Contents

2 Analysis of Variance

2.1 Introduction

We will now move on with Hypothesis testing, and introduce some of the ideas and techniques that allow us to tackle more complex experimental questions. Before studying the mathematics, we will look at the kind of data that may arise.

2.2 Experimental design

Consider the following scientific question: it is required to know whether a planting mix (fertiliser, soil, other additives) is a better growing medium for a particular variety of plant species.

To address this question, we might have 5 particular mixes, and 125 of the plants to test them on. In order for the test to be unbiased, that is, to eliminate other factors, the 125 units would be assigned to the 5 growing media at random, to eliminate other variables. Arranging a test like this is an example of a designed experiment.

The test is organised in this way to try and eliminate the random variation that may influence the results, aside from the variation in the growing media. In particular, it attempts to eliminate, or quantify, the variability between the subjects being used.

This example illustrates some definitions.

2.2.1 Definition – Units/Subjects

The objects upon which measurements are taken are called *experimental units* or *subjects*.

2.2.2 Definition – Factor

Variables or effects completely controlled by the experimenter are called *factors*. The factor is divided into *levels*.

In the example discussed above, the units or subjects were the plants grown in the different growing mixtures. The growing medium makeup, which can be varied from one unit to another, is the factor. There were five levels, that is, five particular mixes.

An experiment may have more than one factor involved, each with its own levels.

2.2.3 Definition – Treatment

A *treatment* is a specific combination of levels of the factors involved in an experiment that affect any given subject in the experiment. If there is only one factor, then the treatments are the levels of that factor.

2.2.4 The Response Variable

In the example above, it must also be clear how exactly the effect of the growing medium is reflected in the subjects. In other words, what exactly is meant by ‘the best medium’? The response can be measured as the yield of fruit, or the height of the plant. In other words, we must decide what variable measures the response of the subjects to the treatments.

The decision about the response of the groups of subjects will be made by comparing the means of the responses grouped by treatments. The Null hypothesis will be that the means are the same.

2.3 Experimental Designs

The first step in putting together an experiment, once the subjects and factors have been decided, is to decide how to arrange the subjects for receiving a treatment.

In the example given, the 125 units all receive a treatment, and they are randomly assigned. This is an example of a *completely randomised design*.

2.3.1 Definition – Completely Randomised Design.

A *completely randomised design* to compare k treatments is one where a group of N relatively homogenous units are randomly divided into k subgroups, of size n_1, n_2, n_3, … n_k, such that

$$n_1 + n_2 + n_3 \ldots + n_k = N.$$

Then all of the units in each subgroup receive the same treatment.

In order to test the results, it is assumed that each sub-group, of size n_i, for $i = 1$ to N, is an independent random selection from a much larger population. It can be argued that applying treatments to units is the same, from a statistical view, as selecting the units from existing populations.

Here is an important definition, which at first looks like the same thing, but will be extended in a different direction.

2.3.2 Definition – One-Way Layout

Say a set of experimental subjects are to be given a treatment with k levels. A *one way layout* to compare k populations is one in which k independent random samples are taken from each of the populations of interest. Each population receives a different level of the treatment.

By extension, in a *two-way layout*, there would be two treatments, and a corresponding number of levels.

Here is what this layout looks like in practice. Say we have 25 subjects, call them S_1, S_2, … S_{25}, and a factor with 4 levels, call them *A* to *D*. The subjects will be assigned as follows:

Treatment:	A	B	C	D
Subjects:	S_1, S_2, S_3, S_4, S_5.	S_6, S_7, S_8, S_9, S_{10}, S_{11}, S_{12}, S_{13}.	S_{14}, S_{15}, S_{16}, S_{17}, S_{18}.	S_{19}, S_{20}, S_{21}, S_{22}, S_{23}, S_{24}, S_{25}.

Note that the numbers in each group are not necessarily the same.

Selecting populations like this is usually done to compare means, which is an extension of our t-test examples, where we looked at samples from two populations. The next step up is where there are more than two groups, leading to these layouts.

Let Y be the response variable, and let μ_A be the mean for group A, let μ_B be the mean for group B and so on. The Null hypothesis in this case will be

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D.$$

In the context of a one-way layout, it may be beneficial to ensure that the same numbers of subjects are given each treatment; this means the variability within groups can be more accurately compared.

