Part II

ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS

The analysis of variance (**ANOVA**) is an statistical procedure which allows us to find possible differences in the expected value of a random dependent variable Y with normal distribution for different groups of experimental units in the same population. These groups are defined by the possible levels of one or more **factors** that can be considered as categorical independent variables (A, B, C,...). Each combination of factor levels is known as a **treatment** in the ANOVA and the purpose is to detect any significant differences between treatments and, if necessary, quantify the magnitude of the differences. Then an **statistical experiment** is a scientifically planned work to meet this goal.

The methodology of experimental design is based on experimentation and its associated variability, because if we repeat an experiment under identical conditions, the obtained results are likely to show some variability. Thus, the **design of experiments** studies how to vary experimental conditions in order to increase the probability of detecting significant changes in the dependent variable, also called response variable.

The reasons for experimentation may be different. For example: to discover possible causes of variation in the response variable; or find experimental conditions under which an optimum value for the response variable is achieved; or compare responses at different levels of observation of controlled variables; or even obtain a statistical-mathematical model to make predictions of future responses.

The results of any experiment are subjected to three types of variability which must be distinguished:

- **Planned systematic variability**. Caused by the different experimental conditions imposed on the design. It is the type of variability that we try to identify with the design.
- **Random variability**. It is an unpredictable and unavoidable variability due to factors beyond our control. If the experiment was well designed it can be measured and used to draw conclusions and make predictions on the response variable.
- **Unplanned systematic variability**. It is due to unknown causes and unplanned. There are two basic strategies to avoid the presence of this type of variability: randomization and blocking.

Next we present the four basic principles in the design of experiments:

1. **Replication**. It is the use of several experimental units for each of the treatments in the experiment. This principle allows us to obtain an estimation of the random variability, which will be necessary in the further analysis, and, moreover, estimate the effect of each treatment more accurately.
2. **Randomization**. It is to assign levels of the factors to the experimental units at random and also the random selection of the order in which measurements of the response variable are made. This principle transforms the unplanned systematic variability in random variability, prevents the occurrence of systematic errors, avoids the dependence between observations ensuring the independence of the errors in the model and, finally, provides unbiased estimates for the random variability and the effects of the treatments.
3. **Blocking**. If there is a great heterogeneity in the experimental units, they should be divided into groups called blocks so that the observations made in each block are under experimental conditions

as similar as possible. Then all the treatments are used in each block. This principle let us to transform the unplanned systematic variability in planned systematic variability.

4. **Factorization**. It is to cross all the levels of the factors in all possible combinations. This principle let us to detect the existence of interactions between different factors and it is a more efficient strategy than the analysis of the influence of a factor with fixed levels for the other factors.

In general, the steps to be followed in the design and analysis of any statistical experiment are as follows:

1. Define the objectives of the experiment and develop a comprehensive list of questions that must be answered.
2. Identify all possible sources of variation: **treatment factors** to be taken into account and other "nuisance" factors that are not directly relevant but are contemplated to reduce unplanned variability. The treatment factors may be qualitative or quantitative and the levels of the factors to be used must also be set. If a factor is quantitative, it is desirable that the levels are equally spaced. Sometimes a **control treatment** is necessary to be used as a reference for assessing the effect of all the other treatments.
3. Define the **experimental units**, that is, the experimental material (individuals, trees, plots, etc.) that apply to the different levels of the factors and on which it will assess the response variable. They should be a representative sample of the target population of the study. As we have said before, if there is a great heterogeneity between them, it may be desirable to add a block factor with homogeneous experimental units in each level.
4. Choose a mapping rule of the experimental units to the treatments. If you choose a standard design (as you will learn in this course) this rule will be defined by the design.
5. Formulate the statistical model by a mathematical equation with the parameters to be estimated, as we will do in each of the designs.
6. Specify the steps in the statistical analysis: the estimates to be calculated, the contrasts to be performed, the confidence intervals to be evaluated, the degree of fit of the model and the compliance of the assumptions set out in the model.
7. Determine the **sample size** for each of the treatments, that is, the number of replicates for each of them. To choose it beforehand, an estimator of the random variability is required, which is not generally available. Therefore sometimes a pilot experiment with a small number of observations is previously executed to obtain this estimator.
8. Run the experiment, randomizing, if it is possible, the order in which the treatments are used and the order in which the measurements of the response variable are made.
9. Perform the **statistical analysis** of the obtained data with the proposed model, and answer the questions previously raised in the experiment.

In the next chapters we will study some ANOVA models and the basic experimental designs that can be analyzed with each of them.

Chapter 1

ONE-WAY ANOVA

Let us suppose that y_{ij} with $i = 1, ..., a$ and $j = 1, ..., n$ are the observed values for a independent random samples of the dependent variable Y. That is, we have a dependent variable Y and a factor A with a levels. Moreover, we suppose that $Y_{ij} \rightsquigarrow N(\mu_i, \sigma)$ and the purpose is to solve a hypothesis test with $H_o : \mu_i = \mu_{i'}$ for all i, i' versus $H_1 : \mu_i \neq \mu_{i'}$ at least for one pair i, i'.

The model can be formulated as $E(Y_{ij}) = \mu_i$ and, therefore, $Y_{ij} = \mu_i + \varepsilon_{ij}$ with $\varepsilon_{ij} \rightsquigarrow N(0, \sigma)$ and ε_{ij} independent of $\varepsilon_{i'j'}$ for any values i, i', j, j'. Taking into account that the null hypothesis H_o does not specify a particular value for the expected values μ_i, we consider a new parameter $\mu = \frac{1}{a} \sum_i \mu_i$ such that, if H_o is true, then $\mu_i = \mu$ for all i. Then, $\alpha_i = \mu_i - \mu$ can be considered as the specific effect for the level i of the factor A on the overall expected value μ, with $\sum_i \alpha_i = 0$. As a consequence, the model can be alternatively formulated as $E(Y_{ij}) = \mu + \alpha_i$ and, therefore, $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ with $\sum_i \alpha_i = 0$ and $\varepsilon_{ij} \rightsquigarrow N(0, \sigma)$. Now, the null hypothesis is $H_o : \alpha_i = 0$ for all i versus $H_1 : \alpha_i \neq 0$ at least for one value i, which is more suitable.

Using the notation $\overline{Y}_{\cdot\cdot} = \frac{1}{an} \sum_{i,j} Y_{ij}$ and $\overline{Y}_{i\cdot} = \frac{1}{n} \sum_j Y_{ij}$, it seems appropriate to estimate the parameters of the model as $\hat{\mu} = \overline{Y}_{\cdot\cdot}$, $\hat{\mu}_i = \overline{Y}_{i\cdot}$ and $\hat{\alpha}_i = \overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}$. Then, the estimated residuals for the model are $\hat{\varepsilon}_{ij} = e_{ij} = Y_{ij} - \hat{\mu}_i$ and it is clear that all these estimators are unbiased with normal distribution. Moreover, it can be easily shown that

$$\sum_{i,j} \left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2 = n \sum_i \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2 + \sum_{i,j} \left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2$$

or, more briefly, $SST = SS_A + SSE$ with $SST = \sum_{i,j} \left(Y_{ij} - \overline{Y}_{\cdot\cdot}\right)^2$, $SS_A = n \sum_i \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2$ and $SSE = \sum_{i,j} \left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2$. Note that, alternatively, we can write $SST = \sum_{i,j} (Y_{ij} - \hat{\mu})^2$, $SSE = \sum_{i,j} (Y_{ij} - \hat{\mu}_i)^2$ and $SS_A = n \sum_i \hat{\alpha}_i^2$.

Using the Fisher's theorem for each of the a samples, is clear that $\frac{\sum_j \left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2}{\sigma^2} \rightsquigarrow \chi^2_{n-1}$ for all i, and they are independent random variables. Then we can ensure that $\dfrac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{a(n-1)}$ and we can estimate the parameter σ^2 as $\widehat{\sigma^2} = \dfrac{SSE}{a(n-1)}$. This ratio is called the mean square error of the model and it is usually denoted by $MSE = \widehat{\sigma^2}$.

In addition, if $\alpha_i = 0$ for all i, it is clear that $\overline{Y}_{i\cdot} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ with $i = 1, ..., a$ and they are a simple random sample of this distribution with $\overline{Y}_{\cdot\cdot} = \frac{1}{n}\sum_i \overline{Y}_{i\cdot}$. Then, using again the Fisher's theorem, we have

$$\frac{\sum_i \left(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}\right)^2}{\sigma^2/n} = \frac{SS_A}{\sigma^2} \rightsquigarrow \chi^2_{a-1}$$

Finally, taking into account that SS_A is independent of SSE, we obtain that, if $H_o : \alpha_i = 0$ for all i is true and $MS_A = \dfrac{SS_A}{a-1}$, then

$$F_A = \frac{SS_A/(a-1)}{SSE/(a(n-1))} = \frac{MS_A}{MSE} \rightsquigarrow F_{a-1,a(n-1)}$$

Therefore, this statistic let us to obtain an appropriate test for H_o vs H_1 reasoning as usual in the hypothesis test theory.

After solving this test, if H_o is rejected, it is interesting to compare each pair of means μ_i and $\mu_{i'}$, that is, to solve individual tests with $H_o : \mu_i = \mu_{i'}$ (or $\alpha_i - \alpha_{i'} = 0$) for each pair i, i'. To do this, we observe that

$$\frac{\left(\overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot}\right) - (\alpha_i - \alpha_{i'})}{\sigma\sqrt{\frac{2}{n}}} \rightsquigarrow N(0,1)$$

and, if $\alpha_i = \alpha_{i'}$, we have

$$\frac{\overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot}}{\sqrt{\frac{2MSE}{n}}} \rightsquigarrow t_{a(n-1)}$$

Then, we can solve individual tests for each pair of means using this statistic. For a fixed level of significance α, the least significant difference (LSD) for each pair of observed means $\overline{Y}_{i\cdot}, \overline{Y}_{i'\cdot}$ is

$$LSD = t_{a(n-1);\alpha/2}\sqrt{\frac{2MSE}{n}}$$

such that each test is significative if $\left|\overline{Y}_{i\cdot} - \overline{Y}_{i'\cdot}\right| \geqslant LSD$. This is the classical LSD Fisher's test for the means in the analysis of variance.

The standard error for the means is defined as $SE = \sqrt{\dfrac{MSE}{n}}$ and the confidence intervals for the expected values μ_i can be evaluated as

$$\overline{Y}_{i\cdot} \pm t_{a(n-1);\alpha/2}\sqrt{\frac{MSE}{n}}$$

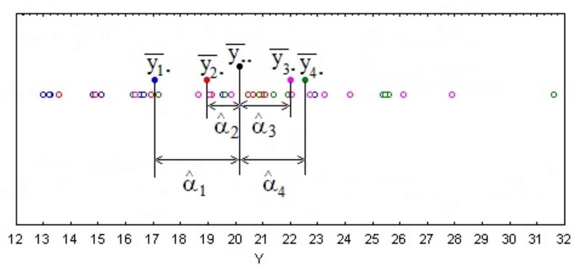
Moreover, for a new independent observation $Y_{i_0 j_0}$, we can ensure that $Y_{i_0 j_0} - \overline{Y}_{i\cdot} \rightsquigarrow N\left(0, \sigma\sqrt{1+\frac{1}{n}}\right)$ and a prediction interval for the observed value is given by

$$\overline{Y}_{i\cdot} \pm t_{a(n-1);\alpha/2}\sqrt{MSE}\sqrt{1+\frac{1}{n}}$$

Example 1: Completely randomized design. To compare the diameter growth in four pine species living on a plantation 15 years old, ten pines of each specie were randomly selected from the plantation and the following data for the diameter (cm) were obtained:

Specie 1 (y_{1j})	16.6	16.7	13.2	13.2	22.9	13.0	15.1	25.5	19.6	14.8	$\sum_j y_{1j} = 170.6$
Specie 2 (y_{2j})	14.9	20.5	20.6	21.1	13.6	21.1	21.1	16.9	20.9	19.1	$\sum_j y_{2j} = 189.8$
Specie 3 (y_{3j})	25.4	16.3	31.6	25.6	20.9	21.9	21.4	17.2	19.6	25.5	$\sum_j y_{3j} = 225.4$
Specie 4 (y_{4j})	27.9	22.0	16.4	19.9	22.7	23.3	19.1	26.1	18.7	24.2	$\sum_j y_{4j} = 220.3$

Denoting by μ_i the expected value for the diameter in the i-specie, we consider the one-way model $E(Y_{ij}) = \mu_i = \mu + \alpha_i + \varepsilon_{ij}$ with $\sum_i \alpha_i = 0$ and $\varepsilon_{ij} \rightsquigarrow N(0, \sigma)$. Taking into account that $\sum_{i,j} y_{ij} = 806.1$ and $\sum_{i,j} y_{ij}^2 = 16988.56$, the estimated parameters for the expected values are: $\hat{\mu} = \bar{y}_{\cdot\cdot} = 20.1525$, $\hat{\mu}_1 = \bar{y}_{1\cdot} = 17.06$, $\hat{\mu}_2 = \bar{y}_{2\cdot} = 18.98$, $\hat{\mu}_3 = \bar{y}_{3\cdot} = 22.54$ and $\hat{\mu}_4 = \bar{y}_{4\cdot} = 22.03$. Therefore, we have $\hat{\alpha}_1 = -3.0925$, $\hat{\alpha}_2 = -1.1725$, $\hat{\alpha}_3 = 2.3875$ and $\hat{\alpha}_4 = 1.8775$. The next graph plots the observed values and the estimated parameters:



Calculating the sum of squares we obtain: SST= $16988.56 - 40\bar{y}_{\cdot\cdot}^2$ =743.630, SS_A=$10\sum_i \hat{\alpha}_i^2$=201.635 and SSE=541.995. Then $MSE = \frac{541.995}{36} = 15.0554$ and the estimated value for the parameter σ is $\hat{\sigma} = \sqrt{MSE} = 3.88$. The F-test with $H_o : \alpha_i = 0$ for all i leads to $F_A = \frac{201.635/3}{15.0554} = 4.46$ and the p-value for the test is $p(F_{3,36} > 4.46) = 0.0091$, which is significative with $\alpha = 0.05$. Therefore there are significant differences among species in the expected diameter growth.

The standard error for the means is $SE = \sqrt{\frac{15.0554}{10}} = 1.227$ and, taking into account that $t_{a(n-1);\alpha/2} = t_{36;0.025} = 2.0281$, the 95% confidence intervals for the expected values are $17.06 \pm 2.49 = (14.57, 19.55)$ for μ_1, $18.98 \pm 2.49 = (16.49, 21.47)$ for μ_2, $22.54 \pm 2.49 = (20.05, 25.03)$ for μ_3 and $22.03 \pm 2.49 = (19.54, 24.52)$ for μ_4. For a significance level $\alpha = 0.05$, the least significant different is $LSD = 2.0281\sqrt{\frac{30.1108}{10}} = 3.52$ and the differences between the observed means are:

Pair	$\mu_1 - \mu_2$	$\mu_1 - \mu_3$	$\mu_1 - \mu_4$	$\mu_2 - \mu_3$	$\mu_2 - \mu_4$	$\mu_3 - \mu_4$
Difference	-1.92	-5.48^*	-4.97^*	-3.56^*	-3.05	0.51

Therefore, there are significant differences between μ_1 and μ_3, μ_1 and μ_4, and μ_2 and μ_3; because the difference is larger in absolute value than the LSD value. The other pairs of means are not significantly different, that is, μ_1 and μ_2, μ_2 and μ_4, and μ_3 and μ_4. The results of this LSD Fishers's test are usually summarized as:

Specie	Mean	
1	17.06	A
2	18.98	AB
4	22.03	BC
3	22.54	C

or, alternatively,

Specie	Mean		
1	17.06	X	
2	18.98	X	X
4	22.03	X	X
3	22.54		X

so that mean values without any common letter are significantly different, and mean values with at least a common letter are not significantly different. Or similarly, mean values together vertically by the sign X are homogeneous subgroups without significant differences, and pair of means which do not appear together in any homogeneous group are significantly different.

Chapter 2

MAIN EFECTS ANOVA

In this model more than one factor is considered and we will use two and three factors to illustrate it. We begin with the case of two factors.

Let us suppose that y_{ijk} with $i = 1, ..., a$, $j = 1, ..., b$ and $k = 1, ..., n$ are the observed values for ab independent random samples of the dependent variable Y. That is, we have a dependent variable Y and two factors: A with a levels and B with b levels, such that all the pairs i, j are tested. Moreover, we suppose that $Y_{ijk} \rightsquigarrow N\left(\mu_{ij}, \sigma\right)$ and $\mu = \frac{1}{ab}\sum_{i,j}\mu_{ij}$ is the global expected value if the factors A and B have no effect on the dependent variable Y_{ijk}. In addition, is considered that the posible effect of the pair i, j on the global expected value is the sum of the main effect α_i for the level i of factor A and the main effect β_j for the level j of factor B. That is, the model is now $E\left(Y_{ij}\right) = \mu_{ij} = \mu + \alpha_i + \beta_j$ and, therefore, $Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \rightsquigarrow N(0, \sigma)$ and ε_{ijk} independent of $\varepsilon_{i'j'k'}$ for any values i, i', j, j', k, k'. The purpose is now to solve two statistical tests: $H_o : \alpha_i = 0$ for all i and $H_o : \beta_j = 0$ for all j, which can be understood as no effect of each of the factors A and B on the global expected value, respectively. As usual, the logical assumptions $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$ are considered. In real practical situations, this model is commonly used with $n = 1$, that is, there is no replica for each of the combinations of the levels of the two factors A and B.

Using notation $\overline{Y}_{...}=\frac{1}{abn}\sum_{i,j,k} Y_{ijk}$, $\overline{Y}_{i..}=\frac{1}{bn}\sum_{j,k} Y_{ijk}$ and $\overline{Y}_{.j.}=\frac{1}{an}\sum_{i,k} Y_{ijk}$, it seems appropriate to estimate the parameters of the model as $\hat{\mu}=\overline{Y}_{...}$, $\hat{\alpha}_i=\overline{Y}_{i..}-\overline{Y}_{...}$, $\hat{\beta}_j=\overline{Y}_{.j.}-\overline{Y}_{...}$, and $\hat{\mu}_{ij}=\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j=\overline{Y}_{i..}+\overline{Y}_{.j.}-\overline{Y}_{...}$. Then, the estimated residuals are $\hat{\varepsilon}_{ijk} = e_{ijk} = Y_{ijk} - \hat{\mu}_{ij}$. As in the previous model, it is clear that all these estimators are unbiased with normal distribution. Moreover, it can be shown that

$$\sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{...}\right)^2 = bn\sum_i\left(\overline{Y}_{i..} - \overline{Y}_{...}\right)^2 + an\sum_j\left(\overline{Y}_{.j.} - \overline{Y}_{...}\right)^2 + \sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)^2$$

or, more briefly, $SST = SS_A + SS_B + SSE$ with $SST = \sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{...}\right)^2$, $SS_A = bn\sum_i\left(\overline{Y}_{i..} - \overline{Y}_{...}\right)^2$, $SS_B = an\sum_j\left(\overline{Y}_{.j.} - \overline{Y}_{...}\right)^2$ and $SSE = \sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}\right)^2$. As in the previous model, we can also write $SST=\sum_{i,j,k}\left(Y_{ijk} - \hat{\mu}\right)^2$, $SSE=\sum_{i,j,k}\left(Y_{ijk} - \hat{\mu}_{ij}\right)^2$ and now $SS_A=bn\sum_i \hat{\alpha}_i^2$ and $SS_B=an\sum_j \hat{\beta}_j^2$.

In this model we have $\frac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{abn-a-b+1}$ and the estimation for σ^2 is $\widehat{\sigma^2} = \frac{SSE}{abn-a-b+1} = MSE$.

Moreover, using $SS_{Model} = SS_A + SS_B$, if $\alpha_i = \beta_j = 0$ for all i, j then

$$F_{Model} = \frac{SS_{Model}/\left(a+b-2\right)}{SSE/\left(abn-a-b+1\right)} \rightsquigarrow F_{a+b-2,abn-a-b+1}$$

and we have a whole model test. The goodness of fit for the model is measured by the determination coefficient, which is defined as $R^2 = \frac{SS_{Model}}{SST} = 1 - \frac{SSE}{SST}$, and the variability of the model is given by the variation coefficient $CV = \frac{\sqrt{MSE}}{\overline{Y}_{\cdots}}$, both commonly expressed as percentages. R^2-values close to 1 (100%) and CV-values close to 0 provide a higher quality of the model and better sensitivity to find significant differences due to the factors.

In a similar way, if $\alpha_i = 0$ for all i then

$$F_A = \frac{SS_A/\left(a-1\right)}{SSE/\left(abn-a-b+1\right)} = \frac{MS_A}{MSE} \rightsquigarrow F_{a-1,abn-a-b+1}$$

And, if $\beta_j = 0$ for all j then

$$F_B = \frac{SS_B/\left(b-1\right)}{SSE/\left(abn-a-b+1\right)} = \frac{MS_B}{MSE} \rightsquigarrow F_{b-1,abn-a-b+1}$$

These statistics can be used for the two F-tests in the ANOVA table of the model, as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\alpha_i = 0$ for all i
Factor B	SS_B	$b-1$	MS_B	F_B	p_B	$\beta_j = 0$ for all j
Error	SSE	$abn-a-b+1$	MSE			
Total	SST	$abn-1$				

Now the least significant difference (LSD) for each pair of means $\overline{Y}_{i\cdot\cdot}, \overline{Y}_{i'\cdot\cdot}$ is

$$LSD = t_{abn-a-b+1;\alpha/2}\sqrt{\frac{2MSE}{bn}}$$

and the least significant difference (LSD) for each pair of means $\overline{Y}_{\cdot j\cdot}, \overline{Y}_{\cdot j'\cdot}$ is

$$LSD = t_{abn-a-b+1;\alpha/2}\sqrt{\frac{2MSE}{an}}$$

Finally, the confidence intervals for the expected values $\mu+\alpha_i$ and $\mu+\beta_j$ are $\overline{Y}_{i\cdot\cdot} \pm t_{abn-a-b+1;\alpha/2}\sqrt{\frac{MSE}{bn}}$ and $\overline{Y}_{\cdot j\cdot} \pm t_{abn-a-b+1;\alpha/2}\sqrt{\frac{MSE}{an}}$, respectively. Note that, using the model, the point estimates for the expected values μ_{ij} are $\hat{\mu}_{ij} = \overline{Y}_{i\cdot\cdot} + \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots}$, which variance is given by (see the note below)

$$Var\left(\hat{\mu}_{ij}\right) = \left(1 - \frac{(a-1)(b-1)}{ab}\right)\left(\frac{\sigma^2}{n}\right)$$

and, therefore, a confidence interval for the expected values μ_{ij} is

$$\hat{\mu}_{ij} \pm t_{abn-a-b+1;\alpha/2}\sqrt{1 - \frac{(a-1)(b-1)}{ab}}\sqrt{\frac{MSE}{n}}$$

In a similar way, a prediction interval for a new independent observation is given by

$$\hat{\mu}_{ij} \pm t_{abn-a-b+1;\alpha/2}\sqrt{1+\frac{1}{n}-\frac{(a-1)(b-1)}{abn}}\sqrt{MSE}$$

The check of the assumptions of the model can be made as in the oneway ANOVA with the estimated residuals e_{ijk}.

Note. For a fixed pair i_o, j_o we have

$$\overline{Y}_{i_o\cdot\cdot}+\overline{Y}_{\cdot j_o\cdot}-\overline{Y}_{\cdots}=\frac{a+b-1}{abn}\sum_k Y_{i_o j_o k}+\frac{a-1}{abn}\sum_{j,k,j\neq j_o} Y_{i_o jk}+\frac{b-1}{abn}\sum_{i,k,i\neq i_o} Y_{ij_o k}-\frac{1}{abn}\sum_{i,j,k,i\neq i_o,j\neq j_o} Y_{ijk}$$

and, taking into account that all Y_{ijk} are independent, we have

$$\begin{aligned}
Var\left(\overline{Y}_{i_o\cdot\cdot}+\overline{Y}_{\cdot j_o\cdot}-\overline{Y}_{\cdots}\right) &= \frac{(a+b-1)^2 n+(a-1)^2(b-1)n+(b-1)^2(a-1)n+(a-1)(b-1)n}{a^2b^2n^2}\sigma^2 \\
&= \frac{(a-1)(b-1)\left[\frac{(a+b-1)^2}{(a-1)(b-1)}+a+b-1\right]}{a^2b^2}\left(\frac{\sigma^2}{n}\right) \\
&= \frac{(a-1)(b-1)(a+b-1)\left[\frac{a+b-1}{(a-1)(b-1)}+1\right]}{a^2b^2}\left(\frac{\sigma^2}{n}\right) \\
&= \frac{a+b-1}{ab}\left(\frac{\sigma^2}{n}\right)=\left(1-\frac{(a-1)(b-1)}{ab}\right)\left(\frac{\sigma^2}{n}\right)
\end{aligned}$$

Example 2: Randomized complete block design. In an experimental study to compare the total production for 10 tomato varieties, a total of 40 experimental small plots were used (10 for each variety). Due to possible heterogeneity among the experimental plots, they were grouped into 4 groups with 10 plots per group, such that the plots within the same group are more homogeneous. Then, for each block, a plot is randomly assigned to each variety and the total production of tomatoes is measured. This experimental design is called randomized complete block design. Let us denote by y_{ij} the total production for the i variety at the j block, with $i = 1, ..., 10$ and $j = 1, 2, 3, 4$. The obtained data (kg/m^2) are included in the following table:

Variety:	1	2	3	4	5	6	7	8	9	10	Mean ($\bar{y}_{\cdot j}$)
Block 1	8.036	6.024	5.336	7.792	12.828	5.620	9.964	7.020	7.540	9.672	7.9832
Block 2	3.132	7.252	7.124	9.356	15.020	13.488	22.520	5.880	8.448	9.320	10.154
Block 3	9.356	3.840	3.176	5.620	10.260	10.300	11.048	10.280	14.060	8.396	8.6336
Block 4	7.344	3.796	3.280	8.696	6.056	3.996	18.428	4.112	10.028	9.812	7.5548
Mean ($\bar{y}_{i\cdot}$)	6.967	5.228	4.729	7.866	11.041	8.351	15.490	6.823	10.019	9.300	$y_{\cdot\cdot} = 8.5814$

For the statistical analysis, we consider two factors (A =variety and B =block) and we use the main effects model $E(Y_{ij}) = \mu_{ij} = \mu + \alpha_i + \beta_j$, that is, $Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ with $\varepsilon_{ij} \rightsquigarrow N(0,\sigma)$ and ε_{ij} independent of $\varepsilon_{i'j'}$ for any values i, i', j, j' (note that we do not use subscript k because there is no replicates in this experiment). Taking into account that $\sum_{i,j} y_{ij} = 343.256$ and $\sum_{i,j} y_{ij}^2 = 3605.25$, the estimated values for the parameters of the model are: $\hat{\mu} = \bar{y}_{\cdot\cdot} = 8.581$, $\hat{\beta}_1 = \bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot} = -0.5982$, $\hat{\beta}_2=\bar{y}_{\cdot 2}-\bar{y}_{\cdot\cdot}=1.5726$, $\hat{\beta}_3=\bar{y}_{\cdot 3}-\bar{y}_{\cdot\cdot}=0.0522$, $\hat{\beta}_4=\bar{y}_{\cdot 4}-\bar{y}_{\cdot\cdot}=-1.0266$ and the following values for $\hat{\alpha}_i=\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot}$:

Variety	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$	-1.6144	-3.3534	-3.8524	-0.7154	2.4596	-0.2304	6.9086	-1.7584	1.4376	0.7186

Calculating the sum of squares we obtain: $SST = 3605.25 - 40\bar{y}_{\cdot\cdot}^2 = 659.6345$, $SS_A = 4\sum_i \hat{\alpha}_i^2 = 354.844$, $SS_B = 10\sum_j \hat{\beta}_j^2 = 38.8755$ and $SSE = 265.915$. Therefore, $MSE = \frac{265.915}{27} = 9.8487$ and the estimated value for the parameter σ is $\hat{\sigma} = \sqrt{MSE} = 3.138$.

The determination coefficient of the model is $R^2 = 1 - \frac{265.915}{659.6345} = 59.7\%$ and the variation coefficient is $CV = \frac{313.8}{8.5814} = 36.6\%$. The whole model test gives $F_{Model} = \frac{(354.844+38.8755)/12}{9.8487} = 3.33$ which leads to p-value$= p\left(F_{12,27} > 3.33\right) = 0.0046$. Therefore the model is significative with $\alpha = 0.05$.

The F-test for the variety factor gives $F_A = \frac{354.844/9}{9.8487} = \frac{67.2116}{15.0551} = 4.00$ with p-value$= p\left(F_{9,27} > 4.00\right) = 0.0024$. That is, the variety factor is significative with $\alpha = 0.05$ and there are differences between varieties.

The F-test for block factor leads to $F_B = \frac{38.8755/3}{9.8487} = \frac{12.9585}{15.0551} = 1.32$ with p-value$= p\left(F_{3,27} > 1.32\right) = 0.2897$. That is, the block factor is not significative with $\alpha = 0.05$ and there are no differences between blocks.

The standard error for the means of the varieties is $SE = \sqrt{\frac{9.8487}{4}} = 1.5691$ and, taking into account that $t_{27;0.025} = 2.0518$, the 95% confidence intervals for the expected values are $\bar{y}_{i\cdot} \pm 3.219$.

