $\alpha_\bullet = \beta_\bullet = 0$. Equating this to zero at $\mu = \hat{\mu}$, we find $\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}$. Then from

$$\begin{aligned} \frac{\partial S}{\partial \alpha_i} &= -2 \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \mu - \alpha_i - \beta_j) \\ &= -2\,(y_{i\bullet\bullet} - bn\mu - bn\alpha_i) \end{aligned}$$

and $\beta_\bullet = 0$, we find $\hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}$. Finally, from

$$\begin{aligned} \frac{\partial S}{\partial \beta_j} &= -2 \sum_{i=1}^{a} \sum_{k=1}^{n} (y_{ijk} - \mu - \alpha_i - \beta_j) \\ &= -2\,(y_{\bullet j\bullet} - an\mu - an\beta_j) \end{aligned}$$

and $\alpha_\bullet = 0$, we find $\hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}$.

It can be shown that residuals satisfy $\mathbb{E}(e_{ijk}) = 0$ and

$$\mathrm{Cov}(e_{ijk}, e_{lst}) = \left(\delta_{il}\delta_{js}\delta_{kt} - \frac{1}{bn}\delta_{il} - \frac{1}{an}\delta_{js} + \frac{1}{abn}\right)\sigma^2.$$

In the same way as in the linear regression, an analysis of normality assumptions for residuals is used for examining the appropriateness of the ANOVA model.

2.3 *Analysis of variance*

The point of interest is to test for systematic differences among the a groups. Hence we want to test the null hypothesis

$$H_0^A : \text{all } \alpha_i = 0 \quad \text{vs.} \quad H_1^A : \text{at least one } \alpha_i \neq 0. \tag{8}$$

Although the block effects β_j are rarely of interest on their own right, and we generally just wish to eliminate them from the experimental error, it is nevertheless possible to formulate an analogous null hypothesis

$$H_0^B : \text{all } \beta_j = 0 \quad \text{vs.} \quad H_1^B : \text{at least one } \beta_j \neq 0. \tag{9}$$

These hypotheses are tested using the AoV method.

The ***total sum of squares*** is defined the usual way, as the sum of differences of observations and the overall mean squared,

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2.$$

The ***between groups sum of squares*** is defined as the sum of differences of group means and the overall mean squared,

$$SS_A = bn\sum_{i=1}^{a}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2.$$

The ***between blocks sum of squares*** is defined as the sum of differences of block means and the overall mean squared,

$$SS_B = an\sum_{j=1}^{b}(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet})^2.$$

The last ingredient is the ***residual (error) sum of squares***,

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2.$$

The AoV identity for the additive two-way design is

$$SS_T = SS_A + SS_B + SS_E.$$

To prove the AoV identity, the group, block and total means, $\bar{y}_{i\bullet\bullet}$, $\bar{y}_{\bullet j\bullet}$ and $\bar{y}_{\bullet\bullet\bullet}$, are added to and subtracted from the difference $y_{ijk} - y_{\bullet\bullet\bullet}$, and grouped in a meaningful way:

$$\begin{aligned} y_{ijk} - \bar{y}_{\bullet\bullet\bullet} &= (\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) + (\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}) \\ &\quad + (y_{ijk} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}). \end{aligned}$$

Squaring both sides of this identity and summing over indices i, j and k, all the "cross" terms vanish, since they all include at least one sum of values about their mean. What remains is the wanted AoV identity.

For computation purposes it is more convenient to use these expressions:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - N\,\bar{y}_{\bullet\bullet\bullet}^2,$$

$$SS_A = bn\sum_{i=1}^{a}\bar{y}_{i\bullet\bullet}^2 - N\,\bar{y}_{\bullet\bullet\bullet}^2,$$

$$SS_B = an\sum_{j=1}^{b}\bar{y}_{\bullet j\bullet}^2 - N\,\bar{y}_{\bullet\bullet\bullet}^2,$$

$$SS_E = SS_T - SS_A - SS_B.$$

Here you should not confuse SS_B with the between groups sum of squares in the one-way design.

Define the ***mean squares***

$$MS_A = \frac{SS_A}{a-1}, \qquad MS_B = \frac{SS_B}{b-1}, \qquad MS_E = \frac{SS_E}{N-a-b+1}.$$

The quantities $\nu_A = a-1$, $\nu_B = b-1$ and $\nu_E = N-b-a-1$ are the degrees of freedom of SS_A, SS_B and SS_E, respectively. It can be shown that

To see that $\nu_E = N-a-b-1$ we need to subtract ν_A and ν_B from $\nu_T = N-1$:

$$\begin{aligned}\nu_E &= \nu_T - \nu_A - \nu_B \\ &= (N-1)-(a-1)-(b-1) \\ &= N-a-b+1.\end{aligned}$$

$$\mathbb{E}(MS_A) = \sigma^2 + \frac{b}{a-1}\sum_{i=1}^{a}\alpha_i^2,$$

$$\mathbb{E}(MS_B) = \sigma^2 + \frac{a}{b-1}\sum_{j=1}^{b}\beta_j^2, \qquad \mathbb{E}(MS_E) = \sigma^2.$$

Given the additive two-way model (7) and the null hypotheses (8) and (9), the following distributional results can be derived.

