

Chapter 3

3.1

In this chapter we will discuss the ANOVA model, but not only the one-factor case, but also we will extend it to the two- and three-factor models. A characteristic of the two-factor ANOVA model is that it considers the effect of two factors or independent variables on a dependent variable. Each factor consists of two or more levels. This gives us a factorial design. There are three reasons for including more than one factor:

Researcher is interested in studying the second factor. By allowing two factors in one analysis we can show the effects of both factors collectively, rather than the effect of each individual factor (main effects). So we can show interaction effects, this shows us whether the factors are operating independent from one another (no interaction exists) or whether they operate together to produce some additional impact (interaction exists). In this test we will have to test three hypotheses: one for each factor or main effect, and one for the interaction effect.

Also including another factor decreases the error (within-group) variation, which is the variation unexplained by the first factor. Because of this a two-factor design is more powerful than a one-factor design.

It also provides a greater generalizability of the results, this will allow for more use of the observations.

In the two-factor ANOVA every level of the first factor is paired with every level of the second factor, this is called a fully crossed design. Individuals are randomly assigned to one of the factors. We consider in this chapter that all the factors are fixed, so a fixed-effect model. Another condition for the factorial ANOVA is that the dependent variable is measured at least at the interval level and the independent variables are categorical (either nominal or ordinal).

So in short, the characteristics of the two-factor ANOVA fixed-effects model are: (a) two independent variables (both are categorical), with each two or more levels, (b) levels of independent variables are fixed, (c) subjects are randomly assigned to only one combination of these levels, (d) two factors are fully crossed, and (e) the dependent variable is measured at least at the interval level.

In the context of experimental design, the two-factor ANOVA is also called the complete randomized factorial design.

We will now discuss the layout of the data. We name each observation Y_{ijk} . Where j tells us what level (or category) of factor A the observation belongs to, the k tells us what level of factor B, and the i tells us the observation or identification number within that combination of factor A and factor B. The ranges for the subscripts are: $i = 1, \dots, n$; $j = 1, \dots, J$; and $k = 1, \dots, K$.

There are total of $JKn = N$ observations, because there are J levels of factor A, K levels of factor B and n subjects in each cell.

The two-factor ANOVA model is a form of the general linear model (GLM). The two-factor ANOVA fixed-effects model can be written in terms of population parameters as:

Where

Y_{ijk} is the observed score on the criterion (i.e. dependent) variable for individual i in the level j of factor A and level k of factor B (or in the jk cell).

μ is the overall or grand population mean

α_j is the main effect for level j of factor A (row effect)

β_k is the main effect for level k of factor B (column effect)

$(\alpha\beta)_{jk}$ is the interaction effect for the combination of level j of factor A and level k of factor B

ε_{ijk} is the random residual error for individual i in cell jk .

This residual error can be due to different thing, for example measurement error, individual differences or factor not under investigation.

The population effects and residual error can be computed as follows:

$$\begin{aligned}\alpha_j &= \mu_{.j} - \mu \\ \beta_k &= \mu_{.k} - \mu \\ (\alpha\beta)_{jk} &= \mu_{jk} - (\mu_{.j} + \mu_{.k} - \mu) \quad \text{or} \quad (\alpha\beta)_{jk} = \mu_{jk} - \alpha_j - \beta_k - \mu \\ \varepsilon_{ijk} &= Y_{ijk} - \mu_{jk}\end{aligned} \quad (38)$$

The row effect is equal to the difference between the population mean of level j of factor A and the overall population mean. The column effect is equal to the difference between the population mean of level k of factor B and the overall population mean. The interaction effect is the effect of being in a certain combination of the levels of factor A and factor B. Whereas the residual error is equal to the difference between an individual's observed score and the population mean of cell jk .

You can also see the row, column and interaction effect as the average effect of being a member of a particular row, column or cell.

The sum of the row effect is equal to 0, and also the sum of the column and interaction effect is equal to 0.