Finally, the least significant difference (LSD) with $\alpha = 0.05$ for each pair of means $\bar{y}_{i\cdot}, \bar{y}_{i'\cdot}$ is $LSD = t_{27;0.025}\sqrt{\frac{2MSE}{4}} = 4.55$. Using this value we obtain the following results for the LSD Fisher's test of the varieties with $\alpha = 0.05$:

Variety	Mean	
3	4.729	A
2	5.228	AB
8	6.823	ABC
1	6.967	ABC
4	7.866	ABC
6	8.351	ABC
10	9.300	BC
9	10.019	C
5	11.041	CD
7	15.490	D

or, alternatively,

Variety	Mean	
3	4.729	X
2	5.228	XX
8	6.823	XXX
1	6.967	XXX
4	7.866	XXX
6	8.351	XXX
10	9.300	XX
9	10.019	X
5	11.041	XX
7	15.490	X

Therefore, with 95% of confidence, the variety with the number 7 has a greater expected value than all the others, except the variety with number 5.

As an extension of the previous model, we will consider a case with three factors.

Example 3: Latin square design. In the randomized block design studied previously we considered a major factor and a control factor or block which is introduced in order to eliminate their influence on the response variable and reduce the experimental error. In this new design we use more than one variable block to reduce experimental error.

Thus, if two block variables are considered simultaneously, a complete randomized block design would be to form a block for every combination of levels of these variables and then apply all levels of the main factor in each of the blocks obtained. For example, suppose an experiment in which we want to study the effect of different types of seed on wheat yield and we believes that this performance can also be influenced by the types of fertilizers and insecticides. To perform this study, it is possible to use a complete randomized block design, where the main factor is the type of seed and the block variables are the types of fertilizers and insecticides. A disadvantage of these designs is sometimes the excessive experimental units required for implementation. A complete block design with a main factor and two block factors, with K_1, K_2 and K_3 levels in each factor, requires $K_1K_2K_3$ experimental units. In our example, if the main factor, type of seed, has 4 levels, the first block variable, type of fertilizer, 5 levels and the second block variable, type of insecticide, 3 levels, it would take 60 experimental units.

Some experiments can have different causes, such as economic in nature, that advise against using many experimental units. In this situation you can use a special type of randomized incomplete block designs. The basic idea of these designs is the fraction, that is, select a part of the whole design so that, under certain general assumptions, we can estimate the effects of interest. One the most important designs with randomized incomplete blocks using a main factor and two block factors is the latin square design. This model assumes the same number of levels for the three factors.

In general, for K levels in each factor, a complete randomized block design uses K^2 blocks, each block being applied in the K levels of the main factor, resulting in a total of K^3 experimental units. Latin square designs reduce the number of experimental units to K^2 by considering the K^2 blocks in the experiment but using only one treatment in each block with a special provision. Specifically, in each block a single treatment is applied so that each treatment must appear with each of the levels of the two control factors applied. Thus, if $K = 4$, a complete block design would need 64 observations, while the latin square design would need only 16 observations.

If we consider a two-way table where rows and columns represent each of the two block factors and the cells represent the levels of the main factor or treatments, the above requirement means that each treatment must appear once and only once in each row and in each column. For example, let us suppose that $a_1, a_2, ..., a_K$ denote the K levels of the first block factor A (rows); $b_1, b_2, ..., b_K$ denote the K levels of the second block factor B (columns) and $c_1, c_2, ..., c_K$ denote the K levels of the main factor C. Then, a latin square design with order K is an arrangement in rows and columns of the K latin letters $c_1, c_2, ..., c_K$, so that each letter appears only once in each row and each column. For the example we have been considering, if we have 4 fertilizers (factor A), 4 insecticides (factor B) and 4 types of seeds (factor C) a latin square with order 4 may have the following distribution:

	Insecticides			
Fertilizers	b_1	b_2	b_3	b_4
a_1	c_3	c_4	c_2	c_1
a_2	c_2	c_1	c_3	c_4
a_3	c_4	c_3	c_1	c_2
a_4	c_1	c_2	c_4	c_3

In a latin square design with order k, we denote by $y_{ij(k)}$, with $i, j, k \in \{1, 2, .., K\}$, the K^2 observed values of the dependent variable Y and, for the statistical analysis, we use the main effects model $E\left(Y_{ij(k)}\right) = \mu_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_k$, that is, $Y_{ij(k)} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij(k)}$ with $\varepsilon_{ij(k)} \rightsquigarrow N(0, \sigma)$ and $\varepsilon_{ij(k)}$ independent of $\varepsilon_{i'j'(k')}$ for any values i, i', j, j', k, k'. Therefore, it is considered that the posible effect of the combination of the levels i, j, k of the factors A, B and C on the global expected value is the sum of the main effect α_i for the level i of factor A, the main effect β_j for the level j of factor B and the main effect γ_k for the level k of factor C. As usual, we suppose that $\sum\limits_i \alpha_i = \sum\limits_j \beta_j = \sum\limits_k \gamma_k = 0$.

Using notation $\overline{Y}_{...} = \frac{1}{K^2} \sum\limits_i \sum\limits_j Y_{ij(k)}$, $\overline{Y}_{i..} = \frac{1}{K} \sum\limits_j Y_{ij(k)}$, $\overline{Y}_{.j.} = \frac{1}{K} \sum\limits_i Y_{ij(k)}$ and $\overline{Y}_{..k} = \frac{1}{K} \sum\limits_{i,j} Y_{ij(k)}$, it seems appropriate to estimate the parameters of the model as $\hat{\mu} = \overline{Y}_{...}$, $\hat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...}$, $\hat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$, $\hat{\gamma}_k = \overline{Y}_{..k} - \overline{Y}_{...}$ and $\hat{\mu}_{ij(k)} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k = \overline{Y}_{i..} + \overline{Y}_{.j.} + \overline{Y}_{..k} - 2\overline{Y}_{...}$. Then, the estimated residuals are $\hat{\varepsilon}_{ij(k)} = e_{ij(k)} = Y_{ij(k)} - \hat{\mu}_{ij(k)}$ and, as always, all these estimators are unbiased with normal distribution. Moreover, it can be shown that:

$$\sum_{i.j} \left(Y_{ij(k)} - \overline{Y}_{...}\right)^2 = K \sum_i \left(\overline{Y}_{i..} - \overline{Y}_{...}\right)^2 + K \sum_j \left(\overline{Y}_{.j.} - \overline{Y}_{...}\right)^2 + K \sum_k \left(\overline{Y}_{..k} - \overline{Y}_{...}\right)^2 + \sum_{i,j} \left(Y_{ij(k)} - \overline{Y}_{i..} - \overline{Y}_{.j.} - \overline{Y}_{..k} + 2\overline{Y}_{...}\right)^2$$

or, more briefly, $SST = SS_A + SS_B + SS_C + SSE$ with $SST = \sum\limits_{i,j} \left(Y_{ij(k)} - \overline{Y}_{...}\right)^2$, $SS_A = K \sum\limits_i \left(\overline{Y}_{i..} - \overline{Y}_{...}\right)^2$, $SS_B = K \sum\limits_j \left(\overline{Y}_{.j.} - \overline{Y}_{...}\right)^2$, $SS_C = K \sum\limits_j \left(\overline{Y}_{..k} - \overline{Y}_{...}\right)$ and $SSE = \sum\limits_{i,j} \left(Y_{ij(k)} - \overline{Y}_{i..} - \overline{Y}_{.j.} - \overline{Y}_{..k} + 2\overline{Y}_{...}\right)^2$. As usual, we can also write $SST = \sum\limits_{i,j} \left(Y_{ij(k)} - \hat{\mu}\right)^2$, $SSE = \sum\limits_{i,j} \left(Y_{ij(k)} - \hat{\mu}_{ij(k)}\right)^2$ and now $SS_A = K \sum\limits_i \hat{\alpha}_i^2$, $SS_B = K \sum\limits_j \hat{\beta}_j^2$ and $SS_C = K \sum\limits_j \hat{\gamma}_k^2$.

In this model we have $\dfrac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{K^2 - 3K + 2}$ and the estimation for σ^2 is $\widehat{\sigma^2} = \dfrac{SSE}{K^2 - 3K + 2} = MSE$ (note that $K^2 - 3K + 2 = (K-1)(K-2)$).

Moreover, using $SS_{Model} = SS_A + SS_B + SS_C$, if $\alpha_i = \beta_j = \gamma_k = 0$ for all i, j, k then

$$F_{Model} = \frac{SS_{Model} / (3K - 3)}{SSE / ((K-1)(K-2))} \rightsquigarrow F_{3K-3,(K-1)(K-2)}$$

and we have a whole model test. The R^2 and CV coefficients can be evaluated as in the previous model.

In this model, if $\alpha_i = 0$ for all i then

$$F_A = \frac{SS_A / (K-1)}{SSE / ((K-1)(K-2))} = \frac{MS_A}{MSE} \rightsquigarrow F_{K-1,(K-1)(K-2)}$$

In a similar way, if $\beta_j = 0$ for all j then

$$F_B = \frac{SS_B/(K-1)}{SSE/((K-1)(K-2))} = \frac{MS_B}{MSE} \rightsquigarrow F_{K-1,(K-1)(K-2)}$$

and, finally, if $\gamma_k = 0$ for all k then

$$F_C = \frac{SS_C/(K-1)}{SSE/((K-1)(K-2))} = \frac{MS_C}{MSE} \rightsquigarrow F_{K-1,(K-1)(K-2)}$$

These statistics can be used for the three F-tests in the ANOVA table, as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor A	SS_A	$K-1$	MS_A	F_A	p_A	$\alpha_i = 0$ for all i
Factor B	SS_B	$K-1$	MS_B	F_B	p_B	$\beta_j = 0$ for all j
Factor C	SS_C	$K-1$	MS_C	F_C	p_C	$\gamma_k = 0$ for all k
Error	SSE	$(K-1)(K-2)$	MSE			
Total	SST	K^2-1				

Now the least significant differences (LSD) for all pair of means $\overline{Y}_{i\cdot\cdot}$ and $\overline{Y}_{i'\cdot\cdot}$, $\overline{Y}_{\cdot j\cdot}$ and $\overline{Y}_{\cdot j'\cdot}$, and $\overline{Y}_{\cdot\cdot k}$ and $\overline{Y}_{\cdot\cdot k'}$, are

$$LSD = t_{(K-1)(K-2);\alpha/2}\sqrt{\frac{2MSE}{K}}$$

and the confidence intervals for the expected values $\mu+\alpha_i$, $\mu+\beta_j$ and $\mu+\gamma_k$ are $\overline{Y}_{i\cdot\cdot} \pm t_{(K-1)(K-2);\alpha/2}\sqrt{\frac{MSE}{K}}$, $\overline{Y}_{\cdot j\cdot} \pm t_{(K-1)(K-2);\alpha/2}\sqrt{\frac{MSE}{K}}$ and $\overline{Y}_{\cdot\cdot k} \pm t_{(K-1)(K-2);\alpha/2}\sqrt{\frac{MSE}{K}}$, respectively.

For the estimators $\hat{\mu}_{ij(k)} = \overline{Y}_{i\cdot\cdot} + \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdot\cdot k} - 2\overline{Y}_{\cdot\cdot\cdot}$ we obtain (see the note below)

$$Var\left(\hat{\mu}_{ij(k)}\right) = \left(1 - \frac{(K-1)(K-2)}{K^2}\right)\sigma^2$$

and, therefore, a confidence interval for the expected values $\mu_{ij(k)}$ is

$$\hat{\mu}_{ij(k)} \pm t_{(K-1)(K-2);\alpha/2}\sqrt{1 - \frac{(K-1)(K-2)}{K^2}}\sqrt{MSE}$$

In a similar way, a prediction interval for a new independent observation is given by

$$\hat{\mu}_{ij(k)} \pm t_{(K-1)(K-2);\alpha/2}\sqrt{2 - \frac{(K-1)(K-2)}{K^2}}\sqrt{MSE}$$

The check of the assumptions of the model can be made as in the previous models with the estimated residuals $e_{ij(k)}$.

Note. For a fixed pair i_o, j_o (with $k = k_o$) we have

$$\begin{aligned}\overline{Y}_{i_o\cdot\cdot} + \overline{Y}_{\cdot j_o\cdot} + \overline{Y}_{\cdot\cdot k_o} - 2\overline{Y}_{\cdot\cdot\cdot} \;=\; & \frac{3K-2}{K^2}Y_{i_oj_o(k_o)} + \frac{K-2}{K^2}\sum_{j\neq j_o}Y_{i_oj(k)} + \frac{K-2}{K^2}\sum_{i\neq i_o}Y_{ij_o(k)} \\ & + \frac{K-2}{K^2}\sum_{i\neq i_o, j\neq j_o}Y_{ij(k_o)} - \frac{2}{K^2}\sum_{i\neq i_o, j\neq j_o, k\neq k_o}Y_{ij(k)}\end{aligned}$$

and, taking into account that all $Y_{ij(k)}$ are independent, we have

$$\begin{aligned}Var\left(\overline{Y}_{i_o\cdot\cdot} + \overline{Y}_{\cdot j_o\cdot} + \overline{Y}_{\cdot\cdot k_o} - 2\overline{Y}_{\cdot\cdot\cdot}\right) &= \frac{(3K-2)^2 + 3(K-2)^2(K-1) + 4(K-1)(K-2)}{K^4}\sigma^2 \\ &= \frac{(3K-2)^2 + (K-1)(K-2)(3K-2)}{K^4}\sigma^2 \\ &= \frac{(3K-2)K^2}{K^4}\sigma^2 = \left(1 - \frac{(K-1)(K-2)}{K^2}\right)\sigma^2\end{aligned}$$

To illustrate this model suppose that, in the example we have been considering, we obtained the following data for the wheat yield with different fertilizers, insecticides and seed types:

	Insecticides					
Fertilizers	b_1	b_2	b_3	b_4	$\bar{y}_{i\cdot\cdot}$	$\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}$
a_1	$y_{11(3)} = 7$	$y_{12(4)} = 8$	$y_{13(2)} = 4$	$y_{14(1)} = 3$	$\bar{y}_{1\cdot\cdot} = \frac{22}{4}$	$\hat{\alpha}_1 = \frac{-115}{16}$
a_2	$y_{21(2)} = 15$	$y_{22(1)} = 16$	$y_{23(3)} = 18$	$y_{24(4)} = 23$	$\bar{y}_{2\cdot\cdot} = \frac{72}{4}$	$\hat{\alpha}_2 = \frac{85}{16}$
a_3	$y_{31(4)} = 18$	$y_{32(3)} = 12$	$y_{33(1)} = 12$	$y_{34(2)} = 10$	$\bar{y}_{3\cdot\cdot} = \frac{52}{4}$	$\hat{\alpha}_3 = \frac{5}{16}$
a_4	$y_{41(1)} = 14$	$y_{42(2)} = 13$	$y_{43(4)} = 16$	$y_{44(3)} = 14$	$\bar{y}_{4\cdot\cdot} = \frac{57}{4}$	$\hat{\alpha}_4 = \frac{25}{16}$
$\bar{y}_{\cdot j\cdot}$	$\bar{y}_{\cdot 1\cdot} = \frac{54}{4}$	$\bar{y}_{\cdot 2\cdot} = \frac{49}{4}$	$\bar{y}_{\cdot 3\cdot} = \frac{50}{4}$	$\bar{y}_{\cdot 4\cdot} = \frac{50}{4}$	$\bar{y}_{\cdot\cdot\cdot} = \frac{203}{16}$	
$\hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}$	$\hat{\beta}_1 = \frac{13}{16}$	$\hat{\beta}_2 = \frac{-7}{16}$	$\hat{\beta}_3 = \frac{-3}{16}$	$\hat{\beta}_4 = \frac{-3}{16}$	$\sum y_{ij(k)}^2 = 3001$	

For the seed types, the means and the estimated parameters are included in the next table:

Seed type:	c_1	c_2	c_3	c_4
$\bar{y}_{\cdot\cdot k}$	$\frac{45}{4}$	$\frac{42}{4}$	$\frac{51}{4}$	$\frac{65}{4}$
$\hat{\gamma}_k = \bar{y}_{\cdot\cdot k} - \bar{y}_{\cdot\cdot\cdot}$	$\frac{-23}{16}$	$\frac{-35}{16}$	$\frac{1}{16}$	$\frac{57}{16}$

Calculating the sum of squares we obtain: $SST = 3001 - 16\bar{y}_{\cdot\cdot}^2 = 425.4375$, $SS_A = 4\sum_i \hat{\alpha}_i^2 = 329.6875$, $SS_B = 4\sum_j \hat{\beta}_j^2 = 3.6875$, $SS_C = 4\sum_k \hat{\gamma}_k^2 = 78.1875$ and $SSE = 13.875$. Therefore, $MSE = \frac{13.875}{6} = 2.3125$ and the estimated value for the parameter σ is $\hat{\sigma} = \sqrt{MSE} = 1.52$.

The determination coefficient of the model is $R^2 = 1 - \frac{13.875}{425.4375} = 96.7\%$ and the variation coefficient is $CV = \frac{1.52}{12.6875} = 12.0\%$. The whole model test gives $F_{Model} = \frac{(329.6875+3.6875+78.1875)/9}{2.3125} = 19.77$ which leads to p-value$= p\left(F_{9,6} > 19.77\right) = 0.0008$. Therefore the model is significative with $\alpha = 0.05$.

The F-test for the fertilizers (first block factor) gives $F_A = \frac{329.6875/3}{2.3125} = 47.52$ with p-value$= p\left(F_{3,6} > 47.52\right) = 0.0001$. That is, this factor is significative with $\alpha = 0.05$ and there are differences between fertillizers.

The F-test for the insecticides (second block factor) leads to $F_B = \frac{3.6875/3}{2.3125} = 0.53$ with p-value$= p\left(F_{3,6} > 0.53\right) = 0.6781$. That is, this factor is not significative with $\alpha = 0.05$ and there are no differences between insecticides.

Finally, for the main factor (seeds type) the F-test leads to $F_C = \frac{78.1875/3}{2.3125} = 11.27$ with p-value$= p\left(F_{3,6} > 11.27\right) = 0.0071$. That is, this factor is significative with $\alpha = 0.05$ and there are differences between seed types.

The standard error for all the means of the fertilizers, insecticides and seed types is $SE = \sqrt{\frac{2.3125}{4}} = 0.7603$ and, taking into account that $t_{6;0.025} = 2.4469$, the 95% confidence intervals for the expected values are $\bar{y}_{i\cdot\cdot} \pm 1.86$, $\bar{y}_{\cdot j\cdot} \pm 1.86$ and $\bar{y}_{\cdot\cdot k} \pm 1.86$.

Finally, the least significant difference (LSD) with $\alpha = 0.05$ for all the pair of means is $LSD = t_{6;0.025}\sqrt{\frac{2MSE}{4}} = 2.63$. Using this value we obtain the following results for the LSD Fisher's test of the three factor with $\alpha = 0.05$:

Fertilizers	Mean	
a_2	18	A
a_4	14.25	B
a_3	13	B
a_1	5.5	C

Insecticides	Mean	
b_1	13.5	A
b_3	12.5	A
b_4	12.5	A
b_2	12.25	A

Seed types	Mean	
c_4	16.25	A
c_3	12.75	B
c_1	11.25	B
c_2	10.5	B

Therefore, the wheat yield is significantly higher with the seed type c_4 and there are not differences for the other three seed types. With respect to the block factors, there are no differences between insecticides and the wheat yield is significantly higher with the fertilizer a_2, significantly lower with the fertilizer a_1 and there are no differences between fertilizers a_3 and a_4.

The 95%-confidence intervals for the expected values $\mu_{ij(k)}$ are $\hat{\mu}_{ij(k)} \pm 2.94$ and the prediction interval for a new independent observation is $\hat{\mu}_{ij(k)} \pm 4.74$. For example, for the best option with fertilizer a_2, insecticide b_1 and seed type c_4, these intervals are 22.375 ± 2.94 and 22.375 ± 4.74, that is, $(19.435, 25.315)$ for the expected value and $(17.635, 27.115)$ for a new independent observation.

Chapter 3

TWO-WAY FACTORIAL ANOVA

With the same assumptions as in the previous model with two factors, is now considered that the posible effect of the pair i, j on the global expected value is not necessarily the sum of the main effect α_i for the level i of factor A and the main effect β_j for the level j of factor B. That is, a posible interaction effect $\alpha\beta_{ij}$ between levels i, j of the factors A and B is allowed. Then, the model is now $E(Y_{ij}) = \mu_{ij} = \mu+\alpha_i+\beta_j+\alpha\beta_{ij}$ and, therefore, $Y_{ijk} = \mu+\alpha_i+\beta_j+\alpha\beta_{ij}+\varepsilon_{ijk}$ with $\varepsilon_{ijk} \rightsquigarrow N(0,\sigma)$ and ε_{ijk} independent of $\varepsilon_{i'j'k'}$ for any values i, i', j, j', k, k'. In addition to the two statistical tests of the previous model, we have now a third test $H_o : \alpha\beta_{ij} = 0$ for all i, j, which can be understood as no interaction effect between factors A and B. Besides the previous conditions $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$, these others are now added: $\sum_j \alpha\beta_{ij} = 0$ for all i and $\sum_i \alpha\beta_{ij} = 0$ for all j. Finally, it should be noted that now $n > 1$ is required in this model, which was not necessary in the previous model with main effects.

Using $\overline{Y}_{\cdots}$, $\overline{Y}_{i\cdot\cdot}$ and $\overline{Y}_{\cdot j\cdot}$ as in the previous model, $\overline{Y}_{ij\cdot}=\frac{1}{n}\sum_k Y_{ijk}$ is now considered and the appropriate estimators for the parameters are $\hat{\mu} = \overline{Y}_{\cdots}$, $\hat{\mu}_{ij} = \overline{Y}_{ij\cdot}$, $\hat{\alpha}_i = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}$, $\hat{\beta}_j = \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots}$ and

$$\widehat{\alpha\beta}_{ij} = (\overline{Y}_{ij\cdot} - \overline{Y}_{\cdots}) - \hat{\alpha}_i - \hat{\beta}_j = \overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{\cdots}$$

Then, the estimated residuals are now $\hat{\varepsilon}_{ijk} = e_{ijk} = Y_{ijk} - \overline{Y}_{ij\cdot}$ and, as always, all these estimators are unbiased with normal distribution. Moreover, it can be shown that

$$\sum_{i,j,k}\left(Y_{ijk}-\overline{Y}_{\cdots}\right)^2 = bn\sum_i\left(\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdots}\right)^2 + an\sum_j\left(\overline{Y}_{\cdot j\cdot}-\overline{Y}_{\cdots}\right)^2 + n\sum_{i,j}\left(Y_{ij\cdot}-\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot j\cdot}+\overline{Y}_{\cdots}\right)^2 + \sum_{i,j,k}\left(Y_{ijk}-\overline{Y}_{ij\cdot}\right)^2$$

or briefly $SST = SS_A + SS_B + SS_{AB} + SSE$ with $SST = \sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{\cdots}\right)^2$, $SS_A = bn\sum_i\left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}\right)^2$, $SS_B = an\sum_j\left(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots}\right)^2$, $SS_{AB} = n\sum_{i,j}\left(Y_{ij\cdot}-\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot j\cdot}+\overline{Y}_{\cdots}\right)^2$ and $SSE = \sum_{i,j,k}\left(Y_{ijk}-\overline{Y}_{ij\cdot}\right)^2$. As usual, we can also write $SST = \sum_{i,j}(Y_{ij} - \hat{\mu})^2$, $SSE = \sum_{i,j,k}\left(Y_{ijk} - \hat{\mu}_{ij}\right)^2$ and now $SS_A=bn\sum_i \hat{\alpha}_i^2$, $SS_B=an\sum_j \hat{\beta}_j^2$ and $SS_{AB}=n\sum_{i,j}\left(\widehat{\alpha\beta}_{ij}\right)^2$.

In this model we have $\dfrac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{ab(n-1)}$ and the estimation for σ^2 is $\widehat{\sigma^2} = MSE = \dfrac{SSE}{ab(n-1)}$.

Using $SS_{Model} = SS_A + SS_B + SS_{AB}$, if $\alpha_i = \beta_j = \alpha\beta_{ij} = 0$ for all i, j then

$$F_{Model} = \frac{SS_{Model}/\,(ab-1)}{SSE/\,(ab(n-1))} \rightsquigarrow F_{ab-1,ab(n-1)}$$

and we have a whole model test. The R^2 and CV coefficients can be evaluated as before.

In a similar way, if $\alpha_i = 0$ for all i then

$$F_A = \frac{SS_A/\,(a-1)}{SSE/\,(ab(n-1))} = \frac{MS_A}{MSE} \rightsquigarrow F_{a-1,ab(n-1)}$$

And, if $\beta_j = 0$ for all j then

$$F_B = \frac{SS_B/\,(b-1)}{SSE/\,(ab(n-1))} = \frac{MS_B}{MSE} \rightsquigarrow F_{b-1,ab(n-1)}$$

And, finally, if $\alpha\beta_{ij} = 0$ for all i, j then

$$F_{AB} = \frac{SS_{AB}/\,((a-1)\,(b-1))}{SSE/\,(ab(n-1))} = \frac{MS_{AB}}{MSE} \rightsquigarrow F_{(a-1)(b-1),ab(n-1)}$$

These statistics can be used for the three F-test in the ANOVA table, as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\alpha_i = 0$ for all i
Factor B	SS_B	$b-1$	MS_B	F_B	p_B	$\beta_j = 0$ for all j
Interaction $A*B$	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_{AB}	p_{AB}	$\alpha\beta_{ij} = 0$ for all i, j
Error	SSE	$ab(n-1)$	MSE			
Total	SST	$abn-1$				

Now the confidence intervals for the expected values $\mu+\alpha_i$, $\mu+\beta_j$ and μ_{ij} are $\overline{Y}_{i\cdot\cdot} \pm t_{ab(n-1);\alpha/2}\sqrt{\frac{MSE}{bn}}$, $\overline{Y}_{\cdot j\cdot} \pm t_{ab(n-1);\alpha/2}\sqrt{\frac{MSE}{an}}$ and $\overline{Y}_{ij\cdot} \pm t_{ab(n-1);\alpha/2}\sqrt{\frac{MSE}{n}}$, respectively.

Finally, the least significant differences (LSD) for each pair of means are: $LSD = t_{ab(n-1);\alpha/2}\sqrt{\frac{2MSE}{bn}}$ for $\overline{Y}_{i\cdot\cdot}, \overline{Y}_{i'\cdot\cdot}$, $LSD = t_{ab(n-1);\alpha/2}\sqrt{\frac{2MSE}{an}}$ for $\overline{Y}_{\cdot j\cdot}, \overline{Y}_{\cdot j'\cdot}$ and $LSD = t_{ab(n-1);\alpha/2}\sqrt{\frac{2MSE}{n}}$ for $\overline{Y}_{ij\cdot}, \overline{Y}_{i'j'\cdot}$.

The check of the assumptions of the model can be made as in the previous models with the estimated residuals e_{ijk}.

Example 4: Two-way factorial design. To study the effect of fertilization with nitrogen and potassium on growth of cauliflower, a fertilization experiment is performed with three nitrogen doses (60, 120 and 180 kg/ha) and three potassium doses (100, 200 and 300 kg/ha). In the area of land available 27 micro plots are plotted and 3 are randomly assigned to each of the 9 possible combinations of fertilization. At the end of the experiment, the production of each micro plot is scored (the data are given in tons per hectare).

60+100	60+200	60+300	120+100	120+200	120+300	180+100	180+200	180+300
18	20	23	21	24	21	23	20	19
18	20	24	23	25	19	22	19	18
16	19	21	20	22	20	24	21	21

Considering a factor A with 3 levels for nitrogen doses and a factor B with three levels for potasium doses, we can use a two-way factorial model $E(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$ and therefore $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \rightsquigarrow N(0, \sigma)$ and ε_{ijk} independent of $\varepsilon_{i'j'k'}$ for any values i, i', j, j', k, k'. For $i, j = 1, 2, 3$, α_i is the main effect for i-dose of nitrogen, β_j is the main effect for j-dose of potasium and $\alpha\beta_{ij}$ is the possible interaction effect between i-dose of nitrogen and j-dose of potasium.

Initially, we calculate $\sum_{i,j,k} y_{ijk} = 561$, $\sum_{i,j,k} y_{ijk}^2 = 11785$ and all the means observed in the experiment, which are included in the following table with the usual notation:

$\bar{y}_{11\cdot} = 52/3$	$\bar{y}_{12\cdot} = 59/3$	$\bar{y}_{13\cdot} = 68/3$	$\bar{y}_{1\cdot\cdot} = 179/9$
$\bar{y}_{21\cdot} = 64/3$	$\bar{y}_{22\cdot} = 71/3$	$\bar{y}_{23\cdot} = 60/3$	$\bar{y}_{2\cdot\cdot} = 195/9$
$\bar{y}_{31\cdot} = 69/3$	$\bar{y}_{32\cdot} = 60/3$	$\bar{y}_{33\cdot} = 58/3$	$\bar{y}_{3\cdot\cdot} = 187/9$
$\bar{y}_{\cdot 1\cdot} = 185/9$	$\bar{y}_{\cdot 2\cdot} = 190/9$	$\bar{y}_{\cdot 3\cdot} = 186/9$	$\bar{y}_{\cdots} = 187/9$

From these data we can obtain the estimated parameters of the model. For example, $\hat{\alpha}_1 = \bar{y}_{1\cdot\cdot} - \bar{y}_{\cdots} = \frac{-8}{9}$, $\hat{\beta}_1 = \bar{y}_{\cdot 1\cdot} - \bar{y}_{\cdots} = \frac{-2}{9}$ and $\widehat{\alpha\beta}_{11} = \bar{y}_{11\cdot} - \bar{y}_{\cdots} - \hat{\alpha}_1 - \hat{\beta}_1 = -\frac{31}{9} + \frac{10}{9} = \frac{-21}{9}$. In a similar way, the estimations of the other parameters are evaluated and they are included in the following table:

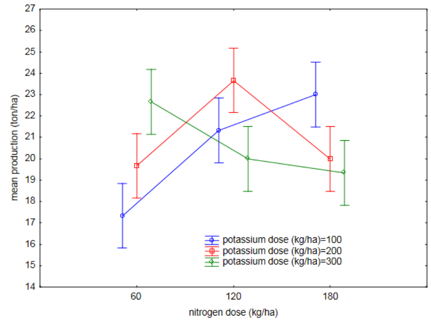
$\widehat{\alpha\beta}_{11} = -21/9$	$\widehat{\alpha\beta}_{12} = -5/9$	$\widehat{\alpha\beta}_{13} = 26/9$	$\hat{\alpha}_1 = -8/9$
$\widehat{\alpha\beta}_{21} = -1/9$	$\widehat{\alpha\beta}_{22} = 15/9$	$\widehat{\alpha\beta}_{23} = -14/9$	$\hat{\alpha}_2 = 8/9$
$\widehat{\alpha\beta}_{31} = 22/9$	$\widehat{\alpha\beta}_{32} = -10/9$	$\widehat{\alpha\beta}_{33} = -12/9$	$\hat{\alpha}_3 = 0$
$\hat{\beta}_1 = -2/9$	$\hat{\beta}_2 = 3/9$	$\hat{\beta}_3 = -1/9$	$\hat{\mu} = \bar{y}_{\cdots} = 187/9$

Calculating the sum of squares we obtain: $SST = 11785 - 27\bar{y}_{\cdots}^2 = \frac{386}{3}$, $SS_A = 9\sum_i \hat{\alpha}_i^2 = \frac{128}{9}$, $SS_B = 9\sum_j \hat{\beta}_j^2 = \frac{14}{9}$, $SS_{AB} = 3\sum_{i,j}\left(\widehat{\alpha\beta}_{ij}\right)^2 = \frac{764}{9}$ and $SSE = \frac{1158-128-14-764}{9} = 28$. Therefore, $MSE = \frac{14}{9} = 1.55$ and the estimated value for the parameter σ is $\hat{\sigma} = \sqrt{MSE} = 1.25$. The whole model test leads to $F_{Model} = \frac{302/24}{28/18} = 8.0893$ with p-value= $p\left(F_{8,18} > 8.0893\right) = 0.0001$. Therefore the model is significative with $\alpha = 0.05$. The determination coefficient of the model is $R^2 = 1 - \frac{28}{386/3} = 78.2\%$ and the variation coefficient is $CV = \frac{\sqrt{14}/3}{187/9} = 6.0\%$.