Under the null hypotheses, the normality and independence of the errors ε_{ijk} imply

$$\frac{SS_A}{\sigma^2} \sim \chi^2_{a-1}, \qquad \frac{SS_B}{\sigma^2} \sim \chi^2_{b-1}, \qquad \frac{SS_E}{\sigma^2} \sim \chi^2_{N-a-b+1}.$$

Moreover, SS_A/σ^2 and SS_E/σ^2, and SS_B/σ^2 and SS_E/σ^2 are pairwise independent random variables. Consequently

$$F_A = \frac{MS_A}{MS_E} \sim F_{a-1,N-a-b+1}, \qquad F_B = \frac{MS_B}{MS_E} \sim F_{b-1,N-a-b+1}$$

are F-distributed random variables.

The ratios F_A and F_B can be used to conduct formal tests of the hypotheses H_0^A and H_0^B in the same manner as all previous F-tests. The null hypothesis in each case is rejected if the calculated ratio exceeds the corresponding critical value at the chosen significance level. The standard additive two-way ANOVA table is shown below.

Source of variation	d.o.f.	SS	MS	F
Between groups	$a-1$	SS_A	MS_A	$F_A = \frac{MS_A}{MS_E}$
Between blocks	$b-1$	SS_B	MS_B	$F_B = \frac{MS_B}{MS_E}$
Residual/Error	$N-a-b+1$	SS_E	MS_E	
Total	$N-1$	SS_T		

Table 2: Two-way ANOVA table. Between groups is also called Factor A, Between blocks is also called Factor B.

2.4 *Linear contrasts*

The between groups sum of squares SS_A can once again be partitioned into linear contrasts among the group totals $y_{i\bullet\bullet}$ representing a single degree of freedom each. The approach used in the one-way design applies exactly in the same way, except that $y_{i\bullet}$ are replaced with $y_{i\bullet\bullet}$ and all n_i are replaced with bn for each D_w.

The partition is now

$$SS_A = \frac{z_1^2}{D_1} + \frac{z_2^2}{D_2} + \ldots + \frac{z_{a-1}^2}{D_{a-1}}$$

where

$$z_w = l_{w1}y_{1\bullet\bullet} + l_{w2}y_{2\bullet\bullet} + \ldots + l_{wa}y_{a\bullet\bullet}$$
$$D_w = bn(l_{w1}^2 + l_{w2}^2 + \ldots + l_{wa}^2)$$

Analogous contrasts could, if desired, be defined for the block totals $y_{\bullet j \bullet}$, and the between blocks sum of squares SS_B could be partitioned into single degree of freedom contrasts. However, in most applications the purpose of incorporating blocks in the analysis is to make comparisons between the groups more precise, and investigation of differences between blocks is rarely of interest in its own right. Consequently partitioning SS_B in this way is rarely undertaken.

If $b = 2$, then SS_B is itself a contrast between the two blocks, and no follow up analysis is needed.

Example 5: Tyre manufacturer (continued)

Suppose that indeed three different cars, A, B, and C, were used for the tyre wear trials:

Position	Car A			Car B			Car C		
FL	20.935	19.013	20.332	17.123	15.919	15.285	29.590	28.092	28.304
FR	18.279	21.200	19.389	14.815	11.280	12.153	19.973	20.096	20.477
RL	28.535	27.998	30.073	37.227	38.853	40.017	30.529	29.177	30.795
RR	20.182	18.792	19.203	34.340	34.707	36.307	29.023	28.176	28.701

We wish to eliminate any systematic "car" effect from comparisons among the four tyre positions, so we must use the randomised block analysis. We have $a = 4$ groups, $b = 3$ blocks, and $n = 3$ trials per each car/position combination.