To estimate the parameter of the model (from formula (37)), we use the least squares method of estimation. They are represented by: \bar{Y} , \bar{a}_j , \bar{b}_k $(ab)_{jk}$ and e_{ijk} . Where the last four are computed as follows:

$$\begin{aligned}\bar{a}_j &= \bar{Y}_{.j} - \bar{Y}_{...} \\ \bar{b}_k &= \bar{Y}_{.k} - \bar{Y}_{...} \\ (ab)_{jk} &= \bar{Y}_{jk} - (\bar{Y}_{.j} + \bar{Y}_{.k} - \bar{Y}_{...}) \\ e_{ijk} &= Y_{ijk} - \bar{Y}_{jk}\end{aligned} \quad (39)$$

$\bar{Y}_{...}$ represents the overall sample mean

$\bar{Y}_{.j}$ represents the sample mean for level j of factor A

$\bar{Y}_{.k}$ represents the sample mean for level k of factor B

\bar{Y}_{jk} represents the sample mean for cell jk

As mentioned before, there are three hypotheses. For each of the main effects and one for the interaction effect. We can write them as:

$$\begin{aligned}H_{01} &= \alpha_1 = \alpha_2 = \dots = \alpha_J = 0 \\ H_{02} &= \beta_1 = \beta_2 = \dots = \beta_K = 0 \\ H_{03} &= (\alpha\beta)_{jk} = 0 \text{ for all } j \text{ and } k\end{aligned}$$

These are all the null hypotheses. 1 and 2 are for the main effects and their alternative hypothesis is that not all the means are equal. The third one is the null hypothesis for the interaction effect. If one of the null hypotheses is rejected, then consider an MCP so as to determine which means, or combination of means, are significantly different.

The exact definition of the main effect of factor A is the effect of factor A, averaged across the levels of factor B, on the dependent variable Y. So it represents the unique effect of factor A on the outcome Y, controlling statistically for factor B.

For the interaction effect the exact definition has different options: An interaction is said to exist if (a) certain combinations of the two factors produce effects beyond the effects of the two factors when those two factors are considered separately; (b) the mean difference among the levels of factor A are not constant across, and thus depend on, the levels of factor B; (c) there is a joint effect of factors A and B on Y; or (d) there is a unique effect that could not be predicted from knowledge of only the main effect.

Profile plots of the two factors in ANOVA can give information about the possible existence of a main effect for A, a main effect for B, and/or an interaction effect. The main effect for factor A (B) can be examined by taking the means for each level of A (B) and averaging them across the levels of B (A). If these marginal means are for each level of A the same or nearly so, this will indicate no main effect for factor A (B).

The interaction effect is determined by whether the cell means for the levels of A are constant across the levels of B (or vice versa). This can be easily checked in the profile plot by determining whether or not the lines are parallel. Parallel lines indicate no interaction, whereas nonparallel lines suggest that an interaction can exist.

If both lines lie horizontal, there is not effect for A. The other way around when they are not horizontal there is a main effect for factor A. When the lines are really close to one another there is not difference in the value of B and so not main effect for factor B. When they are far from each other there is a difference and so a main effect for factor B. When the lines cross each other exactly in the middle it means there are no main effects, because the means for both A and B are the same. (see page 88 for the graphs).

There is made a distinction between the type of interaction shown in the profile plot. An ordinal interaction is said to exist when the lines are not parallel and they do not cross. A disordinal interaction is said to exist when the lines are not parallel and they do cross.

When there is no significant interaction effect, the main effects can be generalized with greater confidence. In this situation the main effects are known as additive effects. However when there is a significant interaction effect, the findings regarding the main effects cannot be generalized with such confidence. In this case the main effects are not additive and also the interaction effect needs to be included in the model. This is something you can see in the profile plots.

The assumptions and effects of violations for the Two-factor ANOVA design:

1. Independence. The effects of assumption violation:

- Increased likelihood of Type I and/or Type II error in F statistic
- Influences standard errors of means and thus inferences about those means

2. Homogeneity of variance. The effects of assumption violation:

- Bias in SS_{with}
- Increased likelihood of a Type I and/or Type II error
- Less effect with balanced or nearly balanced design
- Effect decreased as n increases

3. Normality. The effects of assumption violation:

- Minimal effect with moderate violation
- Minimal effect with balanced or nearly balanced design
- Effect decreases as n increases.