The evaluation of the three F-tests in the ANOVA table leads to:

$$F_A = \frac{64}{14} = 4.57 \text{ with p-value= } p\left(F_{2,18} > 4.57\right) = 0.0248$$
$$F_B = \frac{7}{14} = 0.50 \text{ with p-value= } p\left(F_{2,18} > 0.50\right) = 0.6147$$
$$F_{AB} = \frac{191}{14} = 13.64 \text{ with p-value= } p\left(F_{4,18} > 13.64\right) = 0.0000$$

Therefore the interaction between factor A and factor B is significative with $\alpha = 0.05$, and the effect of the nitrogen doses depends on the potassium doses and upside. Note that the main effect of potassium dose is not significant, but we can not say that potassium fertilization has no effect on growth, because the interaction is significant. To assist in interpreting the interaction effect, it is helpful plotting a graph of the means at each treatment combination. This graph is shown below:



The significant interaction is indicated by the lack of parallelism of the lines. By using the low dose of potassium, the mean of production increases with higher doses of nitrogen. Inversely, with the higher dose of potassium, the mean of production decreases with higher doses of nitrogen. With the intermediate dose of potassium, the mean of production first grows and then decreases. The higher mean of production is obtained with intermediate doses of nitrogen and potassium (120 and 200 kg/ha respectively).

The standard error for the mean of each combination of doses is $SE = \sqrt{\frac{14/9}{3}} = 0.72$ and, taking into account that $t_{18;0.025} = 2.101$, the 95% confidence intervals for the expected values are $\bar{y}_{ij\cdot} \pm 1.51$. These intervals are included in the previous figure. Averaging in doses of potassium, the standard error for the mean of each nitrogen dose is $SE = \sqrt{\frac{14/9}{9}} = 0.42$ and the 95% confidence intervals for the expected values are $\bar{y}_{i\cdot\cdot} \pm 0.87$. In a similar way, averaging in doses of nitrogen, the same standard error is obtained and the 95% confidence intervals for the expected values are $\bar{y}_{\cdot j\cdot} \pm 0.87$.

Finally, the least significant difference (LSD) with $\alpha = 0.05$ for each pair of means $\bar{y}_{ij\cdot}, \bar{y}_{i'j'\cdot}$ is $LSD = t_{18;0.025}\sqrt{\frac{2MSE}{3}} = 2.14$. In a similar way, the least significant difference (LSD) with $\alpha = 0.05$ for each pair of means $\bar{y}_{i\cdot\cdot}, \bar{y}_{i'\cdot\cdot}$ or $\bar{y}_{\cdot j\cdot}, \bar{y}_{\cdot j'\cdot}$ is $LSD = t_{18;0.025}\sqrt{\frac{2MSE}{9}} = 1.24$. Using these values, we obtain the following results for the LSD Fisher's tests of the interaction and the main effects, with $\alpha = 0.05$:

	Potassium= 100	Potassium= 200	Potassium= 300	
Nitrogen= 60	17.33 Bc	19.67 Bb	22.67 Aa	19.89 B
Nitrogen= 120	21.33 Ab	23.67 Aa	20.00 Bb	21.67 A
Nitrogen= 180	23.00 Aa	20.00 Bb	19.33 Bb	20.78 AB
	20.56 a	21.11 a	20.67 a	

where we have used uppercase letters for nitrogen doses and lowercase letters for potassium doses.

The complete LSD Fisher's test with all the nine means leads to:

Fertilization	Mean	
60 – 100	17.33	A
180 – 300	19.33	AB
60 – 200	19.67	B
120 – 300	20.00	B
180 – 200	20.00	B
120 – 100	21.33	BC
60 – 300	22.67	CD
180 – 100	23.00	CD
120 – 200	23.67	D

or, alternatively,

Fertilization	Mean	
60 – 100	17.33	X
180 – 300	19.33	XX
60 – 200	19.67	$\ X$
120 – 300	20.00	$\ X$
180 – 200	20.00	$\ X$
120 – 100	21.33	$\ XX$
60 – 300	22.67	$\ \ XX$
180 – 100	23.00	$\ \ XX$
120 – 200	23.67	$\ \ \ X$

Chapter 4

NESTED ANOVA

In certain multifactor experiments, the levels of one factor (e.g., factor B) occurs in conjunction with only one level of another factor (e.g., A). Then, we say that the levels of factor B are nested within the factor A. Let us suppose that y_{ijk} with $i = 1, ..., a$, $j = 1, ..., b$ and $k = 1, ..., n$ are the observed values for ab independent random samples of the dependent variable Y where a is the number of levels of the factor A and b is the number of levels of the factor B nested whitin each level of the factor A. As a consequence, the total number of levels for the factor B is ab and they have been nested whitin the levels of factor A in a balanced way. That is, we have a dependent variable Y and two factors: A with a levels and B with ab levels balanced way nested in factor A. Moreover, we suppose that $Y_{ijk} \rightsquigarrow N\left(\mu_{j(i)}, \sigma\right)$ and $\mu = \frac{1}{ab}\sum\limits_{i,j}\mu_{j(i)}$ is the global expected value if the factors A and B have no effect on the dependent variable Y_{ijk}.

This model states that $E\left(Y_{ijk}\right) = \mu_{ij} = \mu + \alpha_i + \beta_{j(i)}$, where α_i is the effect due to level i of factor A and $\beta_{j(i)}$ is the effect due to level j of factor B nested within level i of factor A. Therefore, $Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \rightsquigarrow N(0, \sigma)$ and ε_{ijk} independent of $\varepsilon_{i'j'k'}$ for any values i, i', j, j', k, k'. Note that in this model the effect of the nested factor B is measured within each level of factor A and it is supposed that $\sum\limits_j \beta_{j(i)} = 0$ for each value i of the factor A (there are a restrictions for the levels of factor B instead of a unique restriction $\sum\limits_{i,j}\beta_{j(i)} = 0$). Using the same notation as in the previous model, the appropriate estimators for the parameters of the model are $\hat{\mu} = \overline{Y}_{\cdots}$, $\hat{\mu}_{ij} = \overline{Y}_{ij\cdot}$, $\hat{\alpha}_i = \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}$, $\hat{\beta}_{j(i)} = \overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}$, with estimated residuals $\hat{\varepsilon}_{ijk} = e_{ijk} = Y_{ijk} - \hat{\mu}_{ij}$, and the decomposition of sum squares is now

$$\sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{\cdots}\right)^2 = bn\sum_i\left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}\right)^2 + n\sum_{i,j}\left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right)^2 + \sum_{i,j,k}\left(Y_{ijk} - \overline{Y}_{ij\cdot}\right)^2$$

or briefly $SST = SS_A + SS_{B(A)} + SSE$ with $SST = \sum\limits_{i,j,k}\left(Y_{ijk} - \overline{Y}_{\cdots}\right)^2$, $SS_A = bn\sum\limits_i\left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdots}\right)^2$, $SS_{B(A)} = n\sum\limits_{i,j}\left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right)^2$ and $SSE = \sum\limits_{i,j,k}\left(Y_{ijk} - \overline{Y}_{ij\cdot}\right)^2$. As usual, we can also write $SST = \sum\limits_{i,j,k}\left(Y_{ij} - \hat{\mu}\right)^2$, $SSE = \sum\limits_{i,j,k}\left(Y_{ijk} - \hat{\mu}_{ij}\right)^2$ and now $SS_A = bn\sum\limits_i\hat{\alpha}_i^2$ and $SS_{B(A)} = n\sum\limits_{i,j}\hat{\beta}_{j(i)}^2$.

In this model, we also have $\dfrac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{ab(n-1)}$ and the estimation of σ^2 is $\widehat{\sigma^2} = MSE = \dfrac{SSE}{ab(n-1)}$.

Using $SS_{Model} = SS_A + SS_{B(A)}$, if $\alpha_i = \beta_{j(i)} = 0$ for all i, j then

$$F_{Model} = \frac{SS_{Model}/\left(ab-1\right)}{SSE/\left(ab(n-1)\right)} \rightsquigarrow F_{ab-1,ab(n-1)}$$

and we have a whole model test. The R^2 and CV coefficients are evaluated as usual.

In a similar way, if $\alpha_i = 0$ for all i then

$$F_A = \frac{SS_A/\left(a-1\right)}{SSE/\left(ab(n-1)\right)} \rightsquigarrow F_{a-1,ab(n-1)}$$

And, if $\beta_{j(i)} = 0$ for all i, j then

$$F_{B(A)} = \frac{SS_{B(A)}/\left(a(b-1)\right)}{SSE/\left(ab(n-1)\right)} \rightsquigarrow F_{a(b-1),ab(n-1)}$$

These statistics can be used for the two F-test in the ANOVA table, as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\alpha_i = 0$ for all i
Factor $B(A)$	$SS_{B(A)}$	$a(b-1)$	$MS_{B(A)}$	$F_{B(A)}$	$p_{B(A)}$	$\beta_{j(i)} = 0$ for all i, j
Error	SSE	$ab(n-1)$	MSE			
Total	SST	$abn-1$				

Now the confidence intervals for the expected values $\mu + \alpha_i$ and μ_{ij} are $\overline{Y}_{i\cdot\cdot} \pm t_{ab(n-1);\alpha/2}\sqrt{\frac{MSE}{bn}}$ and $\overline{Y}_{ij\cdot} \pm t_{ab(n-1);\alpha/2}\sqrt{\frac{MSE}{n}}$, respectively.

Finally, the least significant differences (LSD) for each pair of means are: $LSD = t_{ab(n-1);\alpha/2}\sqrt{\frac{2MSE}{bn}}$ for $\overline{Y}_{i\cdot\cdot}, \overline{Y}_{i'\cdot\cdot}$ and $LSD = t_{ab(n-1);\alpha/2}\sqrt{\frac{2MSE}{n}}$ for $\overline{Y}_{ij\cdot}, \overline{Y}_{ij'\cdot}$ within each level i of the factor A. Note that we do not initially have a test for the expected values for two levels in the factor B nested within different levels of the factor A, that is, for example, to compare $\beta_{j(i)}$ and $\beta_{j'(i')}$ with $i \neq i'$. This is because if $i = i'$ we have $\hat{\beta}_{j(i)} - \hat{\beta}_{j'(i)} = \overline{Y}_{ij\cdot} - \overline{Y}_{ij'\cdot}$, but this is not true if $i \neq i'$ since now $\hat{\beta}_{j(i)} - \hat{\beta}_{j'(i')} = \left(\overline{Y}_{ij\cdot} - \overline{Y}_{i'j'\cdot}\right) - \left(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}\right)$. Therefore, to carry out this test we need to use $\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}$ versus $\overline{Y}_{i'j'\cdot} - \overline{Y}_{i'\cdot\cdot}$. Taking into account that $\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}$ is independent of $\overline{Y}_{i'j'\cdot} - \overline{Y}_{i'\cdot\cdot}$ and, for fixed values i and j_0, we have $\overline{Y}_{ij_0\cdot} - \overline{Y}_{i\cdot\cdot} = \overline{Y}_{ij_0\cdot} - \frac{1}{b}\sum_j \overline{Y}_{ij\cdot} = \left(1 - \frac{1}{b}\right)\overline{Y}_{ij_0\cdot} - \frac{1}{b}\sum_{j\neq j_0}\overline{Y}_{ij\cdot}$, then $Var\left(\overline{Y}_{ij_0\cdot} - \overline{Y}_{i\cdot\cdot}\right) = \left(1 - \frac{1}{b}\right)^2\left(\frac{\sigma^2}{n}\right) + \frac{b-1}{b^2}\left(\frac{\sigma^2}{n}\right) = \left(1 - \frac{1}{b}\right)\left(\frac{\sigma^2}{n}\right)$ and, we can prove that

$$\frac{\left(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot}\right) - \left(\overline{Y}_{i'j'\cdot} - \overline{Y}_{i'\cdot\cdot}\right) - \left(\beta_{j(i)} - \beta_{j'(i')}\right)}{\sqrt{\left(1 - \frac{1}{b}\right)\frac{2MSE}{n}}} \rightsquigarrow t_{ab(n-1)}$$

This statistic let us to obtain a test for $H_o : \beta_{j(i)} = \beta_{j'(i')}$.

The check of the assumptions of the model can be made as usual with the estimated residuals e_{ijk}.

Example 5: Two nested factors design. We want to study the effect of three fertilizers (factor A) and nine irrigation doses (factor B) on the growth of potted plants. It is recognized that the two factors are independent (no interaction) and we decide to perform five repetitions. Because of the independence between the factors and in order to use a lower number of experimental units, we decide to nest the irrigation doses within the fertilizers and to limit the experiment to 45 pots (instead of the 135 pots required in the factorial design). The irrigation doses nested within each fertilizer were randomly selected. The plant heights at the end of the experiment were recorded (cm).

Fertilizer	Irrigation dose						$\sum_k y_{ijk}$	$\sum_k y_{ijk}^2$	$\sum_{j,k} y_{ijk}$	$\sum_{j,k} y_{ijk}^2$
1	1 (1)	39	43	41	46	48	217	9471		
1	2 (2)	52	58	55	50	55	270	14618	813	45425
1	3 (3)	69	62	71	61	63	326	21336		
2	1 (4)	72	65	69	62	71	339	23055		
2	2 (5)	61	63	55	52	60	291	17019	921	57229
2	3 (6)	55	50	55	69	62	291	17155		
3	1 (7)	65	69	62	71	61	328	21592		
3	2 (8)	63	55	48	52	58	276	15366	895	54113
3	3 (9)	55	50	55	69	62	291	17155		

For the statistical analysis we use the two-factors nested model $E(Y_{ij}) = \mu_{ij} = \mu + \alpha_i + \beta_{j(i)}$, where α_i is the effect due to the i-fertilizer and $\beta_{j(i)}$ is the effect due to j-dose of irrigation nested within the i-fertilizer, with $i = 1, 2, 3$ and $j = 1, 2, 3$. Therefore, $Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \rightsquigarrow N(0, \sigma)$ and ε_{ijk} independent of $\varepsilon_{i'j'k'}$ for any values i, i', j, j', k, k' $(k = 1, ..., 5)$.

Initially, we calculate $\sum_{i,j,k} y_{ijk} = 2629$, $\sum_{i,j,k} y_{ijk}^2 = 156767$ and all the means observed in the experiment, which are included in the following table with the usual notation:

$\bar{y}_{11\cdot}$	$\bar{y}_{12\cdot}$	$\bar{y}_{13\cdot}$	$\bar{y}_{21\cdot}$	$\bar{y}_{22\cdot}$	$\bar{y}_{23\cdot}$	$\bar{y}_{31\cdot}$	$\bar{y}_{32\cdot}$	$\bar{y}_{33\cdot}$	$\bar{y}_{1\cdot\cdot}$	$\bar{y}_{2\cdot\cdot}$	$\bar{y}_{3\cdot\cdot}$	$\bar{y}_{\cdot\cdot\cdot}$
$\frac{217}{5}$	$\frac{270}{5}$	$\frac{326}{5}$	$\frac{339}{5}$	$\frac{291}{5}$	$\frac{291}{5}$	$\frac{328}{5}$	$\frac{276}{5}$	$\frac{291}{5}$	$\frac{813}{15}$	$\frac{921}{15}$	$\frac{895}{15}$	$\frac{2629}{45}$

From these data we can obtain the estimated parameters of the model. For example, $\hat{\alpha}_1 = \bar{y}_{1\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot} = \frac{-190}{45}$ and $\hat{\beta}_{1(1)} = \bar{y}_{11\cdot} - \bar{y}_{1\cdot\cdot} = \frac{-162}{15}$. In a similar way, the estimations of the other parameters are evaluated and they are included in the following table:

$\hat{\beta}_{1(1)}$	$\hat{\beta}_{2(1)}$	$\hat{\beta}_{3(1)}$	$\hat{\beta}_{1(2)}$	$\hat{\beta}_{2(2)}$	$\hat{\beta}_{3(2)}$	$\hat{\beta}_{1(3)}$	$\hat{\beta}_{2(3)}$	$\hat{\beta}_{3(3)}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\mu} = \bar{y}_{\cdot\cdot\cdot}$
$\frac{-162}{15}$	$\frac{-3}{15}$	$\frac{165}{15}$	$\frac{96}{15}$	$\frac{-48}{15}$	$\frac{-48}{15}$	$\frac{89}{15}$	$\frac{-67}{15}$	$\frac{-22}{15}$	$\frac{-190}{45}$	$\frac{134}{45}$	$\frac{56}{45}$	$\frac{2629}{45}$

Calculating the sum of squares we obtain: $SST = 156767 - 45\bar{y}_{\cdot\cdot\cdot}^2 = \frac{142874}{45}$, $SS_A = 15\sum_i \hat{\alpha}_i^2 = \frac{19064}{45}$, $SS_{B(A)} = 5\sum_{i,j} \hat{\beta}_{j(i)}^2 = \frac{80196}{45}$ and $SSE = \frac{142874-19064-80196}{45} = \frac{43614}{45}$. Therefore, $MSE = \frac{43614}{1620} = 26.92$ and the estimated value for the parameter σ is $\hat{\sigma} = \sqrt{MSE} = 5.19$. The whole model test leads to $F_{Model} = \frac{99260/360}{43614/1620} = 10.24$ with p-value= $p(F_{8,36} > 10.24) = 0.0000$. Therefore the model is significative with $\alpha = 0.05$. The determination coefficient of the model is $R^2 = \frac{99260}{142874} = 69.5\%$ and the variation coefficient is $CV = \frac{5.19}{2629/45} = 8.9\%$.

The evaluation of the two F-tests in the ANOVA table leads to:

$$F_A = \frac{19064/90}{43614/1620} = 7.87 \text{ with p-value} = p(F_{2,36} > 7.87) = 0.0015$$
$$F_{B(A)} = \frac{80196/270}{43614/1620} = 11.03 \text{ with p-value} = p(F_{2,36} > 11.03) = 0.0000$$

Therefore, both factors in the experiment are significative with $\alpha = 0.05$.

The standard error for the means of the nine treatments is $SE = \sqrt{\frac{MSE}{5}} = 2.32$ and, taking into account that $t_{36;0.025} = 2.028$, the 95% confidence intervals for the expected values μ_{ij} are $\bar{y}_{ij\cdot} \pm 4.71$. The least significant difference (LSD) with $\alpha = 0.05$ for the comparison of the estimations of μ_{ij}, that is, for each pair of means $\bar{y}_{ij\cdot}, \bar{y}_{i'j'\cdot}$, is $LSD = t_{36;0.025}\sqrt{\frac{2MSE}{5}} = 6.66$. Using this value, we obtain the following results for the LSD Fisher's tests of the nine treatments, with $\alpha = 0.05$:

Fertilizer	Irrigation dose	Mean ($\bar{y}_{ij\cdot}$)	
1	1 (1)	43.4	A
1	2 (2)	54.0	B
3	2 (8)	55.2	B
2	2 (5)	58.2	B
2	3 (6)	58.2	B
3	3 (9)	58.2	B
1	3 (3)	65.2	C
3	1 (7)	65.6	C
2	1 (4)	67.8	C

Therefore, the best growth is obtained with the dose 1 within the fertilizer 2 (that is, the fourth dose) but no significative differences with the dose 1 within the fertilizer 3 (that is, the seventh dose), and with the dose 3 within the fertilizer 1 (that is, the third dose). All the other treatments lead to a lower expected growth.

Nevertheless, as we said before, the comparison of the estimated expected values (that is, $\hat{\mu}_{ij} - \hat{\mu}_{i'j'} = \bar{y}_{ij\cdot} - \bar{y}_{i'j'\cdot}$) is different of the comparison of the estimated effects of the irrigation doses (that is, $\hat{\beta}_{j(i)} - \hat{\beta}_{j'(i')} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{i'j'\cdot} + \bar{y}_{i'\cdot\cdot}$), if $i \neq i'$. In this case, the least significant difference (LSD) with $\alpha = 0.05$ for the comparison of the estimated effects is $LSD = t_{36;0.025}\sqrt{\left(1 - \frac{1}{3}\right)\frac{2MSE}{5}} = 5.43$. Then the comparison of estimated effects with $i \neq i'$ leads to the following results:

$\hat{\beta}_{1(1)}-\hat{\beta}_{1(2)}$	$\hat{\beta}_{1(1)}-\hat{\beta}_{2(2)}$	$\hat{\beta}_{1(1)}-\hat{\beta}_{3(2)}$	$\hat{\beta}_{1(1)}-\hat{\beta}_{1(3)}$	$\hat{\beta}_{1(1)}-\hat{\beta}_{2(3)}$	$\hat{\beta}_{1(1)}-\hat{\beta}_{3(3)}$	$\hat{\beta}_{2(1)}-\hat{\beta}_{1(2)}$	$\hat{\beta}_{2(1)}-\hat{\beta}_{2(2)}$	$\hat{\beta}_{2(1)}-\hat{\beta}_{3(2)}$
-17.2^*	-7.6^*	-7.6^*	-16.7^*	-6.3^*	-9.3^*	-6.6^*	3	3

$\hat{\beta}_{2(1)}-\hat{\beta}_{1(3)}$	$\hat{\beta}_{2(1)}-\hat{\beta}_{2(3)}$	$\hat{\beta}_{2(1)}-\hat{\beta}_{3(3)}$	$\hat{\beta}_{3(1)}-\hat{\beta}_{1(2)}$	$\hat{\beta}_{3(1)}-\hat{\beta}_{2(2)}$	$\hat{\beta}_{3(1)}-\hat{\beta}_{3(2)}$	$\hat{\beta}_{3(1)}-\hat{\beta}_{1(3)}$	$\hat{\beta}_{3(1)}-\hat{\beta}_{2(3)}$	$\hat{\beta}_{3(1)}-\hat{\beta}_{3(3)}$
-6.1^*	4.3	1.3	4.6	14.2^*	14.2^*	5.1	15.5^*	12.5^*

$\hat{\beta}_{1(2)}-\hat{\beta}_{1(3)}$	$\hat{\beta}_{1(2)}-\hat{\beta}_{2(3)}$	$\hat{\beta}_{1(2)}-\hat{\beta}_{3(3)}$	$\hat{\beta}_{2(2)}-\hat{\beta}_{1(3)}$	$\hat{\beta}_{2(2)}-\hat{\beta}_{2(3)}$	$\hat{\beta}_{2(2)}-\hat{\beta}_{3(3)}$	$\hat{\beta}_{3(2)}-\hat{\beta}_{1(3)}$	$\hat{\beta}_{3(2)}-\hat{\beta}_{2(3)}$	$\hat{\beta}_{3(2)}-\hat{\beta}_{3(3)}$
0.5	10.9^*	7.9^*	-9.1^*	1.3	-1.7	-9.1^*	1.3	-1.7

where the differences marked with an asterisk are significant with $\alpha = 0.05$. With this significance level we have not observed any difference between this results and the others obtained for the comparisons of treatments. But if we suppose $\alpha = 0.22$, as $t_{36;0.11} = 1.25$, the least significant difference for the comparison of treatments is $LSD = t_{36;0.05}\sqrt{\frac{2MSE}{5}} = 4.10$ and therefore μ_{12} (second dose with the fertilizer 1) is significantly different of μ_{33} (ninth dose with the fertilizer 3), because the observed difference is 4.2. But nevertheless, the least significant difference for the comparison of effects nested within different fertilizers $LSD = t_{36;0.11}\sqrt{\left(1 - \frac{1}{3}\right)\frac{2MSE}{5}} = 3.35$ and therefore $\beta_{2(1)}$ (effect of the second dose, which is nested within the fertilizer 1) is not significantly different of $\beta_{3(3)}$ (effect of the ninth dose, which is nested within the fertilizer 3), because $\hat{\beta}_{2(1)} - \hat{\beta}_{3(3)} = 1.3$.

Finally, for the fertilizers, the standard error of the means is $SE = \sqrt{\frac{MSE}{15}} = 1.34$ and the 95% confidence intervals for the expected values $\mu + \alpha_i$ are $\bar{y}_{i\cdot\cdot} \pm 2.72$. The least significant difference (LSD) with $\alpha = 0.05$ for the comparison of the estimations of $\mu + \alpha_i$, that is, for each pair of means $\bar{y}_{i\cdot\cdot}, \bar{y}_{i'\cdot\cdot}$, is $LSD = t_{36;0.025}\sqrt{\frac{2MSE}{5}} = 3.84$. Using this value, we obtain the following results for the LSD Fisher's tests of the three fertilizers, with $\alpha = 0.05$:

Fertilizer	Mean	
1	54.20	*A*
3	59.67	*B*
2	61.40	*B*

and we can say that the expected growth is lower with the fertilizer 1, with no significant differences between the fertilizers 2 and 3.

Chapter 5

CROSSED-NESTED ANOVA

Occasionally in a multifactor experiment, some factors are arranged in a factorial layout and other factors are nested. This type of designs are usually called crossed–nested designs. The statistical analysis of one such design with three factors is now illustrated. Let us suppose that y_{ijkl} are the observed values for abc independent random samples of size n from a dependent variable Y where a is the number of levels of factor A, b is the number of levels of factor B and c is the number of levels of factor C nested whitin the levels of the factor B in a balanced way (as a consequence, bc is the total number of levels in factor C). As usual, we suppose that $Y_{ijkl} \rightsquigarrow N\left(\mu_{ijk}, \sigma\right)$ and $\mu = \frac{1}{abc} \sum\limits_{i,j,k} \mu_{ijk}$ is the global expected value if none of the factors A, B and C have any effect on the dependent variable Y_{ijkl}.

In this model, we state that $E\left(Y_{ijk}\right) = \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{k(j)} + \alpha\beta_{ij} + \alpha\gamma_{ik(j)}$, where α_i is the effect due to level i of factor A, β_j is the effect due to level j of factor B, $\gamma_{k(j)}$ is the effect due to level k of factor C nested within level j of factor B and $\alpha\beta_{ij}$, $\alpha\gamma_{ik(j)}$ are the interaction effects between pair of factors A, B and A, C respectively. Note that we can not consider an interaction effect between factors B and C due to nesting used in these factors. In addition, we have now many restrictions on the parameters, namely: $\sum\limits_i \alpha_i = 0$, $\sum\limits_j \beta_j = 0$, $\sum\limits_k \gamma_{k(j)} = 0$ for all j, $\sum\limits_j \alpha\beta_{ij} = 0$ for all i, $\sum\limits_i \alpha\beta_{ij} = 0$ for all j, $\sum\limits_k \alpha\gamma_{ik(j)} = 0$ for all i, j and $\sum\limits_i \alpha\gamma_{ik(j)} = 0$ for all j, k. Therefore, the model is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_{k(j)} + \alpha\beta_{ij} + \alpha\gamma_{ik(j)} + \varepsilon_{ijkl}$$

with $\varepsilon_{ijkl} \rightsquigarrow N(0, \sigma)$ and ε_{ijkl} independent of $\varepsilon_{i'j'k'l'}$ for any values $i, i', j, j', k, k', l, l'$.