The total sum of squares, $SS_T = 2112.132$, and the between groups sum of squares $SS_A = 1189.113$ (in one-way design it was denoted SS_B) remain the same as well. The new calculations needed for the two-way analysis is the between blocks sum of squares,

$$SS_B = 12 \times (21.994^2 + 25.669^2 + 26.911^2) - 36 \times 24.858^2 = 156.900,$$

and the residual sum of squares

$$SS_E = 2112.132 - 1189.113 - 156.900 = 766.119.$$

The resulting two-way ANOVA table is:

Source of variation	d.o.f.	SS	MS	F
Between groups	3	1189.113	396.371	15.521
Between blocks	2	156.900	78.450	3.072
Residual/Error	30	766.119	25.537	
Total	35	2112.132		

Conclusions:

- Taking into account the block effects has reduced MS_E from 28.844 to 25.537, and this has sharpened up the F-test for differences between positions by increasing the calculated ratio F_A from 13.742 to 15.521.
- Two degrees of freedom attributable to blocks have been removed from those available for error, so the latter have fallen from the previous 32 to 30.

- The critical values of the $F_{2,30}$ distribution are 2.49 (for 10% significance) and 3.32 (for 5% significance). Thus the calculated ratio $F_B = 3.072$ indicates that there is some effect due to the different cars used, albeit not a very strong one. Nevertheless, there are sufficient systematic differences in tyre wear between the different cars to justify the randomised blocks analysis.
- Since twelve observations make up each block, the standard error of a block mean is $\sqrt{25.537/12}$ and that of a treatment mean is $\sqrt{25.537/9}$. The sum of squares between groups breaks down into single degree of freedom contrasts exactly in the same way as before, but the new error mean square enhances the F-ratios for these contrasts slightly.

The R code for this example is:

```
> # read raw data: tyre.csv
> tyres.raw <- read.csv(file.choose())
> # inspect the raw data: 9 obs. of 5 variables
> str(tyres.raw)
> # stack all but first column, and name the resulting two columns
> tyres.data <- stack(tyres.raw[,-1])
> names(tyres.data) <- c("wear", "position")
> # create a new column "car" by replicating the Car column 4 times
> tyres.data$car <- rep(tyres.raw$Car,4)
> # inspect the final data: 36 obs. of 3 variables
> str(tyres.data)
> # fit an additive two-way anova model
> tyres.aov <- aov(wear~position+car, data=tyres.data)
> summary(tyres.aov)
```

2.5 *The interacting fixed-effects model*

The fundamental assumption in the additive two-way model is that each block has the same effect on every group. However, such a situation is often an oversimplification.

For example, consider grouping subjects in a psychological experiment by age. Suppose that the experiment involves reacting to various stimuli (e.g. different coloured lights), the response variable is the speed of the subject's reaction, and we treat the different age groups as blocks for the purpose of the analysis. It is certainly true that different age groups may react differently to the stimuli (generally, the older the subject the slower will be their reaction), so blocking subjects by age is a sensible measure that will improve precision of analysis. However, there is no reason to expect that the "age differences" will be constant over all stimuli. Some colours will be more visible than others, and the slowing reaction with age will be less pronounced for these colours than for the others.

In similar vein, we might expect differential effects of litters in responses to diets in animal feed trials, differential effects of panellists in food-tasting experiments, and so on. In order to take such situations into accounts we write the "general" model,

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \overset{ind}{\sim} N(0, \sigma^2),$$

where μ_{ij} is the unknown mean of responses in the (i,j)-th cell: the j-th block of the i-th group. If group and block effects are independent, then

$$\mu_{ij} = \mu + \alpha_i + \beta_j,$$

and we obtain the already known additive two-way model. Plotting such μ_{ij} will yield ***parallel lines***, as shown in Figure 2.1 (a), where $a = 3$ and $b = 3$.

However, if the effects are not independent, then the plots will depart from parallelism. Two possible types of departure are divergence of lines, as in Figure 2.1 (b) where the differences between groups are always in the same direction but increase progressively from block 1 to block 3, or crossing–over of lines, as in Figure 2.1 (c) where the pattern of differences between groups changes between blocks. For such cases we write

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}\,,$$

where γ_{ij}, often denoted by $(\alpha\beta)_{ij}$, is called ***interaction effect***. It measures the extent of departure from additivity. We say that the groups ***interact*** with the blocks, or that there is a ***group by block interaction***.

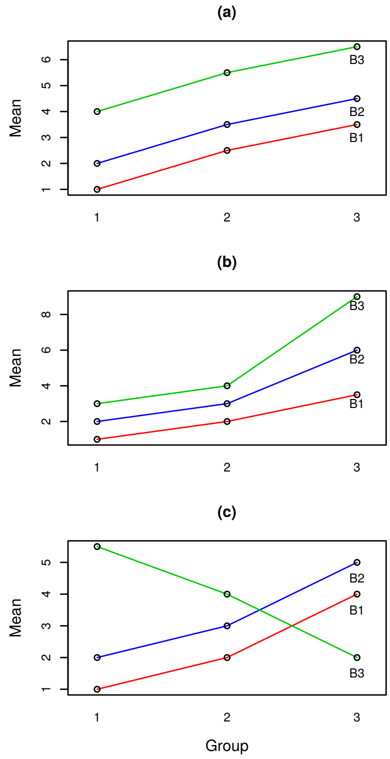


Figure 2.1: Possible patters of means of responses between groups over blocks.

