Analysis of Variance

Damian Cox

March 24, 2019

1 Analysis of Variance

1.1 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.

1.2 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, the 125 units would be assigned to the 5 growing media at random, to eliminate other variables.

Arranging a test like this is a simple version of a designed experiment. The test is run 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. The example illustrates some definitions.

1.2.1 Definition: Units/Subjects

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

1.2.2 Definition: Factor

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

The example involved a factor, the growing medium makeup, which can be varied from one unit to another. There were five levels, that is, five particular mixes.

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

1.2.3 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.

1.3 Designs

The first step in putting together an experiment, once you have decided on subjects and factors, is to decide how you need to arrange your subjects for how they will receive 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.

1.3.1 Definition: Completely Randomised Design

A completely randomised design to compare k treatments is one where a group of N relatively homogeneous units are randomly divided into k subgroups, of size

$$n_1, n_2, n_3, \ldots n_k,$$

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.

1.3.2 Definition: One-Way Layout

Say a set of experimental subjects are to be given a factor 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 factor.

By extension, in a two-way layout, there would be two factor, 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}	

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.

The t-test 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.

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.

1.3.3 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.

Here is a Randomised Block Design for a one way layout, with 4 treatments and 20 subjects. Crucially, the number of treatments divides evenly into the number of subjects available.

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}

1.3.4 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.

1.3.5 Within/Between Subjects

In the example just quoted, applying each treatment to the same number of subjects does not, however, 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.

This is summarised in a definition:

1.3.6 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.

1.4 The F-test and Analysis of Variance

The analysis of the experimental designs discussed here is carried out by using the variances of the four groups to compare their means. This is done by looking at a ratio of averaged sums of squares.

The distribution needed for tests on the is known as the F-test, and the method used is called the Analysis of Variance, ANOVA for short.

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 F distribution is the ratio of variances and so the sample value is a ratio of two χ-square distributions.

This then 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 Two-way Analysis of Variance.

1.5 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.

Investigate whether the data suggests 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 denote 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

$$\sum n_i = n_1 + n_2 + \ldots n_k = N.$$

Let $\bar{X}$ 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:

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

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.

1.5.1 The Sums of Squares

Some summations necessary for the test statistic need to be defined first. We will look at the general terms and then specify them for this case.

The total sum of squares, denoted S_1, is

$$\sum_i (X_i - \bar{X})^2.$$

This is essentially the 'top line' of a variance calculation. It can be written as

$$S_1 = \sum_i X_i^2 - N\bar{X}^2 = \sum_i X_i^2 - Nm^2.$$

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

$$\sum_i X_i = 1,779 \Rightarrow m = 77.348.$$

Next, the total sum of squares is

$$\sum X_i^2 = 139,511.$$

The value of S_1 is

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

This sum of squares is now 'partitioned':

- 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 word partitioned means

$$S_1 = S_T + S_E.$$

The ratio of the means of these sums of squares, in effect variances, will give the required test.

The sum of squares S_T is given by the following summation over the groups:

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

This can be shown to be equal to

$$S_T = \sum_j \frac{\left(\sum_{\text{group } j} X_i\right)^2}{n_j} - Nm^2.$$

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 $454^2/6 = 34,352.667$.
- 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; this is:

$$\sum_j \sum_i (X_i - m_j)^2.$$

But recall that the sum of squares due to error is

$$S_1 = S_T + S_E.$$

This means that

$$S_E = S_1 - S_T = 1909.2\text{–}712.57 = 1196.63.$$

1.5.2 The Mean Sums of Squares and the Test

The means of these sums of squares are now calculated, divided by the relevant degrees of freedom, to convert them into variances. Bear in mind that the total degrees of freedom will be $N - 1$.

- The degrees of freedom for S_T is the number of groups less 1, so define
$$M_T = \frac{S_T}{k-1},$$
where k is the number of groups.

- For the error, the remaining degrees of freedom is $(N-1) - (k-1) = N - k$. Then
$$M_E = \frac{S_E}{N-k},$$

- With these variances, the sample value of statistic F is then
$$F = \frac{M_T}{M_T},$$
with $k-1$ as the numerator degrees of freedom and $N-k$ as the denominator degrees of freedom.

In this instance, the test statistic is $F = M_T/M_E = 3.77$. The two degrees of freedom values are 3 and 19.

Let us summarise this work in the form of a NHBST.

The null hypothesis is that the four groups, organised by treatment, have the same means.

The results of the calculations done above are summarised in an ANOVA table:

Source of variance	*d.o.f.*	*Sum Sq.*	*Mean Sum Sq.*	*F*
Treatments	3	712.6	237.5	3.77
Error	19	1196.6	63.0	
Total	22	1909.2		

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, that is, the four methods of production.

1.6 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 first set up the concept behind the test.

1.6.1 Partition for Two-way ANOVA

In a two way Analysis of Variance, the total sum of squares, mentioned already, is partitioned into sums of squares for the two factors.

- Let N be the total number of sample values.
- Let A and B be the two factors, with numbers of levels a and b respectively.
- As before, let m be the grand mean.
- Let m_i be the mean across a given value i of the treatment A. Thus it is the mean of the groups defined by factor A, ignoring B.
- Let m_j be the mean across a given value j of the factor B. Thus it is the mean of the groups defined by the levels of factor B, ignoring A.
- Let $m_{i,j}$ be the mean across a given value i of the factor A and a given value j of the factor B. Thus it is the mean of the groups defined by each combination of levels of the factors A and B, in other words, each treatment.

The total sum of squares S_1 remains as before:

$$S_1 = \sum_i (X_i - m)^2.$$

The sum of squares for treatment A is

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

The sum of squares for treatment B is

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

The error term is

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

As before, this is the sum of squares of the difference of each element in the data with the mean of the group it is in, that group being defined by the two levels of the factors acting on $X_{i,j,k}$.

It is not the case that these sums of squares add up to S_1. It can be shown that

$$S_1 - S_A - S_B - S_E = \frac{N}{ab} \sum_i (m_{i,j} - m_i - m_j + m)^2.$$

This is known as the variance due to an interaction effect between the two factors A and B. It is labelled S_{AB}. Thus

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

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

1.6.2 Sample Conclusions of a Two-way ANOVA

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 a combined 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 of variance	*Sum Sq.*	*d.o.f.*	*Mean Sum Sq.*	*F*
Method	2.542	3	0.847	1.853
Plant	6.609	2	3.305	7.232
Interaction effect	3.571	6	0.595	1.424
Error	21.958	48	0.457	
Total	31.680	59		

Here is the summary of the terms used.

- The factors 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 interaction-effect 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 s_1, which is 59, less

the 2 + 3 + 6 above, leaving 48.

The table now has the mean sums of squares calculated by dividing the partitioned sums of squares by the corresponding 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 $F = M_M/M_E = 1.853$.
- 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 not rejected 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 $F = M_P/M_E = 7.232$.
- The number of degrees of freedom are 2 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 no difference.

Now test for a interaction effect;

- In this case, the test statistic F is given by $F = M_I/M_E = 1.424$.
- The number of degrees of freedom are 6 and 48, so the critical value for 0.05 is $F_{0.05} = 2.32$.
- Since the value found does not exceed this, the null hypothesis is not rejected and there is no interaction effect.

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