With the usual notation, the appropriate estimators for the parameters of the model are $\hat{\mu} = \overline{Y}_{\cdot\cdot\cdot\cdot}$, $\hat{\mu}_{ijk} = \overline{Y}_{ijk\cdot}$, $\hat{\alpha}_i = \overline{Y}_{i\cdot\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot\cdot}$, $\hat{\beta}_j = \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot\cdot}$, $\widehat{\alpha\beta}_{ij} = \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i\cdot\cdot\cdot} - \overline{Y}_{\cdot j\cdot\cdot} + \overline{Y}_{\cdot\cdot\cdot\cdot}$, $\hat{\gamma}_{k(j)} = \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j\cdot\cdot}$ and $\widehat{\alpha\gamma}_{ik(j)} = \overline{Y}_{ijk\cdot} - \overline{Y}_{ij\cdot} - \overline{Y}_{\cdot jk\cdot} + \overline{Y}_{\cdot j\cdot\cdot}$, with estimated residuales $\hat{\varepsilon}_{ijkl} = e_{ijkl} = Y_{ijkl} - \hat{\mu}_{ijk}$. The decomposition of sum squares is now

$$SST = SS_A + SS_B + SS_{C(B)} + SS_{AB} + SS_{AC(B)} + SSE$$

where $SST = \sum\limits_{i,j,k,l} \left(Y_{ijkl} - \overline{Y}_{\cdot\cdot\cdot\cdot}\right)^2$, $SS_A = bcn \sum\limits_i \left(\overline{Y}_{i\cdot\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot\cdot}\right)^2$, $SS_B = acn \sum\limits_j \left(\overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot\cdot}\right)^2$, $SS_{C(B)} = an \sum\limits_{j,k} \left(\overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j\cdot\cdot}\right)^2$, $SS_{AB} = cn \sum\limits_{i,j} \left(\overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i\cdot\cdot\cdot} - \overline{Y}_{\cdot j\cdot\cdot} + \overline{Y}_{\cdot\cdot\cdot\cdot}\right)^2$, $SS_{AC(B)} = n \sum\limits_{i,j,k} \left(\overline{Y}_{ijk\cdot} - \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{\cdot jk\cdot} + \overline{Y}_{\cdot j\cdot\cdot}\right)^2$

and $SSE = \sum_{i,j,k,l}\left(Y_{ijkl} - \overline{Y}_{ijk\cdot}\right)^2$. As usual, we can also write $SST = \sum_{i,j,k,l}\left(Y_{ij} - \hat{\mu}\right)^2$, $SSE = \sum_{i,j,k,l}\left(Y_{ijkl} - \hat{\mu}_{ijk}\right)^2$ and now $SS_A = bcn\sum_i \hat{\alpha}_i^2$, $SS_B = acn\sum_j \hat{\beta}_j^2$, $SS_{C(B)} = an\sum_{j,k}\hat{\gamma}_{k(j)}^2$, $SS_{AB} = cn\sum_{i,j}\left(\widehat{\alpha\beta}_{ij}\right)^2$ and $SS_{AC(B)} = n\sum_{i,j,k}\left(\widehat{\alpha\gamma}_{ik(j)}\right)^2$.

In this model, we have $\frac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{abc(n-1)}$ and the estimation of σ^2 is $\widehat{\sigma^2} = MSE = \frac{SSE}{abc(n-1)}$. Then, using $SS_{Model} = SS_A + SS_B + SS_{C(B)} + SS_{AB} + SS_{AC(B)}$, the whole model test with the null hypothesis $\alpha_i = \beta_j = \gamma_{k(j)} = \alpha\beta_{ij} = \alpha\gamma_{ik(j)} = 0$ for all i, j, k is given by

$$F_{Model} = \frac{SS_{Model}/\left(abc-1\right)}{SSE/\left(abc(n-1)\right)} \rightsquigarrow F_{abc-1,abc(n-1)}$$

and the R^2 and CV coefficients are evaluated as usual.

The appropriate statistics for the F-tests in the ANOVA table with their respective probability distributions are:

$$\begin{aligned}
F_A &= \frac{SS_A/\left(a-1\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_A}{MSE} \rightsquigarrow F_{a-1,abc(n-1)} && \text{if } \alpha_i = 0 \text{ for all } i\\
F_B &= \frac{SS_B/\left(b-1\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_B}{MSE} \rightsquigarrow F_{b-1,abc(n-1)} && \text{if } \beta_j = 0 \text{ for all } j\\
F_{C(B)} &= \frac{SS_{C(B)}/\left(b(c-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{C(B)}}{MSE} \rightsquigarrow F_{b(c-1),abc(n-1)} && \text{if } \gamma_{k(j)} = 0 \text{ for all } j,k\\
F_{AB} &= \frac{SS_{AB}/\left((a-1)(b-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{AB}}{MSE} \rightsquigarrow F_{(a-1)(b-1),abc(n-1)} && \text{if } \alpha\beta_{ij} = 0 \text{ for all } i,j\\
F_{AC(B)} &= \frac{SS_{AC(B)}/\left(b(a-1)(c-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{AC(B)}}{MSE} \rightsquigarrow F_{b(a-1)(c-1),abc(n-1)} && \text{if } \alpha\gamma_{ik(j)} = 0 \text{ for all } i,j,k
\end{aligned}$$

The ANOVA table for this model is as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\alpha_i = 0$ for all i
Factor B	SS_B	$b-1$	MS_B	F_B	p_B	$\beta_j = 0$ for all j
Factor $C(B)$	$SS_{C(B)}$	$b(c-1)$	$MS_{C(B)}$	$F_{C(B)}$	$p_{C(B)}$	$\gamma_{k(j)} = 0$ for all j,k
Int. $A*B$	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_{AB}	p_{AB}	$\alpha\beta_{ij} = 0$ for all i,j
Int. $A*C(B)$	$SS_{AC(B)}$	$b(a-1)(c-1)$	$MS_{AC(B)}$	$F_{AC(B)}$	$p_{AC(B)}$	$\alpha\gamma_{ik(j)} = 0$ for all i,j,k
Error	SSE	$abc(n-1)$	MSE			
Total	SST	$abcn-1$				

Now the confidence intervals for the marginal expected values are:

$$\begin{aligned}
&\overline{Y}_{i\cdot\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{bcn}} && \text{for } \mu + \alpha_i\\
&\overline{Y}_{\cdot j\cdot\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{acn}} && \text{for } \mu + \beta_j\\
&\overline{Y}_{ij\cdot\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{cn}} && \text{for } \mu_{ij\cdot} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}\\
&\overline{Y}_{\cdot jk\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{an}} && \text{for } \mu_{\cdot k(j)} = \mu + \beta_j + \gamma_{k(j)}\\
&\overline{Y}_{ijk\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{n}} && \text{for } \mu_{ijk}
\end{aligned}$$

Finally, the least significant differences (LSD) for each pair of means are:

$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{bcn}} \quad \text{for } \overline{Y}_{i\cdots} - \overline{Y}_{i'\cdots}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{acn}} \quad \text{for } \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot j'\cdot\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{cn}} \quad \text{for } \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i'j'\cdot\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{an}} \quad \text{for } \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot jk'\cdot} \text{ with fixed } j$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{n}} \quad \text{for } \overline{Y}_{ijk\cdot\cdot} - \overline{Y}_{i'jk'\cdot} \text{ with fixed } j$$

Arguing as in the previous model, we obtain that the appropriate statistic for the test $\gamma_{k(j)} = \gamma_{k'(j')}$ with $j \neq j'$ is

$$\frac{\left(\overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j\cdot\cdot}\right) - \left(\overline{Y}_{\cdot j'k'\cdot} - \overline{Y}_{\cdot j'\cdot\cdot}\right) - \left(\gamma_{k(j)} - \gamma_{k'(j')}\right)}{\sqrt{\left(1-\frac{1}{c}\right)\frac{2MSE}{an}}} \rightsquigarrow t_{abc(n-1)}$$

In a similar way, the appropriate statistic for the test $\gamma_{ik(j)} = \gamma_{i'k'(j')}$ with $j \neq j'$ is

$$\frac{\left(\overline{Y}_{ijk\cdot} - \overline{Y}_{ij\cdot\cdot}\right) - \left(\overline{Y}_{i'j'k'\cdot} - \overline{Y}_{i'j'\cdot\cdot}\right) - \left(\gamma_{ik(j)} - \gamma_{i'k'(j')}\right)}{\sqrt{\left(1-\frac{1}{c}\right)\frac{2MSE}{n}}} \rightsquigarrow t_{abc(n-1)}$$

An special case of this type of models is the factorial design augmented with an additional control treatment defined as no application of neither of the factors. Next we illustrate this model using a factorial design with two factors and a control treatment. Let us suppose that we have two factors, A with a levels and B with b levels in a complete factorial design with n replicates, and an additional control treatment T, also with n replicates, defined by none of the levels of the factors A and B. Then we consider two levels for factor T: 0 for the control (no treatment) and 1 for the ab treatments in the ab factorial design. Let us denote the observed values by y_{ijkl} with $i = 0, 1$; $j = 0, 1, ..., a$; $k = 0, 1, ..., b$ and $l = 1, ..., n$ where y_{000l} represent the values for the control treatment and y_{1jkl} with $j = 1, ..., a$ and $k = 1, ..., b$ the values for the ab treatments in the model. Note that the factors A and B are nested in factor T with an unbalanced way (level 0 nested within $i = 0$ and the other levels nested within $i = 1$) and the total number of observed values is $n(1 + ab)$. As usual we suppose that $Y_{ijkl} \rightsquigarrow N\left(\mu_{ijk}, \sigma\right)$ and $\mu = \frac{\mu_{000} + \sum_{j,k} \mu_{1jk}}{1+ab}$ is the global expected value if the treatments defined by the factors A and B have no effect on the dependent variable Y_{ijkl}. Therefore we have $1 + ab$ independent random samples of the dependent variable.

The model states that $E\left(Y_{ijk}\right) = \mu_{ijk} = \mu + \tau_i + \alpha_{j(i)} + \beta_{k(i)} + \alpha\beta_{jk(i)}$, and therefore $Y_{ijkl} = \mu + \tau_i + \alpha_{j(i)} + \beta_{k(i)} + \alpha\beta_{jk(i)} + \varepsilon_{ijkl}$ with $\varepsilon_{ijkl} \rightsquigarrow N(0, \sigma)$ and ε_{ijkl} independent of $\varepsilon_{i'j'k'l'}$ for any values $i, i', j, j', k, k', l, l'$, where τ_o is the effect of the control treatment on the global mean μ, τ_1 is the main effect of the use of factors A and B, and the remaining parameters are defined as in a factorial design. The restrictions on the parameters are now: $\tau_o + ab\tau_1 = 0$; $\alpha_{0(0)} = \beta_{0(0)} = \alpha\beta_{j0(0)} = \alpha\beta_{0k(0)} = 0$ for all j, k; $\sum_j \alpha_{j(1)} = 0$, $\sum_k \beta_{k(1)} = 0$, $\sum_k \alpha\beta_{jk(1)} = 0$ for all j and $\sum_j \alpha\beta_{jk(1)} = 0$ for all k (note that we use $\tau_o + ab\tau_1 = 0$ instead of $\tau_o + \tau_1 = 0$ by the unbalanced nesting).

Now we define $\overline{Y}_{....} = \frac{\sum_{l=1}^{n} Y_{000l} + \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{l=1}^{n} Y_{1jkl}}{n(1+ab)}$, $\overline{Y}_{000\cdot} = \frac{1}{n}\sum_{l=1}^{n} Y_{000l}$, $\overline{Y}_{1...} = \frac{1}{abn}\sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{l=1}^{n} Y_{1jkl}$, $\overline{Y}_{1j..} = \frac{1}{bn}\sum_{k=1}^{b}\sum_{l=1}^{n} Y_{1jkl}$, $\overline{Y}_{1\cdot k\cdot} = \frac{1}{an}\sum_{j=1}^{a}\sum_{l=1}^{n} Y_{1jkl}$, $\overline{Y}_{1jk\cdot} = \frac{1}{n}\sum_{l=1}^{n} Y_{1jkl}$. The appropriate estimators for the parameters of the model are $\hat{\mu} = \overline{Y}_{....}$, $\hat{\tau}_o = \overline{Y}_{000\cdot} - \overline{Y}_{....}$, $\hat{\tau}_1 = \overline{Y}_{1...} - \overline{Y}_{....}$, $\hat{\alpha}_{j(1)} = \overline{Y}_{1j..} - \overline{Y}_{1...}$, $\hat{\beta}_{k(1)} = \overline{Y}_{1\cdot k\cdot} - \overline{Y}_{1...}$, $\widehat{\alpha\beta}_{jk(1)} = \overline{Y}_{1jk\cdot} - \overline{Y}_{1j..} - \overline{Y}_{1\cdot k\cdot} + \overline{Y}_{1...}$, $\hat{\mu}_{000} = \hat{\mu} + \hat{\tau}_o = \overline{Y}_{000\cdot}$ and $\hat{\mu}_{1jk} = \hat{\mu} + \hat{\tau}_1 + \hat{\alpha}_{j(1)} + \hat{\beta}_{k(1)} + \widehat{\alpha\beta}_{jk(1)} = \overline{Y}_{1jk\cdot}$, with estimated residuals $\hat{\varepsilon}_{ijkl} = e_{ijkl} = Y_{ijkl} - \hat{\mu}_{ijk}$. The decomposition of sum squares is now

$$SST = SS_T + SS_{A(T)} + SS_{B(T)} + SS_{AB(T)} + SSE$$

where $SST = \sum_{l}\left(Y_{000l} - \overline{Y}_{....}\right)^2 + \sum_{j,k,l}\left(Y_{1jkl} - \overline{Y}_{....}\right)^2$, $SS_T = n\left(\overline{Y}_{000\cdot} - \overline{Y}_{....}\right)^2 + abn\left(\overline{Y}_{1...} - \overline{Y}_{....}\right)^2$, $SS_{A(T)} = bn\sum_{j}\left(\overline{Y}_{1j..} - \overline{Y}_{1...}\right)^2$, $SS_{B(T)} = an\sum_{k}\left(\overline{Y}_{1\cdot k\cdot} - \overline{Y}_{1...}\right)^2$, $SS_{AB(T)} = n\sum_{j,k}\left(\overline{Y}_{1jk\cdot} - \overline{Y}_{1j..} - \overline{Y}_{1\cdot k\cdot} + \overline{Y}_{1...}\right)^2$, and $SSE = \sum_{l}\left(Y_{000l} - \overline{Y}_{000\cdot}\right)^2 + \sum_{j,k,l}\left(Y_{1jkl} - \overline{Y}_{1jk\cdot}\right)^2$ (note that in all these latter summations the subscript 0 is not included). In this case, we also can write $SST = \sum_{i,j,k,l}(Y_{ijkl} - \hat{\mu})^2$, $SSE = \sum_{i,j,k,l}\left(Y_{ijkl} - \hat{\mu}_{ijk}\right)^2$ and now $SS_T = n\hat{\tau}_0^2 + abn\hat{\tau}_1^2$, $SS_{A(T)} = bn\sum_{j}\hat{\alpha}_{j(1)}^2$, $SS_{B(T)} = an\sum_{k}\hat{\beta}_{k(1)}^2$ and $SS_{AB(T)} = n\sum_{j,k}\left(\widehat{\alpha\beta}_{jk(1)}\right)^2$.

In this model, we have $\frac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{(1+ab)(n-1)}$ and the estimation of σ^2 is $\widehat{\sigma^2} = MSE = \frac{SSE}{(1+ab)(n-1)}$. Then, $SS_{Model} = SS_T + SS_{A(T)} + SS_{B(T)} + SS_{AB(T)}$ and the whole model test with the null hypothesis $\tau_i = \alpha_{j(1)} = \beta_{k(1)} = \alpha\beta_{jk(1)} = 0$ for all i, j, k is given by

$$F_{Model} = \frac{SS_{Model}/\,(ab)}{SSE/\,((1+ab)(n-1))} \rightsquigarrow F_{ab,(1+ab)(n-1)}$$

The R^2 and CV coefficients are evaluated as before.

The appropriate statistics for the F-tests in the ANOVA table with their respective probability distributions are:

$$F_T = \frac{SS_T}{SSE/\,((1+ab)(n-1))} = \frac{MS_T}{MSE} \rightsquigarrow F_{1,(1+ab)(n-1)} \qquad \text{if } \tau_0 = \tau_1 = 0$$

$$F_{A(T)} = \frac{SS_{A(T)}/\,(a-1)}{SSE/\,((1+ab)(n-1))} = \frac{MS_{A(T)}}{MSE} \rightsquigarrow F_{a-1,(1+ab)(n-1)} \qquad \text{if } \alpha_{j(1)} = 0 \text{ for all } j$$

$$F_{B(T)} = \frac{SS_{B(T)}/\,(b-1)}{SSE/\,((1+ab)(n-1))} = \frac{MS_{B(T)}}{MSE} \rightsquigarrow F_{b-1,(1+ab)(n-1)} \qquad \text{if } \beta_{k(1)} = 0 \text{ for all } k$$

$$F_{AB(T)} = \frac{SS_{AB(T)}/\,((a-1)(b-1))}{SSE/\,((1+ab)(n-1))} = \frac{MS_{AB(T)}}{MSE} \rightsquigarrow F_{(a-1)(b-1),(1+ab)(n-1)} \qquad \text{if } \alpha\beta_{jk(1)} = 0 \text{ for all } j,k$$

The ANOVA table for this model is as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor T	SS_T	1	MS_T	F_T	p_T	$\tau_0 = \tau_1 = 0$
Factor $A(T)$	$SS_{A(T)}$	$a-1$	$MS_{A(T)}$	$F_{A(T)}$	$p_{A(T)}$	$\alpha_{j(1)} = 0$ for all j
Factor $B(T)$	$SS_{B(T)}$	$b-1$	$MS_{B(T)}$	$F_{B(T)}$	$p_{B(T)}$	$\beta_{k(1)} = 0$ for all k
Int. $A{*}B(T)$	$SS_{AB(T)}$	$(a-1)(b-1)$	$MS_{AB(T)}$	$F_{AB(TB)}$	$p_{AB(T)}$	$\alpha\beta_{jk(1)} = 0$ for all j, k
Error	SSE	$(1+ab)(n-1)$	MSE			
Total	SST	$(1+ab)n-1$				

The confidence intervals for the marginal expected values are:

$$\overline{Y}_{000\cdot} \pm t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{MSE}{n}} \quad \text{for } \mu_{000} = \mu + \tau_0$$
$$\overline{Y}_{1\cdots} \pm t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{MSE}{abn}} \quad \text{for } \mu_{1\cdot\cdot} = \frac{1}{ab}\sum_{j,k}\mu_{1jk} = \mu + \tau_1$$
$$\overline{Y}_{1j\cdot\cdot} \pm t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{MSE}{bn}} \quad \text{for } \mu_{1j\cdot} = \mu + \tau_1 + \alpha_{j(1)}$$
$$\overline{Y}_{1\cdot k\cdot} \pm t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{MSE}{an}} \quad \text{for } \mu_{1\cdot k} = \mu + \tau_1 + \beta_{k(1)}$$
$$\overline{Y}_{1jk\cdot} \pm t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{MSE}{n}} \quad \text{for } \mu_{1jk} = \mu + \tau_1 + \alpha_{j(1)} + \beta_{k(1)} + \alpha\beta_{jk(1)}$$

Finally, the least significant differences (LSD) for each pair of means are:

$$LSD = t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{(1+ab)MSE}{abn}} \quad \text{for } \overline{Y}_{000\cdot} - \overline{Y}_{1\cdots}$$
$$LSD = t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{2MSE}{bn}} \quad \text{for } \overline{Y}_{1j\cdot\cdot} - \overline{Y}_{1j'\cdot\cdot}$$
$$LSD = t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{(1+b)MSE}{bn}} \quad \text{for } \overline{Y}_{000\cdot} - \overline{Y}_{1j\cdot\cdot}$$
$$LSD = t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{2MSE}{an}} \quad \text{for } \overline{Y}_{1\cdot k\cdot} - \overline{Y}_{1\cdot k'\cdot}$$
$$LSD = t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{(1+a)MSE}{an}} \quad \text{for } \overline{Y}_{000\cdot} - \overline{Y}_{1\cdot k\cdot}$$
$$LSD = t_{(1+ab)(n-1);\alpha/2}\sqrt{\frac{2MSE}{n}} \quad \text{for } \overline{Y}_{1jk\cdot} - \overline{Y}_{1j'k'\cdot} \text{ or } \overline{Y}_{000\cdot} - \overline{Y}_{1jk\cdot}$$

The check of the assumptions of the model can be made as usual with the estimated residuals e_{ijkl}.

Example 6: Two-way factorial design augmented with an additional control treatment. In a research study on bread making is proposed fortifying flour with proteins derived from different products. A complete factorial experiment using five sources of protein (factor A) and two doses (factor B), with four replications was designed. Control treatment (T) defined by the non-utilization of protein in bread making was also used. Protein levels for factor A were: no protein (0), gluten (1), pea (2), egg (3), milk (4) and soy (5). Doses levels for factor B were: no protein (0), 5% (1) and 10% (2). The T factor was coded as 0 (no protein) and 1 for all the ten treatments with protein. In this example, the bread volume Y (hundred of cm^3) is used as the dependent variable.

For the analysis we use the crossed-nested model

$$E\left(Y_{ijk}\right)=\mu_{ijk}=\mu+\tau_i+\alpha_{j(i)}+\beta_{k(i)}+\alpha\beta_{jk(i)}$$

with i=0, 1; j=0, 1, 2, 3, 4, 5; k=0, 1, 2 and l=1, 2, 3, 4. That is, just the previous model with $a=5$, $b=2$ and $n=4$.

The obtained data are shown in the following table:

T	A	B	Replications				$\sum_l y_{ijkl}$	$\sum_l y_{ijkl}^2$
0	0	0	9.76	9.88	14.32	15.52	49.48	638.8048
1	1	1	8.32	7.80	7.36	7.68	31.16	243.2144
1	1	2	8.36	7.44	9.60	9.56	34.96	308.7968
1	2	1	14.32	14.12	12.68	12.76	53.88	728.0368
1	2	2	11.60	10.44	10.04	10.64	42.72	457.5648
1	3	1	12.88	12.56	15.84	15.08	56.36	801.9600
1	3	2	12.00	11.92	11.48	11.20	46.60	543.3168
1	4	1	13.12	12.64	15.32	16.68	57.76	844.8288
1	4	2	15.00	14.48	17.92	17.12	64.52	1048.8912
1	5	1	13.28	13.20	12.96	12.08	51.52	664.4864
1	5	2	11.68	12.08	11.04	10.48	45.28	514.0608
			Total sums				$\sum_{i,j,k,l} y_{ijkl}=534.24$	$\sum_{i,j,k,l} y_{ijkl}^2=6793.9616$

Initially, all the observed means are included in the following table with the usual notation:

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	
k=1	$\bar{y}_{111\cdot}$=7.79	$\bar{y}_{121\cdot}$=13.47	$\bar{y}_{131\cdot}$=14.09	$\bar{y}_{141\cdot}$=14.44	$\bar{y}_{151\cdot}$=12.88	$\bar{y}_{1\cdot1\cdot}$=12.534
k=2	$\bar{y}_{112\cdot}$=8.74	$\bar{y}_{122\cdot}$=10.68	$\bar{y}_{132\cdot}$=11.65	$\bar{y}_{142\cdot}$=16.13	$\bar{y}_{152\cdot}$=11.32	$\bar{y}_{1\cdot2\cdot}$=11.704
	$\bar{y}_{11\cdot\cdot}$=8.265	$\bar{y}_{12\cdot\cdot}$=12.075	$\bar{y}_{13\cdot\cdot}$=12.87	$\bar{y}_{14\cdot\cdot}$=15.285	$\bar{y}_{15\cdot\cdot}$=12.1	$\bar{y}_{1\cdots}$=12.119
	$i=j=k=0$		$\bar{y}_{000\cdot}$=12.37			$\bar{y}_{\cdots\cdot}=\frac{133.56}{11}$

From these data we can obtain the estimated parameters of the model. For example, $\hat{\alpha}_{1(1)}$=$\bar{y}_{11\cdot\cdot}-\bar{y}_{1\cdots}$= -3.854, $\hat{\beta}_{1(1)}$=$\bar{y}_{1\cdot1\cdot}-\bar{y}_{1\cdots}$=0.415, $\widehat{\alpha\beta}_{11(1)}$=$\bar{y}_{111\cdot}$–$\bar{y}_{1\cdots}$–$\hat{\alpha}_{1(1)}$–$\hat{\beta}_{1(1)}$=$-0.89$, $\hat{\mu}$=$\bar{y}_{\cdots\cdot}$, $\hat{\tau}_o$= $\bar{y}_{000\cdot}-\bar{y}_{\cdots\cdot}$=0.228, $\hat{\tau}_1$=$\bar{y}_{1\cdots}-\bar{y}_{\cdots\cdot}$=$-0.0228$. In a similar way, the estimations of the other parameters are evaluated and they are included in the following table:

$\widehat{\alpha\beta}_{11(1)}$=$-0.89$	$\widehat{\alpha\beta}_{21(1)}$=0.98	$\widehat{\alpha\beta}_{31(1)}$=0.805	$\widehat{\alpha\beta}_{41(1)}$=$-1.26$	$\widehat{\alpha\beta}_{51(1)}$=0.365	$\hat{\beta}_{1(1)}$=0.415
$\widehat{\alpha\beta}_{12(1)}$=0.89	$\widehat{\alpha\beta}_{22(1)}$=$-0.98$	$\widehat{\alpha\beta}_{32(1)}$=$-0.805$	$\widehat{\alpha\beta}_{42(1)}$=1.26	$\widehat{\alpha\beta}_{52(1)}$=$-0.365$	$\hat{\beta}_{2(1)}$=-0.415
$\hat{\alpha}_{1(1)}$=-3.854	$\hat{\alpha}_{2(1)}$=-0.044	$\hat{\alpha}_{3(1)}$=0.751	$\hat{\alpha}_{4(1)}$=3.166	$\hat{\alpha}_{5(1)}$=-0.019	

Calculating the sum of squares we obtain: $SST=6793.9616-44\bar{y}_{\cdots\cdot}^2=307.3166$, $SS_T=n\hat{\tau}_0^2+abn\hat{\tau}_1^2=0.2291$, $SS_{A(T)}=8\sum_i\hat{\alpha}_i^2=203.5454$, $SS_{B(T)}=20\sum_j\hat{\beta}_j^2=6.889$, $SS_{AB(T)}$=$4\sum_{j,k}\left(\widehat{\alpha\beta}_{jk(1)}\right)^2=32.9708$ and $SSE=307.3166-243.6343=63.6823$. Therefore, $MSE=\frac{63.6823}{33}=1.93$ and the estimated value for the parameter σ is $\hat{\sigma}=\sqrt{MSE}=1.39$. The whole model test leads to $F_{Model}=\frac{243.6343/10}{63.6823/33}=12.63$ with p-value= $p\left(F_{10,33}>12.63\right)=0.0000$. Therefore the model is significative with $\alpha=0.05$. The determination coefficient of the model is $R^2=\frac{243.6343}{307.3166}=79.3\%$ and the variation coefficient is $CV=\frac{1.39}{133.56/11}=11.4\%$.

The evaluation of the F-tests in the ANOVA table leads to:

$$F_T = \frac{0.2291}{63.6823/33} = 0.12 \text{ with p-value} = p\left(F_{1,33} > 0.12\right) = 0.7326$$
$$F_{A(T)} = \frac{203.5454/4}{63.6823/33} = 26.37 \text{ with p-value} = p\left(F_{4,33} > 26.37\right) = 0.0000$$
$$F_{B(T)} = \frac{6.889}{63.6823/33} = 3.57 \text{ with p-value} = p\left(F_{1,33} > 3.57\right) = 0.0677$$
$$F_{AB(T)} = \frac{32.9708/4}{63.6823/33} = 4.27 \text{ with p-value} = p\left(F_{4,33} > 4.27\right) = 0.0068$$

Therefore, the first F-test shows that, averaging in the ten treatments with protein, there is no significative difference with the control treatment (no protein) with $\alpha = 0.05$. In addition, from the last F-test, there is a significant interaction between factors A and B with $\alpha = 0.05$. The $F_{B(T)}$-test is not significative and there is no difference between doses if we average the proteins. Finally, $F_{A(T)}$-test is significative and there are significative differences between proteins if we average the doses.

The standard error for the means of the ten treatments with protein and the control treatment without protein is $SE = \sqrt{\frac{MSE}{4}} = 0.69$ and, taking into account that $t_{33;0.025} = 2.035$, the 95% confidence intervals for the expected values μ_{ijk} are $\bar{y}_{1jk\cdot} \pm 1.41$ and $\bar{y}_{000\cdot} \pm 1.41$. The least significant difference (LSD) with $\alpha = 0.05$ for the comparison of the estimations of μ_{ijk}, that is, for each pair of means $\bar{y}_{ijk\cdot}, \bar{y}_{i'j'k'\cdot}$, is $LSD = t_{33;0.025}\sqrt{\frac{MSE}{2}} = 2.00$. Using this value, we obtain the following results for the LSD Fisher's tests of the ten treatments with protein and the control treatment, with $\alpha = 0.05$:

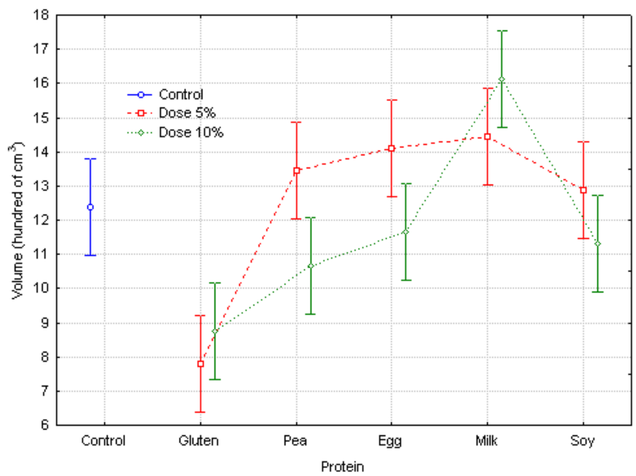
Protein	Source	Dose	Mean ($\bar{y}_{ij\cdot}$)	
1	1 (gluten)	1(5%)	7.79	A
1	1 (gluten)	2(10%)	8.74	AB
1	2 (pea)	2(10%)	10.68	BC
1	5 (soy)	2(10%)	11.32	CD
1	3 (egg)	2(10%)	11.65	CDE
0	0 (no protein)	0	12.37	$CDEF$
1	5 (soy)	1(5%)	12.88	$DEFG$
1	2 (pea)	1(5%)	13.47	EFG
1	3 (egg)	1(5%)	14.09	FG
1	4 (milk)	1(5%)	14.44	GH
1	4 (milk)	2(10%)	16.13	H

Therefore we obtain the following conclusions:

1. With respect to the control treatment, using protein from pea, soy or egg has not a significant effect in the expected value for bread volume.
2. With respect to the control treatment, using protein from gluten leads to a significative lower expected value for bread volume. Generally, this value seems to be also lower than in the other treatments with protein.
3. With respect to the control treatment, using protein from milk leads to a significative higher expected value for bread volume. Generally, this value seems to be also higher than in the other treatments with protein.
4. Using protein from gluten, soy or milk, there are no significative differences between the two doses. Nevertheless, with protein from egg or pea, using the 5%-dose leads to a significative higher expected value for bread volume.