Definition 2.2:

The interacting fixed-effects two-way ANOVA model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \overset{ind}{\sim} N(0, \sigma^2) \qquad (10)$$

where $i = 1, \ldots, a$, $j = 1, \ldots, b$ and $k = 1, \ldots, n$.

The least-squares estimates of μ, α_i, β_j and γ_{ij} are

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}, \qquad \hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \qquad \hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet},$$
$$\hat{\gamma}_{ij} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}.$$

The fitted value of the observation y_{ijk} is

$$\hat{y}_{ijk} = \bar{y}_{ij\bullet}.$$

The residuals are

$$e_{ijk} = y_{ijk} - \bar{y}_{ij\bullet}.$$

It is possible to show that

$$\mathbb{E}(e_{ijk}) = 0\,, \qquad \operatorname{Cov}(e_{ijk}, e_{lst}) = \delta_{il}\delta_{js}\left(\delta_{kt} - \frac{1}{n}\right)\sigma^2.$$

Note that

$$\sum_{i=1}^{a} \alpha_i = 0, \qquad \sum_{i=j}^{b} \alpha_i = 0,$$
$$\sum_{i=1}^{a} \gamma_{ij} = 0, \qquad \sum_{j=1}^{b} \gamma_{ij} = 0.$$

This was already shown for α_i and β_j. The γ_{ij} are defined as

$$\gamma_{ij} = \mu_{ij} - \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet j} + \mu.$$

We have

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

Therefore

$$\sum_{i=1}^{a} \gamma_{ij} = \sum_{i=1}^{a} (\mu_{ij} - \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet j} + \mu)$$
$$= \mu_{\bullet j} - a\bar{\mu}_{\bullet\bullet} - a\bar{\mu}_{\bullet j} + a\mu = 0.$$

In a similar way,

$$\sum_{j=1}^{b} \gamma_{ij} = \mu_{i\bullet} - b\bar{\mu}_{i\bullet} - b\bar{\mu}_{\bullet\bullet} + b\mu = 0\,.$$

2.6 *Analysis of variance in the presence of an interaction*

Plotting the $\hat{\mu}_{ij}$ in the manner of Figure 2.1 may show up the presence of an interaction. However, sampling variance will now play a part, and considerable visual deviation from parallelism of lines can still be consistent with additivity of effects. We cannot therefore rely on a simple visual inspection, and need a formal test for the presence of an interaction.

We will apply the analysis of variance method to test the null hypothesis

$$H_0^{AB} : \text{all } \gamma_{ij} = 0 \quad \text{vs.} \quad H_0^{AB} : \text{at least one } \gamma_{ij} \neq 0. \qquad (11)$$

Assuming that the number of trials in each cell is $n > 1$, we can write the analysis of variance identity as

See ***Kutner et al.***, Section 20 for $n = 1$ case

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E,$$

where SS_T, SS_A and SS_B are given by the same formulas as before, the SS_E is the residual sum of squares in the presence of an interaction,

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij\bullet})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij\bullet}^2$$

The remaining component is the interaction sum of squares,

$$\begin{aligned} SS_{AB} &= SS_T - SS_A - SS_B - SS_E \\ &= SS_E^{add} - SS_E \\ &= n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet})^2 \end{aligned}$$

For computational purposes it is more convenient to use this expression:

$$SS_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}_{ij\bullet}^2 - N\bar{y}_{\bullet\bullet\bullet}^2 - SS_A - SS_B$$

where SS_E^{add} is the residual sum of squares in the additive two-way model, (2.3).

Define the ***mean squares***

$$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}, \qquad MS_E = \frac{SS_E}{N - ab}.$$

The quantities $\nu_{AB} = (a-1)(b-1)$ and $\nu_E = N - ab$ are degrees of freedom of SS_{AB} and SS_E, respectively. It can be shown that, in the presence of an interaction,

The ν_E is the number of linearly independent residuals in the model. The ν_{AB} is obtained by

$$\begin{aligned} \nu_{AB} &= \nu_T - \nu_A - \nu_B - \nu_E \\ &= N - 1 - (a-1) - (b-1) - (N - ab) \\ &= (a-1)(b-1). \end{aligned}$$

$$\mathbb{E}(MS_A) = \sigma^2 + \frac{nb}{a-1}\sum_{i=1}^{a}\alpha_i^2, \qquad \mathbb{E}(MS_B) = \sigma^2 + \frac{na}{b-1}\sum_{i=1}^{a}\beta_j^2,$$

$$\mathbb{E}(MS_{AB}) = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\gamma_{ij}^2, \qquad \mathbb{E}(MS_E) = \sigma^2.$$

Given the interacting two-way model (10) and the null hypotheses (8), (9), and (11), the following distributional results can be derived.

