One-Way Independent ANOVA by Hand

The Basic Idea

The *t*-test is limited to situations in which there are only two levels of the independent variable (i.e. two experimental groups). It is common to run experiments in which there are three, four or even five levels of the independent variable and in these cases the *t*-test is inappropriate. Instead, a technique called analysis of variance (or ANOVA to its friends) is used. Also, *t*-tests can be used when only one independent variable has been measured. However, ANOVA has the advantage that it can be used to analyze situations in which there are several independent variables. In these situations, ANOVA tells us how these independent variables interact with each other and what effects these interactions have on the dependent variable.

When we perform a *t*-test, we test the hypothesis that the two samples have the same mean. Similarly, ANOVA tells us whether three or more means are the same, so, it tests the hypothesis that all group means are equal. An ANOVA produces an *F-statistic* or *F-ratio*, which is similar to the *t*-statistic in that it compares the amount of systematic variance in the data to the amount of unsystematic variance. However, ANOVA is an *omnibus* test, which means that it tests for an overall experimental effect: so, there are things that an ANOVA cannot tell us. Although ANOVA tells us whether the experimental manipulation was generally successful, it does not provide specific information about which groups were affected. Assuming an experiment was conducted with three different groups, the *F*-ratio simply tells us that the means of these three samples are not equal (i.e. that $\bar{X}_1 = \bar{X}_2 = \bar{X}_3$ is *not* true). However, there are a number of ways in which the means can differ. The first possibility is that all three sample means are significantly different ($\bar{X}_1 \neq \bar{X}_2 \neq \bar{X}_3$). A second possibility is that the means of group 1 and 2 are the same but group 3 has a significantly different mean from both of the other groups ($\bar{X}_1 = \bar{X}_2 \neq \bar{X}_3$). Another possibility is that groups 2 and 3 have similar means but group 1 has a significantly different mean ($\bar{X}_1 \neq \bar{X}_2 = \bar{X}_3$). Finally, groups 1 and 3 could have similar means but group 2 has a significantly different mean from both ($\bar{X}_1 = \bar{X}_3 \neq \bar{X}_2$). So, the *F*-ratio tells us only that the experimental manipulation has had some effect, but it doesn't tell us specifically what the effect was.

Calculating ANOVA by Hand

Example: A few years ago there was a lot of controversy about whether the drug Viagra (a sexual stimulant) was a useful treatment for impotence. Suppose we tested the belief that Viagra is useful as a treatment for sexual dysfunction by taking three groups of people and administering one group with a placebo (such as a sugar pill), one group with a low dose of Viagra and one with a high dose. The dependent variable was an objective measure of libido (I will tell you only that it was measured over the course of a week—the rest I shall leave to your own imagination). The data can be found in Table 1.

Table 1: Data in **Viagra.sav**

	Placebo	Low Dose	High Dose
	3	5	7
	2	2	4
	1	4	5
	1	2	3
	4	3	6
$\bar{X}$	2.20	3.20	5.00
s	1.30	1.30	1.58
s^2	1.70	1.70	2.50
Grand Mean = 3.467 Grand SD = 1.767 Grand Variance = 3.124			

Like the *t*-test, ANOVA is a way of comparing the ratio of systematic variance to unsystematic variance in an experimental study. The ratio of these variances is known as the *F*-ratio. We've come across the *F*-ratio a few weeks ago when we saw that it was used to assess how well a regression model predicts an outcome compared to the error within that model. The *F*-ratio we're learning about now is exactly the same as in regression: it compares the ratio of variance explained by the model to the error in the model, only now the model fitted to the data relates to systematic experimental manipulations across groups rather than naturally occurring variables. In fact, ANOVA can even be represented by a multiple regression equation in which the number of predictors is one less than the number of categories of the independent variable (see Field, 2009 for details).

