



## The Theory of ANOVA

### Using a linear model to compare means

We saw in last week’s lecture that if we include a predictor variable containing two categories into the linear model then the resulting  $b$  for that predictor compares the difference between the mean score for the two categories. We also saw that if we want to include a categorical predictor that contains more than two categories, this can be achieved by recoding that variable into several categorical predictors each of which has only two categories (dummy coding). We can flip this idea on its head to ask how we can use a linear model to compare differences between the means of more than two groups. The answer is the same: we use dummy coding to represent the groups and stick them in a linear model. ANOVA and regression are often taught as though they are completely unrelated tests. However, we test the fit of a regression model with an ANOVA (the  $F$ -test).

Let’s take an example. Viagra is a sexual stimulant (used to treat impotence). In the psychology literature sexual performance issues have been linked to a loss of libido (Hawton, 1989). Suppose we tested this belief by taking three groups of participants 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. The data can be found in the file **Viagra.sav** (which is described in Field (2013)) and are 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
$\bar{X}$	<b>2.20</b>	<b>3.20</b>	<b>5.00</b>
$s$	<b>1.30</b>	<b>1.30</b>	<b>1.58</b>
$s^2$	<b>1.70</b>	<b>1.70</b>	<b>2.50</b>
Grand Mean = <b>3.467</b> Grand SD = <b>1.767</b> Grand Variance = <b>3.124</b>			

If we want to predict levels of libido from the different levels of Viagra then we can use:

$$\text{outcome}_i = (\text{model}) + \text{error}_i$$

If we want to use a linear model, then when there are only two groups we could replace the ‘model’ in this equation with a linear regression equation with one dummy variable to describe two groups (see Field, 2013, Chapter 9). This dummy variable was a categorical variable with two numeric codes (0 for one group and 1 for the other). With three groups, we extend this idea and use a multiple regression model with two dummy variables. We can extend the model to any number of groups and the number of dummy variables needed will be one less than the number of categories of the independent variable (see Field, 2013, Chapters 10 and 11). As with the two-group case, we need a base category and you should choose the condition to which you intend to compare the other groups. Usually this category will be the control group. In unbalanced designs (in which the group sizes are unequal) it is important that the base category contains a fairly large number of cases to ensure that the estimates of the regression coefficients are reliable. In the Viagra example, we can take the placebo group as the base category because this group was a placebo control. If the placebo group is the base category then the two dummy variables that we have to create represent the other two conditions: so, we should have one dummy variable called High and the other one called Low). The resulting equation is described as:



$$\text{Libido}_i = b_0 + b_2\text{High}_i + b_1\text{Low}_i + \varepsilon_i \quad \text{Eq. 1}$$

In Eq. 1 a person's libido can be predicted from knowing their group code (i.e., the code for the High and Low dummy variables) and the intercept ( $b_0$ ) of the model. The dummy variables in Eq. 1 can be coded in several ways, but the simplest way is as we did earlier on this module. The base category is coded as 0; If a participant was given a high dose of Viagra then they are coded with a 1 for the High dummy variable and 0 for all other variables. If a participant was given a low dose of Viagra then they are coded with the value 1 for the Low dummy variable and coded with 0 for all other variables. Using this coding scheme we can express each group by combining the codes of the two dummy variables (see Table 2).

Table 2: Dummy coding for the three-group experimental design

Group	Dummy Variable 1 (High)	Dummy Variable 2 (Low)
Placebo	0	0
Low Dose Viagra	0	1
High Dose Viagra	1	0

When the predictor is made up of groups, the predicted values (the value of libido in Eq. 1) will be the group mean. Knowing this we can look at the model for each group.

**Placebo Group:** In the placebo group both the High and Low dummy variables are coded as 0. The predicted value for the model will be the mean of the placebo group. If we ignore the error term ( $\varepsilon_i$ ), the regression equation becomes:

$$\text{Libido}_i = b_0 + (b_2 \times 0) + (b_1 \times 0)$$

$$\text{Libido}_i = b_0$$

$$\bar{X}_{\text{Placebo}} = b_0$$

We are looking at predicting the level of libido when both doses of Viagra are ignored, and so the predicted value will be the mean of the placebo group (because this group is the only one included in the model). Hence, the intercept of the regression model,  $b_0$ , is always the mean of the base category (in this case the mean of the placebo group).