5. The treatment with protein from milk and 10%-dose leads to a significative higher expected value for bread volumen than all the treatments with other protein.

The next plot includes all the 95% confidence intervals for the treatments in the experiment:



If we average in the doses, the standard error for the means $\bar{y}_{1j\cdot\cdot}$ is $SE = \sqrt{\frac{MSE}{8}} = 0.49$ and the 95% confidence intervals for the expected values $\mu_{1j\cdot}$ are $\bar{y}_{1j\cdot\cdot} \pm 0.999$. The least significant difference with $\alpha = 0.05$ for the comparison of the estimations of $\mu_{1j\cdot}$, that is, for each pair of means $\bar{y}_{1j\cdot\cdot}, \bar{y}_{1j'\cdot\cdot}$, is $LSD = t_{33;0.025}\sqrt{\frac{MSE}{4}} = 1.41$. For the comparisons of $\bar{y}_{000\cdot}, \bar{y}_{1j\cdot\cdot}$, the least significant difference is $LSD = t_{33;0.025}\sqrt{\frac{3MSE}{8}} = 1.73$. Using these values, we obtain the following results for the LSD Fisher's tests of proteins if we average in the two doses, with $\alpha = 0.05$:

Protein	Mean	
1 (gluten)	8.265	*A*
2 (pea)	12.075	*B*
5 (soy)	12.1	*B*
0 (no protein)	12.37	*B*
3 (egg)	12.87	*B*
4 (milk)	15.285	*C*

Chapter 6

THREE-WAY FACTORIAL ANOVA

The two-way factorial model studied previously can be extended with three o more factors in a factorial design. Next we illustrate this case with three factors: A with a levels, B with b levels and C with c levels. Let us suppose that all the possible abc treatments have been experimented with n replicates and y_{ijkl} denotes the observed values with $i = 1, ..., a$; $j = 1, ..., b$; $k = 1, ..., c$ and $l = 1, ..., n$. As usual, we suppose that $Y_{ijkl} \rightsquigarrow N\left(\mu_{ijk}, \sigma\right)$ and $\mu = \frac{1}{abc} \sum\limits_{i,j,k} \mu_{ijk}$ is the global expected value if none of the factors A, B and C have any effect on the dependent variable Y_{ijkl}.

In this model we state that $E\left(Y_{ijk}\right) = \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$, where α_i is the effect due to level i of factor A, β_j is the effect due to level j of factor B, γ_k is the effect due to level k of factor C, $\alpha\beta_{ij}$, $\alpha\gamma_{ik}$ and $\beta\gamma_{jk}$ are the interaction effects between pairs of factors (A, B), (A, C) and (B, C) respectively, and $\alpha\beta\gamma_{ijk}$ is a possible triple interaction between the three factors A, B and C. In addition, we have now many restrictions on the parameters, namely: $\sum\limits_i \alpha_i = 0$, $\sum\limits_j \beta_j = 0$, $\sum\limits_k \gamma_k = 0$, $\sum\limits_j \alpha\beta_{ij} = 0$ for all i, $\sum\limits_i \alpha\beta_{ij} = 0$ for all j, $\sum\limits_k \alpha\gamma_{ik} = 0$ for all i, $\sum\limits_i \alpha\gamma_{ik} = 0$ for all k, $\sum\limits_k \beta\gamma_{jk} = 0$ for all j, $\sum\limits_j \beta\gamma_{jk} = 0$ for all k, $\sum\limits_k \alpha\beta\gamma_{ijk} = 0$ for all i, j, $\sum\limits_j \alpha\beta\gamma_{ijk} = 0$ for all i, k and $\sum\limits_i \alpha\beta\gamma_{ijk} = 0$ for all j, k. Therefore, the model is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijkl}$$

with $\varepsilon_{ijkl} \rightsquigarrow N(0, \sigma)$ and ε_{ijkl} independent of $\varepsilon_{i'j'k'l'}$ for any values $i, i', j, j', k, k', l, l'$.

With the usual notation, the appropriate estimators for the parameters of the model are $\hat{\mu} = \overline{Y}_{....}$, $\hat{\mu}_{ijk} = \overline{Y}_{ijk\cdot}$, $\hat{\alpha}_i = \overline{Y}_{i...} - \overline{Y}_{....}$, $\hat{\beta}_j = \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{....}$, $\hat{\gamma}_k = \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{....}$, $\widehat{\alpha\beta}_{ij} = \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i...} - \overline{Y}_{\cdot j\cdot\cdot} + \overline{Y}_{....}$, $\widehat{\alpha\gamma}_{ik} = \overline{Y}_{i\cdot k\cdot} - \overline{Y}_{i...} - \overline{Y}_{\cdot\cdot k\cdot} + \overline{Y}_{....}$, $\widehat{\beta\gamma}_{jk} = \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot\cdot k\cdot} + \overline{Y}_{....}$ and finally

$$\begin{aligned} \widehat{\alpha\beta\gamma}_{ijk} &= (\overline{Y}_{ijk\cdot} - \overline{Y}_{....}) - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - \widehat{\alpha\beta}_{ij} - \widehat{\alpha\gamma}_{ik} - \widehat{\beta\gamma}_{jk} \\ &= \overline{Y}_{ijk\cdot} - \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i\cdot k\cdot} - \overline{Y}_{\cdot jk\cdot} + \overline{Y}_{i...} + \overline{Y}_{\cdot j\cdot\cdot} + \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{....} \end{aligned}$$

The estimated residuals for this model are $\hat{\varepsilon}_{ijkl} = e_{ijkl} = Y_{ijkl} - \overline{Y}_{ijk\cdot}$ and the decomposition of sum squares is now

$$SST = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SSE$$

where $SST = \sum_{i,j,k,l} \left(Y_{ijkl} - \overline{Y}_{....}\right)^2$, $SS_A = bcn \sum_i \left(\overline{Y}_{i...} - \overline{Y}_{....}\right)^2$, $SS_B = acn \sum_j \left(\overline{Y}_{.j..} - \overline{Y}_{....}\right)^2$, $SS_C = abn \sum_k \left(\overline{Y}_{..k.} - \overline{Y}_{....}\right)^2$, $SS_{AB} = cn \sum_{i,j} \left(\overline{Y}_{ij..} - \overline{Y}_{i...} - \overline{Y}_{.j..} + \overline{Y}_{....}\right)^2$, $SS_{AC} = bn \sum_{i,k} \left(\overline{Y}_{i.k.} - \overline{Y}_{i...} - \overline{Y}_{..k.} + \overline{Y}_{....}\right)^2$, $SS_{BC} = an \sum_{j,k} \left(\overline{Y}_{.jk.} - \overline{Y}_{.j..} - \overline{Y}_{..k.} + \overline{Y}_{....}\right)^2$, $SS_{ABC} = n \sum_{i,j,k} \left(\overline{Y}_{ijk.} - \overline{Y}_{ij..} - \overline{Y}_{i.k.} - \overline{Y}_{.jk.} + \overline{Y}_{i...} + \overline{Y}_{.j..} + \overline{Y}_{..k.} - \overline{Y}_{....}\right)^2$ and $SSE = \sum_{i,j,k,l} \left(Y_{ijkl} - \overline{Y}_{ijk.}\right)^2$. As usual, we can also write $SST = \sum_{i,j,k,l} \left(Y_{ij} - \hat{\mu}\right)^2$, $SSE = \sum_{i,j,k,l} \left(Y_{ijkl} - \hat{\mu}_{ijk}\right)^2$ and now $SS_A = bcn \sum_i \hat{\alpha}_i^2$, $SS_B = acn \sum_j \hat{\beta}_j^2$, $SS_C = abn \sum_k \hat{\gamma}_k^2$, $SS_{AB} = cn \sum_{i,j} \left(\widehat{\alpha\beta}_{ij}\right)^2$, $SS_{AC} = bn \sum_{i,k} \left(\widehat{\alpha\gamma}_{ik}\right)^2$, $SS_{BC} = an \sum_{j,k} \left(\widehat{\gamma\beta}_{jk}\right)^2$ and $SS_{ABC} = n \sum_{i,j,k} \left(\widehat{\alpha\beta\gamma}_{ijk}\right)^2$.

In this model, we have $\frac{SSE}{\sigma^2} \rightsquigarrow \chi^2_{abc(n-1)}$ and the estimation of σ^2 is $\widehat{\sigma^2} = MSE = \frac{SSE}{abc(n-1)}$. Then, $SS_{Model} = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC}$ and the whole model test with the null hypothesis $\alpha_i = \beta_j = \gamma_k = \alpha\beta_{ij} = \alpha\gamma_{ik} = \beta\gamma_{jk} = \alpha\beta\gamma_{ijk} = 0$ for all i, j, k is given by

$$F_{Model} = \frac{SS_{Model}/\left(abc-1\right)}{SSE/\left(abc(n-1)\right)} \rightsquigarrow F_{abc-1,abc(n-1)}$$

with the usual R^2 and CV coefficients.

The appropriate statistics for the F-tests in the ANOVA table with their respective probability distributions are:

$$
\begin{aligned}
F_A &= \frac{SS_A/\left(a-1\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_A}{MSE} \rightsquigarrow F_{a-1,abc(n-1)} && \text{if } \alpha_i = 0 \text{ for all } i \\
F_B &= \frac{SS_B/\left(b-1\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_B}{MSE} \rightsquigarrow F_{b-1,abc(n-1)} && \text{if } \beta_j = 0 \text{ for all } j \\
F_C &= \frac{SS_C/\left(c-1\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_C}{MSE} \rightsquigarrow F_{c-1,abc(n-1)} && \text{if } \gamma_k = 0 \text{ for all } k \\
F_{AB} &= \frac{SS_{AB}/\left((a-1)(b-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{AB}}{MSE} \rightsquigarrow F_{(a-1)(b-1),abc(n-1)} && \text{if } \alpha\beta_{ij} = 0 \text{ for all } i,j \\
F_{AC} &= \frac{SS_{AC}/\left((a-1)(c-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{AC}}{MSE} \rightsquigarrow F_{(a-1)(c-1),abc(n-1)} && \text{if } \alpha\gamma_{ik} = 0 \text{ for all } i,k \\
F_{BC} &= \frac{SS_{BC}/\left((b-1)(c-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{BC}}{MSE} \rightsquigarrow F_{(b-1)(c-1),abc(n-1)} && \text{if } \beta\gamma_{jk} = 0 \text{ for all } j,k \\
F_{ABC} &= \frac{SS_{ABC}/\left((a-1)(b-1)(c-1)\right)}{SSE/\left(abc(n-1)\right)} = \frac{MS_{ABC}}{MSE} \rightsquigarrow F_{(a-1)(b-1)(c-1),abc(n-1)} && \text{if } \alpha\beta\gamma_{ijk} = 0 \text{ for all } i,j,k
\end{aligned}
$$

The ANOVA table for this model is as follows:

	SS	DF	MS	F-value	p-value	H_0
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\alpha_i = 0$ for all i
Factor B	SS_B	$b-1$	MS_B	F_B	p_B	$\beta_j = 0$ for all j
Factor C	SS_C	$c-1$	MS_C	F_C	p_C	$\gamma_k = 0$ for all k
Int. $A*B$	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_{AB}	p_{AB}	$\alpha\beta_{ij} = 0$ for all i,j
Int. $A*C$	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	F_{AC}	p_{AC}	$\alpha\gamma_{ik} = 0$ for all i,k
Int. $B*C$	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	F_{BC}	p_{BC}	$\beta\gamma_{jk} = 0$ for all j,k
Int. $A*B*C$	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	F_{ABC}	p_{ABC}	$\alpha\beta\gamma_{ijk} = 0$ for all i,j,k
Error	SSE	$abc(n-1)$	MSE			
Total	SST	$abcn-1$				

The confidence intervals for the marginal expected values are:

$$\overline{Y}_{i\cdots} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{bcn}} \quad \text{for } \mu_{i\cdot\cdot} = \mu + \alpha_i$$
$$\overline{Y}_{\cdot j\cdot\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{acn}} \quad \text{for } \mu_{\cdot j\cdot} = \mu + \beta_j$$
$$\overline{Y}_{\cdot\cdot k\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{abn}} \quad \text{for } \mu_{\cdot\cdot k} = \mu + \gamma_k$$
$$\overline{Y}_{ij\cdot\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{cn}} \quad \text{for } \mu_{ij\cdot} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$$
$$\overline{Y}_{i\cdot k\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{bn}} \quad \text{for } \mu_{i\cdot k} = \mu + \alpha_i + \gamma_k + \alpha\gamma_{ik}$$
$$\overline{Y}_{\cdot jk\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{an}} \quad \text{for } \mu_{\cdot jk} = \mu + \beta_j + \gamma_k + \beta\gamma_{jk}$$
$$\overline{Y}_{ijk\cdot} \pm t_{abc(n-1);\alpha/2}\sqrt{\frac{MSE}{n}} \quad \text{or } \mu_{ijk}$$

Finally, the least significant differences (LSD) for each pair of means are:

$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{bcn}} \quad \text{for } \overline{Y}_{i\cdots} - \overline{Y}_{i'\cdots}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{acn}} \quad \text{for } \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot j'\cdot\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{abn}} \quad \text{for } \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{\cdot\cdot k'\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{cn}} \quad \text{for } \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i'j'\cdot\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{bn}} \quad \text{for } \overline{Y}_{i\cdot k\cdot} - \overline{Y}_{i'\cdot k'\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{an}} \quad \text{for } \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j'k'\cdot}$$
$$LSD = t_{abc(n-1);\alpha/2}\sqrt{\frac{2MSE}{n}} \quad \text{for } \overline{Y}_{ijk\cdot} - \overline{Y}_{i'j'k'\cdot}$$

The check of the assumptions of the model can be made as usual with the estimated residuals e_{ijkl}.

Example 7: Three-way factorial design. In a study on the germination rate of black pine (*Pinus pinaster* Ait.) under different conditions of water stress and cold, seeds from three provenances with different ecological properties (Serranía de Cuenca, Sierra de Gredos and Northwest) were used. Four levels of water potential (0, -4, -6 and -8 bars, achieved with different concentrations of polyethylene glycol 6000) and two different temperature conditions (F2=$20°C$ for one week, 4^oC for one day, 0^oC for the next day, 4^oC for the next day and $20°C$ during the remainder of the experiment, F3=$20°C$ throughout the experiment) were tested for 48 days in a factorial design with four replications per treatment. The experiment was performed in a growth chamber and, for each treatment, 100 seeds placed on four petri dishes of 10 cm diameter with 25 seeds each were used. At the end of the experiment, the percentage of germinated seeds were scored. To formulate the problem, three factors were considered: A for provenance with three levels, B for water stress with four levels and C for cold conditions with two levels. In addition, we will use a significance level $\alpha = 0.10$. For the statistical analysis we use the three-way factorial model $E(Y_{ijk})=\mu_{ijk}=\mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}$ with i=1, 2, 3; j=1, 2, 3, 4 and k=1, 2. That is, just the model previously studied with $a = 3$, $b = 4$, $c = 2$ and $n = 4$. The obtained data are shown in the following table:

A	B	$C=1$				$\sum_l y_{ij1l}$	$\sum_l y^2_{ij1l}$	$C=2$				$\sum_l y_{ij2l}$	$\sum_l y^2_{ij2l}$
1	1	84	96	96	92	368	33952	92	84	92	76	344	29760
1	2	72	76	92	80	320	25824	88	80	88	88	344	29632
1	3	72	72	80	32	256	17792	84	100	72	72	328	27424
1	4	72	80	80	72	304	23168	88	68	88	68	312	24736
2	1	92	92	96	96	376	35630	88	96	88	96	368	33920
2	2	84	92	76	100	352	31296	92	92	92	68	344	30016
2	3	80	88	60	52	280	20448	84	88	96	76	344	29792
2	4	80	80	56	72	288	21120	100	84	72	72	328	27424
3	1	72	96	68	76	312	24800	84	80	80	100	344	29856
3	2	80	72	64	72	288	20864	72	60	72	68	272	18592
3	3	60	68	72	32	232	14432	80	64	72	48	264	17984
3	4	68	60	52	68	248	15552	32	56	36	52	176	8160
		$\sum_{i,j,l} y_{ij1l} = 3624$				$\sum_{i,j,l} y^2_{ij1l} = 284608$		$\sum_{i,j,l} y_{ij2l} = 3778$				$\sum_{i,j,l} y^2_{ij2l} = 307296$	

The total sums are $\sum_{i,j,k,l} y_{ijkl} = 7392$ and $\sum_{i,j,k,l} y^2_{ijkl} = 591904$ and we can evaluate all the observed means, which are shown in the following tables with the usual notation:

	$i=1\ k=1$	$i=2\ k=1$	$i=3\ k=1$	$i=1\ k=2$	$i=2\ k=2$	$i=3\ k=2$
$j=1$	$\bar{y}_{111\cdot}=92$	$\bar{y}_{211\cdot}=94$	$\bar{y}_{311\cdot}=78$	$\bar{y}_{112\cdot}=86$	$\bar{y}_{212\cdot}=92$	$\bar{y}_{312\cdot}=86$
$j=2$	$\bar{y}_{121\cdot}=80$	$\bar{y}_{221\cdot}=88$	$\bar{y}_{321\cdot}=72$	$\bar{y}_{122\cdot}=86$	$\bar{y}_{222\cdot}=86$	$\bar{y}_{322\cdot}=68$
$j=3$	$\bar{y}_{131\cdot}=64$	$\bar{y}_{231\cdot}=70$	$\bar{y}_{331\cdot}=58$	$\bar{y}_{132\cdot}=82$	$\bar{y}_{232\cdot}=86$	$\bar{y}_{332\cdot}=66$
$j=4$	$\bar{y}_{141\cdot}=76$	$\bar{y}_{241\cdot}=72$	$\bar{y}_{341\cdot}=62$	$\bar{y}_{142\cdot}=78$	$\bar{y}_{242\cdot}=82$	$\bar{y}_{342\cdot}=44$
	$\bar{y}_{1\cdot1\cdot}=78$	$\bar{y}_{2\cdot1\cdot}=81$	$\bar{y}_{3\cdot1\cdot}=67.5$	$\bar{y}_{1\cdot2\cdot}=83$	$\bar{y}_{2\cdot2\cdot}=86.5$	$\bar{y}_{3\cdot2\cdot}=66$

	$i=1$	$i=2$	$i=3$	
$j=1$	$\bar{y}_{11\cdot\cdot}=89$	$\bar{y}_{21\cdot\cdot}=93$	$\bar{y}_{31\cdot\cdot}=82$	$\bar{y}_{\cdot1\cdot\cdot}=88$
$j=2$	$\bar{y}_{12\cdot\cdot}=83$	$\bar{y}_{22\cdot\cdot}=87$	$\bar{y}_{32\cdot\cdot}=70$	$\bar{y}_{\cdot2\cdot\cdot}=80$
$j=3$	$\bar{y}_{13\cdot\cdot}=73$	$\bar{y}_{23\cdot\cdot}=78$	$\bar{y}_{33\cdot\cdot}=62$	$\bar{y}_{\cdot3\cdot\cdot}=71$
$j=4$	$\bar{y}_{14\cdot\cdot}=77$	$\bar{y}_{24\cdot\cdot}=77$	$\bar{y}_{34\cdot\cdot}=53$	$\bar{y}_{\cdot4\cdot\cdot}=69$
	$\bar{y}_{1\cdots}=80.5$	$\bar{y}_{2\cdots}=83.75$	$\bar{y}_{3\cdots}=66.75$	$\bar{y}_{\cdots\cdot}=77$

	$k=1$	$k=2$
$j=1$	$\bar{y}_{\cdot11\cdot}=88$	$\bar{y}_{\cdot12\cdot}=88$
$j=2$	$\bar{y}_{\cdot21\cdot}=80$	$\bar{y}_{\cdot22\cdot}=80$
$j=3$	$\bar{y}_{\cdot31\cdot}=64$	$\bar{y}_{\cdot32\cdot}=78$
$j=4$	$\bar{y}_{\cdot41\cdot}=70$	$\bar{y}_{\cdot42\cdot}=68$
	$\bar{y}_{\cdot\cdot1\cdot}=75.5$	$\bar{y}_{\cdot\cdot2\cdot}=78.5$

From these data we can obtain the estimated parameters of the model. First of all, from the previous table and by reasoning as in the two-way factorial model for the main effects and the interactions AB and BC we have:

$\widehat{\alpha\beta}_{11}=-2.5$	$\widehat{\alpha\beta}_{21}=-1.75$	$\widehat{\alpha\beta}_{31}=4.25$	$\hat{\beta}_1=11$
$\widehat{\alpha\beta}_{12}=-0.5$	$\widehat{\alpha\beta}_{22}=0.25$	$\widehat{\alpha\beta}_{32}=0.25$	$\hat{\beta}_2=3$
$\widehat{\alpha\beta}_{13}=-1.5$	$\widehat{\alpha\beta}_{23}=0.25$	$\widehat{\alpha\beta}_{33}=1.25$	$\hat{\beta}_3=-6$
$\widehat{\alpha\beta}_{14}=4.5$	$\widehat{\alpha\beta}_{24}=1.25$	$\widehat{\alpha\beta}_{34}=-5.75$	$\hat{\beta}_4=-8$
$\hat{\alpha}_1=3.5$	$\hat{\alpha}_2=6.75$	$\hat{\alpha}_3=-10.25$	$\hat{\mu}=77$

$\widehat{\beta\gamma}_{11}=1.5$	$\widehat{\beta\gamma}_{12}=-1.5$
$\widehat{\beta\gamma}_{21}=1.5$	$\widehat{\beta\gamma}_{22}=-1.5$
$\widehat{\beta\gamma}_{31}=-5.5$	$\widehat{\beta\gamma}_{32}=5.5$
$\widehat{\beta\gamma}_{41}=2.5$	$\widehat{\beta\gamma}_{42}=-2.5$
$\hat{\gamma}_1=-1.5$	$\hat{\gamma}_2=1.5$

Finally, for the triple interaction effects we have, for example,

$$\widehat{\alpha\beta\gamma}_{111} = \bar{y}_{111\cdot} - \bar{y}_{\cdots\cdot} - \hat{\alpha}_1 - \hat{\beta}_1 - \hat{\gamma}_1 - \widehat{\alpha\beta}_{11} - \widehat{\beta\gamma}_{11} - \widehat{\alpha\gamma}_{11} = 15 - 3.5 - 11 + 1.5 + 2.5 - 1.5 + 1 = 4$$

In a similar way, the estimations of the other triple interaction effects are evaluated and they are included in the following table together with the interaction AC:

$\widehat{\alpha\beta\gamma}_{111}=4$	$\widehat{\alpha\beta\gamma}_{211}=2.25$	$\widehat{\alpha\beta\gamma}_{311}=-6.25$	$\widehat{\alpha\beta\gamma}_{112}=-4$	$\widehat{\alpha\beta\gamma}_{212}=-2.25$	$\widehat{\alpha\beta\gamma}_{312}=6.25$
$\widehat{\alpha\beta\gamma}_{121}=-2$	$\widehat{\alpha\beta\gamma}_{221}=2.25$	$\widehat{\alpha\beta\gamma}_{321}=-0.25$	$\widehat{\alpha\beta\gamma}_{122}=2$	$\widehat{\alpha\beta\gamma}_{222}=-2.25$	$\widehat{\alpha\beta\gamma}_{322}=0.25$
$\widehat{\alpha\beta\gamma}_{131}=-1$	$\widehat{\alpha\beta\gamma}_{231}=0.25$	$\widehat{\alpha\beta\gamma}_{331}=0.75$	$\widehat{\alpha\beta\gamma}_{132}=1$	$\widehat{\alpha\beta\gamma}_{232}=-0.25$	$\widehat{\alpha\beta\gamma}_{332}=-0.75$
$\widehat{\alpha\beta\gamma}_{141}=-1$	$\widehat{\alpha\beta\gamma}_{241}=-4.75$	$\widehat{\alpha\beta\gamma}_{341}=5.75$	$\widehat{\alpha\beta\gamma}_{142}=1$	$\widehat{\alpha\beta\gamma}_{242}=4.75$	$\widehat{\alpha\beta\gamma}_{342}=-5.75$
$\widehat{\alpha\gamma}_{11}=-1$	$\widehat{\alpha\gamma}_{21}=-1.25$	$\widehat{\alpha\gamma}_{31}=2.25$	$\widehat{\alpha\gamma}_{12}=1$	$\widehat{\alpha\gamma}_{22}=1.25$	$\widehat{\alpha\gamma}_{32}=-2.25$

Calculating the sum of squares we obtain: $SST = 591904 - 96\hat{\mu}^2 = 22720$, $SS_A = 32\sum_i \hat{\alpha}_i^2 = 5212$, $SS_B = 24\sum_j \hat{\beta}_j^2 = 5520$, $SS_C = 48\sum_k \hat{\gamma}_k^2 = 216$, $SS_{AB} = 8\sum_{i,j}\left(\widehat{\alpha\beta}_{ij}\right)^2 = 692$, $SS_{AC} = 16\sum_{i,k}\left(\widehat{\alpha\gamma}_{ik}\right)^2 = 244$, $SS_{BC} = 12\sum_{j,k}\left(\widehat{\beta\gamma}_{jk}\right)^2 = 984$, $SS_{ABC} = 4\sum_{i,j,k}\left(\widehat{\alpha\beta\gamma}_{ijk}\right)^2 = 1020$ and $SSE = 22720 - 13888 = 8832$. Therefore, $MSE = \frac{8832}{72} = \frac{368}{3} = 122.67$ and the estimated value for the parameter σ is $\hat{\sigma} = \sqrt{MSE} = 11.08$. The whole model test leads to $F_{Model} = \frac{13888/23}{368/3} = 4.92$ with p-value$= p\left(F_{23,72} > 4.92\right) = 0.0000$. Therefore the model is significative with $\alpha = 0.10$. The determination coefficient of the model is $R^2 = \frac{13888}{22720} = 61.1\%$ and the variation coefficient is $CV = \frac{11.08}{77} = 14.4\%$.

The evaluation of the F-tests in the ANOVA table leads to:

$$F_A = \frac{5212/2}{368/3} = 21.24 \text{ with p-value}= p\left(F_{2,72} > 21.24\right) = 0.0000$$
$$F_B = \frac{5520/3}{368/3} = 15 \text{ with p-value}= p\left(F_{3,72} > 15\right) = 0.0000$$
$$F_C = \frac{216}{368/3} = 1.76 \text{ with p-value}= p\left(F_{1,72} > 1.76\right) = 0.1887$$
$$F_{AB} = \frac{692/6}{368/3} = 0.94 \text{ with p-value}= p\left(F_{6,72} > 0.94\right) = 0.4719$$
$$F_{AC} = \frac{244/2}{368/3} = 0.99 \text{ with p-value}= p\left(F_{2,72} > 0.99\right) = 0.3749$$
$$F_{BC} = \frac{984/3}{368/3} = 2.67 \text{ with p-value}= p\left(F_{3,72} > 2.67\right) = 0.0536$$
$$F_{ABC} = \frac{1020/6}{368/3} = 1.39 \text{ with p-value}= p\left(F_{6,72} > 1.39\right) = 0.2320$$

Therefore, the main effect of the factor A and the interaction between factors B and C are significant with $\alpha = 0.10$. The standard error for the means $\bar{y}_{i\cdots}$ is $SE = \sqrt{\frac{MSE}{32}} = 1.96$ and, taking into account that $t_{72;0.05} = 1.67$, the 90% confidence intervals are $\bar{y}_{i\cdots} \pm 3.26$. The least significant difference with $\alpha = 0.10$ is $LSD = 3.26\sqrt{2} = 4.61$ and the LSD Fisher's test with $\alpha = 0.10$ leads to:

Provenance	Mean	
3 (Northwest)	66.75	A
1 (Serranía de Cuenca)	80.5	B
2 (Sierra de Gredos)	83.75	B

Therefore, the expected value of the germination rate is significantly lower for the Northwest provenance and there is no significant differences between the other two provenances.

For the interaction between factors B and C, the standard errors for the means are: $SE = \sqrt{\frac{MSE}{12}} = 3.20$ for $\bar{y}_{\cdot jk\cdot}$, $SE = \sqrt{\frac{MSE}{24}} = 2.26$ for $\bar{y}_{\cdot j\cdot\cdot}$ and $SE = \sqrt{\frac{MSE}{48}} = 1.60$ for $\bar{y}_{\cdot\cdot k\cdot}$. The 90% confidence intervals are $\bar{y}_{\cdot jk\cdot} \pm 5.33$, $\bar{y}_{\cdot j\cdot\cdot} \pm 3.77$ and $\bar{y}_{\cdot\cdot k\cdot} \pm 2.66$. The least significant differences with $\alpha = 0.10$ are $LSD = 5.33\sqrt{2} = 7.53$ for $\bar{y}_{\cdot jk\cdot}$, $LSD = 3.77\sqrt{2} = 5.33$ for $\bar{y}_{\cdot j\cdot\cdot}$ and $LSD = 2.66\sqrt{2} = 3.77$ for $\bar{y}_{\cdot\cdot k\cdot}$. Then we obtain the following results for the comparison of the water stress (B) and cold conditions (C) with $\alpha = 0.10$:

	$B=1$ (0 bar)	$B=1$ (-4 bar)	$B=2$ (-6 bar)	$B=3$ (-8 bar)	
$C=1$ (F2)	88 Ca	80 Ba	64 Aa	70 Aa	75.5 a
$C=2$ (F3)	88 Ca	80 Ba	78 Bb	68 Aa	78.5 a
	88 C	80 B	71 A	69 A	

where we have used uppercase letters to compare the levels of the water stress and lowercase letters to compare the levels of cold conditions. Then we observe that the highest values for the expected germination rate are obtained with 0 bar, without significant difference between cold conditions F2 and F3. With F3 cold conditions, the expected value for the germination rate is higher for -6 bar than for -8 bar, but there is no difference between -6 bar and -8 bar with F2 cold conditions (in fact, we have even obtained higher values with -8 bar). This seems to be the reason that the interaction BC is significant.