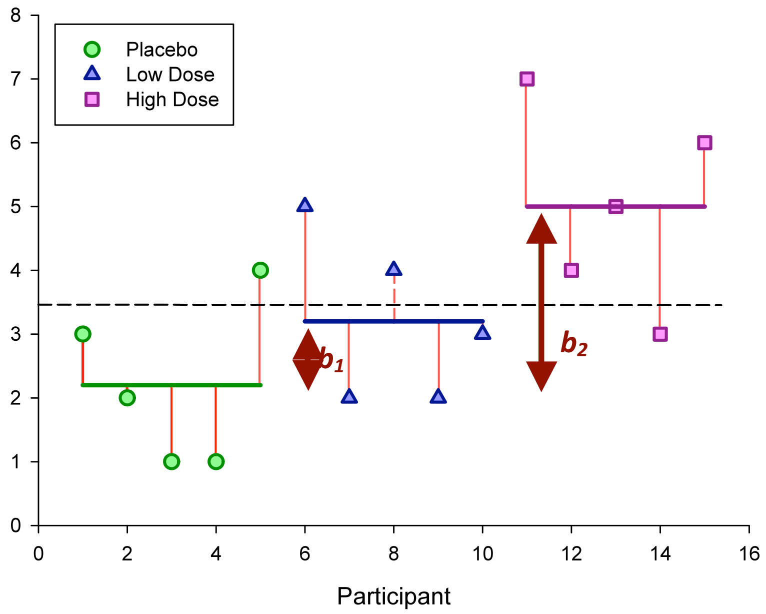

Figure 1: The Viagra data in graphical form. The coloured horizontal lines represent the mean libido of each group. The shapes represent the libido of individual participants (different shapes indicate different experimental groups). The dashed horizontal line is the average libido of all participants

Figure 1 shows the Viagra data in graphical form and includes the group means and the grand mean (i.e. the mean of all participants when you ignore the group to which they belong) and the difference between each case and the group mean. In this example, there were three groups; therefore, we want to test the hypothesis that the means of three groups are different (so, the null hypothesis is that the group means are the same). If the group means were all the same, then we would not expect the placebo group to differ from the low dose group or the high dose group, and we would not expect the low dose group to differ from the high dose group. Therefore, on the diagram, the three dashed lines would be in the same vertical position (the exact position would be the grand mean). We can see from the diagram that the group means are actually different because the dashed lines (the group means) are in different vertical positions.

The logic of ANOVA follows from what we understand about regression:

- The simplest model we can fit to a set of data is the grand mean (the mean of the outcome variable). This basic model represents 'no effect' or 'no relationship between the predictor variable and the outcome'.
- We can fit a different model to the data collected that represents our hypotheses. If this model fits the data well then it must be better than using the grand mean. Sometimes we fit a linear model (the line of best fit) but in experimental research we often fit a model based on the means of different conditions.
- Once a model is fitted to the data, we can calculate the total variability (SS_T), the error in the model (SS_R) and the improvement due to the model (SS_M).
- The model that we fit is based on the group means from different experimental conditions (because if we know someone was in group 1, then we can predict that the value of the outcome for that person will be the mean for group 1). If the group means are very different, then the model will be better at accurately predicting a person's score on the outcome variable (because the groups are well-discriminated from each other).
- We can test how much better the model fits the data than the grand mean by comparing the variance explained by the model (i.e. the improvement due to using the model to predict outcomes instead of using the grand mean) against the error in the model. This involves comparing SS_M against SS_R.
- The result is the *F*-ratio.

Step 1: Calculate the Total Sum of Squares (SS_T)

As in regression, we find the total amount of variation within our data by calculating the difference between each observed data point and the grand mean. We then square these differences and add them together to give us the total sum of squares (SS_T).

$$\mathrm{SS_T} = \sum \left(x_i - \bar{x}_{\mathrm{grand}}\right)^2 \qquad (1)$$

The variance and the sums of squares are related such that variance, $s^2 = SS/(N-1)$, where N is the number of observations. Therefore, we can calculate the total sums of squares from the variance of all observations (the *grand variance*) by rearranging the relationship ($SS = s^2(N-1)$). The grand variance is the variation between all scores, regardless of the experimental condition from which the scores come; it is the squared distance between each person's score and the grand mean (which if you think back to lecture 1 is simply the sum of squares associated with the overall variance of all scores). The grand variance for the Viagra data is given in Table 1, and if we count the number of observations we find that there were 15 in all. Therefore, SS_T is calculated as follows:

$$\begin{aligned} SS_T &= s^2_{grand}(N-1) \\ &= 3.124(15-1) \\ &= 3.124 \times 14 \\ &= 43.74 \end{aligned}$$