The null hypotheses are

$$H_0^A : \text{all } \alpha_i = 0, \qquad H_0^B : \text{all } \beta_j = 0, \qquad H_0^{AB} : \text{all } \gamma_{ij} = 0.$$

Under the null hypotheses, the normality and independence of the errors ε_{ijk} imply that

$$\frac{SS_A}{\sigma^2} \sim \chi^2_{a-1}, \qquad \frac{SS_B}{\sigma^2} \sim \chi^2_{b-1},$$
$$\frac{SS_{AB}}{\sigma^2} \sim \chi^2_{(a-1)(b-1)}, \qquad \frac{SS_E}{\sigma^2} \sim \chi^2_{N-ab}.$$

Moreover, SS_A/σ^2 and SS_E/σ^2, and SS_B/σ^2 and SS_E/σ^2, and SS_{AB}/σ^2 and SS_E/σ^2 are pairwise independent random variables. Consequently

$$F_A = \frac{MS_A}{MS_E} \sim F_{a-1,N-ab} \qquad F_B = \frac{MS_B}{MS_E} \sim F_{b-1,N-ab}$$
$$F_{AB} = \frac{MS_{AB}}{MS_E} \sim F_{(a-1)(b-1),N-ab}$$

are F-distributed random variables.

The ratios F_A, F_B, and F_{AB} can thus be used to conduct formal tests of the hypotheses H_0^A, H_0^B, and H_0^{AB} in the usual way. Rejecting the H_0^{AB} establishes non-additivity of effects. In this case the group means vary with blocks, so it does not make sense to average over blocks when reporting group effects and we need to report conclusions separately for the different blocks. The resulting ANOVA table is shown below.

Source of variation	d.o.f.	SS	MS	F
Between groups	$a-1$	SS_A	MS_A	$F_A = \frac{MS_A}{MS_E}$
Between blocks	$b-1$	SS_B	MS_B	$F_B = \frac{MS_B}{MS_E}$
Interaction	$(a-1)(b-1)$	SS_{AB}	MS_{AB}	$F_{AB} = \frac{MS_{AB}}{MS_E}$
Residual/Error	$N-ab$	SS_E	MS_E	
Total	$N-1$	SS_T		

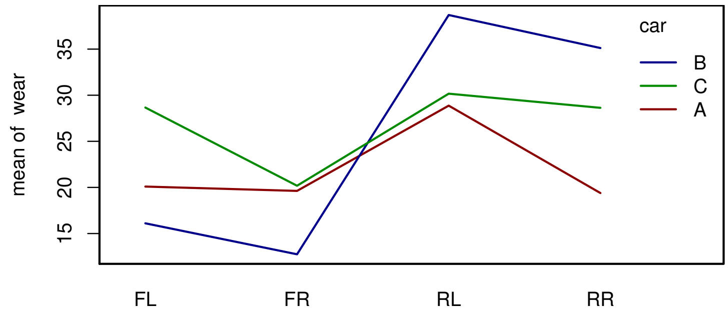
Table 3: Two-way ANOVA table in the presence of an interaction.

Example 6: Tyre manufacturer (continued)

The following are the *means* $\hat{\mu}_{ij}$ of the three observations in each position/car combination:

Position	Car A	Car B	Car C
FL	20.093	16.109	28.662
FR	19.623	12.749	20.182
RL	28.869	38.699	30.167
RR	19.392	35.118	28.633

The interaction plot below shows a strong evidence for an interaction between groups and blocks, and we should conduct a two-way analysis with an interaction.