**High-dose group:** If we examine the high-dose group, the dummy variable High will be coded as 1 and the dummy variable Low will be coded as 0. If we replace the values of these codes into Eq. 1 the model becomes:

$$\text{Libido}_i = b_0 + (b_2 \times 1) + (b_1 \times 0)$$

$$\text{Libido}_i = b_0 + b_2$$

We know already that  $b_0$  is the mean of the placebo group. If we are interested in only the high-dose group then the model should predict that the value of Libido for a given participant equals the mean of the high-dose group:

$$\text{Libido}_i = b_0 + b_2$$

$$\bar{X}_{\text{High}} = \bar{X}_{\text{Placebo}} + b_2$$

$$b_2 = \bar{X}_{\text{High}} - \bar{X}_{\text{Placebo}}$$

Hence,  $b_2$  represents the difference between the means of the high-dose and placebo groups.

**Low-dose group:** Finally, if we look at the model when a low dose of Viagra has been taken, the dummy variable Low is coded as 1 (and hence High is coded as 0). Therefore, the regression equation becomes:

$$\text{Libido}_i = b_0 + (b_2 \times 0) + (b_1 \times 1)$$

$$\text{Libido}_i = b_0 + b_1$$

We know that the intercept is equal to the mean of the base category and that for the low-dose group the predicted value should be the mean libido for a low dose. Therefore, the model reduces down to:



$$\text{Libido}_i = b_0 + b_1$$

$$\bar{X}_{\text{Low}} = \bar{X}_{\text{placebo}} + b_1$$

$$b_1 = \bar{X}_{\text{Low}} - \bar{X}_{\text{placebo}}$$

Hence,  $b_1$  represents the difference between the means of the low-dose group and the placebo group. This form of dummy variable coding is the simplest form, but as we will see later, there are other ways in which variables can be coded to test specific hypotheses. These alternative coding schemes are known as *contrasts* (see your next lecture).

## Logic of the F-ratio

Figure 1 shows the Viagra data in graphical form (including the group means, the overall mean and the difference between each case and the group mean). 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, Figure 1 the three coloured lines would be in the same vertical position (the exact position would be the grand mean—the solid horizontal line in the figure). We can see from the diagram that the group means are different because the coloured lines (the group means) are in different vertical positions. We have just found out that in the regression model,  $b_2$  represents the difference between the means of the placebo and the high-dose group, and  $b_1$  represents the difference in means between the low-dose and placebo groups. These two distances are represented in Figure 1 by the vertical arrows. If the null hypothesis is true and all the groups have the same means, then these  $b$  coefficients should be zero (because if the group means are equal then the difference between them will be zero).

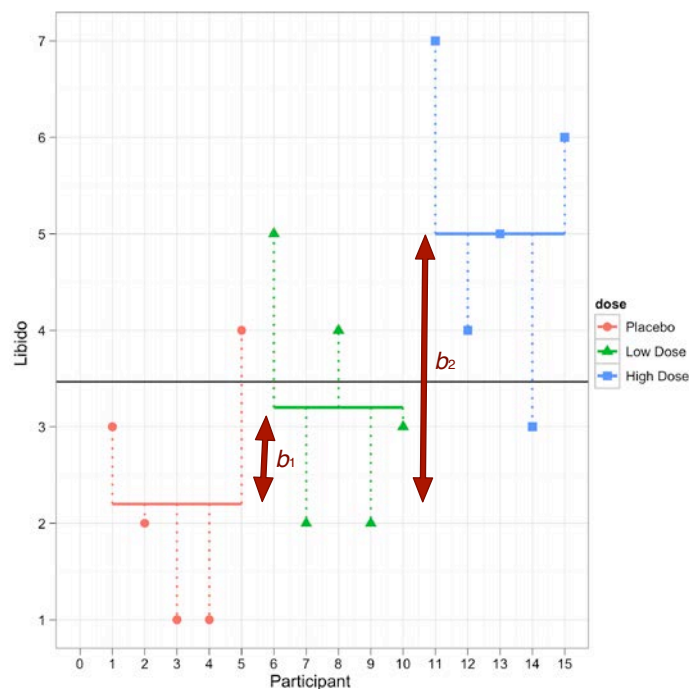


Figure 1: The Viagra data in graphical form.