We now need to have a quick sidetrack to remind ourselves what degrees of freedom are; remember, we encountered them in week 1. In statistical terms the degrees of freedom relates to the number of observations that are free to vary. If we take a sample of four observations from a population, then these four scores can be any value. However, if we then use this sample of four observations to calculate the standard deviation of the population, we have to use the mean of the sample as an estimate of the population's mean. Thus we hold one parameter constant. Say the mean of the sample was 10, then we assume that the population mean is 10 also. With this parameter fixed, can all four scores from our sample vary? The answer is no because to keep the mean constant only three values are free to vary. For example, if the values in the sample were 8, 9, 11, 12 (mean = 10) and we changed three of these values to 7, 15 and 8, then the final value *must* be 10 to keep the mean constant. Therefore, if we hold one parameter constant then the degrees of freedom must be one less than the sample size. This fact explains why when we use a sample to estimate the standard deviation of a population (as we did in lecture 1), we have to divide the sums of squares by $N - 1$ rather than N alone. So, **for sums of squares the degrees are freedom are one less than the number of things you have used to calculate them**. For SS_T, we used the entire sample to calculate the sums of squares (i.e. 15 scores) and so the total degrees of freedom (df_T) are one less than the total number of observations ($N - 1$). For the Viagra data, this value is 14:

$$df_T = N - 1 = 14$$

Step 2: Calculate the Model Sum of Squares (SS_M)

So far, we know that the total amount of variation within the data is 43.74 units. We now need to know how much of this variation the model can explain. In the ANOVA scenario, the model is based upon the group means from our experiment and so the model sums of squares tell us how much of the total variation can be explained by the fact that different data points come from different groups. In short, *the model represents the effect of the experimental manipulation*.

When we learnt about regression (lecture 2) we saw that the model sum of squares is calculated by taking the difference between the values predicted by the model, and the grand mean. In ANOVA, the values predicted by the model are the group means, therefore, for each subject the value predicted by the model is the mean for the group to which the subject belongs. As such, the improvement due to the model will be the difference between the grand mean and the mean of each group for every subject. These differences are squared and summed. In the Viagra example, the predicted value for the five people in the placebo group will be 2.2, for the five people in the low dose condition it will be 3.2, and the five people in the high dose condition it will be 5. The model sum of squares requires us to calculate the differences between each subject's predicted value and the grand mean. These differences are then squared and added together (for reasons that should be clear in your mind by now). We know that the predicted value for people in a particular group is the mean of that group. Therefore, the easiest way to calculate SS_M is to:

1. Calculate the difference between the mean of each group and the grand mean.
2. Square each of these differences.
3. Multiply each result by the number of people within that group (ni).
4. Add the values for each group together.

$$SS_M = \sum n_i \left(\bar{x}_i - \bar{x}_{grand}\right)^2 \qquad (2)$$

Using the means from the Viagra data, we can calculate SS_M as follows:

$$\begin{aligned} SS_M &= 5(2.200 - 3.467)^2 + 5(3.200 - 3.467)^2 + 5(5.000 - 3.467)^2 \\ &= 5(-1.267)^2 + 5(-0.267)^2 + 5(1.533)^2 \\ &= 8.025 + 0.355 + 11.755 \\ &= 20.135 \end{aligned}$$

For SS_M, the degrees of freedom (df_M) is again one less than the number of things used to calculate the sum of squares. For the model sums of squares we calculated the sum of squared errors between the **three** means and the grand mean. Hence, we used three things (the group means) to calculate these sums of squares. So, the degrees of freedom will be 2 (because the calculation of the sums of squares was based on the group means—two of which will be free to vary in the population if the third is held constant). So, the model degrees of freedom is always the number of *groups* (*k*) minus 1:

$$df_M = k - 1 = 2$$

Step 3: Calculate the Residual Sum of Squares (SS_R)

We now know that there are 43.74 units of variation to be explained in our data, and that our model can explain 20.14 of these units (nearly half). The final sum of squares is the residual sum of squares (SS_R), which tells us how much of the variation cannot be explained by the model. This value is the amount of variation caused by extraneous factors such as individual differences in weight, testosterone or whatever. Knowing SS_T and SS_M already, the simplest way to calculate SS_R is to subtract SS_M from SS_T ($SS_R = SS_T - SS_M$); however, doing this provides little insight into what is being calculated.