These assumptions look like the assumptions of the one-factor ANOVA. However there are two differences. First, the effect of heterogeneity is small with balanced designs or nearly balanced designs, and/or with larger n 's. Second, the effect of nonnormality seems to be the same as heterogeneity.

We looked with the one-factor ANOVA at the partitioning the sums of squares. For the two-factor ANOVA we can partition SS_{total} into:

$$SS_{\text{total}} = SSA + SSB + SSAB + SS_{\text{with}} \quad (40)$$

SS_{total} denotes the total sum of squares in Y . This represents the amount of total variation among all of the observations without regard to row, column or cell membership. The SSA is the variation between the levels of factor A. SSB denotes the variation between the levels of factor B. The variation due to interaction between the levels of factors A and B is denoted as $SSAB$. The variation within the cells combined across cells is denoted as SS_{with} . The ANOVA summary table for the two-factor ANOVA is:

Source	SS	df	MS	F
A	SSA	J-1	MSA	MSA/ MS_{with}
B	SSB	K-1	MSB	MSB/ MS_{with}
AB	SSAB	(J-1)(K-1)	MSAB	MSAB/ MS_{with}
Within	SS_{with}	N-JK	MS_{with}	
Total	SS_{total}	N-1		

These different columns are also explained in the summary table for the one-factor ANOVA. Again the MS (mean squares terms) are generated by dividing SS by the degrees of freedom.

The last column, the F statistics is the value with which we will test the hypotheses. Each of the F-values will be compared to the F critical value from table A.4 to make conclusions. The null hypothesis is rejected if the F test statistic exceeds the F critical value.

But when the F statistic does exceed the critical value, and there are more than one degree of freedom for the source being tested, then it is not clear precisely why the null hypothesis was rejected. In this case some MCP should be used to determine where the mean differences are.

We will now extend the MCP procedures to the two-factor ANOVA model. We will now examine the contrasts of both the main and interaction effects. We will begin with the main effects. When the effect for factor A is significant, and there are more than two levels of factor A, we can form contrasts that compare the levels of factor A, ignoring factor B. Considering each factor separately is advised.

For contrasts involving the interaction, the idea is to begin with a complex interaction contrast if there are more than four cells in the model. An example for such contrast is as follows

[where e.g. $(\bar{Y}_{.11} + \bar{Y}_{.21} + \bar{Y}_{.31} + \bar{Y}_{.41})$ is the sum of the cell means of each level of factor A for level 1 of factor B and $(\bar{Y}_{.12} + \bar{Y}_{.22} + \bar{Y}_{.32} + \bar{Y}_{.42})$ is the sum of the cell means of each level of factor A for level 2 of factor B]:

$$\Psi' = \frac{\bar{Y}_{.11} + \bar{Y}_{.21} + \bar{Y}_{.31} + \bar{Y}_{.41}}{4} - \frac{\bar{Y}_{.12} + \bar{Y}_{.22} + \bar{Y}_{.32} + \bar{Y}_{.42}}{4} \quad (41)$$

With a standard error of:

$$s_{\Psi'} = \sqrt{MS_{\text{within}} \left(\sum_{j=1}^J \sum_{k=1}^K \frac{c_{jk}^2}{n_{jk}} \right)} \quad (42)$$

where n_{jk} is the number of observations in cell jk . This contrast would examine the interaction between the four methods of factor A and the first two of factor B. If the complex interaction contrast is significant, then you can follow this up with a simpler interaction contrast that involves only four cell means. This is a single degree of freedom contrast, because it involves only two levels of each factor (tetrad difference). Example of such contrast is:

$$\Psi' = (\bar{Y}_{.11} - \bar{Y}_{.21}) - (\bar{Y}_{.12} - \bar{Y}_{.22}) \quad (43)$$