We already know SS_A, SS_B and SS_T, which remain the same. We only need to find the "new" $SS_E = 27.584$ and $SS_{AB} = 738.535$ and draw up the ANOVA table:

Source of variation	d.o.f.	SS	MS	F
Between groups	3	1189.113	396.371	344.970
Between blocks	2	156.900	78.450	68.277
Interaction	6	738.535	123.089	107.127
Residual/Error	24	27.584	1.149	
Total	35	2112.132		

Conclusions:

- There has been a dramatic reduction in the residual sum of squares, MS_E, from its value in the additive case (25.537). This is because most of the previously unexplained residual sum of squares is now attributed to the interaction sum of squares.
- All three calculated F-ratios are very large and the corresponding effects ("Tyre Positions", "Cars" and "Position by Car interaction") are highly significant.
- The highly significant interaction means that the average responses vary considerably between the 12 position/car combinations, as can be seen clearly in the table of totals.
- This ANOVA table is thus the appropriate summary to quote for any conclusions, as it would be rather misleading to average such diverse values over either positions or cars.

The R code for this example is (continuing from Example 5):

```
> # interaction plot. syntax: group, block, response
> with(tyres, interaction.plot(position, car, wear, lty=1, col = c(2,3,4)))
> # 2-way anova with interaction
> tyres.aov.int <- aov(wear~position*car, data=tyres)
> summary(tyres.aov.int)
```

*For the random-effects two-way model see **Kutner et al.**, Section 25.*

Please let me know if you spot any typos or mistakes.

3 Nested two-way design

In the two-way design considered so far every group appears in a combination with each block. We say that the two factors, groups and blocks, are crossed. A different situation occurs when a subsampling is involved: each group has its own blocks that have nothing to do with blocks from the other groups. In such a situation we say that ***blocks are nested within groups***.

For example:

- A company that purchases its raw material from three different suppliers, wishes to determine whether the purity of the material is the same from each supplier. There are four batches of raw material available from each supplier, and three determinations of purity are taken from each batch.
- A soft drink producer wishes to determine whether the three different bottling machines used in a bottling plant give the same output. Twelve operators are employed on the plant. Four operators are assigned to each machine and work six-hour shifts each. The production data is then collected over a week.

We will focus on the ***balanced nested design*** when the number of replications is equal for all blocks, and the number of blocks is equal for all groups. In general, these numbers can be different, resulting in a more involved calculations that are beyond the scope of the present course.

3.1 The nested fixed-effects model

Let us assume that:

- there are a treatment groups forming the primary focus of the investigation;
- there are b blocks nested within each group, so that the total number of blocks is ab;
- there are n trials per each block, so that the total number of trials is $N = abn$;
- y_{ijk} denotes the observation on the k-th individual in the j-th block within the i-th group.

The observational data can thus be put into the following table:

Group 1						Group a		
Block 1	...	Block b	...	...	...	Block 1	...	Block b
y_{111}	$\cdots$	y_{1b1}	$\cdots$	$\cdots$	$\cdots$	y_{a11}	$\cdots$	y_{ab1}
$\vdots$		$\vdots$				$\vdots$		$\vdots$
y_{11n}	$\cdots$	y_{1bn}	$\cdots$	$\cdots$	$\cdots$	y_{a1n}	$\cdots$	y_{abn}

Lecture 11

Suggested Reading:

- ***W. J. Krzanowski***, *An Introduction to Statistical Modelling, Wiley, 2008* Section 4.4
- ***M. H. Kutner et al.***, *Applied Linear Statistical Models, 5th Edition, McGraw-Hill Irwin, 2005* Chapter 26
- ***D. C. Montgomery***, *Design and Analysis of Experiments, 9th Edition, Wiley, 2017* Chapter 14

There are two types of effects in a nested model:

- each group has a ***group effect*** $\boldsymbol{\alpha_i}$, and
- each block has a ***block effect*** $\boldsymbol{\beta_{j(i)}}$ nested within the i-th group.

The notation $j(i)$ is used to emphasize that blocks are nested within groups. In a similar fashion, the error term will be denoted by $\varepsilon_{(ij)k}$. There are no interaction effects since blocks are nested within groups, not crossed with them.

Definition 3.1:

The nested fixed-effects two-way ANOVA model is

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{(ij)k}, \qquad \varepsilon_{(ij)k} \overset{ind}{\sim} N(0, \sigma^2), \qquad (12)$$

where $i = 1, \dots, a$, $j = 1, \dots, b$, and $k = 1, \dots, n$.