The complete LSD Fisher's test with $\alpha = 0.10$ for all the eight means leads to:

Treatment	Mean	
-6 bar + F2	64	*A*
-8 bar + F3	68	*A*
-8 bar + F2	70	*A*
-6 bar + F3	78	*B*
-4 bar + F2	80	*B*
-4 bar + F3	80	*B*
0 bar + F2	88	*C*
0 bar + F3	88	*C*

Chapter 7

SPLIT-PLOT ANOVA

The split-plot designs arose from agricultural field trials. These often occur when there are two factors of interest and one factor requires larger experimental units than the other. For example, suppose that we are interested in studying irrigation amount and fertilizer type on the growth of a particular plant. Because of the equipment involved, different amounts of irrigation can only be done on a large scale, while different fertilizers can be applied much more locally. For this situation, two type of experimental units are considered: large plots for the levels of irrigation (whole-plots) and small plots within each plot for the fertilizers (split-plots). Levels of irrigation are assigned to the whole-plots in a completely randomized design or a randomized complete block design, and each of the whole plots is divided in split-plots with one fertilizer randomly assigned to each of them.

Additionally, these types of designs are often used in other situations. For example an industrial experiment is used to study the freshness of milk. If the two factors are pasteurization process and type of container, we would need to pasteurize an entire batch, but we could use different types of containers in a particular batch. Similarly, this type of design is applicable in many real practical situations.

To illustrate this type of analysis of variance, an experimental design is considered with a factor A with a levels assigned to the whole-plots in a randomized block design with r blocks (ra is the total number of whole-plots), and a factor B with b levels randomly assigned to the split-plots within each plot (rab is the total number of split-plots). The mathematical formulation for the model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ij} + \gamma_k + \beta\gamma_{jk} + \varepsilon_{ijk}$$

where μ is the global mean effect; α_i is the block effect ($i = 1, .., r$), β_j is the main effect of factor A in the whole-plots ($j = 1, .., a$); γ_k is the main effect of factor B in the split-plots ($k = 1, .., b$); $\beta\gamma_{jk}$ is the interaction effect between factors A and B; ξ_{ij} is the random error for the whole-plots with $\xi_{ij} \rightsquigarrow N(0, \sigma_1)$ and ε_{ijk} is the random error for the split-plots whitin the whole-plots with $\varepsilon_{ijk} \rightsquigarrow N(0, \sigma_2)$. The parameters σ_1^2 and σ_2^2 are the variance parameters for the whole-plots and the split-plots, respectively, and we suppose that all ξ_{ij} and ε_{ijk} are independent random variables. The restrictions on the parameters are: $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_k \gamma_k = 0$, $\sum_k \beta\gamma_{jk} = 0$ for all j, and $\sum_j \beta\gamma_{jk} = 0$ for all k.

Note that a two-way factorial randomized block design would have all the combinations of levels within each block and a unique random error term, while a split-plot model has the A factor applied only one time to each block with factor B applied multiple times in each block, and it has two random error terms.

Therefore, for this model we have $E\left(Y_{ijk}\right)=\mu_{ijk}=\mu+\alpha_i+\beta_j+\gamma_k+\beta\gamma_{jk}$ and, with the usual notation, the appropriate estimators for the parameters of the model are $\hat{\mu}=\overline{Y}_{\cdots}$, $\hat{\alpha}_i=\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdots}$, $\hat{\beta}_j=\overline{Y}_{\cdot j\cdot}-\overline{Y}_{\cdots}$, $\hat{\gamma}_k=\overline{Y}_{\cdot\cdot k}-\overline{Y}_{\cdots}$, $\widehat{\beta\gamma}_{jk}=\overline{Y}_{\cdot jk}-\overline{Y}_{\cdot j\cdot}-\overline{Y}_{\cdot\cdot k}+\overline{Y}_{\cdots}$ and $\hat{\mu}_{ijk}=\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j+\hat{\gamma}_k+\widehat{\beta\gamma}_{jk}=\overline{Y}_{i\cdot\cdot}+\overline{Y}_{\cdot jk}-\overline{Y}_{\cdots}$. The estimated residuals for this model are defined as $\hat{\varepsilon}_{ijk}=e_{ijk}=Y_{ijk}-\hat{\mu}_{ijk}$. In addition, this model has two sources of random variation, due to the existence of whole-plots (for factor A) and split-plots (for factor B). This is the reason that, first, we divide the total variation in two parts, as follows: $SST=SST_1+SST_2$, with $SST=\sum_{i,j,k}\left(Y_{ijk}-\overline{Y}_{\cdots}\right)^2$, $SST_1=b\sum_{i,j}\left(\overline{Y}_{ij\cdot}-\overline{Y}_{\cdots}\right)^2$ and $SST_2=\sum_{i,j,k}\left(\overline{Y}_{ijk}-\overline{Y}_{ij\cdot}\right)^2$. The degrees of freedom for these sum of squares are: $rab-1$ for SST, $ra-1$ for SST_1 and $ra(b-1)$ for SST_2, so that $rab-1=ra-1+ra(b-1)$.

The total variation SST_1 due to the whole-plots can be divided in three sources of variation: blocks, main effect of factor A and a random variation of the whole-plots, as follows

$$SS_{Block}=ab\sum_i\left(\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdots}\right)^2=ab\sum_i\hat{\alpha}_i^2 \text{ with } d.f.=r-1$$
$$SS_A=rb\sum_j\left(\overline{Y}_{\cdot j\cdot}-\overline{Y}_{\cdots}\right)^2=rb\sum_j\hat{\beta}_j^2 \text{ with } d.f.=a-1$$
$$\text{Whole-plot error}=SSE_1=SST_1-SS_{Block}-SS_A= \text{ with } d.f.=(a-1)(r-1)$$

It is easily seen that SSE_1 can be alternatively written as $SSE_1=b\sum_{i,j}\left(\overline{Y}_{ij\cdot}-\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot j\cdot}+\overline{Y}_{\cdots}\right)^2$, that is, as an hypothetical interaction between blocks and factor A. Moreover, it verifies that $E\left(SSE_1\right)=(a-1)(r-1)\left(\sigma_2^2+b\sigma_1^2\right)$ (see the note at the final of this experimental design).

In a similar way, the total variation SST_2 due to the split-plots can be divided in three sources of variation: main effect of factor B, interaction effect between factors A and B, and a random variation of the split-plots, as follows

$$SS_B=ra\sum_k\left(\overline{Y}_{\cdot\cdot k}-\overline{Y}_{\cdots}\right)^2=ra\sum_k\hat{\gamma}_k^2 \text{ with } d.f.=b-1$$
$$SS_{AB}=r\sum_{j,k}\left(\overline{Y}_{\cdot jk}-\overline{Y}_{\cdot j\cdot}-\overline{Y}_{\cdot\cdot k}+\overline{Y}_{\cdots}\right)^2=r\sum_{j,k}\left(\widehat{\beta\gamma}_{jk}\right)^2 \text{ with } d.f.=(a-1)(b-1)$$
$$\text{Split-plot error}=SSE_2=SST_2-SS_B-SS_{AB}= \text{ with } d.f.=a(b-1)(r-1)$$

Moreover, SSE_2 can be alternatively written as $SSE_2=\sum_{i,j,k}\left(\overline{Y}_{ijk}-\overline{Y}_{ij\cdot}-\overline{Y}_{\cdot jk}+\overline{Y}_{\cdot j\cdot}\right)^2$ and it verifies that $E\left(SSE_2\right)=a(b-1)(r-1)\sigma_2^2$ (see the note at the final of this experimental design).

Therefore, taking into account the expressions for $E\left(SSE_2\right)$ and $E\left(SSE_1\right)$, it seems logical to estimate the variance parameters of the model as $\widehat{\sigma_2^2}=\dfrac{SSE_2}{a(b-1)(r-1)}=MSE_2$ and $\widehat{\sigma_1^2}=\dfrac{SSE_1}{(a-1)(r-1)b}-\dfrac{MSE_2}{b}=\dfrac{MSE_1-MSE_2}{b}$ with $MSE_1=\dfrac{SSE_1}{(a-1)(r-1)}$ (this is the Type I estimation method).

Now, the appropriate statistics for the F-tests in the ANOVA table with their respective probability distributions are:

$$F_{Block} = \frac{SS_{Block}/(r-1)}{MSE_1} = \frac{MS_{Block}}{MSE_1} \rightsquigarrow F_{r-1,(a-1)(r-1)} \qquad \text{if } \alpha_i = 0 \text{ for all } i$$
$$F_A = \frac{SS_A/\,(a-1)}{MSE_1} = \frac{MS_A}{MSE_1} \rightsquigarrow F_{a-1,(a-1)(r-1)} \qquad \text{if } \beta_j = 0 \text{ for all } j$$
$$F_B = \frac{SS_B/\,(b-1)}{MSE_2} = \frac{MS_B}{MSE_2} \rightsquigarrow F_{b-1,a(b-1)(r-1)} \qquad \text{if } \gamma_k = 0 \text{ for all } k$$
$$F_{AB} = \frac{SS_{AB}/\,((a-1)(b-1))}{MSE_2} = \frac{MS_{AB}}{MSE_2} \rightsquigarrow F_{(a-1)(b-1),a(b-1)(r-1)} \qquad \text{if } \beta\gamma_{jk} = 0 \text{ for all } j,k$$

The ANOVA table for this model is as follows:

	SS	DF	MS	F-value	p-value	H_0
Blocks	SS_{Block}	$r-1$	MS_{Block}	F_{Block}	p_{Block}	$\alpha_i = 0$ for all i
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\beta_j = 0$ for all j
Plot error	SSE_1	$(a-1)(r-1)$	MSE_1			
Total plots	SST_1	$ar-1$				
Factor B	SS_B	$b-1$	MS_B	F_B	p_B	$\gamma_k = 0$ for all k
Int. $A*B$	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_{AB}	p_{AB}	$\beta\gamma_{jk} = 0$ for all j,k
Split-plot error	SSE_2	$a(b–1)(r–1)$	MSE_2			
Total split-plots	SST_2	$ar(b-1)$				
Total	SST	$abr-1$				

The confidence intervals for the marginal expected values are:

$$\overline{Y}_{i\cdot\cdot} \pm t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{MSE_1}{ab}} \qquad \text{for } \mu_{i\cdot\cdot} = \mu + \alpha_i$$
$$\overline{Y}_{\cdot j\cdot} \pm t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{MSE_1}{rb}} \qquad \text{for } \mu_{\cdot j\cdot} = \mu + \beta_j$$
$$\overline{Y}_{\cdot\cdot k} \pm t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{MSE_2}{ra}} \qquad \text{for } \mu_{\cdot\cdot k} = \mu + \gamma_k$$
$$\overline{Y}_{\cdot jk} \pm t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{MSE_2}{r}} \qquad \text{for } \mu_{\cdot jk} = \mu + \beta_j + \gamma_k + \beta\gamma_{jk}$$

Finally, the least significant differences (LSD) for each pair of means are:

$$LSD = t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_1}{ab}} \qquad \text{for } \overline{Y}_{i\cdot\cdot} - \overline{Y}_{i'\cdot\cdot}$$
$$LSD = t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_1}{rb}} \qquad \text{for } \overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdot j'\cdot}$$
$$LSD = t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_2}{ra}} \qquad \text{for } \overline{Y}_{\cdot\cdot k} - \overline{Y}_{\cdot\cdot k'}$$
$$LSD = t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_2}{r}} \qquad \text{for } \overline{Y}_{\cdot jk} - \overline{Y}_{\cdot j'k'}$$

The check of the assumptions of the model can be made as usual with the estimated residuals e_{ijk}.

It should be noted that if there are no blocks in the model then the parameters α_i and the sum SS_{Block} do not exist and now

$$\text{Whole-plot error} = SSE_1 = b\sum_{i,j}\left(Y_{ij\cdot} - \overline{Y}_{\cdot j\cdot}\right)^2 \text{ with } d.f. = a(r-1)$$

and $\hat{\mu}_{jk} = \overline{Y}_{\cdot jk}$ with residuals $\hat{\varepsilon}_{ijk} = e_{ijk} = Y_{ijk} - \hat{\mu}_{jk}$, but everything else being equal with replicates instead of blocks.

Note. To prove that $E(SSE_1) = (a-1)(r-1)\left(\sigma_2^2 + b\sigma_1^2\right)$ first of all we observe that, for fixed i_0 and j_0, we have $E\left(\overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{i_0\cdot\cdot\cdot} - \overline{Y}_{\cdot j_0\cdot\cdot} + \overline{Y}_{\cdot\cdot\cdot\cdot}\right) = 0$ and therefore

$$E\left(\left(\overline{Y}_{i_0j_0\cdot} - \overline{Y}_{i_0\cdot\cdot} - \overline{Y}_{\cdot j_0\cdot} + \overline{Y}_{\cdot\cdot\cdot}\right)^2\right) = Var\left(\overline{Y}_{i_0j_0\cdot} - \overline{Y}_{i_0\cdot\cdot} - \overline{Y}_{\cdot j_0\cdot} + \overline{Y}_{\cdot\cdot\cdot}\right)$$

$$= Var\left(\frac{1}{b}\sum_k Y_{i_0j_0k} - \frac{1}{ab}\sum_{j,k} Y_{i_0jk} - \frac{1}{rb}\sum_{i,k} Y_{ij_0k} + \frac{1}{rab}\sum_{i,j,k} Y_{ijk}\right)$$

$$= Var\left(\begin{array}{c} \left(\frac{1}{b} - \frac{1}{ab} - \frac{1}{rb} + \frac{1}{rab}\right)\sum_k Y_{i_0j_0k} - \left(\frac{1}{ab} - \frac{1}{rab}\right)\sum_{j,k,j\neq j_0} Y_{i_0jk} \\ -\left(\frac{1}{rb} - \frac{1}{rab}\right)\sum_{i,k,i\neq i_0} Y_{ij_0kl} + \frac{1}{rab}\sum_{i,j,k,i\neq i_0,j\neq j_0} Y_{ijk} \end{array}\right)$$

$$= \frac{(a-1)^2(r-1)^2}{r^2a^2b^2} Var\left[b\xi_{i_0j_0} + \sum_k \varepsilon_{i_0j_0k}\right] + \frac{(r-1)^2}{r^2a^2b^2} Var\left[b\sum_{j,j\neq j_0}\xi_{i_0j} + \sum_{j,k,j\neq j_0}\varepsilon_{i_0jk}\right]$$

$$+\frac{(a-1)^2}{r^2a^2b^2} Var\left[b\sum_{i,i\neq i_0}\xi_{ij_0} + \sum_{i,k,i\neq i_0}\varepsilon_{ij_0kl}\right] + \frac{1}{r^2a^2b^2} Var\left[b\sum_{i,j,i\neq i_0,j\neq j_0}\xi_{ij} + \sum_{i,j,k,i\neq i_0,j\neq j_0}\varepsilon_{ijk}\right]$$

$$= \frac{(a-1)^2(r-1)^2}{r^2a^2b^2}\left[b^2\sigma_1^2 + b\sigma_2^2\right] + \frac{(r-1)^2}{r^2a^2b^2}\left[b^2(a-1)\sigma_1^2 + b(a-1)\sigma_2^2\right]$$

$$+\frac{(a-1)^2}{r^2a^2b^2}\left[b^2(r-1)\sigma_1^2 + b(r-1)\sigma_2^2\right] + \frac{1}{r^2a^2b^2}\left[b^2(r-1)(a-1)\sigma_1^2 + b(r-1)(a-1)\sigma_2^2\right]$$

$$= \frac{(a-1)^2(r-1)^2}{r^2a^2b}\left[b\sigma_1^2 + \sigma_2^2\right] + \frac{(r-1)^2(a-1)}{r^2a^2b}\left[b\sigma_1^2 + \sigma_2^2\right]$$

$$+\frac{(a-1)^2(r-1)}{r^2a^2b}\left[b\sigma_1^2 + \sigma_2^2\right] + \frac{(a-1)(r-1)}{r^2a^2b}\left[b\sigma_1^2 + \sigma_2^2\right]$$

$$= \frac{(a-1)(r-1)}{r^2a^2b}\left[(a-1)(r-1) + r - 1 + a - 1 + 1\right]\left[b\sigma_1^2 + \sigma_2^2\right] = \frac{(a-1)(r-1)}{rab}\left[b\sigma_1^2 + \sigma_2^2\right]$$

As consequence, we have $E(SSE_1) = b\sum_{i,j}\frac{(a-1)(r-1)}{rab}\left[b\sigma_1^2 + \sigma_2^2\right] = (a-1)(r-1)\left(\sigma_2^2 + b\sigma_1^2\right)$.

In a similar way, to prove that $E(SSE_2) = a(b-1)(r-1)\sigma_2^2$ first of all we observe that, for fixed i_0, j_0 and k_0, the expression $\overline{Y}_{i_0j_0k_0} - \overline{Y}_{i_0j_0\cdot} - \overline{Y}_{\cdot j_0k_0} + \overline{Y}_{\cdot j_0\cdot}$ does not depend on the random variables ξ_{ij_0} and we have $E\left(\overline{Y}_{i_0j_0k_0} - \overline{Y}_{i_0j_0\cdot} - \overline{Y}_{\cdot j_0k_0} + \overline{Y}_{\cdot j_0\cdot}\right) = 0$. Therefore

$$E\left(\left(\overline{Y}_{i_0j_0k_0} - \overline{Y}_{i_0j_0\cdot} - \overline{Y}_{\cdot j_0k_0} + \overline{Y}_{\cdot j_0\cdot}\right)^2\right) = Var\left(\overline{Y}_{i_0j_0k_0} - \overline{Y}_{i_0j_0\cdot} - \overline{Y}_{\cdot j_0k_0} + \overline{Y}_{\cdot j_0\cdot}\right)$$

$$= Var\left(Y_{i_0j_0k_0} - \frac{1}{b}\sum_k Y_{i_0j_0k} - \frac{1}{r}\sum_i Y_{ij_0k_0} + \frac{1}{rb}\sum_{i,k} Y_{ij_0k}\right)$$

$$= Var\left(\begin{array}{c} \left(1 - \frac{1}{b} - \frac{1}{r} + \frac{1}{rb}\right)Y_{i_0j_0k_0} - \left(\frac{1}{b} - \frac{1}{rb}\right)\sum_{k,k\neq k_0} Y_{i_0j_0k} \\ -\left(\frac{1}{r} - \frac{1}{rb}\right)\sum_{i,i\neq i_0} Y_{ij_0k_0} + \frac{1}{rb}\sum_{i,k,i\neq i_0,k\neq k_0} Y_{ij_0k} \end{array}\right)$$

$$= \frac{(b-1)^2(r-1)^2}{r^2b^2} Var\left(\varepsilon_{i_0j_0k_0}\right) + \frac{(r-1)^2}{r^2b^2}\sum_{k,k\neq k_0} Var\left(\varepsilon_{i_0j_0k}\right)$$

$$+\frac{(b-1)^2}{r^2b^2}\sum_{i,i\neq i_0} Var\left(\varepsilon_{ij_0k_0}\right) + \frac{1}{r^2b^2}\sum_{i,k,i\neq i_0,k\neq k_0} Var\left(\varepsilon_{ij_0k}\right)$$

$$= \frac{(b-1)^2 (r-1)^2}{r^2 b^2}\sigma_2^2 + \frac{(r-1)^2(b-1)}{r^2b^2}\sigma_2^2 + \frac{(b-1)^2(r-1)}{r^2b^2}\sigma_2^2 + \frac{(r-1)(b-1)}{r^2b^2}\sigma_2^2$$
$$= \frac{(b-1)(r-1)[(b-1)(r-1)+(r-1)+(b-1)+1]}{r^2b^2}\sigma_2^2 = \frac{(b-1)(r-1)}{rb}\sigma_2^2$$

As consequence, we have $E\left(SSE_2\right) = \sum_{i,j,k} \frac{(b-1)(r-1)}{rb}\sigma_2^2 = a(b-1)(r-1)\sigma_2^2$.

Example 8: Split-plot design. An experimental design is used to test the effect of four crops of compost (barley, vetch, barley-vetch growing together and fallow, which is the control level) with two levels of nitrogen fertilization (120 pounds of nitrogen per acre and no nitrogen application, which is the control level) on the subsequent production of sugar beet. The experimental area of land is divided into three blocks because of the possible influence of the soil characteristics. At first, it was assumed that the sugar beet would respond in different ways depending on the level nitrogen fertilization; therefore the aim was to compare as precisely as possible the effect of plant fertilizers in each level of nitrogen. Thus each block is divided into two main plots randomly assigning to each one of the two nitrogen levels. After each main plot was subdivided into four subplots to which were randomized each of the four types of plant fertilizer. At the final of the experiment, the total production of sugar beet in tons per acre was annotated for each of the 24 subplots. Then we have two types of experimental units: 6 plots associated with nitrogen levels (2 levels with 3 replicates, one for each block) and 24 subplots associated with the plant fertilizers (4 levels with 6 replicates). To formulate the problem, three factors were considered: blocks, A for nitrogen fertilization and B for plant fertilizers. For the statistical analysis we use the split-plot model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \xi_{ij} + \gamma_k + \beta\gamma_{jk} + \varepsilon_{ijk}$$

which leads to $E\left(Y_{ijk}\right) = \mu + \alpha_i + \beta_j + \gamma_k + \beta\gamma_{jk}$ with $i=1,2,3$ for the blocks; $j=1,2$ for the nitrogen fertilization and $k=1,2,3,4$ for the plant fertilizers. That is, just the model previously studied with $r=3$, $a=2$ and $b=4$. The obtained results are given in the following table:

	No nitrogen ($j=1$)				Whole	Nitrogen=120 ($j=2$)				Whole	Total
	Fallow	Barley	Vetch	Bar.-Vet.	plot	Fallow	Barley	Vetch	Bar.-Vet.	plot	block
Block	$k=1$	$k=2$	$k=3$	$k=4$	$\bar{y}_{i1\cdot}$	$k=1$	$k=2$	$k=3$	$k=4$	$\bar{y}_{i2\cdot}$	$\bar{y}_{i\cdot\cdot}$
$i=1$	27.6	31.0	42.0	37.8	34.60	38.6	44.4	50.6	51.8	46.35	40.475
$i=2$	27.0	30.0	45.4	36.6	34.75	36.0	48.4	49.6	53.4	46.85	40.8
$i=3$	26.4	30.4	44.6	39.2	35.15	41.0	50.8	56.8	55.2	50.95	43.05

The total sums are $\sum_{i,j,k} y_{ijk} = 994.6$, $\sum_{i,j,k} y_{ijk}^2 = 43282.36$ and, therefore, $SST = 43282.36 - \frac{994.6^2}{24} = \frac{6193.435}{3}$. In a similar way, for the whole-plots, the total sums are $\sum_{i,j} \bar{y}_{ij\cdot} = 248.65$, $\sum_{i,j} \bar{y}_{ij\cdot}^2 = 10579.3925$ and, therefore, $SST_1 = 4\left(10579.3925 - \frac{248.65^2}{6}\right) = \frac{3299.065}{3}$. Now we evaluate the observed means for all the treatments, which are shown in the following table with the usual notation:

	$k=1$	$k=2$	$k=3$	$k=4$	
$j=1$	$\bar{y}_{\cdot 11} = \frac{81}{3}$	$\bar{y}_{\cdot 12} = \frac{91.4}{3}$	$\bar{y}_{\cdot 13} = \frac{132}{3}$	$\bar{y}_{\cdot 14} = \frac{113.6}{3}$	$\bar{y}_{\cdot 1\cdot} = \frac{104.5}{3}$
$j=2$	$\bar{y}_{\cdot 21} = \frac{115.6}{3}$	$\bar{y}_{\cdot 22} = \frac{143.6}{3}$	$\bar{y}_{\cdot 23} = \frac{157}{3}$	$\bar{y}_{\cdot 24} = \frac{160.4}{3}$	$\bar{y}_{\cdot 2\cdot} = \frac{144.15}{3}$
	$\bar{y}_{\cdot\cdot 1} = \frac{98.3}{3}$	$\bar{y}_{\cdot\cdot 2} = \frac{117.5}{3}$	$\bar{y}_{\cdot\cdot 3} = \frac{144.5}{3}$	$\bar{y}_{\cdot\cdot 4} = \frac{137}{3}$	$\bar{y}_{\cdot\cdot\cdot} = \frac{124.325}{3}$

The estimated parameters of the model are: $\hat{\mu} = \frac{124.325}{3}$, $\hat{\alpha}_1 = 40.475 - \frac{124.325}{3} = -\frac{2.9}{3}$, $\hat{\alpha}_2 = 40.8 - \frac{124.325}{3} = -\frac{1.925}{3}$, $\hat{\alpha}_3 = 43.05 - \frac{124.325}{3} = \frac{4.825}{3}$, and, for example, $\hat{\beta}_1 = \frac{104.5}{3} - \frac{124.325}{3} = -\frac{19.825}{3}$, $\hat{\gamma}_1 = \frac{98.3}{3} - \frac{124.325}{3} = -\frac{26.025}{3}$, $\widehat{\beta\gamma}_{11} = \frac{81-124.325}{3} + \frac{19.825}{3} + \frac{26.025}{3} = \frac{2.525}{3}$. In a similar way, the other estimated parameters are evaluated and they are included in the following table:

	$k=1$	$k=2$	$k=3$	$k=4$	
$j=1$	$\widehat{\beta\gamma}_{11} = \frac{2.525}{3}$	$\widehat{\beta\gamma}_{12} = -\frac{6.275}{3}$	$\widehat{\beta\gamma}_{13} = \frac{7.325}{3}$	$\widehat{\beta\gamma}_{14} = -\frac{3.575}{3}$	$\hat{\beta}_1 = -\frac{19.825}{3}$
$j=2$	$\widehat{\beta\gamma}_{21} = -\frac{2.525}{3}$	$\widehat{\beta\gamma}_{22} = \frac{6.275}{3}$	$\widehat{\beta\gamma}_{23} = -\frac{7.325}{3}$	$\widehat{\beta\gamma}_{24} = \frac{3.575}{3}$	$\hat{\beta}_2 = \frac{19.825}{3}$
	$\hat{\gamma}_1 = -\frac{26.025}{3}$	$\hat{\gamma}_2 = -\frac{6.825}{3}$	$\hat{\gamma}_3 = \frac{20.175}{3}$	$\hat{\gamma}_4 = \frac{12.675}{3}$	

Now we can evaluate all the sums of squares for the model: $SS_{Block} = 8\sum_i \hat{\alpha}_i^2 = \frac{94.39}{3}$, $SS_A = 12\sum_j \hat{\beta}_j^2 = \frac{3144.245}{3}$, $SSE_1 = SST_1 - SS_{Block} - SS_A = \frac{3299.065-94.39-3144.245}{3} = \frac{60.43}{3}$, $SS_B = 6\sum_k \hat{\gamma}_k^2 = \frac{2583.135}{3}$, $SS_{AB} = 3\sum_{j,k} \left(\widehat{\beta\gamma}_{jk}\right)^2 = \frac{224.375}{3}$ and

$$SSE_2 = SST - SST_1 - SS_B - SS_{AB} = \frac{6193.435 - 3299.065 - 2583.135 - 224.375}{3} = \frac{86.86}{3}$$

The estimations for the variance parameters of the model are $\widehat{\sigma_2^2} = MSE_2 = \dfrac{SSE_2}{12} = 2.41$ and $\widehat{\sigma_1^2} = \dfrac{MSE_1 - MSE_2}{4} = \frac{\frac{60.43}{6} - \frac{86.86}{36}}{4} = \frac{68.93}{38} = 1.81$.

The evaluation of the F-tests in the ANOVA table leads to:

$F_{Block} = \frac{94.39/6}{60.43/6} = 1.56$ with p-value$= p\left(F_{2,2} > 1.56\right) = 0.3906$

$F_A = \frac{3144.245/3}{60.43/6} = 104.06$ with p-value$= p\left(F_{1,2} > 104.06\right) = 0.0095$

$F_B = \frac{2583.135/9}{86.86/36} = 118.96$ with p-value$= p\left(F_{3,12} > 118.96\right) = 0.0000$

$F_{AB} = \frac{224.375/9}{86.86/36} = 10.33$ with p-value$= p\left(F_{3,12} > 10.33\right) = 0.0012$

Therefore, the two factors are significative and, moreover, there is an interaction between them. The standard error for the means $\bar{y}_{\cdot j \cdot}$ is $SE = \sqrt{\dfrac{MSE_1}{12}} = 0.92$ and, taking into account that $t_{2;0.025} = 4.30$, the 95% confidence intervals are $\bar{y}_{\cdot j \cdot} \pm 3.96$. The least significant difference with $\alpha = 0.05$ is $LSD = 3.96\sqrt{2} = 5.60$. For the means $\bar{y}_{\cdot\cdot k}$, the standard error is $SE = \sqrt{\dfrac{86.86}{216}} = 0.63$ and, taking into account that $t_{12;0.025} = 2.18$, the 95% confidence intervals are $\bar{y}_{\cdot\cdot k} \pm 1.37$. The least significant difference with $\alpha = 0.05$ is $LSD = 1.37\sqrt{2} = 1.94$. Finally, for the means $\bar{y}_{\cdot jk}$, the standard error is $SE = \sqrt{\dfrac{86.86}{108}} = 0.87$ and the 95% confidence intervals are $\bar{y}_{\cdot jk} \pm 1.90$. The least significant difference with $\alpha = 0.05$ is $LSD = 1.90\sqrt{2} = 2.69$.