The error term is the same as formula (42). There are various measures of effect size, as discussed in chapter 1. For the two-factor ANOVA we will discuss two commonly used measures, which assume equal variances across the cells. The first is the partial eta squared, η^2 . This measure shows the portion of variation in Y, which is explained by the effect of interest. This effect of interest can be factor A, factor B or the AB interaction. We determine partial η^2 as follows:

$$\begin{aligned} \text{Partial } \eta^2_A &= \frac{SS_A}{SS_A + SS_{\text{within}}} \\ \text{Partial } \eta^2_B &= \frac{SS_B}{SS_B + SS_{\text{within}}} \\ \text{Partial } \eta^2_{AB} &= \frac{SS_{AB}}{SS_{AB} + SS_{\text{within}}} \end{aligned} \quad (44)$$

Another measure of effect size is omega squared statistic, ω^2 :

$$\begin{aligned} \omega^2_A &= \frac{SS_A - (J-1)MS_{\text{within}}}{SS_{\text{total}} + MS_{\text{within}}} \\ \omega^2_B &= \frac{SS_B - (K-1)MS_{\text{within}}}{SS_{\text{total}} + MS_{\text{within}}} \\ \omega^2_{AB} &= \frac{SS_{AB} - (J-1)(K-1)MS_{\text{within}}}{SS_{\text{total}} + MS_{\text{within}}} \end{aligned} \quad (45)$$

These measures can be interpreted as follows: η^2 or $\omega^2 = .01$ is a small effect. Medium effect is a η^2 or $\omega^2 = .06$ and a large effect is η^2 or $\omega^2 = .14$.

As mentioned before CIs can be used for providing interval estimates of a population mean or mean difference. For the two-factor ANOVA we can form CIs for row means, column means, cell means, the overall mean and any possible contrast formed through an MCP.

3.1.10 gives an example of all the theory explained in chapter 3.1

As learned for the one-factor ANOVA the expected mean square for a particular source of variation represents the average mean square value for that source obtained if the same study were to be replicated an infinite number of times. For the two-factor ANOVA there are two situations. The first is that the H_0 is actually being true. In this case there is no main effect or interaction effect and the expected mean squares are:

$$\begin{aligned} E(MS_A) &= \sigma_\varepsilon^2 \\ E(MS_B) &= \sigma_\varepsilon^2 \\ E(MS_{AB}) &= \sigma_\varepsilon^2 \\ E(MS_{\text{within}}) &= \sigma_\varepsilon^2 \end{aligned} \quad (46)$$

So when using MS_{within} as the error term will produce F values around 1.

The second situation is that the H_0 is actually false, so there are main effects and an interaction effect. Then the expected mean squares are as follows:

$$\begin{aligned} E(MS_A) &= \sigma_\varepsilon^2 + (nK \sum_{j=1}^J \alpha_j^2) / (J - 1) \\ E(MS_B) &= \sigma_\varepsilon^2 + (nJ \sum_{k=1}^K \beta_k^2) / (K - 1) \\ E(MS_{AB}) &= \sigma_\varepsilon^2 + [n \sum_{j=1}^J \sum_{k=1}^K (\alpha\beta)_{jk}^2] / (J - 1)(K - 1) \\ E(MS_{\text{within}}) &= \sigma_\varepsilon^2 \end{aligned} \quad (47)$$

In this case, using MS_{within} as error term will yield F values > 1 .

All the information together will give us that in general the F ratio represents:

$$F = (\text{systematic variability} + \text{error variability}) / (\text{error variability}) \quad (48)$$

Where for the two-factor ANOVA the systematic variability is due to the main or interaction effects. The error variability is variability within.

3.2

We will now take a look at the three factor- and higher-order ANOVA model. All of the characteristics discussed for the two-factor ANOVA model also apply for the three-factor model. However there is one exception, there are now three factors instead of two. This will result in other interactions, namely three main effects (one for each factor, A, B and C), there are now three interaction effects (AB, BC, and AC), and there will be a three-way interaction (ABC). This three-way interaction is stated as: Is the AB interaction constant across all levels of factor C? (or AC across levels of B, or BC across levels of A).

The model for the three-factor design is:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_l + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} + \varepsilon_{ijk} \quad (49)$$

As you can see this model is the same as the two-factor model (formula (37)), but now the third factor is added and the new interaction effect and three-way interaction. In this model: Y_{ijk} is the observed score on the criterion (i.e. dependent) variable for individual i in the level j of factor A and level k of factor B, and level l of the factor C (or in the ijk cell).