When we looked at regression we saw that the residual sum of squares is the difference between what the model predicts and what was actually observed. We already know that for a given person, the model predicts the mean of the group to which that person belongs. Therefore, SS_R is calculated by looking at the difference between the score obtained by a person and the mean of the group to which the person belongs. As such it represents individual differences between people in an experimental group. The distances between each data point and the group mean are squared and then added together to give the residual sum of squares, SS_R:

$$SS_R = \sum (x_i - \bar{x}_i)^2 \quad (3)$$
or
$$SS_R = s^2_{group1}(n_1 - 1) + s^2_{group2}(n_2 - 1) + s^2_{group3}(n_3 - 1)$$

Now, the sum of squares for each group represents the sum of squared differences between each subject's score in that group and the group mean. Therefore, we can express SS_R as $SS_R = SS_{group1} + SS_{group2} + SS_{group3}$ … and so on. Given that we know the relationship between the variance and the sums of squares, we can use the variances for each group of the Viagra data to create an equation like we did for the total sum of squares. As such, SS_R can be expressed as:

$$\begin{aligned} SS_R &= s^2_{group1}(n_1 - 1) + s^2_{group2}(n_2 - 1) + s^2_{group3}(n_3 - 1) \\ &= (1.70)(5-1) + (1.70)(5-1) + (2.50)(5-1) \\ &= (1.70 \times 4) + (1.70 \times 4) + (2.50 \times 4) \\ &= 6.8 + 6.8 + 10 \\ &= 23.60 \end{aligned}$$

SS_R was based on adding together the individual sums of squares for each group. As such we need to calculate the degrees of freedom for each of these sums of squares. For each group, the sums of squares was based on 5 observations, so the degrees of freedom in each group will be one less than this value (i.e. 4). The degrees of freedom for SS_R (df_R) is obtained by simply adding the degrees of freedom for each group:

$$\begin{aligned} df_R &= df_{group1} + df_{group2} + df_{group3} \\ &= (n_1 - 1) + (n_2 - 1) + (n_3 - 1) \\ &= 4 + 4 + 4 \\ &= 12 \end{aligned}$$

Double Check Your Calculations: At this stage it is useful to double check your calculations. First, we know that the total sums of squares are made up of the residual and model sums of squares. As such, if we add SS_M and SS_R we should get a value the same as SS_T:

$$\begin{aligned} SS_T &= SS_M + SS_R \\ 43.74 &= 20.14 + 23.60 \\ 43.74 &= 43.74 \end{aligned}$$

Likewise, the degrees of freedom should add up in the same way:

$$\begin{aligned} df_T &= df_M + df_R \\ 14 &= 2 + 12 \\ 14 &= 14 \end{aligned}$$

If these values do not add up then you know that you've made an error in your calculations. Therefore, this is a very useful way to check that you've done everything correctly (which can be handy in the exam!).

Step 4: Calculate the Mean Squares

SS_M tells us how much variation the model (e.g. the experimental manipulation) explains and SS_R tells us how much variation is due to extraneous factors. However, because both of these values are summed values they are influenced by the number of scores that were summed (for example, SS_M used the sum of only 3 different values (the group means) compared to SS_R and SS_T, which used the sum of 14 different values). To eliminate this bias we can calculate the average sum of squares (known as the *mean squares*, MS), which is simply the sum of squares divided by the degrees of freedom. The reason why we divide by the degrees of freedom rather than the number of parameters used to calculate the SS is because we are trying to extrapolate to a population and so some parameters within that populations will be held constant. So, for the Viagra data we find the following mean squares:

$$MS_M = \frac{SS_M}{df_M} = \frac{20.135}{2} = 10.067$$

$$MS_R = \frac{SS_R}{df_R} = \frac{23.60}{12} = 1.967$$

MS_M represents the average amount of variation explained by the model (e.g. the systematic variation), whereas MS_R is a gauge of the average amount of variation explained by extraneous variables (the unsystematic variation).