Note that

$$\sum_{i=1}^{a} \alpha_i = 0, \qquad \sum_{j=1}^{b} \beta_{j(i)} = 0.$$

The least-squares estimates of μ, α_i, and $\beta_{j(i)}$ are

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}, \qquad \hat{\alpha}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}, \qquad \hat{\beta}_{j(i)} = \bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet}.$$

The fitted values are

$$\hat{y}_{ijk} = \bar{y}_{ij\bullet}.$$

The residuals are

$$e_{ijk} = y_{ijk} - \bar{y}_{ij\bullet}.$$

It can be shown that residuals satisfy

$$\mathbb{E}(e_{ijk}) = 0, \qquad \mathrm{Cov}(e_{ijk}, e_{lst}) = \delta_{il}\delta_{js}\left(\delta_{kt} - \frac{1}{n}\right)\sigma^2.$$

3.2 *Analysis of variance*

We want to test the null hypothesis

$$H_0^A : \text{all } \alpha_i = 0 \quad \text{vs.} \quad H_1^A : \text{at least one } \alpha_i \neq 0. \qquad (13)$$

The effects $\beta_{j(i)}$ are not of primary interest, nevertheless it is possible test the null hypothesis

$$H_0^{B(A)} : \text{all } \beta_{j(i)} = 0 \quad \text{vs.} \quad H_1^{B(A)} : \text{at least one } \beta_{j(i)} \neq 0. \qquad (14)$$

The nested two-way design can be viewed as a series of one-way investigations nested within groups. Hence the null hypothesis

$$H_0^B : \text{all } \beta_{j(i)} = 0 \text{ for a fixed } i$$

can be tested using the methods studied in Lecture 9.

As usual, these hypotheses are tested using the AoV method.

The ***total sum of squares*** is defined the usual way, as the sum of differences of observations and the overall mean squared,

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r} (y_{ijk} - \bar{y}_{\bullet\bullet\bullet})^2.$$

The ***between groups sum of squares*** is also is defined in the usual way, as the sum of differences of group means and the overall mean squared,

$$SS_A = bn\sum_{i=1}^{a}(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})^2.$$

The ***between nested blocks sum of squares*** is defined as the sum of differences of block means within a group and the group means squared

$$SS_{B(A)} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet})^2.$$

The ***the residual (error) sum of squares*** is defined by

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij\bullet})^2.$$

The AoV identity for the nested design is

$$SS_T = SS_A + SS_{B(A)} + SS_E,$$

Define the ***mean squares***

$$MS_A = \frac{SS_A}{a-1}, \quad MS_{B(A)} = \frac{SS_{B(A)}}{a(b-1)}, \quad MS_E = \frac{SS_E}{ab(n-1)}.$$

The quantities $\nu_A = a - 1$, $\nu_{B(A)} = a(b-1)$, and $\nu_E = ab(n-1)$ are the degrees of freedom of SS_A, $SS_{B(A)}$, and SS_E, respectively. It can be shown that

$$\mathbb{E}(MS_A) = \sigma^2 + \frac{bn}{a-1}\sum_{i=1}^{a}\alpha_i^2,$$

$$\mathbb{E}(MS_{B(A)}) = \sigma^2 + \frac{n}{b(a-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}^2, \qquad \mathbb{E}(MS_E) = \sigma^2.$$

Given the nested two-way model (12) and the null hypotheses (13) and (14), the following distributional results can be derived.

Under the null hypotheses, the normality and independence of the errors $\varepsilon_{(ij)k}$ imply

$$\frac{SS_A}{\sigma^2} \sim \chi^2_{a-1}, \qquad \frac{SS_{B(A)}}{\sigma^2} \sim \chi^2_{a(b-1)}, \qquad \frac{SS_E}{\sigma^2} \sim \chi^2_{ab(n-1)}.$$

Moreover, SS_A/σ^2 and SS_E/σ^2, and $SS_{B(A)}/\sigma^2$ and SS_E/σ^2 are pairwise independent random variables. Consequently

$$F_A = \frac{MS_A}{MS_E} \sim F_{a-1,\,ab(n-1)},$$

$$F_{B(A)} = \frac{MS_{B(A)}}{MS_E} \sim F_{a(b-1),\,ab(n-1)}$$

are F-distributed random variables.

To prove the AoV identity we add and subtract the group means, $y_{i\bullet\bullet}$, and block means, $y_{ij\bullet}$, to the total deviation, and suitably bracketing the resulting expression,

$$\underbrace{y_{ijk} - \bar{y}_{\bullet\bullet\bullet}}_{\text{total deviation}} = \underbrace{(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet})}_{A\text{ main effect}} + \underbrace{(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet})}_{\substack{\text{specific } B \text{ effect}\\ \text{when } A \text{ is}\\ \text{at the } i\text{-th level}}} + \underbrace{(y_{ijk} - \bar{y}_{ij\bullet})}_{\text{residual}},$$

Squaring both sides of the identity and summing over all observations, all cross products vanish because they contain at least one sum of values about their mean. What remains is the wanted AoV identity.