Using the three previous values for the least significant differences LSD we obtain the following results for the comparisons of means in the two factors:

	Fallow	Barley	Vetch	Barley-Vetch	
No nitrogen	$\frac{81}{3}$ Aa	$\frac{91.4}{3}$ Ba	$\frac{132}{3}$ Da	$\frac{113.6}{3}$ Ca	$\frac{104.5}{3}$ a
Nitrogen=120	$\frac{115.6}{3}$ Ab	$\frac{143.6}{3}$ Bb	$\frac{157}{3}$ Cb	$\frac{160.4}{3}$ Cb	$\frac{144.15}{3}$ b
	$\frac{98.3}{3}$ A	$\frac{117.5}{3}$ B	$\frac{144.5}{3}$ D	$\frac{137}{3}$ C	

where uppercase letters are used to compare the plant fertilizers and lowercase letters to compare the levels of nitrogen fertilization. Then we observe that, for all the plant fertilizers, the expected value for the production of sugar beet is higher if we use nitrogen fertilization. If we do not use nitrogen fertilization, the expected value is higher with vetch, then with barley-vetch, then with barley and finally with fallow; being the differences significant in all cases. If we use nitrogen fertilization, the same order is maintained but there is not significative differences between vetch and barley-vetch. This seems to be the reason that the interaction between the two factors is significant.

The complete LSD Fisher's test with $\alpha = 0.05$ for all the eight means leads to:

Treatment	Mean	
Fallow with no nitrogen (Control)	$\frac{81}{3}$	A
Barley with no nitrogen	$\frac{91.4}{3}$	B
Barley-Vetch with no nitrogen	$\frac{113.6}{3}$	C
Fallow with nitrogen=120	$\frac{115.6}{3}$	C
Vetch with no nitrogen	$\frac{132}{3}$	D
Barley with nitrogen=120	$\frac{143.6}{3}$	E
Vetch with nitrogen=120	$\frac{157}{3}$	F
Barley-Vetch with nitrogen=120	$\frac{160.4}{3}$	F

Chapter 8

SPLIT-SPLIT-PLOT ANOVA

The model just discussed can be generalized to the case of three factors with different sizes of experimental units for each of them. That is, three plot sizes corresponding to the three factors; namely, the whole-plots for the main factor (factor A with a levels), the intermediate size plot for the split-plot factor (factor B with b levels), and the smallest plot for the split-split-plot factor (factor C with c levels). In this case, there are three levels of precision with the whole-plot factor receiving the lowest precision, and the split-split factor receiving the highest precision. The levels of the factor A are randomly assigned to the whole-plots in a completely randomized design or a randomized complete block design; then each of the whole-plots is divided in b split-plots with one of them randomly assigned for each of the levels of the factor B, and finally, each of the split-plots is divided in c split-split-plots with one of them randomly assigned for each of the levels of the factor C.

To illustrate this type of analysis of variance, we consider a experimental design with a factor A with a levels assigned to the whole-plots in a randomized block design with r blocks (ra is the total number of whole-plots), a factor B with b levels randomly assigned to the split-plots within each plot (rab is the total number of split-plots) and a factor C with c levels randomly assigned to the split-split-plots within each split-plot ($rabc$ is the total number of split-split-plots, which are the smallest experimental units). The mathematical formulation for the model is

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \xi_{ij} + \gamma_k + \beta\gamma_{jk} + \eta_{ijk} + \delta_l + \beta\delta_{jl} + \gamma\delta_{kl} + \beta\gamma\delta_{jkl} + \varepsilon_{ijkl}$$

where μ is the global mean effect; α_i is the block effect ($i = 1, .., r$), β_j is the main effect of factor A in the whole-plots ($j = 1, .., a$); γ_k is the main effect of factor B in the split-plots ($k = 1, .., b$); δ_l is the main effect of factor C in the split-split-plots ($l = 1, .., c$); $\beta\gamma_{jk}$, $\beta\delta_{jl}$, $\gamma\delta_{kl}$ and $\beta\gamma\delta_{jkl}$ are the interaction effects as in the three-way factorial ANOVA; ξ_{ij} is a random effect of the whole-plots with $\xi_{ij} \rightsquigarrow N(0, \sigma_1)$; η_{ijk} is a random effect of the split-plots with $\eta_{ijk} \rightsquigarrow N(0, \sigma_2)$ and ε_{ijkl} is the random error for the split-split-plots in the model with $\varepsilon_{ijkl} \rightsquigarrow N(0, \sigma_3)$. As usual, we suppose that all ξ_{ij}, η_{ijk} and ε_{ijkl} are independent random variables. The parameters σ_1^2, σ_2^2 and σ_3^2 are the variance parameters for the whole-plot, the split-plots and the split-split-plots, respectively. Now, the restrictions on the parameters are: $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_k \gamma_k = 0$, $\sum_l \delta_l = 0$, $\sum_k \beta\gamma_{jk} = 0$ for all j, $\sum_j \beta\gamma_{jk} = 0$ for all k, $\sum_l \beta\delta_{jl} = 0$

for all j, $\sum_j \beta\delta_{jl} = 0$ for all l, $\sum_l \gamma\delta_{kl} = 0$ for all k, $\sum_k \gamma\delta_{kl} = 0$ for all l, $\sum_l \beta\gamma\delta_{jkl} = 0$ for all j,k, $\sum_k \beta\gamma\delta_{jkl} = 0$ for all j,l and $\sum_j \beta\gamma\delta_{jkl} = 0$ for all k,l.

Therefore, in this model we have

$$E\left(Y_{ijkl}\right) = \mu_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \beta\gamma_{jk} + \delta_l + \beta\delta_{jl} + \gamma\delta_{kl} + \beta\gamma\delta_{jkl}$$

and, with the usual notation, the appropriate estimators for the parameters of the model are $\hat{\mu} = \overline{Y}_{....}$, $\hat{\alpha}_i = \overline{Y}_{i...} - \overline{Y}_{....}$, $\hat{\beta}_j = \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{....}$, $\hat{\gamma}_k = \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{....}$, $\hat{\delta}_l = \overline{Y}_{...l} - \overline{Y}_{....}$, $\widehat{\beta\gamma}_{jk} = \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot\cdot k\cdot} + \overline{Y}_{....}$, $\widehat{\beta\delta}_{jl} = \overline{Y}_{\cdot j\cdot l} - \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{...l} + \overline{Y}_{....}$, $\widehat{\gamma\delta}_{kl} = \overline{Y}_{\cdot\cdot kl} - \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{...l} + \overline{Y}_{....}$ and finally

$$\begin{aligned} \widehat{\beta\gamma\delta}_{jkl} &= \left(\overline{Y}_{\cdot jkl} - \overline{Y}_{....}\right) - \hat{\beta}_j - \hat{\gamma}_k - \hat{\delta}_l - \widehat{\beta\gamma}_{jk} - \widehat{\beta\delta}_{jl} - \widehat{\gamma\delta}_{kl} \\ &= \overline{Y}_{\cdot jkl} - \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot\cdot kl} - \overline{Y}_{\cdot j\cdot l} + \overline{Y}_{\cdot j\cdot\cdot} + \overline{Y}_{\cdot\cdot k\cdot} + \overline{Y}_{...l} - \overline{Y}_{....} \end{aligned}$$

With these estimators we have $\hat{\mu}_{ijkl} = \overline{Y}_{i...} + \overline{Y}_{\cdot jkl} - \overline{Y}_{....}$ and the residuals are $\hat{\varepsilon}_{ijkl} = e_{ijkl} = Y_{ijkl} - \hat{\mu}_{ijkl}$.

This model has three sources of random variation, due to the existence of whole-plots (for factor A), split-plots (for factor B) and split-split-plots (for factor C). This is the reason that, first, we divide the total variation in the three parts, as follows: $SST = SST_1 + SST_2 + SST_3$, with $SST = \sum_{i,j,k,l} \left(Y_{ijkl} - \overline{Y}_{....}\right)^2$, $SST_1 = bc\sum_{i,j} \left(\overline{Y}_{ij\cdot\cdot} - \overline{Y}_{....}\right)^2$, $SST_2 = c\sum_{i,j,k} \left(\overline{Y}_{ijk\cdot} - \overline{Y}_{ij\cdot\cdot}\right)^2$ and $SST_3 = \sum_{i,j,k,l} \left(Y_{ijkl} - \overline{Y}_{ijk\cdot}\right)^2$. The degrees of freedom for these sum of squares are: $rabc - 1$ for SST, $ra - 1$ for SST_1, $ra(b-1)$ for SST_2 and $rab(c-1)$ for SST_3, so that $rabc - 1 = ra - 1 + ra(b-1) + rab(c-1)$.

The total variation SST_1 due to the whole-plots can be divided in three sources of variation: blocks, main effect of factor A and a random variation of the whole-plots, as follows

$$SS_{Block} = abc\sum_i \left(\overline{Y}_{i...} - \overline{Y}_{....}\right)^2 = abc\sum_i \hat{\alpha}_i^2 \text{ with } d.f. = r - 1$$

$$SS_A = rbc\sum_j \left(\overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{....}\right)^2 = rbc\sum_j \hat{\beta}_j^2 \text{ with } d.f. = a - 1$$

$$\text{Whole-plot error} = SSE_1 = SST_1 - SS_{Block} - SS_A = \text{ with } d.f. = (a-1)(r-1)$$

It is easily seen that SSE_1 can be alternatively written as $SSE_1 = bc\sum_{i,j} \left(\overline{Y}_{ij\cdot\cdot} - \overline{Y}_{i...} - \overline{Y}_{\cdot j\cdot\cdot} + \overline{Y}_{....}\right)^2$, that is, as an hypothetical interaction between blocks and factor A. Moreover, it verifies that $E\left(SSE_1\right) = (a-1)(r-1)\left(\sigma_3^2 + c\sigma_2^2 + bc\sigma_1^2\right)$ (see the note at the final of this experimental design).

In a similar way, the total variation SST_2 due to the split-plots can be divided in three sources of variation: main effect of factor B, interaction effect between factor A and B, and a random variation of the split-plots, as follows

$$SS_B = rac\sum_k \left(\overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{....}\right)^2 = rac\sum_k \hat{\gamma}_k^2 \text{ with } d.f. = b - 1$$

$$SS_{AB} = rc\sum_{j,k} \left(\overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot\cdot k\cdot} + \overline{Y}_{....}\right)^2 = rc\sum_{j,k} \left(\widehat{\beta\gamma}_{jk}\right)^2 \text{ with } d.f. = (a-1)(b-1)$$

$$\text{Split-plot error} = SSE_2 = SST_2 - SS_B - SS_{AB} = \text{ with } d.f. = a(b-1)(r-1)$$

Now, SSE_2 can be alternatively written as $SSE_2 = c \sum_{i,j,k} \left(\overline{Y}_{ijk\cdot} - \overline{Y}_{ij\cdot\cdot} - \overline{Y}_{\cdot jk\cdot} + \overline{Y}_{\cdot j\cdot\cdot}\right)^2$, that is, as an hypothetical interaction between blocks and factor B within each level of factor A. Moreover, it verifies that $E\left(SSE_2\right) = a(b-1)(r-1)\left(\sigma_3^2 + c\sigma_2^2\right)$ (see the note at the final of this experimental design).

Finally, the total variation SST_3 due to the split–split-plots can be divided in five sources of variation: main effect of factor C, interaction effects between factor C and factors A and B, triple interaction effect between factors A, B and C, and a random variation of the split–split-plots, as follows

$$SS_C = rab \sum_l \left(\overline{Y}_{\cdots l} - \overline{Y}_{\cdots\cdot}\right)^2 = rab \sum_l \hat{\delta}_l^2 \text{ with } d.f. = c-1$$

$$SS_{AC} = rb \sum_{j,l} \left(\overline{Y}_{\cdot j\cdot l} - \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdots l} + \overline{Y}_{\cdots}\right)^2 = rb \sum_{j,l} \left(\widehat{\beta\delta}_{jl}\right)^2 \text{ with } d.f. = (a-1)(c-1)$$

$$SS_{BC} = ra \sum_{k,l} \left(\overline{Y}_{\cdot\cdot kl} - \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{\cdots l} + \overline{Y}_{\cdots}\right)^2 = ra \sum_{k,l} \left(\widehat{\gamma\delta}_{kl}\right)^2 \text{ with } d.f. = (b-1)(c-1)$$

$$\begin{aligned} SS_{ABC} &= r \sum_{j,k,l} \left(\overline{Y}_{\cdot jkl} - \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot\cdot kl} - \overline{Y}_{\cdot j\cdot l} + \overline{Y}_{\cdot j\cdots} + \overline{Y}_{\cdot\cdot k\cdot} + \overline{Y}_{\cdots l} - \overline{Y}_{\cdots\cdot}\right)^2 \\ &= r \sum_{j,k,l} \left(\widehat{\beta\gamma\delta}_{jkl}\right)^2 \text{ with } d.f. = (a-1)(b-1)(c-1) \end{aligned}$$

Split–split-plot error $= SSE_3 = SST_3 - SS_C - SS_{AC} - SS_{BC} - SS_{ABC} =$ with $d.f. = ab(c-1)(r-1)$

Note that SSE_3 can be alternatively written as $SSE_3 = \sum_{i,j,k,l} \left(Y_{ijkl} - \overline{Y}_{ijk\cdot} - \overline{Y}_{\cdot jkl} + \overline{Y}_{\cdot jk\cdot}\right)^2$, that is, as an hypothetical interaction between blocks and factor C within each pair of levels from factors A and B. Moreover, it verifies that $E\left(SSE_3\right) = ab(c-1)(r-1)\sigma_3^2$ (see the note at the final of this experimental design).

Using the Type I estimation method (other methods are available), the estimations for the variance parameters are $\widehat{\sigma_3^2} = MSE_3 = \dfrac{SSE_3}{ab(c-1)(r-1)}$, $\widehat{\sigma_2^2} = \dfrac{MSE_2 - MSE_3}{c}$ with $MSE_2 = \dfrac{SSE_2}{a(b-1)(r-1)}$, and $\widehat{\sigma_1^2} = \dfrac{MSE_1 - MSE_2}{bc}$ with $MSE_1 = \dfrac{SSE_1}{(a-1)(r-1)}$.

In addition, the appropriate statistics for the F-tests in the ANOVA table with their respective probability distributions are:

$$F_{Block} = \frac{SS_{Block}/(r-1)}{MSE_1} \rightsquigarrow F_{r-1,(a-1)(r-1)} \quad \text{if } \alpha_i = 0 \text{ for all } i$$

$$F_A = \frac{SS_A/\left(a-1\right)}{MSE_1} \rightsquigarrow F_{a-1,(a-1)(r-1)} \quad \text{if } \beta_j = 0 \text{ for all } j$$

$$F_B = \frac{SS_B/\left(b-1\right)}{MSE_2} \rightsquigarrow F_{b-1,a(b-1)(r-1)} \quad \text{if } \gamma_k = 0 \text{ for all } k$$

$$F_{AB} = \frac{SS_{AB}/\left((a-1)(b-1)\right)}{MSE_2} \rightsquigarrow F_{(a-1)(b-1),a(b-1)(r-1)} \quad \text{if } \beta\gamma_{jk} = 0 \text{ for all } j,k$$

$$F_C = \frac{SS_C/\left(c-1\right)}{MSE_3} \rightsquigarrow F_{c-1,ab(c-1)(r-1)} \quad \text{if } \delta_l = 0 \text{ for all } l$$

$$F_{AC} = \frac{SS_{AC}/\left((a-1)(c-1)\right)}{MSE_3} \rightsquigarrow F_{(a-1)(c-1),ab(c-1)(r-1)} \quad \text{if } \beta\delta_{jl} = 0 \text{ for all } j,l$$

$$F_{BC} = \frac{SS_{BC}/\left((b-1)(c-1)\right)}{MSE_3} \rightsquigarrow F_{(b-1)(c-1),ab(c-1)(r-1)} \quad \text{if } \gamma\delta_{kl} = 0 \text{ for all } k,l$$

$$F_{ABC} = \frac{SS_{ABC}/\left((a-1)(b-1)(c-1)\right)}{MSE_3} = \frac{MS_{ABC}}{MSE_3} \rightsquigarrow F_{(a-1)(b-1)(c-1),ab(c-1)(r-1)} \quad \text{if } \beta\gamma\delta_{jkl} = 0 \text{ for all } j,k,l$$

The ANOVA table for this model is as follows:

	SS	DF	MS	F-value	p-value	H_0
Blocks	SS_{Block}	$r-1$	MS_{Block}	F_{Block}	p_{Block}	$\alpha_i = 0$ for all i
Factor A	SS_A	$a-1$	MS_A	F_A	p_A	$\beta_j = 0$ for all j
Plot error	SSE_1	$(a-1)(r-1)$	MSE_1			
Total plots	SST_1	$ar-1$				
Factor B	SS_B	$b-1$	MS_B	F_B	p_B	$\gamma_k = 0$ for all k
Int. $A*B$	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_{AB}	p_{AB}	$\beta\gamma_{jk} = 0$ for all j,k
Split-plot error	SSE_2	$a(b-1)(r-1)$	MSE_2			
Total split-plots	SST_2	$ar(b-1)$				
Factor C	SS_C	$c-1$	MS_C	F_C	p_C	$\delta_l = 0$ for all l
Int. $A*C$	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	F_{AC}	p_{AC}	$\beta\delta_{jl} = 0$ for all j,l
Int. $B*C$	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	F_{BC}	p_{BC}	$\gamma\delta_{kl} = 0$ for all k,l
Int. $A*B*C$	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	F_{ABC}	p_{ABC}	$\beta\gamma\delta_{jkl} = 0$ for all j,k,l
Ssplit-plot error	SSE_3	$ab(c-1)(r-1)$	MSE_3			
Total ssplit-plots	SST_3	$abr(c-1)$				
Total	SST	$abcr-1$				

The confidence intervals for the marginal expected values are:

$$\overline{Y}_{i\cdots} \pm t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{MSE_1}{abc}} \quad \text{for } \mu_{i\cdots} = \mu + \alpha_i$$

$$\overline{Y}_{\cdot j\cdot\cdot} \pm t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{MSE_1}{rbc}} \quad \text{for } \mu_{\cdot j\cdot\cdot} = \mu + \beta_j$$

$$\overline{Y}_{\cdot\cdot k\cdot} \pm t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{MSE_2}{rac}} \quad \text{for } \mu_{\cdot\cdot k\cdot} = \mu + \gamma_k$$

$$\overline{Y}_{\cdots l} \pm t_{ab(c-1)(r-1);\alpha/2}\sqrt{\frac{MSE_3}{rab}} \quad \text{for } \mu_{\cdots l} = \mu + \delta_l$$

$$\overline{Y}_{\cdot jk\cdot} \pm t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{MSE_2}{rc}} \quad \text{for } \mu_{\cdot jk\cdot} = \mu + \beta_j + \gamma_k + \beta\gamma_{jk}$$

$$\overline{Y}_{\cdot j\cdot l} \pm t_{ab(c-1)(r-1);\alpha/2}\sqrt{\frac{MSE_3}{rb}} \quad \text{for } \mu_{\cdot j\cdot l} = \mu + \beta_j + \delta_l + \beta\delta_{jl}$$

$$\overline{Y}_{\cdot\cdot kl} \pm t_{ab(c-1)(r-1);\alpha/2}\sqrt{\frac{MSE_3}{ra}} \quad \text{for } \mu_{\cdot\cdot kl} = \mu + \gamma_k + \delta_l + \gamma\delta_{kl}$$

$$\overline{Y}_{\cdot jkl} \pm t_{ab(c-1)(r-1);\alpha/2}\sqrt{\frac{MSE_3}{r}} \quad \text{for } \mu_{\cdot jkl} = \mu + \beta_j + \gamma_k + \beta\gamma_{jk} + \delta_l + \beta\delta_{jl} + \gamma\delta_{kl} + \beta\gamma\delta_{jkl}$$

Finally, the least significant differences (LSD) for each pair of means are:

$$LSD = t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_1}{abc}} \quad \text{for } \overline{Y}_{i\cdots} - \overline{Y}_{i'\cdots}$$
$$LSD = t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_1}{rbc}} \quad \text{for } \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdot j'\cdot\cdot}$$
$$LSD = t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_2}{rac}} \quad \text{for } \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{\cdot\cdot k'\cdot}$$
$$LSD = t_{ab(c-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_3}{rab}} \quad \text{for } \overline{Y}_{\cdots l} - \overline{Y}_{\cdots l'}$$
$$LSD = t_{(a-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_2}{rc}} \quad \text{for } \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdot j'k'\cdot}$$
$$LSD = t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_3}{rb}} \quad \text{for } \overline{Y}_{\cdot j\cdot l} - \overline{Y}_{\cdot j'\cdot l'}$$
$$LSD = t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_3}{ra}} \quad \text{for } \overline{Y}_{\cdot\cdot kl} - \overline{Y}_{\cdot\cdot k'l'}$$
$$LSD = t_{a(b-1)(r-1);\alpha/2}\sqrt{\frac{2MSE_3}{r}} \quad \text{for } \overline{Y}_{\cdot jkl} - \overline{Y}_{\cdot j'k'l'}$$

The check of the assumptions of the model can be made as usual with the estimated residuals e_{ijkl}.

It should be noted that if there are no blocks in the model then the parameters α_i and the sum SS_{Block} do not exist and now

$$\text{Whole-plot error} = SSE_1 = bc\sum_{i,j}\left(Y_{ij\cdot\cdot} - \overline{Y}_{\cdot j\cdot\cdot}\right)^2 \text{ with } d.f. = a(r-1)$$

and $\hat{\mu}_{jkl} = \overline{Y}_{\cdot jkl}$ with residuals $\hat{\varepsilon}_{ijkl} = e_{ijkl} = Y_{ijkl} - \overline{Y}_{\cdot jkl}$, but everything else being equal with replicates instead of blocks.

Note. To prove that $E(SSE_1) = (a-1)(r-1)\left(bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right)$, first of all we observe that, for fixed i_0 and j_0, we have $E\left(\overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{i_0\cdots} - \overline{Y}_{\cdot j_0\cdot\cdot} + \overline{Y}_{\cdots\cdot}\right) = 0$ and therefore, for fixed values i_0 and j_0, it verifies that:

$$E\left(\left(\overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{i_0\cdots} - \overline{Y}_{\cdot j_0\cdot\cdot} + \overline{Y}_{\cdots\cdot}\right)^2\right) = Var\left(\overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{i_0\cdots} - \overline{Y}_{\cdot j_0\cdot\cdot} + \overline{Y}_{\cdots\cdot}\right)$$
$$= Var\left(\frac{1}{bc}\sum_{k,l}Y_{i_0j_0kl} - \frac{1}{abc}\sum_{j,k,l}Y_{i_0jkl} - \frac{1}{rbc}\sum_{i,k,l}Y_{ij_0kl} + \frac{1}{rabc}\sum_{i,j,k,l}Y_{ijkl}\right)$$
$$= Var\left(\begin{array}{c}\left(\frac{1}{bc} - \frac{1}{abc} - \frac{1}{rbc} + \frac{1}{rabc}\right)\sum_{k,l}Y_{i_0j_0kl} - \left(\frac{1}{abc} - \frac{1}{rabc}\right)\sum_{j,k,l,j\neq j_0}Y_{i_0jkl} \\ -\left(\frac{1}{rbc} - \frac{1}{rabc}\right)\sum_{i,k,l,i\neq i_0}Y_{ij_0kl} + \frac{1}{rabc}\sum_{i,j,k,l,i\neq i_0,j\neq j_0}Y_{ijkl}\end{array}\right)$$
$$= \frac{(a-1)^2(r-1)^2}{r^2a^2b^2c^2}Var\left[bc\xi_{i_0j_0} + c\sum_k \eta_{i_0j_0k} + \sum_{k,l}\varepsilon_{i_0j_0kl}\right]$$
$$+\frac{(r-1)^2}{r^2a^2b^2c^2}Var\left[bc\sum_{j,j\neq j_0}\xi_{i_0j} + c\sum_{j,k,j\neq j_0}\eta_{i_0jk} + \sum_{j,k,l,j\neq j_0}\varepsilon_{i_0jkl}\right]$$
$$+\frac{(a-1)^2}{r^2a^2b^2c^2}Var\left[bc\sum_{i,i\neq i_0}\xi_{ij_0} + c\sum_{i,k,i\neq i_0}\eta_{ij_0k} + \sum_{i,k,l,i\neq i_0}\varepsilon_{ij_0kl}\right]$$
$$+\frac{1}{r^2a^2b^2c^2}Var\left[bc\sum_{i,j,i\neq i_0,j\neq j_0}\xi_{ij} + c\sum_{i,j,k,i\neq i_0,j\neq j_0}\eta_{ijk} + \sum_{i,j,k,l,i\neq i_0,j\neq j_0}\varepsilon_{ijkl}\right]$$

$$= \frac{(a-1)^2 (r-1)^2}{r^2a^2b^2c^2} \left[b^2c^2\sigma_1^2 + c^2b\sigma_2^2 + bc\sigma_3^2\right]$$
$$+\frac{(r-1)^2}{r^2a^2b^2c^2} \left[b^2c^2(a-1)\sigma_1^2 + c^2b(a-1)\sigma_2^2 + cb(a-1)\sigma_1^2\right]$$
$$+\frac{(a-1)^2}{r^2a^2b^2c^2} \left[b^2c^2(r-1)\sigma_1^2 + c^2b(r-1)\sigma_2^2 + cb(r-1)\sigma_1^2\right]$$
$$+\frac{1}{r^2a^2b^2c^2} \left[b^2c^2(r-1)(a-1)\sigma_1^2 + c^2b(r-1)(a-1)\sigma_2^2 + cb(r-1)(a-1)\sigma_3^2\right]$$
$$= \frac{(a-1)^2 (r-1)^2}{r^2a^2bc} \left[bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right] + \frac{(r-1)^2(a-1)}{r^2a^2bc} \left[bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right]$$
$$+\frac{(a-1)^2(r-1)}{r^2a^2bc} \left[bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right] + \frac{(a-1)(r-1)}{r^2a^2bc} \left[bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right]$$
$$= \frac{(a-1)(r-1)}{r^2a^2bc} \left[(a-1)(r-1) + r - 1 + a - 1 + 1\right] \left[bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right]$$
$$= \frac{(a-1)(r-1)}{rabc} \left[bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right]$$

Therefore. we have

$$E\left(SSE_1\right) = bc \sum_{i,j} \frac{(a-1)(r-1)}{rabc} \left(bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right) = (a-1)\left(r-1\right)\left(bc\sigma_1^2 + c\sigma_2^2 + \sigma_3^2\right)$$

In a similar way, to prove that $E\left(SSE_2\right) = a\left(b-1\right)\left(r-1\right)\left(c\sigma_2^2 + \sigma_3^2\right)$, we observe that, for fixed values i_0, j_0 and k_0, the expression $\overline{Y}_{i_0j_0k_0\cdot} - \overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{\cdot j_0k_0\cdot} + \overline{Y}_{\cdot j_0\cdot\cdot}$ does not depend on the values ξ_{ij_0} (each of them appears two times, one with positive sign and one with negative sign). Then we have:

$$E\left(\left(\overline{Y}_{i_0j_0k_0\cdot} - \overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{\cdot j_0k_0\cdot} + \overline{Y}_{\cdot j_0\cdot\cdot}\right)^2\right) = Var\left(\overline{Y}_{i_0j_0k_0\cdot} - \overline{Y}_{i_0j_0\cdot\cdot} - \overline{Y}_{\cdot j_0k_0\cdot} + \overline{Y}_{\cdot j_0\cdot\cdot}\right)$$
$$= Var\left(\frac{1}{c}\sum_l Y_{i_0j_0k_0l} - \frac{1}{bc}\sum_{k,l} Y_{i_0j_0kl} - \frac{1}{rc}\sum_{i,l} Y_{ij_0k_0l} + \frac{1}{rbc}\sum_{i,k,l} Y_{ij_0kl}\right)$$
$$= Var\left(\begin{array}{c} \left(\frac{1}{c} - \frac{1}{bc} - \frac{1}{rc} + \frac{1}{rbc}\right)\sum_l Y_{i_0j_0k_0l} - \left(\frac{1}{bc} - \frac{1}{rbc}\right)\sum_{k,l,k\neq k_0} Y_{i_0j_0kl} \\ -\left(\frac{1}{rc} - \frac{1}{rbc}\right)\sum_{i,l,i\neq i_0} Y_{ij_0k_0l} + \frac{1}{rbc}\sum_{i,k,l,i\neq i_0,k\neq k_0} Y_{ij_0kl} \end{array}\right)$$
$$= \frac{(b-1)^2(r-1)^2}{r^2b^2c^2}\left[Var\left(c\eta_{i_0j_0k_0}\right) + Var\left(\sum_l \varepsilon_{i_0j_0k_0l}\right)\right]$$
$$+\frac{(r-1)^2}{r^2b^2c^2}\left[\sum_{k,k\neq k_0} Var\left(c\eta_{i_0j_0k}\right) + Var\left(\sum_{k,l,k\neq k_0}\varepsilon_{i_0j_0k_0l}\right)\right]$$
$$+\frac{(b-1)^2}{r^2b^2c^2}\left[\sum_{i,i\neq i_0} Var\left(c\eta_{ij_0k_0}\right) + Var\left(\sum_{i,l,i\neq i_0}\varepsilon_{i_0j_0k_0l}\right)\right]$$
$$+\frac{1}{r^2b^2c^2}\left[\sum_{i,k,i\neq i_0,k\neq k_0} Var\left(c\eta_{ij_0k}\right) + Var\left(\sum_{i,l,i\neq i_0}\varepsilon_{ij_0kl}\right)\right]$$
$$= \frac{(b-1)^2(r-1)^2}{r^2b^2c^2}(c^2\sigma_2^2 + c\sigma_3^2) + \frac{(r-1)^2}{r^2b^2c^2}(b-1)(c^2\sigma_2^2 + c\sigma_3^2)$$
$$+\frac{(b-1)^2}{r^2b^2c^2}(r-1)(c^2\sigma_2^2 + c\sigma_3^2) + \frac{1}{r^2b^2c^2}(r-1)(b-1)(c^2\sigma_2^2 + c\sigma_3^2)$$

$$= \frac{(b-1)(r-1)[(b-1)(r-1)+(r-1)+(b-1)+1]}{r^2b^2c}(c\sigma_2^2+\sigma_3^2)$$
$$= \frac{(b-1)(r-1)}{rbc}(c\sigma_2^2+\sigma_3^2)$$

