One-way between-subjects Analysis of Variance

By the end of this lecture you will be able to perform the calculations by hand. SPSS will in fact perform the calculations for you in practice, but you will be examined on all the concepts explained in the lecture. In the exam at the end of the year you will be expected to be able to provide explanations of the sort given in this lecture. Please re-read these notes before the session where we use SPSS to analyze your data so that you understand the SPSS output.

The problem.

This module we will be looking at the simplest type of Analysis of Variance (ANOVA), namely the one-way between-subjects ANOVA.

"One-way" means one independent variable. ("Two-way" is two independent variables, and we consider that next module.) Remember, the independent variable is the variable that the experimenter can set the values of – he knows what value the subject has even before the subject walks in the room. The dependent variable is the variable the experimenter has to measure to know the value for a subject, and the point of doing the experiment is to see what values you get.

"Between-subjects" means each subject contributes to only one condition. "Within-subjects" means each subject contributes to each condition.

Consider the following experiment. We want to know whether ginseng affects reaction time. So we have three conditions, people are either given 0mg, 100mg, or 200mg of ginseng, and their reaction time is measured (in milliseconds) on a task. We have ten people in each condition and each person provides a single reaction time score. The independent variable is amount of ginseng, that is set by the experimenter. The dependent variable is reaction time. The design is between subjects because each subject has just one dose of ginseng.

If we ran the experiment, we could calculate a mean and standard deviation for each condition. Let us say the means for the different conditions were different. Would this answer our question, whether ginseng affects reaction time?

What we are really interested in is not the overall reaction time scores for the particular subjects we ran in each condition. We are interested in knowing whether ginseng affects reaction time for people in general. The set of all people we are in principle interested in is called the population. We take a sample of ten people from this population. We are not really interested in this sample per se, only in so far as it tells us something about the population. We are not interested in just whether our sample means differ; we want to know if this tells us whether the population means differ. Drawing inferences from samples about populations relies on inferential statistics.

Last year you learnt an inferential statistical technique that could be applied to the ginseng data: Kruskal-Wallis. One-way between-subjects ANOVA is the parametric equivalent of Kruskal-Wallis. Parametric statistics make more assumptions about the population distributions than non-parametric statistics. We will be returning to the particular assumptions of ANOVA later. If the assumptions seem to be satisfied, it is good to use parametric tests, because the gain you get from having to make the assumptions is a more sensitive test.

Logic of ANOVA: Central limit theorem

One-way between subjects ANOVA will answer the question: Can we conclude that that the population mean for any one of our conditions is different from the population mean for any other?

But if analysis of variance analyzes variances, how can it tell us something about means?

A variance is a measure of spread of scores about the mean, m. If we want to measure spread, a good place to start would be the difference between each score x and the mean, (x –

m), and in fact for variance we take the square of this difference, $(x-m)^2$. The squared difference is summed over all scores, $\Sigma (x-m)^2$ and then we take a sort of average by dividing by (n-1), where n is the number of scores.
Variance = $\Sigma (x-m)^2/(n-1)$. If we divided by n that would be finding the variance of our scores per se; if we want to use our scores as a sample to estimate population variance, then we must divided by (n-1). The top part of the variance formula ($\Sigma (x-m)^2$) is called a Sum of Squares (SS); the bottom part (n-1) is called the degree of freedom (df). Variance = SS/df. SS/df is also called Mean Square (MS) in ANOVA terminology. The standard deviation is another measure of spread; it is equal to the square root of the variance.

Imagine we have a population with a certain mean and variance, σ^2. (By convention, we use Greek symbols for populations; σ (sigma) is a population standard deviation, so σ^2 is the population variance.) We take a random sample of 10 people and find the mean of the sample. It would be unlikely for the sample mean to be exactly equal to the population mean; it will probably be a little bit more or a little bit less than the population mean. Now I randomly take another sample. I get a different sample mean. I repeat this process of drawing samples and finding the means of my samples. I end up with a large set of numbers, each number being a sample mean. I can find the distribution of these sample means – they will have a mean and a variance themselves. The mean of all the sample means is expected to be the population mean. But what about the variance of my distribution of sample means? (call that Vs, for variance of sample means.) How does Vs relate to the variance of the population (σ^2) from which I sampled?

First question. As the variance of my population increases, what happens to the variance of my sample means, Vs? Does Vs increase, decrease, or stay the same?