μ is the overall or grand population mean

α_j is the effect for level j of factor A

β_k is the effect for level k of factor B

γ_l is the effect for level l of factor C

$(\alpha\beta)_{jk}$ is the interaction effect for the combination of level j of factor A and level k of factor B

$(\alpha\gamma)_{jl}$ is the interaction effect for the combination of level j of factor A and level l of factor C

$(\beta\gamma)_{kl}$ is the interaction effect for the combination of level k of factor B and level l of factor C

$(\alpha\beta\gamma)_{jkl}$ is the interaction effect for the combination of the level j of factor A, level k of factor B, and level l of factor C (three-way interaction)

ε_{ijk} is the random residual error for individual i in cell ijk .

The hypotheses are the same as the ones for the two-way model, however they are expanded with the new main effect, interaction effect and the three-way interaction.

Three-factor ANOVA summary table

Source	SS	Df	MS	F
A	SSA	J-1	MSA	MSA/MSwith
B	SSB	K-1	MSB	MSB/MSwith
C	SSC	L-1	MSc	MSC/MSwith
AB	SSAB	(J-1)(K-1)	MSAB	MSAB/MSwith
AC	SSAC	(J-1)(L-1)	MSAC	MSAC/MSwith
BC	SSBC	(K-1)(L-1)	MSBC	MSBC/MSwith
ABC	SSABC	(J-1)(K-1)(L-1)	MSABC	MSABC/MSwith
Within	SSwith	N-JKL	MSwith	
Total	SStotal	N-1		

The row within can be named in SPSS as error. This shows the within group sum of squares that tells us how much variation there is within the cells combined across the cells. The row total can be labelled in SPSS as “corrected total” that is the sum of squares total.

All the other things that are used and assumed for the two-factor model are the same for the three-factor model. The only new thing is the three-way interaction. If this interaction is significant, then this means that the two-way interaction is different across the levels of the third factor.

The inclusion of extra factor will result most of the time in a lower MSwith, which is a good thing. However, adding extra factors, also gives more risks. You need to take into account the possibility of significant higher-order interactions. This is mostly hard to interpret and deal with it.

3.3

Until now we have assumed that we will deal with equal n's or balanced case, this makes the equations easier. However this is not always the case, and we will look at the unequal n's case. When there is an unequal n's case, the main effects and interaction effects are not orthogonal. This means that the sums of squares cannot be partitioned into independent effects (as in formula (40)). Thus the individual SS do not necessarily add up to the SS_{total}.

There are three modern approaches for the unequal n's case. Each of these approaches uses different hypotheses and thus may not have the same results:

Sequential approach (hierarchical sums of squares approach):

In this approach the effects being tested are as follows:

$$\begin{array}{l} \alpha | \mu \\ \beta | \mu, \alpha \\ \alpha\beta | \mu, \alpha, \beta \end{array}$$

For example this means, that the effect for factor B (β) is adjusted or controls for (denoted by the vertical line) the overall mean (μ) and the main effect due to the factor A (α). (In SPSS this is the Type I sum of squares method). Here the interaction is not taken into account in estimating the main effects.

Partially sequential approach (partially hierarchical, or experimental design, or method of fitting constants approach).

In this method the effects are being tested as follows:

$$\begin{array}{l} \alpha | \mu, \beta \\ \beta | \mu, \alpha \\ \alpha\beta | \mu, \alpha, \beta \end{array}$$

The difference with the first method is that in this method every main effect controls for the other main effect, but not for the interaction effect. (in SPSS this is called the Type II sum of squared method). This is the only method where the sum of the squares will be equal to the total sum of squares.

Also in this method is the interaction effect not included in the determination of the main effects.

Regression approach (marginal means or unique approach). In this method the effects are being tested as follows:

$$\begin{array}{l} \alpha | \mu, \beta, \alpha\beta \\ \beta | \mu, \alpha, \alpha\beta \\ \alpha\beta | \mu, \alpha, \beta \end{array}$$

In this method, every effect controls for all the other effects. A difference with the first two methods is that this method also includes the interaction effect in the determination of the main effects. (in SPSS this is called the Type III sum of squares method). This method is the most used, because it looks most like the traditional ANOVA.