Step 5: Calculate The F-Ratio

The *F*-ratio is a measure of the ratio of the variation explained by the model and the variation explained by unsystematic factors. It can be calculated by dividing the model mean squares by the residual mean squares.

$$F = \frac{MS_M}{MS_R} \tag{4}$$

As with the independent *t*-test, the *F*-ratio is, therefore, a measure of the ratio of systematic variation to unsystematic variation. As such, it is the ratio of the experimental effect to the individual differences in performance. An interesting point about the *F*-ratio is that because it is the ratio of systematic variance to unsystematic variance, if its value is less than 1 then it must, by definition, represent a non-significant effect. The reason why this statement is true is because if the *F*-ratio is less than 1 it means that MS_R is greater than MS_M, which in real terms means that there is more unsystematic than systematic variance. You can think of this in terms of the effect of natural differences in ability being greater than differences bought about by the experiment. In this scenario, we can, therefore, be sure that our experimental manipulation has been unsuccessful (because it has bought about less change than if we left our participants alone!). For the Viagra data, the *F*-ratio is:

$$F = \frac{MS_M}{MS_R} = \frac{10.067}{1.967} = 5.12$$

This value is greater than 1, which indicates that the experimental manipulation had some effect above and beyond the effect of individual differences in performance. However, it doesn't yet tell us whether the *F*-ratio is large enough to not be a chance result.

The degrees of freedom associated with this F-value are the degrees of freedom used to calculate the two mean squares. These are 2 (degrees of freedom for the model sum of squares, SS_M) and 12 (degrees of freedom for the residual sums of squares, SS_R).

Step 6: Look up the Critical value of F

Table 2 shows the critical values for *F* for different combinations of degrees of freedom. These values represent the size of *F* that we could expect to find by chance alone for different combinations of degrees of freedom. The degrees of freedom we use are those associated with the mean squares used to calculate *F* (that is, the model degrees of freedom and the residual degrees of freedom). That is, 2 and 12. Using Table 2, we need to read across to 2 degrees of freedom (model degrees of freedom) and then read down to 12 (residual degrees of freedom). The resulting values are highlighted in grey. One of these values represents the critical value of *F* when using a .05 level of probability (the convention in psychology). This value, 3.89, is the value of *F* that we would expect to get by chance for 5% of tests. So, if we took 100 sets of data in which there was no effect (i.e. we carried out no experimental manipulation so the group means were relatively similar) and calculated the *F*-ratio, then we would get one as big as 3.89 in only 5 of these datasets. The second value is for the stricter criterion of .01. This value, 6.93, is the value of *F* that we would expect to get by chance for 1% of tests. So, if we took 100 sets of data in which there was no effect and calculated the *F*-ratio, then we would get one as big as 6.93 in only 1 dataset.

As such, if our experimental manipulation has been successful (in other words, if our experimental manipulation has caused some groups to behave differently to others) we expect the means to be very different. If the group means are different then the model we've fitted will be an improvement on using the grand mean and so we'd expect *F* to be greater than the value we'd get by chance alone. The value of *F* we calculated was 5.12, and because this is bigger than the critical value of 3.89, we can say that Viagra had a significant effect on libido. However, note that because the observed value is smaller than 6.93 (the critical value for a .01 significance value) we know that this *F*-value is not significant at the stricter .01 level. We can report this by saying 'The dose of Viagra had a significant effect on a person's libido [$F(2, 12) = 5.12$, $p < .05$]'. Note that we quote the degrees of freedom associated with the *F*-ratio and the probability (*p*) of obtaining that *F*-value by chance.

Assumptions of ANOVA

The assumptions under which ANOVA is reliable are the same as for all parametric tests. Namely, data should have a normally distributed sampling distribution, be from a normally distributed population, the variances in each experimental condition are fairly similar, observations should be independent and the dependent variable should be measured on at least an interval scale. These assumptions are not completely inflexible. If you want a fairly detailed overview of the situations in which these assumptions matter most then see Field (2009) section 10.2.10 and Jane Superbrain box 5.1. The short summary is that when group sizes are equal the *F* in ANOVA is fairly robust to (i.e. unaffected by) violations in normality and when variances are unequal, but when group sizes are unequal then all hell can break loose if these two assumptions are not met. Violations of the assumption of independence can have very serious consequences indeed, so whatever you do don't break this assumption.