For computation purposes it is more convenient to use these expressions:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - N\bar{y}^2_{\bullet\bullet\bullet},$$

$$SS_A = bn\sum_{i=1}^{a}\bar{y}^2_{i\bullet\bullet} - N\bar{y}^2_{\bullet\bullet\bullet},$$

$$SS_{B(A)} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}^2_{ij\bullet} - bn\sum_{i=1}^{a}\bar{y}^2_{i\bullet\bullet},$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - n\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{y}^2_{ij\bullet}.$$

The SS_A has its usual number of degrees of freedom, $\nu_A = a - 1$. The $SS_{B(A)}$ is a sum of a separate sums of squares, each having $b-1$ degrees of freedom, giving $\nu_{B(A)} = a(b-1)$ in total. There are $\nu_T = abn - 1$ degrees of freedom in total. Hence

$$\nu_E = \nu_T - \nu_A - \nu_{B(A)} = ab(n-1)$$

are the residual degrees of freedom.

The ratios F_A and $F_{B(A)}$ can be used to test the null hypotheses H_0^A and $H_0^{B(A)}$ in the same manner as all previous F-tests. The nested ANOVA table is shown below.

Source of variation	d.o.f.	SS	MS	F
Between groups	$a-1$	SS_A	MS_A	$F_A = \frac{MS_A}{MS_E}$
Between nested blocks	$a(b-1)$	$SS_{B(A)}$	$MS_{B(A)}$	$F_{B(A)} = \frac{MS_{B(A)}}{MS_E}$
Residual/Error	$ab(n-1)$	SS_E	MS_E	
Total	$N-1$	SS_T		

Table 4: Nested ANOVA table.

If it is concluded that group effects are present, it is often desired to ascertain whether they are present in all groups or only in some. (In some cases one may wish to proceed immediately to this analysis.) Suitable linear contrasts can be employed for such a follow up analysis.

Example 7: Training schools

A large manufacturing company operates three regional training schools for mechanics, one in each of its operating districts. The schools have two instructors each, who teach classes of about 15 mechanics in three-week sessions. The company was concerned about the effect of school (group) and instructor (block) on the learning achieved. To investigate these effects, classes in each district were formed in the usual way and then randomly assigned to one of the two instructors in the school. This was done for two sessions, and at the end of each session a suitable summary measure of learning for the class was obtained. The results are presented in the table below:

School	Instructor 1		Instructor 2	
Leeds	25	29	14	11
Manchester	11	6	22	18
Liverpool	17	20	5	2

The layout of this table appears identical to an ordinary two-factor investigation, with two observations per cell. However, the study is not an ordinary two-factor study. The reason is that the instructors in the Leeds school did not also teach in the other two schools, and similarly for the other instructors. In fact, six different instructors were involved:

School	Instructor 1		Instructor 2		Instructor 3		Instructor 4		Instructor 5		Instructor 6	
Leeds	25	29	14	11								
Manchester					11	6	22	18				
Liverpool									17	20	5	2

It is clear from this table that instructors are nested within schools. To analyse the instructor and school effects formally, we begin by computing the relevant sums of squares:

$$SS_T = 766.00, \qquad SS_A = 156.50, \qquad SS_{B(A)} = 567.50, \qquad SS_E = 42.00.$$

The resulting nested ANOVA table is:

Source of variation	d.o.f.	SS	MS	F
Between groups	2	156.50	78.25	11.18
Between nested blocks	3	567.50	189.17	27.02
Residual/Error	6	42.00		
Total	11	766.00		

Assuming significance level $\alpha = 5\%$, the critical F-ratios are $F_{2,6,0.05} = 5.14$ and $F_{3,6,0.05} = 4.76$, for F_A and $F_{B(A)}$, respectively. Since the calculated F-ratios are greater that the critical values, we conclude that the three schools differ in mean learning, and that instructors within at least one school differ in terms of mean learning.

The R code for this example is:

```
> # load data manipulation libraries
> library(plyr)
> library(dplyr)
> # experimental data
> y <- c(25,29,11,6,17,20,14,11,22,18,5,2)
> a <- as.factor( c(1,1,2,2,3,3,1,1,2,2,3,3) ) %>% revalue(c("1"="Leeds", "2"=
    "Manchester", "3"="Liverpool"))
> b <- as.factor( c(rep(1,6), rep(2,6)) ) %>% revalue(c("1"="Instructor 1", "2
    "="Instructor 2"))
> # dataframe
> training <- data.frame(learning = y, school = a, instructor = b)
> # inspect the datframe
> str(training)
> # fit a nested two-way model: the "/" defines the nested variable
> training.2way.nested <- aov(learning~school/instructor, data=training)
> summary(training.2way.nested)
```

*For the nested random-effects two-way model see **Kutner et al.**, Section 26.*

Please let me know if you spot any typos or mistakes.