If this is done, we are dealing with the idea of blocks and a randomised block design, laid out in the following definition.

2.3.3 Definition – Randomised Block Design

A *randomised block design* to compare k treatments is one in which b blocks of experimental units are set up, with each block having k units so that each treatment appears exactly once in the block.

The above definition means that the total number of subjects N, the number of blocks b and the number of treatments k must follow the equation $N = bk$. This also mean the number of treatments divides into the number of available subjects evenly.

Here is a Randomised Block Design for a one-way layout, with 4 treatments and 20 subjects. A block in this case is a selection of subjects across each row of the table.

Randomised Block Design:

Treatment:	A	B	C	D
	S_1,	S_6,	S_{11},	S_{16},
Subjects:	S_2,	S_7,	S_{12},	S_{17},
	S_3,	S_8,	S_{13},	S_{18},

	S_4, S_5.	S_9, S_{10}.	S_{14}, S_{15}.	S_{19}, S_{20}.

2.3.4 An Example of a RBD

A sample of 16 components is available to test switching times of a component under 4 different treatments, where each treatment is a particular combination of temperature and humidity.

In this example, the first, and simplest method would be to assign the 4 treatments to the 16 units (at random), so that each treatment is applied to 4 of the components. This would be a case of the randomised block design.

2.3.5 Within/Between Subjects

In the example just quoted, applying each treatment to the same number of subjects does not, however, completely rule out the variability of the components themselves, so an alternative would be, if it is possible, to carry out each treatment on each one of the 16 components, giving 64 results.

This would also be a randomised block design, but now it is referred to as a *within-subjects design*, referring to the fact that repeated measures are taken on a given subject. In this context, the original version was a *between-subjects design*.

The t-test we saw before on differences was a ‘before and after’ case, where the means of two groups were compared. Each unit in the ‘before’ group was matched with one in the ‘after’ group (they were the same unit), leading to the name *matched pairs* for this test.

These ideas are summarised in a definition:

2.3.6 Definition – Between/Within Subjects Designs

An Experimental Design is referred to as *Between-subjects* if the treatments are applied only once to a given subject. In a *Within-subjects design*, the treatments are applied repeatedly to the subjects.

2.4 The *F*-test and Analysis of Variance

In the experiments we have discussed, we are comparing the means of the relevant groups. This analysis is carried out by using the variances of the four groups, specifically by looking at a ratio of averaged sums of squares. The statistic and distribution needed for tests on the means is known as the F variable, and the method used is called the Analysis of Variance, ANOVA for short. The test on the statistic is the F-test. In the next section, we will look at the exact meaning and definitions of these quantities.

The F distribution has two 'degrees of freedom' parameters. This happens because the sample F statistic is calculated from a ratio of variances, and so the distribution is the ratio of two chi-squared distributions. This means one degree of freedom arises for the denominator and the other for the numerator.

We will illustrate the calculations involved and the test that is done with examples in each case. We will start with a One-way Analysis of Variance, then a block design of the same problem, and finally a Two-way Analysis of Variance.

2.5 An Example of a One-Way ANOVA

A component manufacturer wishes to measure the efficiency of four different methods of production of components. A random sample of components was taken from a given days production and the components tested to destruction, with the time noted in hours and the method of manufacture for each component noted.

Carry out a test to see if the data suggest one method is better than the others.

The data is presented in the following table.

Method:	1	2	3	4
Lifetimes:	65	75	59	94
	87	69	78	89
	73	83	67	80
	79	81	62	88
	81	72	83	
	69	79	76	
		90		

In this experiment, the subjects are the components and the treatment is the method of manufacture. The symbol X will be the random variable of the lifetimes of the components, and we have measured an overall total of N samples.

Set k to be the number of groups of values. In this case $k = 4$.

Let n_i be the number of values on group i, with an overall total of N, so that $\Sigma\, n_i = n_1 + n_2 + \ldots\, n_k = N$.

Let X with the usual bar symbol denote the overall mean. Because we will be using a lot of symbols for means, we will now also use the symbol m for the overall mean, and m_j to be the mean for group j:

$$m = \overline{X} = \frac{\sum X_i}{N} \text{ and } m_j = \overline{X}_j = \frac{\sum_{\text{group } j} X_i}{n_j}.$$

The experiment is a between-subjects one-way layout, since the components are samples taken from outside populations, and there is only one factor involved.