Table 2: Critical values for *F* taken from Field (2009).

	p	df (Numerator)																
df (Denominator)		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50	1000
1	.05	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01	249.26	250.10	251.14	251.77	254.19
	.01	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6157.31	6208.74	6239.83	6260.65	6286.79	6302.52	6362.70
2	.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.45	19.46	19.46	19.47	19.48	19.49
	.01	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.43	99.45	99.46	99.47	99.47	99.48	99.50
3	.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58	8.53
	.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	26.87	26.69	26.58	26.50	26.41	26.35	26.14
4	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70	5.63
	.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.20	14.02	13.91	13.84	13.75	13.69	13.47
5	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44	4.37
	.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.72	9.55	9.45	9.38	9.29	9.24	9.03
6	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75	3.67
	.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.30	7.23	7.14	7.09	6.89
7	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32	3.23
	.01	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	6.06	5.99	5.91	5.86	5.66
8	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02	2.93
	.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.26	5.20	5.12	5.07	4.87
9	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80	2.71
	.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.71	4.65	4.57	4.52	4.32
10	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64	2.54
	.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.31	4.25	4.17	4.12	3.92
11	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51	2.41
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.25	4.10	4.01	3.94	3.86	3.81	3.61
12	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40	2.30
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.86	3.76	3.70	3.62	3.57	3.37
13	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31	2.21
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.82	3.66	3.57	3.51	3.43	3.38	3.18
14	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24	2.14
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.66	3.51	3.41	3.35	3.27	3.22	3.02
15	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18	2.07
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.28	3.21	3.13	3.08	2.88
16	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12	2.02
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.41	3.26	3.16	3.10	3.02	2.97	2.76
17	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08	1.97
	.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.31	3.16	3.07	3.00	2.92	2.87	2.66
18	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04	1.92
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.23	3.08	2.98	2.92	2.84	2.78	2.58
19	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00	1.88
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.15	3.00	2.91	2.84	2.76	2.71	2.50
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97	1.85
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.84	2.78	2.69	2.64	2.43
22	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91	1.79
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.98	2.83	2.73	2.67	2.58	2.53	2.32

Continued over page ….

	p	df (Numerator)																
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50	1000
24	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89	1.86	1.74
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.89	2.74	2.64	2.58	2.49	2.44	2.22
26	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85	1.82	1.70
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.81	2.66	2.57	2.50	2.42	2.36	2.14
28	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82	1.79	1.66
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.75	2.60	2.51	2.44	2.35	2.30	2.08
30	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79	1.76	1.63
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.55	2.45	2.39	2.30	2.25	2.02
35	.05	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	1.96	1.88	1.82	1.79	1.74	1.70	1.57
	.01	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.60	2.44	2.35	2.28	2.19	2.14	1.90
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69	1.66	1.52
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.37	2.27	2.20	2.11	2.06	1.82
45	.05	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.89	1.81	1.75	1.71	1.66	1.63	1.48
	.01	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.46	2.31	2.21	2.14	2.05	2.00	1.75
50	.05	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.73	1.69	1.63	1.60	1.45
	.01	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.42	2.27	2.17	2.10	2.01	1.95	1.70
60	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.69	1.65	1.59	1.56	1.40
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.35	2.20	2.10	2.03	1.94	1.88	1.62
80	.05	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.79	1.70	1.64	1.60	1.54	1.51	1.34
	.01	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.27	2.12	2.01	1.94	1.85	1.79	1.51
100	.05	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.77	1.68	1.62	1.57	1.52	1.48	1.30
	.01	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.22	2.07	1.97	1.89	1.80	1.74	1.45
150	.05	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.73	1.64	1.58	1.54	1.48	1.44	1.24
	.01	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.16	2.00	1.90	1.83	1.73	1.66	1.35
300	.05	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.70	1.61	1.54	1.50	1.43	1.39	1.17
	.01	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.10	1.94	1.84	1.76	1.66	1.59	1.25
500	.05	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.53	1.48	1.42	1.38	1.14
	.01	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.07	1.92	1.81	1.74	1.63	1.57	1.20
1000	.05	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.68	1.58	1.52	1.47	1.41	1.36	1.11
	.01	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.06	1.90	1.79	1.72	1.61	1.54	1.16

Another Example:

An experiment was done to see how food/water deprivation influences learning. Students were split into four groups: ad lib control (students ate and drank normally), Food deprived, water deprived and food and water deprived. Students were then given an SPSS class on One Way ANOVA and were subsequently tested on how well they could perform such an analysis. The dependent measure was the mark given for the analysis (out of 24). The data are below, calculate the *F*-ratio for these data and find out whether it is significant [answers are on the following pages].