Imagine the population is the set of everyone in this room, and the scores are people's heights. If everyone had the same height – 5'8" – then each and every sample would have the same mean. Everyone in the sample would have a height of 5'8", and so the sample mean would be 5'8". Because this is true for every sample, there is no variance in the sample means. No population variance, no variance in sample means. Now imagine the room has people with different heights in – there are some dwarves and some giants. There is a lot of variance in the population scores. Sometimes, just by chance, I will sample just dwarves and get a very low sample mean. Sometimes, just by chance, I will sample just giants, and I will get a very high sample mean. My sample means will vary a lot between themselves. First conclusion: As the population variance (σ^2) increase, the variance (Vs) of my sample means also increases.

Second question. If the size, n, of my sample increases, what happens to the variance of my sample means, Vs? Does Vs increase, decrease, or stay the same?

Once again, a good way to think about the form of relationships is to consider the extreme cases. If my sample size n is 1, the samples are just individual people. The variance of my sample means would be expected to be just the variance of my population. Now if my sample size was the size of the population my sample mean would be the population mean. Every sample of that size is just the population, so every sample would give me the same mean, the population mean. There would be no variance in the sample means. Second conclusion: As my sample size increases, the variance of my sample means decreases. Note that this makes good sense. The bigger my sample is, the better I can estimate the population mean. So with a big sample, I have a good estimate. Although it is unlikely to be exactly the population mean, it will only be a little bit more or a little bit less than the population mean. With a small sample, my sample mean could be a lot less than the population mean or a lot more.

Now we put our two conclusions together. We can express them in an equation:
$Vs \approx \sigma^2/n$ (central limit theorem)

$\approx$ means approximately equal to. (They would be exactly equal if we took an infinite number of samples.) You won't know why it is exactly this equation that is true. But I want you to notice that it satisfies our two conclusions: As σ^2 increases, Vs increases (conclusion

one); as n increases, Vs decreases (conclusion two). And it happens to be about the simplest way of satisfying the two conclusions. If you understand what that equation means, and it makes intuitive sense to you, then the hard work is done. You are now ready to understand how analysis of variance works.

Logic of ANOVA: answering our research problem

Now let's return to our problem: Does ginseng affect reaction time? Imagine what would happen if ginseng does not affect reaction time (this hypothesis is called the null hypothesis, Ho). That is, our population mean RT after taking 0mg, is the same as after taking 100mg, is the same as after taking 200mg of ginseng. We will also assume that the variances for each of these population distributions is the same. So sampling from these three populations, is just like sampling from one population with a certain mean (μ, mu) and variance (σ^2). From the experiment, we end up with a sample mean m1 for the 0 mg condition, mean m2 from the 100 mg condition, and m3 from the 200 mg condition. We also have sample variances V1, V2 and V3 for the three conditions.

We have two independent ways of estimating the population variance, σ^2. The first way is by using our sample variances. V1 is an estimate of σ^2. That's just what a variance calculated from a sample with a denominator of (n-1) is: An estimate of the variance of the population from which it was drawn. V2 is also an estimate of σ^2, and so is V3. Assuming each sample has an equal n, then an even better estimate is an average of these three estimates, namely (V1 + V2 + V3)/3. This average variance is called Mean Square Error, or MSe. It does not mean you have made any errors. It is just the variance between subjects you can't account for in terms of what else you have measured in the experiment. There is another way of calculating MSe. Remember that MS = SS/df. So V1 = SS1/df1 and so on for V2 and V3. Sum of Squares error (SSe) = SS1 + SS2 + SS3. Degrees of freedom error (dfe) = df1 + df2 + df3. MSe = SSe/dfe. MSe is one estimate of population variance, σ^2.

There is another estimate of population variance. Assuming the null hypothesis, each sample is just a sample drawn from the same population. Our three sample means are just three numbers and we can find their variance, just like we can find the variance of any set of numbers. First we find the mean of all our means, call it M. Then we can find the squared difference of each mean m from M and sum: $\Sigma (m-M)^2$. Finally we divide by the number of means (call it k) minus 1: $\Sigma (m-M)^2/(k-1)$. This is the variance, Vs, of our sample means. Now we know the relationship between Vs and σ^2 from the last section: $Vs \approx \sigma^2/n$
Re-arranging: $\sigma^2 \approx Vs*n$

So another estimate of the population variance is $n*\Sigma (m-M)^2/(k-1)$, or n times the variance between the means. This estimate of the population variance is called Mean Square Treatment, or MSt. The top part (numerator) $n*\Sigma (m-M)^2$ is called Sums of Squares treatment (SSt). The bottom part (denominator) (k-1) is called degrees of freedom treatment, dft. MSt = SSt/dft.