2.5.1 The Sums of Squares

Some terms need to be defined first. We will look at the general terms, and then specify them for this case. Each of the terms we will define here will be a sum of squares of differences around a mean, in other words an expression resembling the top line of a variance calculation. To pursue this analogy, we will see that they are divided by their respective degrees of freedom, so they are effectively variances.

The *total sum of squares*, denoted *SS*, is

$$S_1 = \sum \left(X_i - \overline{X}\right)^2 = \sum (X_i - m)^2 .$$

That is, it is the sum of the difference of every value with the overall mean. This is essentially the overall variance S^2 of the population, but without the last step of dividing it by $N - 1$. It can be written as

$$S_1 = \sum X_i^{\ 2} - N\overline{X}^{\ 2} = \sum X_i^{\ 2} - Nm^2 .$$

In the case we are studying, case, the total is 1,779, so the grand mean is

$$m = 1{,}779/23 = 77.348.$$

Next, the overall sum of squares of the sample values is

$$\Sigma X_i^2 = 139{,}511.$$

The value of S_1 is

$$S_1 = \Sigma X_i^2 - Nm^2 = 139{,}511 - 137{,}601.8 = 1909.2.$$

This sum of squares, which loosely speaking measures the variance of all the data, is now split into two parts,

- The sum of squares arising between the groups and their treatment, called S_T, and
- The sum of squares within each group, called the sum of squares due to error, denoted S_E.

The ratio of the means of these sums of squares, in effect variances, will give the required test. 'Spilt into two parts' means

$$S_1 = S_T + S_E.$$

The sum of squares SS_T is given by the following expression:

$$S_T = \sum_j n_j \left(m_j - m\right)^2 .$$

This may sometimes be seen in the form

$$S_T = \sum_i \frac{\left(\sum_{\text{group}\,i} X_j\right)^2}{n_i} - N\left(\overline{X}\right)^2 .$$

Using the second version of the equations. The first sum, above the line, in this equation is the sum of the values within each group. Here are the calculations:

- For the first group, the value is $454^2/6 = 34{,}352.667$.
- For the second $549^2/7 = 43{,}057.285$.
- The third is $425^2/6 = 30{,}104.166$.
- The fourth is $351^2/4 = 30{,}800.25$.

Adding these up gives 138,314.37.

So $S_T = 138{,}314.37 - 137{,}601.8 = 712.57$.

The next term is the sum of squares due to error; in this context, the word ‘error’ should be taken as meaning the natural variation in the data. It is defined as is:

$$S_E = \sum_j \sum_{\text{each gp}} \left(X_i - m_j\right)^2 .$$

Again, it has an alternative form:

$$S_E = \sum_j \left(\sum_{\text{each gp}} X_i^{\,2} - n_j \overline{X}_j^{\,2} \right).$$

But recall that the sum of squares due to error is $S_E = S_1 - S_T$.

This means that

$$S_E = S_1 - S_T = 1909.2 - 712.57 = 1196.63.$$

2.5.2 The Mean Sums of Squares and the Test

These squares are now divided by the relevant degrees of freedom, to convert them into variances. They are:

- The degrees of freedom of the treatments is $k - 1$, where k is the number of groups. The variance in this case is therefore $MS_T = SS_T/(k - 1)$.
- The remaining degrees of freedom for the error comes from the total degrees of freedom $N - 1$, less the value $k - 1$. Therefore the degrees of freedom is $(N - 1) - (k - 1) = N - k$. Therefore the variance for the error is $MS_E = SS_E/(N - k)$.

With these variances, the value of F is then

$$F = MS_T/MS_E,$$

with $k - 1$ as the numerator degrees of freedom and $N - k$ as the denominator degrees of freedom. The sample value here is:

$$F = MS_T/MS_E = 3.77.$$

The results are summarized in an ANOVA table:

Source	*D.f.*	*Sum of squares*	*Mean SS*	*F*
Treatments	3	712.6	237.5	3.77
Error	19	1196.6	63.0	
Total	*22*	*1909.2*		

This test will now be recast as a formal Hypothesis test:

The null hypothesis is that the four treatments have the same means. This is measured by the F statistic, on the test value found for the treatments, in other words the ratio of the mean squares for the treatment and the mean squares for the error.