When you would use one of these three methods for the balanced n's case, all of the approaches will give the same hypotheses and results.

3.4

Steps for conducting a factorial ANOVA in SPSS:

- Go to “Analyze” and select “General Linear Model”, and then select “Univariate”.
- Move the dependent variable in the “dependent variable” box. Move first and second independent variable in the “Fixed Factors” box and click on “options”.
- Select in “Options” which tests you will do, mostly we select “Descriptive statistics”, “Estimates of effect size”, “Observed power”, “Homogeneity tests” and “Spread vs. level plot”. Then click on “continue”.
- From the “Univariate” box, click on “Plots” to obtain a profile plot of the means. Click the independent variable in the “Horizontal Axis” box. (Most easy to plot the independent variable with the most categories or levels on the horizontal axis, this will make it easier to interpret the graph). Then click the second independent variable in the “Separate Lines” box. Then click “Add” to move the variable into the “Plots” box. Then click on “continue”.
- From the “Univariate” box click on “Post Hoc” to select various post hoc MCPs or click on “contrasts” to select various planned MCPs. From the “Post Hoc Multiple Comparisons for Observed Means” dialog box, click on the names of the independent variables in the “Factor(s)” list and move them to “Post Hoc Tests for” box. Check an appropriate MCP. (Here we use “Tukey”). Click on “continue”.
- From “Univariate” box click on “save” and select elements you want to save. (In our case, we save the unstandardized residuals that will be used later to examine the extent to which normality and independence are met). Then click on “Ok” to generate output.

Interpreting the output is important. A lot of the results we already have discussed, but I will add some extra points.

The F test that is named in the Levene’s Test of Equality of Error Variances is examined to determine if equal variances can be assumed.

The R squared is listed as a footnote underneath the test results. This shows the ratio of sum of squares between (i.e. combined SS for main effects and for the interaction) divided by sum of squares total.

The observed power tells us if the test is powerful enough to detect mean differences if they really exist. For example, power of 1.000 indicates the maximum probability of rejecting the null hypothesis if it is really false (very strong power).

“Sig”-column shows the p values and provides the results of the contrasts.

The Spread vs. level plots are plots of the dependent variable standard deviations (or variances) against the cell means. These are used to determine what to do when homogeneity of variance assumption has been violated. Homogeneity is suggested when the plot provides a random display of points. If the plot suggests linear relationships, transforming the data by taking the log of the dependent variable values may be a solution to the heterogeneity. Taking the log is a solution because logarithms require positive values. Another solution can be transforming the data by taking the square root of the dependent variable values.

We will now look at how we need to examine data for the different assumptions:

Assumption of normality

We use the residuals to examine the extent to which the normality is met. We use explore to examine the extent to which the assumption of normality is met. You can look at a histogram to see if the assumption is met. Also other statistics can be used to examine this assumption. The formal test of normality, the Shapiro-Wilk (S-W) provides evidence of the extent to which our sample distribution is statistically different from a normal distribution.

Also the Q-Q plots are used. If the points fall on or close to the diagonal line, there is evidence of normality.

The boxplot can also be used to examine normality. There is evidence for normality if there is a relatively normal distributional shape of residuals and no outliers.

Assumption of Independence

We can plot the residuals against levels of our independent variables in a scatterplot to get an idea of whether or not there are patterns in the data and thereby provide an indication of whether we have met this assumption. We have in this case more independent variables, so we will split the scatterplot by levels of one independent variable and then generate a bivariate scatterplot for the other independent variable by residual.

In examining the scatterplots for evidence of independence, the points should fall relatively randomly above and below a horizontal line at 0.

We will now look at the Post Hoc power and the Priori power of the factorial ANOVA. First we will discuss the Post Hoc power. When there are multiple independent variables, G*Power must be calculated for each main effect and for each interaction. This is already explained in chapter 2.

Second, we look at the priori power. We can determine the total sample size needed for the main effects and/or interactions given an estimated effect size, f , alpha level, desired power, numerator degrees of freedom and number of groups or cells. Also see chapter 2 for the explanation.