	Control	Food Deprived	Water Deprived	Food/Water Deprived
	18	6	15	12
	20	9	10	11
	21	8	14	8
	16	6	12	6
	15	6	14	13
$\bar{X}$	**18**	**7**	**13**	**10**
s				
s^2				
Grand Mean = 12.00 Grand Variance =				

Answer to set Question

Complete Data Table

First step, calculate the overall variance and group variances in the data table (using calculators). The answers should be:

	Control	Food Deprived	Water Deprived	Food/Water Deprived
	18	6	15	12
	20	9	10	11
	21	8	14	8
	16	6	12	6
	15	6	14	13
$\bar{X}$	18	7	13	10
s	2.55	1.41	2.00	2.92
s^2	6.50	2.00	4.00	8.50
Grand Mean = 12.00 Grand SD = 4.67 Grand Variance = 21.79				

Step 1: SST

$$\begin{aligned} SS_T &= s^2_{\text{grand}}(n-1) \\ &= 21.79(20-1) \\ &= 21.79 \times 19 \\ &= 414 \end{aligned}$$

$$df_T = N-1 = 20-1 = 19$$

Step 2: SS_M

$$SS_M = \sum n_i \left(\bar{x}_i - \bar{x}_{\text{grand}}\right)^2$$

$$\begin{aligned} SS_M &= 5(18-12)^2 + 5(7-12)^2 + 5(13-12)^2 + 5(10-12)^2 \\ &= 5(6)^2 + 5(-5)^2 + 5(1)^2 + 5(-2)^2 \\ &= 180 + 125 + 5 + 20 \\ &= 330 \end{aligned}$$

$$df_M = k-1 = 3$$

Step 3: SS_R

$$\begin{aligned} SS_R &= s^2_{\text{group1}}(n_1-1) + s^2_{\text{group2}}(n_2-1) + s^2_{\text{group3}}(n_3-1) + s^2_{\text{group4}}(n_4-1) \\ &= (6.50)(5-1) + (2.00)(5-1) + (4.00)(5-1) + (8.50)(5-1) \\ &= (6.50 \times 4) + (2.00 \times 4) + (4.00 \times 4) + (8.50 \times 4) \\ &= 26 + 8 + 16 + 34 \\ &= 84 \end{aligned}$$

$$\begin{aligned} df_R &= df_{group1} + df_{group2} + df_{group3} + df_{group4} \\ &= (n_1-1) + (n_2-1) + (n_3-1) + (n_4-1) \\ &= 4+4+4+4 \\ &= 16 \end{aligned}$$

Double-check the calculations:

$$SS_T = SS_M + SS_R$$
$$414 = 330 + 84$$
$$414 = 414$$

$$df_T = df_M + df_R$$
$$19 = 3 + 16$$
$$19 = 19$$

Step 4: Calculate Mean Squares

$$\mathrm{MS_M} = \frac{\mathrm{SS_M}}{\mathrm{df_M}} = \frac{330}{3} = 110.00$$
$$\mathrm{MS_R} = \frac{\mathrm{SS_R}}{\mathrm{df_R}} = \frac{84}{16} = 5.25$$

Step 5: Calculate the F-ratio

$$F = \frac{\mathrm{MS_M}}{\mathrm{MS_R}} = \frac{110}{5.25} = 20.95$$

Step 6: Look up Critical F

From Table 2, the critical values for 3 and 16 degrees of freedom are 3.24 (α = .05) and 5.29 (α = .01). Our observed value of 20.95 is bigger than both of these values and so we can say that the observed *F*-ratio is significant at the .01 level. We could write that the type of deprivation (none, water, food or water and food) had a significant effect on performance on the SPSS exam, $F(3, 16) = 20.95$, $p < .01$. However, at this stage we don't know anything more specific than this.

SPSS summarises the information for this example as follows:

ANOVA

SPSS Mark

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	330.000	3	110.000	20.952	.000
Within Groups	84.000	16	5.250		
Total	414.000	19			

This table provides the exact significance of the observed *F*-ratio. This significance value is much less than .05, .01 or even .001 which tells us that this result is highly significant. By providing this information SPSS saves us the trouble of having to look up the critical value of *F*, instead it tells us how likely it is that we would get an *F* as large as the one we observed by chance alone. In this case, the probability is very small (zero to three decimal places), but the important thing is that it's less than .05.

This handout contains material from:

Field, A. P. (2009). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll (3rd Edition)*. London: Sage.