- The degrees of freedom is 3 and 19.
- The critical value for $\alpha = 0.05$ is $F_{0.05} = 3.13$.

Therefore the null hypothesis is rejected; the groups have different means.

This means there was a significant difference in the means of the four treatments, i.e. the four methods of production.

2.5.3 A One-way Block Design

The calculations just done are naturally simplified if the data is organised so that there are the same number of values in each group; in other words if it is a Block Design.

The *total sum of squares,* denoted S_1, is the same:

$$S_1 = \sum \left(X_i - \overline{X}\right)^2 = \sum (X_i - m)^2 .$$

It can be written as

$$S_1 = \sum X_i^{\ 2} - N\overline{X}^2 = \sum X_i^{\ 2} - Nm^2 .$$

This sum of squares is now split into two parts, as before

- The sum of squares arising between the groups and their treatment, called S_T, and
- The sum of squares within each group, called the sum of squares due to error, denoted S_E.

Let b be the number of blocks. This is the same as the number of subjects in each of the k treatment groups. So $N = bk$.

Now in this instance, the sum of squares S_T is given by the following expression:

$$S_T = b\sum_j (m_j - m)^2 = \frac{1}{b}\sum_i \left(\sum_{\text{group}\,i} X_j\right)^2 - N\left(\overline{X}\right)^2 .$$

The next term is the sum of squares due to error; this is:

$$S_E = \sum_{i,j} (X_i - m_j)^2 = \sum_i X_i^{\,2} - \frac{1}{b}\sum_i \left(\sum_{\text{group}\,i} X_j\right)^2 .$$

But recall that the sum of squares due to error is $S_E = S_1 - S_T$.

We then proceed as before, with one exception.

We can now measure the variation within each block, giving a measure of the natural variability of the data.

This can be done with an alternative sum of squares:

$$S_B = k\sum_j \left(m'_j - m\right)^2 = \frac{1}{k}\sum_j \left(\sum_{\text{block } j} X_i \right)^2 - N\left(\overline{X}\right)^2 .$$

Here is an example.

2.5.4 A Block Design One-way ANOVA

A component manufacturer wishes to measure the efficiency of four different methods of production of components in its manufacturing plants. An experiment to test this was laid out in a randomised block design, where the response being measured was the length of time taken to manufacture a certain number of working components. The day on which the measurements were taken is also recorded as Monday (M), Wednesday (W) or Friday (F).

The relevant sums of squares were calculated for an Analysis of Variance; this information is shown in the following table:

Source	*D.f.*	*Sum of squares*	*Mean SS*	*F*
Blocks	*2*	*2.542*	*1.271*	*3.894*
Treatments	3	6.609	2.203	6.75
Error	8	1.958	0.326	
Total	11	8.567		

Carry out an appropriate test to decide whether there is a difference in efficiency in the four methods of production.

The subjects in the randomised block design were the factories in which the manufacturing methods were applied. The blocks are the days on which production was observed, the treatments were the manufacturing methods.

- The number of degrees of freedom for the blocks is 3 – 1 = 2.
- The number of degrees of freedom for the treatment is 4 – 1 = 3.
- The number of degrees of freedom for the error is then the degrees of freedom for the overall sum of squares S_1, which is 11, less the 3 above, leaving 8.

The tests are carried out as follows:

Blocks:

Take the null hypothesis that no difference exists between the blocks, i.e. the days.

In this case, the test statistic F is given by $F = MS_B/MS_E = 3.894$. The number of degrees of freedom are 2 and 8, so the critical value for 0.05 is $F_{0.05} = 4.458$. Since the value found does not exceed this, the null hypothesis is not rejected and there is no difference.

It is now possible to test for the null hypothesis that there is no difference between treatments.

Treatments:

The number of degrees of freedom are 3 and 8, so the critical value for 0.05 is $F_{0.05}$ = 4.066. The sample value found is 6.75, which exceeds this, therefore the null hypothesis is rejected.

Looking at the mean values, calculated from the sums for each method already found, the first method is clearly higher than the other three.

2.6 A Two-way ANOVA

A component manufacturer wishes to measure the efficiency of four different methods of production of components in its three manufacturing plants.

An experiment to test this was laid out in a Factorial between-subjects Analysis of Variance, where the response being measured was the length of time taken to manufacture a certain number of working components, on 60 days.