Therefore, we have

$$E(SSE_2) = c\sum_{i,j,k}\frac{(b-1)(r-1)}{rbc}(c\sigma_2^2+\sigma_3^2) = a(b-1)(r-1)\left(c\sigma_2^2+\sigma_3^2\right)$$

Finally, to prove that $E(SSE_3) = ab(c-1)(r-1)\sigma_3^2$, we observe that, for fixed values i_0, j_0,k_0 and l_0, the expression $Y_{i_0j_0k_0l_0} - \overline{Y}_{i_0j_0k_0\cdot} - \overline{Y}_{\cdot j_0k_0l_0} + \overline{Y}_{\cdot j_0k_0\cdot}$ does not depend on ξ_{ij_0} or $\eta_{ij_0k_0}$, because of the same reason as before. Then we have:

$$E\left(\left(Y_{i_0j_0k_0l_0} - \overline{Y}_{i_0j_0k_0\cdot} - \overline{Y}_{\cdot j_0k_0l_0} + \overline{Y}_{\cdot j_0k_0\cdot}\right)^2\right) = Var\left(Y_{i_0j_0k_0l_0} - \overline{Y}_{i_0j_0k_0\cdot} - \overline{Y}_{\cdot j_0k_0l_0} + \overline{Y}_{\cdot j_0k_0\cdot}\right)$$
$$= Var\left(Y_{i_0j_0k_0l_0} - \frac{1}{c}\sum_{l}Y_{i_0j_0k_0l} - \frac{1}{r}\sum_{i}Y_{ij_0k_0l_0} + \frac{1}{rc}\sum_{i,l}Y_{ij_0k_0l}\right)$$
$$= Var\left(\left(1-\frac{1}{c}-\frac{1}{r}+\frac{1}{rc}\right)Y_{i_0j_0k_0l_0} - \left(\frac{1}{c}-\frac{1}{rc}\right)\sum_{l,l\neq l_0}Y_{i_0j_0k_0l} - \left(\frac{1}{r}-\frac{1}{rc}\right)\sum_{i,i\neq i_0}Y_{ij_0k_0l_0} + \frac{1}{rc}\sum_{i,l,i\neq i_0,l\neq l_0}Y_{ij_0k_0l}\right)$$
$$= \frac{(c-1)^2(r-1)^2}{r^2c^2}Var\left(\varepsilon_{i_0j_0k_0l_0}\right) + \frac{(r-1)^2}{r^2c^2}\sum_{l,l\neq l_0}Var\left(\varepsilon_{i_0j_0k_0l}\right)$$
$$+\frac{(c-1)^2}{r^2c^2}\sum_{i,i\neq i_0}Var\left(\varepsilon_{ij_0k_0l_0}\right) + \frac{1}{r^2c^2}\sum_{i,l,i\neq i_0,l\neq l_0}Var\left(\varepsilon_{ij_0k_0l}\right)$$
$$= \frac{(c-1)^2(r-1)^2}{r^2c^2}\sigma_3^2 + \frac{(r-1)^2}{r^2c^2}(c-1)\sigma_3^2 + \frac{(c-1)^2}{r^2c^2}(r-1)\sigma_3^2 + \frac{1}{r^2c^2}(r-1)(c-1)\sigma_3^2$$
$$= \frac{(c-1)(r-1)[(c-1)(r-1)+(r-1)+(c-1)+1]}{r^2c^2}\sigma_3^2$$
$$= \frac{(c-1)(r-1)}{rc}\sigma_3^2$$

Therefore, we have

$$E(SSE_3) = \sum_{i,j,k,l}\frac{(c-1)(r-1)}{r}\sigma_3^2 = ab(c-1)(r-1)\sigma_3^2$$

Example 9: Split–split-plot design. We want to evaluate the effect of planting date (factor A), aphid control (factor B) and the date of harvesting (factor C) on the production of sugar beet. Three planting dates were chosen: March 2 (level 1), April 2 (level 2) and May 2 (level 3). Two levels were established for factor B: applying a treatment for the aphid (level 2) and not applying any treatment (level 1, which is the control). Finally, three harvest dates were chosen for factor C: August 27 (level 1), September 24 (level 2) and October 22 (level 3). The available area of land for the experiment was divided in four blocks because of the possible differences in physical and chemical soil characteristics. Three whole-plots in each block are established and a planting date is randomly assigned to each one. Then, each of the whole-plots was divided into two split-plots, one of which is sprayed against aphids and other not, at random. Finally, each of the split-plots is divided into three split–split-plots to each of them, at random, it is applied to a harvest date. At the end of the experiment, the production of each of the 72 split-split-plots expressed in tons per acre was recorded. We have, therefore, three types of experimental units: the 12 whole-plots associated with planting dates (3 levels with 4 repetitions, one per block), 24 split-plots associated with aphid control (2 levels with 12 repetitions, 3 per block) and 72 split-split-plots associated with the harvest date (3 levels with 24 repetitions, 6 per block).

For the statistical analysis we use the split–split-plot model

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \xi_{ij} + \gamma_k + \beta\gamma_{jk} + \eta_{ijk} + \delta_l + \beta\delta_{jl} + \gamma\delta_{kl} + \beta\gamma\delta_{jkl} + \varepsilon_{ijkl}$$

which leads to $E\left(Y_{ijkl}\right) = \mu + \alpha_i + \beta_j + \gamma_k + \beta\gamma_{jk} + \delta_l + \beta\delta_{jl} + \gamma\delta_{kl} + \beta\gamma\delta_{jkl}$ with i=1, 2, 3, 4 for the blocks; j=1, 2, 3 for the planting date, k=1, 2 for the aphid control and $l = 1, 2, 3$ for the harvest date. That is, just the model previously studied with $r = 4$, $a = 3$, $b = 2$ and $c = 3$. The observed data and some of the means in the model are given in the following table:

Treatments			Blocks				Treatment
A_j	B_k	C_l	$i = 1$	$i = 2$	$i = 3$	$i = 4$	totals $\bar{y}_{\cdot jkl}$
1	1	1	25.7	25.4	23.8	22.0	$\bar{y}_{\cdot 111} = \frac{96.9}{4}$
1	1	2	31.8	29.5	28.7	26.4	$\bar{y}_{\cdot 112} = \frac{116.4}{4}$
1	1	3	34.6	37.2	29.1	23.7	$\bar{y}_{\cdot 113} = \frac{124.6}{4}$
Split-plot $\bar{y}_{i11\cdot}$			$\bar{y}_{111\cdot} = \frac{92.1}{3}$	$\bar{y}_{211\cdot} = \frac{92.1}{3}$	$\bar{y}_{311\cdot} = \frac{81.6}{3}$	$\bar{y}_{411\cdot} = \frac{72.1}{3}$	$\bar{y}_{\cdot 11\cdot} = \frac{337.9}{12}$
1	2	1	27.7	30.3	30.2	33.2	$\bar{y}_{\cdot 121} = \frac{121.4}{4}$
1	2	2	38.0	40.6	34.6	31.0	$\bar{y}_{\cdot 122} = \frac{144.2}{4}$
1	2	3	42.1	43.6	44.6	42.7	$\bar{y}_{\cdot 123} = \frac{173.0}{4}$
Split-plot $\bar{y}_{i12\cdot}$			$\bar{y}_{112\cdot} = \frac{107.8}{3}$	$\bar{y}_{212\cdot} = \frac{114.5}{3}$	$\bar{y}_{312\cdot} = \frac{109.4}{3}$	$\bar{y}_{412\cdot} = \frac{106.9}{3}$	$\bar{y}_{\cdot 12\cdot} = \frac{438.6}{12}$
Whole-plot $\bar{y}_{i1\cdot\cdot}$			$\bar{y}_{11\cdot\cdot} = \frac{199.9}{6}$	$\bar{y}_{21\cdot\cdot} = \frac{206.6}{6}$	$\bar{y}_{31\cdot\cdot} = \frac{191.0}{6}$	$\bar{y}_{41\cdot\cdot} = \frac{179.0}{6}$	$\bar{y}_{\cdot 1\cdot\cdot} = \frac{776.5}{24}$
2	1	1	28.9	24.7	27.8	23.4	$\bar{y}_{\cdot 211} = \frac{104.8}{4}$
2	1	2	37.5	31.5	31.0	27.8	$\bar{y}_{\cdot 212} = \frac{127.8}{4}$
2	1	3	38.4	32.5	31.2	29.8	$\bar{y}_{\cdot 213} = \frac{131.9}{4}$
Split-plot $\bar{y}_{i21\cdot}$			$\bar{y}_{121\cdot} = \frac{104.8}{3}$	$\bar{y}_{221\cdot} = \frac{88.7}{3}$	$\bar{y}_{321\cdot} = \frac{90.0}{3}$	$\bar{y}_{421\cdot} = \frac{81.0}{3}$	$\bar{y}_{\cdot 21\cdot} = \frac{364.5}{12}$
2	2	1	38.0	31.0	29.5	30.7	$\bar{y}_{\cdot 221} = \frac{129.2}{4}$
2	2	2	36.9	31.9	31.5	35.9	$\bar{y}_{\cdot 222} = \frac{136.2}{4}$
2	2	3	44.2	41.6	38.9	37.6	$\bar{y}_{\cdot 223} = \frac{162.3}{4}$
Split-plot $\bar{y}_{i22\cdot}$			$\bar{y}_{122\cdot} = \frac{119.1}{3}$	$\bar{y}_{222\cdot} = \frac{104.5}{3}$	$\bar{y}_{322\cdot} = \frac{99.9}{3}$	$\bar{y}_{422\cdot} = \frac{104.2}{3}$	$\bar{y}_{\cdot 22\cdot} = \frac{427.7}{12}$
Whole-plot $\bar{y}_{i2\cdot\cdot}$			$\bar{y}_{12\cdot\cdot} = \frac{223.9}{6}$	$\bar{y}_{22\cdot\cdot} = \frac{193.2}{6}$	$\bar{y}_{32\cdot\cdot} = \frac{189.9}{6}$	$\bar{y}_{42\cdot\cdot} = \frac{185.2}{6}$	$\bar{y}_{\cdot 2\cdot\cdot} = \frac{792.2}{24}$
3	1	1	23.4	24.2	21.2	20.9	$\bar{y}_{\cdot 311} = \frac{89.7}{4}$
3	1	2	25.3	27.7	23.7	24.3	$\bar{y}_{\cdot 312} = \frac{101.0}{4}$
3	1	3	29.8	29.9	24.3	23.8	$\bar{y}_{\cdot 313} = \frac{107.8}{4}$
Split-plot $\bar{y}_{i31\cdot}$			$\bar{y}_{131\cdot} = \frac{78.5}{3}$	$\bar{y}_{231\cdot} = \frac{81.8}{3}$	$\bar{y}_{331\cdot} = \frac{69.2}{3}$	$\bar{y}_{431\cdot} = \frac{69.0}{3}$	$\bar{y}_{\cdot 31\cdot} = \frac{298.5}{12}$
3	2	1	20.8	23.0	25.2	23.1	$\bar{y}_{\cdot 321} = \frac{92.1}{4}$
3	2	2	29.0	32.0	26.5	31.2	$\bar{y}_{\cdot 322} = \frac{118.7}{4}$
3	2	3	36.6	37.8	34.8	40.2	$\bar{y}_{\cdot 323} = \frac{149.4}{4}$
Split-plot $\bar{y}_{i32\cdot}$			$\bar{y}_{132\cdot} = \frac{86.4}{3}$	$\bar{y}_{232\cdot} = \frac{92.8}{3}$	$\bar{y}_{332\cdot} = \frac{86.5}{3}$	$\bar{y}_{432\cdot} = \frac{94.5}{3}$	$\bar{y}_{\cdot 32\cdot} = \frac{360.2}{12}$
Whole-plot $\bar{y}_{i3\cdot\cdot}$			$\bar{y}_{13\cdot\cdot} = \frac{164.9}{6}$	$\bar{y}_{23\cdot\cdot} = \frac{174.6}{6}$	$\bar{y}_{33\cdot\cdot} = \frac{155.7}{6}$	$\bar{y}_{43\cdot\cdot} \frac{163.5}{6}$	$\bar{y}_{\cdot 3\cdot\cdot} = \frac{658.7}{24}$
Total block $\bar{y}_{i\cdots}$			$\bar{y}_{1\cdots} = \frac{588.7}{18}$	$\bar{y}_{2\cdots} = \frac{574.4}{18}$	$\bar{y}_{3\cdots} = \frac{536.6}{18}$	$\bar{y}_{4\cdots} = \frac{527.7}{18}$	$\bar{y}_{\cdots\cdot} = \frac{2227.4}{72}$

The rest of the means are given in the following table:

$\bar{y}_{\cdot 1 \cdot 1} = \frac{1964.7}{72}$	$\bar{y}_{\cdot 2 \cdot 1} = \frac{2106}{72}$	$\bar{y}_{\cdot 3 \cdot 1} = \frac{1636.2}{72}$	$\bar{y}_{\cdot\cdot 11} = \frac{1748.4}{72}$	$\bar{y}_{\cdot\cdot 21} = \frac{2056.2}{72}$	$\bar{y}_{\cdot\cdot\cdot 1} = \frac{1902.3}{72}$
$\bar{y}_{\cdot 1 \cdot 2} = \frac{2345.4}{72}$	$\bar{y}_{\cdot 2 \cdot 2} = \frac{2376}{72}$	$\bar{y}_{\cdot 3 \cdot 2} = \frac{1977.3}{72}$	$\bar{y}_{\cdot\cdot 12} = \frac{2071.2}{72}$	$\bar{y}_{\cdot\cdot 22} = \frac{2394.6}{72}$	$\bar{y}_{\cdot\cdot\cdot 2} = \frac{2232.9}{72}$
$\bar{y}_{\cdot 1 \cdot 3} = \frac{2678.4}{72}$	$\bar{y}_{\cdot 2 \cdot 3} = \frac{2647.8}{72}$	$\bar{y}_{\cdot 3 \cdot 3} = \frac{2314.8}{72}$	$\bar{y}_{\cdot\cdot 13} = \frac{2185.8}{72}$	$\bar{y}_{\cdot\cdot 23} = \frac{2908.2}{72}$	$\bar{y}_{\cdot\cdot\cdot 3} = \frac{2547}{72}$
$\bar{y}_{\cdot 1 \cdot\cdot} = \frac{2329.5}{72}$	$\bar{y}_{\cdot 2 \cdot\cdot} = \frac{2376.6}{72}$	$\bar{y}_{\cdot 3 \cdot\cdot} = \frac{1976.1}{72}$	$\bar{y}_{\cdot\cdot 1 \cdot} = \frac{2001.8}{72}$	$\bar{y}_{\cdot\cdot 2 \cdot} = \frac{2453}{72}$	$\bar{y}_{\cdot\cdot\cdot\cdot} = \frac{2227.4}{72}$

The total sums are $\sum\limits_{i,j,k,l} y_{ijkl} = 2227.4$, $\sum\limits_{i,j,k,l} y_{ijkl}^2 = 71747.7$ and, therefore, $SST = 71747.7 - \frac{2227.4^2}{72} = \frac{204523.64}{72}$. In a similar way, for the whole-plots, the total sums are $\sum\limits_{i,j} \bar{y}_{ij\cdot\cdot} = \frac{2227.4}{6}$, $\sum\limits_{i,j} \bar{y}_{ij\cdot\cdot}^2 = \frac{417635.98}{36}$ and, therefore, $SST_1 = 6\left(\frac{417635.98}{36} - \frac{\left(\frac{2227.4}{6}\right)^2}{12}\right) = \frac{50321}{72}$. Finally, for the split-plots, we have $\sum\limits_{i,j,k} \bar{y}_{ijk\cdot}^2 = \frac{211295.72}{9}$, $\sum\limits_{i,j,k} \bar{y}_{ij\cdot\cdot}^2 = \frac{417635.98}{18}$ and, therefore, $SST_2 = 3\left(\frac{211295.72}{9} - \frac{417635.98}{18}\right) = 825.91$.

First of all, the estimated parameters $\hat{\mu} = \bar{y}_{\cdots\cdot}$ and $\hat{\alpha}_i = \bar{y}_{i\cdots} - \bar{y}_{\cdots\cdot}$ are included in the following table:

$\hat{\mu}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$
$\frac{2227.4}{72}$	$\frac{127.4}{72}$	$\frac{70.2}{72}$	$-\frac{81}{72}$	$-\frac{116.6}{72}$

In a similar way, the estimated parameters $\hat{\beta}_j = \overline{Y}_{\cdot j\cdot\cdot} - \overline{Y}_{\cdots\cdot}$, $\hat{\gamma}_k = \overline{Y}_{\cdot\cdot k\cdot} - \overline{Y}_{\cdots\cdot}$, $\hat{\delta}_l = \overline{Y}_{\cdots l} - \overline{Y}_{\cdots\cdot}$, $\widehat{\beta\delta}_{jl} = \overline{Y}_{\cdot j\cdot l} - \overline{Y}_{\cdots\cdot} - \hat{\beta}_j - \hat{\delta}_l$ and $\widehat{\gamma\delta}_{kl} = \overline{Y}_{\cdot\cdot kl} - \overline{Y}_{\cdots\cdot} - \hat{\gamma}_k - \hat{\delta}_l$ are included in the following tables:

$\widehat{\beta\delta}_{11} = -\frac{39.7}{72}$	$\widehat{\beta\delta}_{21} = \frac{54.5}{72}$	$\widehat{\beta\delta}_{31} = -\frac{14.8}{72}$	$\hat{\delta}_1 = -\frac{325.1}{72}$
$\widehat{\beta\delta}_{12} = \frac{10.4}{72}$	$\widehat{\beta\delta}_{22} = -\frac{6.1}{72}$	$\widehat{\beta\delta}_{32} = -\frac{4.3}{72}$	$\hat{\delta}_2 = \frac{5.5}{72}$
$\widehat{\beta\delta}_{13} = \frac{29.3}{72}$	$\widehat{\beta\delta}_{23} = -\frac{48.4}{72}$	$\widehat{\beta\delta}_{33} = \frac{19.1}{72}$	$\hat{\delta}_3 = \frac{319.6}{72}$
$\hat{\beta}_1 = \frac{102.1}{72}$	$\hat{\beta}_2 = \frac{149.2}{72}$	$\hat{\beta}_3 = \frac{-251.3}{72}$	

$\widehat{\gamma\delta}_{11} = \frac{71.7}{72}$	$\widehat{\gamma\delta}_{21} = -\frac{71.7}{72}$
$\widehat{\gamma\delta}_{12} = \frac{63.9}{72}$	$\widehat{\gamma\delta}_{22} = -\frac{63.9}{72}$
$\widehat{\gamma\delta}_{13} = -\frac{135.6}{72}$	$\widehat{\gamma\delta}_{23} = \frac{135.6}{72}$
$\hat{\gamma}_1 = -\frac{225.6}{72}$	$\hat{\gamma}_2 = \frac{225.6}{72}$

Finally, the estimated parameters $\widehat{\beta\gamma}_{jk} = \overline{Y}_{\cdot jk\cdot} - \overline{Y}_{\cdots\cdot} - \hat{\beta}_j - \hat{\gamma}_k$ and

$$\widehat{\beta\gamma\delta}_{jkl} = \overline{Y}_{\cdot jkl} - \overline{Y}_{\cdots\cdot} - \hat{\beta}_j - \hat{\gamma}_k - \hat{\delta}_l - \widehat{\beta\gamma}_{jk} - \widehat{\beta\delta}_{jl} - \widehat{\gamma\delta}_{kl}$$

are included in the following table:

$\widehat{\beta\gamma\delta}_{111} = \frac{9.9}{72}$	$\widehat{\beta\gamma\delta}_{211} = -\frac{101.7}{72}$	$\widehat{\beta\gamma\delta}_{311} = \frac{91.8}{72}$	$\widehat{\beta\gamma\delta}_{121} = -\frac{9.9}{72}$	$\widehat{\beta\gamma\delta}_{221} = \frac{101.7}{72}$	$\widehat{\beta\gamma\delta}_{321} = -\frac{91.8}{72}$
$\widehat{\beta\gamma\delta}_{112} = -\frac{12}{72}$	$\widehat{\beta\gamma\delta}_{212} = \frac{50.1}{72}$	$\widehat{\beta\gamma\delta}_{312} = -\frac{38.1}{72}$	$\widehat{\beta\gamma\delta}_{122} = \frac{12}{72}$	$\widehat{\beta\gamma\delta}_{222} = -\frac{50.1}{72}$	$\widehat{\beta\gamma\delta}_{322} = \frac{38.1}{72}$
$\widehat{\beta\gamma\delta}_{113} = \frac{2.1}{72}$	$\widehat{\beta\gamma\delta}_{213} = \frac{51.6}{72}$	$\widehat{\beta\gamma\delta}_{313} = -\frac{53.7}{72}$	$\widehat{\beta\gamma\delta}_{123} = -\frac{2.1}{72}$	$\widehat{\beta\gamma\delta}_{223} = -\frac{51.6}{72}$	$\widehat{\beta\gamma\delta}_{323} = \frac{53.7}{72}$
$\widehat{\beta\gamma}_{11} = -\frac{76.5}{72}$	$\widehat{\beta\gamma}_{21} = \frac{36}{72}$	$\widehat{\beta\gamma}_{31} = \frac{40.5}{72}$	$\widehat{\beta\gamma}_{12} = \frac{76.5}{72}$	$\widehat{\beta\gamma}_{22} = -\frac{36}{72}$	$\widehat{\beta\gamma}_{32} = -\frac{40.5}{72}$

Now we can evaluate all the sums of squares for the model: $SS_{Block} = 18\sum\limits_i \hat{\alpha}_i^2 = \frac{10328.84}{72}$, $SS_A = 24\sum\limits_j \hat{\beta}_j^2 = \frac{31945.58}{72}$, $SSE_1 = SST_1 - SS_{Block} - SS_A = \frac{50321 - 10328.84 - 31945.58}{72} = \frac{8046.58}{72}$, $SS_B = 36\sum\limits_k \hat{\gamma}_k^2 = \frac{50895.36}{72}$, $SS_{AB} = 12\sum\limits_{j,k}\left(\widehat{\beta\gamma}_{jk}\right)^2 = \frac{2929.5}{72}$, $SSE_2 = SST_2 - SS_B - SS_{AB} = \frac{59465.52 - 50895.36 - 2929.5}{72} = \frac{5640.66}{72}$, $SS_C = 24\sum\limits_l \hat{\delta}_l^2 = \frac{69288.14}{72}$, $SS_{AC} = 8\sum\limits_{j,l}\left(\widehat{\beta\delta}_{jl}\right)^2 = \frac{943.9}{72}$, $SS_{BC} = 12\sum\limits_{k,l}\left(\widehat{\gamma\delta}_{kl}\right)^2 = \frac{9203.82}{72}$,

$SS_{ABC} = 4 \sum_{j,k,l} \left(\widehat{\beta\gamma\delta}_{jkl}\right)^2 = \frac{3169.38}{72}$, and finally

$$\begin{aligned} SSE_3 &= SST - SST_1 - SST_2 - SS_C - SS_{AC} - SS_{BC} - SS_{ABC} = \\ &= \frac{204523.64-50321-59465.52-69288.14-943.9-9203.82-3169.38}{72} = \frac{12131.88}{72} \end{aligned}$$

Then we have $MSE_3 = \dfrac{SSE_3}{36} = 4.68$, $MSE_2 = \dfrac{SSE_2}{9} = 8.70$, $MSE_1 = \dfrac{SSE_1}{6} = 18.63$, and the estimations for the variance parameters of the model are $\widehat{\sigma_3^2}=MSE_3=4.68$, $\widehat{\sigma_2^2}=\dfrac{MSE_2 - MSE_3}{3}=1.34$ and $\widehat{\sigma_1^2}=\dfrac{MSE_1 - MSE_2}{6}=1.66$.

The evaluation of the F-tests in the ANOVA table leads to:

$$\begin{aligned}
F_{Block} &= \frac{SS_{Block}/3}{MSE_1} = \frac{10328.84/216}{18.63} = 2.57 \text{ with p-value= } p\left(F_{3,6} > 2.57\right) = 0.1502 \\
F_A &= \frac{SS_A/2}{MSE_1} = \frac{31945.58/144}{18.63} = 11.91 \text{ with p-value= } p\left(F_{2,6} > 11.91\right) = 0.0081 \\
F_B &= \frac{SS_B}{MSE_2} = \frac{50895.36/72}{8.70} = 81.21 \text{ with p-value= } p\left(F_{1,9} > 81.21\right) = 0.0000 \\
F_{AB} &= \frac{SS_{AB}/2}{MSE_2} = \frac{2929.5/144}{8.70} = 2.34 \text{ with p-value= } p\left(F_{2,9} > 2.34\right) = 0.1522 \\
F_C &= \frac{SS_C/2}{MSE_3} = \frac{69288.14/144}{4.68} = 102.80 \text{ with p-value= } p\left(F_{2,36} > 102.80\right) = 0.0000 \\
F_{AC} &= \frac{SS_{AC}/4}{MSE_3} = \frac{943.9/288}{4.68} = 0.70 \text{ with p-value= } p\left(F_{4,36} > 0.70\right) = 0.5969 \\
F_{BC} &= \frac{SS_{BC}/2}{MSE_3} = \frac{9203.82/144}{4.68} = 13.66 \text{ with p-value= } p\left(F_{2,36} > 13.66\right) = 0.0000 \\
F_{ABC} &= \frac{SS_{ABC}/4}{MSE_3} = \frac{3169.38/288}{4.68} = 2.35 \text{ with p-value= } p\left(F_{4,36} > 2.35\right) = 0.0725
\end{aligned}$$

Therefore, there are a significant effect of factor A (planting date) and a significative interaction between factors B and C (aphid control and harvest date). For the planting date, the standard error for the means $\bar{y}_{\cdot j \cdot\cdot}$ is $\sqrt{\dfrac{MSE_1}{24}} = 0.88$ and, taking into account that $t_{6;0.025} = 2.45$, the 95%-confidence intervals for the means are $\bar{y}_{\cdot j \cdot\cdot} \pm 2.16$. The least significant difference (LSD) for these means is $LSD_A = 2.16\sqrt{2} = 3.05$ and the LSD-Fisher test leads to:

Planting date	Mean	
May 2	$\frac{1976.1}{72} = 27.45$	A
March 2	$\frac{2329.5}{72} = 32.35$	B
April 2	$\frac{2376.6}{72} = 33.01$	B

that is, planting at May cause a significantly decrease on the average production and there is no significant difference between March and April.

For the interaction effect between the factors B and C, the standard error for the means $\bar{y}_{\cdot\cdot kl}$ is $\sqrt{\dfrac{MSE_3}{12}} = 0.62$ and, taking into account that $t_{36;0.025} = 2.03$, the 95%-confidence intervals for the means are $\bar{y}_{\cdot\cdot kl} \pm 1.26$. The least significant difference (LSD) for these means is $LSD_{BC} = 1.26\sqrt{2} = 1.78$. In a similar way, the standard error for the means $\bar{y}_{\cdot\cdot\cdot l}$ is $\sqrt{\dfrac{MSE_3}{24}} = 0.44$, the 95%-confidence intervals for the means are $\bar{y}_{\cdot\cdot\cdot l} \pm 0.89$ and the least significant difference (LSD) for these means is $LSD_C =$

$0.89\sqrt{2} = 1.26$. Finally, the standard error for the means $\bar{y}_{\cdot\cdot k\cdot}$ is $\sqrt{\frac{MSE_2}{36}} = 0.49$ and, taking into account that $t_{9;0.025} = 2.26$, the 95%-confidence intervals for the means are $\bar{y}_{\cdot\cdot k\cdot} \pm 1.11$ and the least significant difference (LSD) for these means is $LSD_B = 1.11\sqrt{2} = 1.57$.

Using the three previous values for the least significant differences LSD we obtain the following results for the comparisons of means in the two factors:

	Harvest date			
	August 27	September 24	October 22	
No aphid control	$\frac{1748.4}{72} = 24.28$ Aa	$\frac{2071.2}{72} = 28.77$ Ba	$\frac{2185.8}{72} = 30.36$ Ba	$\frac{2001.8}{72} = 27.80$ a
Treatment for aphid	$\frac{2056.2}{72} = 28.56$ Ab	$\frac{2394.6}{72} = 33.26$ Bb	$\frac{2908.2}{72} = 40.39$ Cb	$\frac{2453}{72} = 34.07$ b
	$\frac{1902.3}{72} = 26.42$ A	$\frac{2232.9}{72} = 31.01$ B	$\frac{2547}{72} = 35.38$ C	

where uppercase letters are used to compare the harvest dates and lowercase letters to compare the treatment for the aphid. Therefore, we observe that the average production is always higher with the treatment for the aphid and, in general, it is better late harvesting. However, with no aphid control, there is no significant differences between September 24 and October 22. This seems to be the reason that the interaction between the two factors is significant.

The complete LSD Fisher's test with $\alpha = 0.05$ for all the means in the interaction BC leads to:

Treatment	Mean	
No aphid control with harvest date August 27	24.28	A
Treatment for the aphid with harvest date August 27	28.56	B
No aphid control with harvest date September 24	28.77	BC
No aphid control with harvest date October 22	30.36	C
Treatment for the aphid with harvest date September 24	33.26	D
Treatment for the aphid with harvest date October 22	40.39	E

It seems clear that the best option would be to plant in March or April, applying treatment for the aphid, and to harvest in October.