We have two estimates of population variance, MSe and MSt. What happens to them if the null hypothesis is false? Let us assume the population means pull apart but the population variance for each condition remains the same. What happens to MSe?

MSe consists of each sample variance and each sample variance is an estimate of its population variance. By assumption, the population variances have remained the same. So MSe is NOT affected by the population means becoming different – no matter how different the means, the population variances remain the same.

What happens to MSt if the null hypothesis is false? MSt is the variability between the sample means. The further apart the population means, the further apart one would expect the sample means to be. That is, if the null hypothesis is false, MSt would be expected to bigger than if the null hypothesis were true.

At last we have a means of deciding if the null hypothesis is true or not. Divide MSt by MSe. This ratio is called F. F = MSt/MSe. If the null hypothesis is true, F should be about 1. MSt and MSe both estimate the population variance so they should be about equal. Of

course they are unlikely to be exactly equal. Because of the vagaries of random sampling, F will probably be a little bit more than 1 or a little bit less than 1, but you expect it to be about 1. If the null hypothesis is false MSt is expected to be greater than MSe so F is expected to be more than 1.

MSt is the between group variability. MSe is the within group variability. F compares the between group to the within group variability. If the null hypothesis is true, the variabilities are calculated so that they should be about the same. If the null hypothesis is false, the between group variability should be greater than the within group variability.

There is another way of looking at this. Statistics is about trying to see a signal through noise. People, and biological systems generally, are always noisy – whatever their tendencies, there is always variability around those tendencies. The population mean differences are the signal we are trying to detect. MSt is our estimate of the signal. The noise through which we are trying to see the signal is the within group variability – this is noise because it is by virtue of there being within-group variability that our sample means will differ even when there is no difference between population means. The noise creates apparent signal even when there is not one really. Our estimate of the noise is MSe. We are only willing so say that there is a signal there is our estimate of the signal (MSt) is large relative to the noise (MSe).

Now we follow the logic of inferential statistics you learnt about last year. Assuming Ho, F can be any value but it is unlikely to be very high. We can calculate (or rather the computer will do so for you) the value of F such that a value that extreme or more extreme would be obtained only 5% of the time if Ho were true. This is the critical value of F at the 5% level. (The critical value depends on both dfe and dft.) If we obtain an F that high or higher we will say assuming the null hypothesis, it is very unlikely we would have obtained results as extreme as this or more extreme. Since our data seems implausible given our assumptions, there must be something wrong with our assumptions. So we will reject the null hypothesis.

In fact, SPSS will calculate what it calls p – the probability of obtaining an F as extreme or more extreme than you obtained (assuming Ho). If $p < .05$ you say your results are significant at the 5% level and you entitled to reject the null hypothesis; you can accept that ginseng affects reaction time.

There are two assumptions made in calculating the p values. First, remember we assumed that the variances for each population were the same. Second, the calculations also assume that the populations are roughly normally distributed.

<u>Postscript: Note on the meaning of p</u>

p is a probability of obtaining a certain type of data, D, given the Ho, p(D/Ho). (The "/" sign is read as "given".) This is NOT the probability of Ho given the data, p(Ho/D). You cannot conclude from p alone what the probability of Ho is, or how likely it is you have made a wrong decision in this case. Ho for any particular experiment is either true or false – the objective probability of Ho, if it is anything at all, is 0 (false) or 1 (true).

The probability that you will die soon given you have a brain tumor is very high, close to 1. But if you take all the people who are about to die soon, the probability of having a brain tumor is very low, almost 0. P(brain tumor/die soon) does NOT equal p(die soon/brain tumor). Just so p(D/Ho) does not equal p(Ho/D).

If p is very large – say $p = 0.85$- that does NOT mean the null hypothesis is very likely to be true. If Ho in fact is false – ginseng does affect RT – then the probability of Ho is 0, whatever the p value for the experiment. The p value you actually obtain depends not only on whether Ho is false but also (amongst other things) on how many subjects you run, n. There may really be a difference between population means, but if you run too few subjects, then the experiment may be insignificant. $p > .05$ just means you are not entitled to reject Ho. Whether you decide to accept Ho or not, depends on how sensitive you thought your experiment was.

Do not worry if this issue is not so clear now: We will return to it later in the course. In the meantime, feel free to ask questions about it.

Zoltan Dienes