This case is now qualitatively different; we have two factors (hence the use of the word ‘factorial’.

We will carry out an appropriate test to decide whether there is a difference in efficiency in the production methods, the plants and whether there is an interaction effect between the two variables. This case is now qualitatively different; we have two factors (hence the use of the word ‘factorial’).

The total sum of squares, mentioned already, is partitioned into sums of squares for the two separate treatments, and a mixed effect, that is, the interaction between the two factors. The parameters needed are:

- Let N be the total number of sample values.
- Let a be the number of levels of treatment A.
- Let b be the number of levels of treatment B.
- As before, let m be the grand mean.
- Let m_i be the mean across a given value i of the treatment A.
- Let m_j be the mean across a given value j of the treatment B.
- Let m_{ij} be the mean across a given value i of the treatment A and across a given value j of the treatment B.

The quantity *total sum of squares*, denoted S_1, is

$$S_1 = \sum \left(X_i - \overline{X} \right)^2 = \sum \left(X_i - m \right)^2 = \sum X_i^{\,2} - Nm^2 .$$

The sum of squares for treatment A, S_A, is given by the following expression:

$$SS_A = \frac{N}{a} \sum_i \left(m_i - m \right)^2 .$$

The sum of squares for treatment B, S_B, is given by the following expression:

$$S_B = \frac{N}{b}\sum_j \left(m_j - m\right)^2 .$$

The interaction effect term, is:

$$S_{AB} = \frac{N}{ab}\sum_{i,j} \left(m_{ij} - m_i - m_j + m\right)^2 .$$

The error is then

$$S_E = \sum_{i,j,k} \left(X_{i,j,k} - m_{i,j}\right)^2 .$$

The sums of squares of differences shown in these equations can all be set up by analogy with those we had from the one way analysis, except the interaction term. This includes the error term defined last. It is to ensure that the various sum of squares of differences terms add up to the overall sum of squares of differences that the interaction term is found, so that:

$$S_1 = S_A + S_B + S_{AB} + S_E$$

We will carry out an appropriate test to decide whether there is a difference in efficiency in the production methods, the plants and whether there is an interaction effect between the two variables.

The relevant sums of squares were calculated for an Analysis of Variance; this information is shown in the following table.

Source	*Sum of squares*	*d.o.f*	*Mean SS*	*F*
Method	2.542	3	0.847	1.853

Plant	6.609	2	3.305	7.232
Mixed effect	3.571	6	0.595	1.424
Error	21.958	48	0.457	
Total	*31.680*	*59*		

The treatments were the manufacturing methods and the plants.

- The number of degrees of freedom for the plants is 3 – 1 = 2.
- The number of degrees of freedom for the production methods is 4 – 1 = 3.
- The number of degrees of freedom for the mixed-effect element is (3 – 1)(4 – 1) = 6
- The degrees of freedom for the error is then the degrees of freedom for the overall sum of squares SS, which is 59, less the 2 + 3 + 6 above, leaving 60 – 12 = 48.

The table now has the mean sums of squares calculated by dividing the partitioned sums of squares by their degrees of freedom.

Now carry out an appropriate test to decide whether there is a difference in efficiency in the factors.

Take the null hypothesis that no difference exists between the production methods.

- In this case, the test statistic F is given by $F = MS_{(M)}/MS_{(E)} =$ 1.853.

- The number of degrees of freedom are 3 and 48, so the critical value for 0.05 is $F_{0.05} = 2.84$.
- Since the value found does not exceed this, the null hypothesis is accepted and there is no difference.

Take the null hypothesis that no difference exists between the plants.

- In this case, the test statistic F is given by $F = MS_{(P)}/MS_{(E)} = 7.232$.
- The number of degrees of freedom are 3 and 48, so the critical value for 0.05 is $F_{0.05} = 6.60$.
- Since the value found does exceed this, the null hypothesis is rejected and there is a difference.

Now test for the interaction effect;

- The sum of squares here was 3.751
- With 6 degrees of freedom, this is a mean sum of squares of 0.595.
- The ratio with the error term is 1.424.
- The number of degrees of freedom are 6 and 48, so the critical value for 0.05 is $F_{0.05} = 2.32$, therefore the null hypothesis is accepted, there are no mixed effects.

Thus the only positive result is a different in the plants.