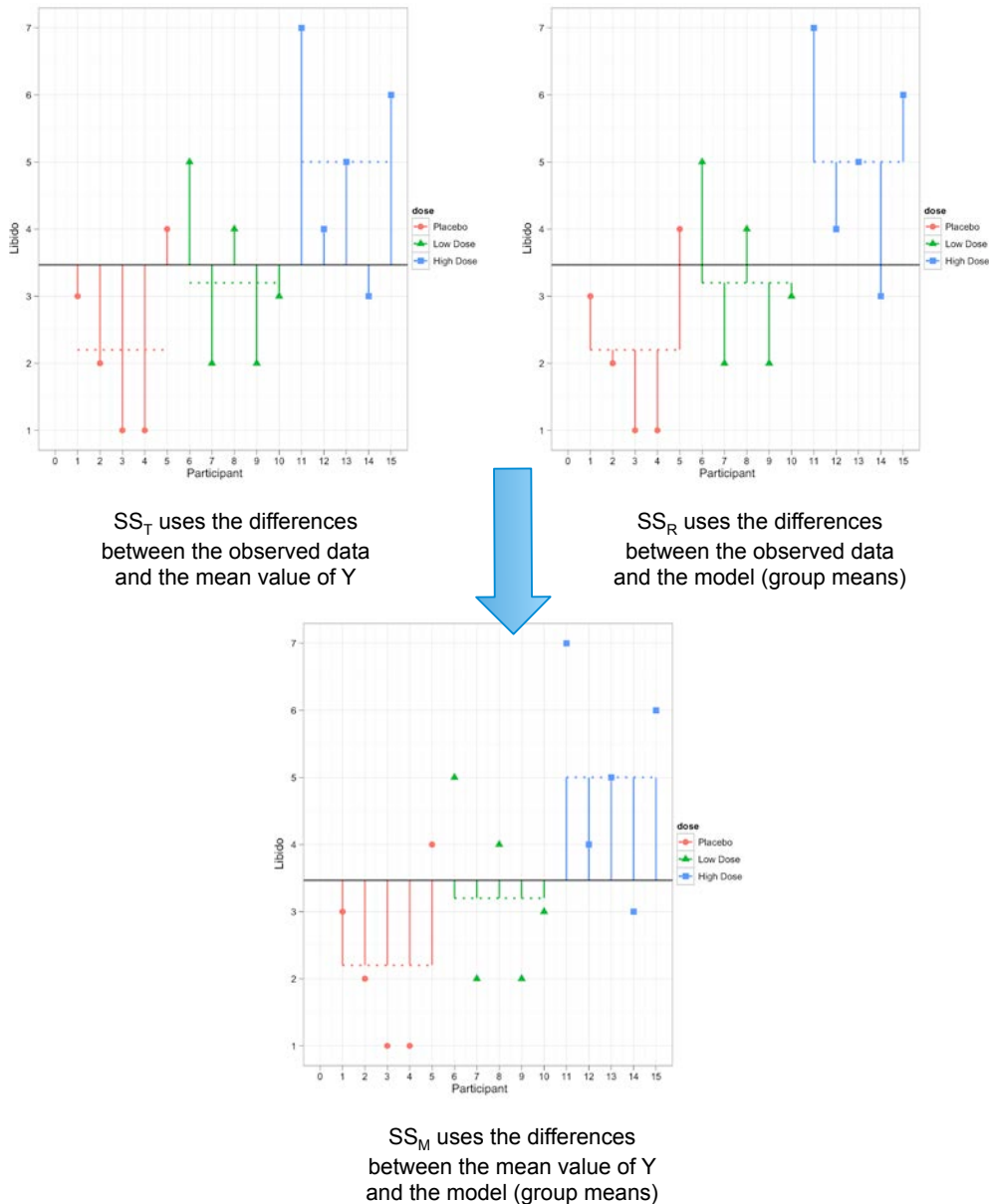


Figure 2: Graphical representation of the different sums of squares in ANOVA designs

The logic of ANOVA follows from what we already know about linear models:

- 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.
- The intercept and one or more parameters ( $b$ ) describe the model.
- The parameters determine the shape of the model that we have fitted; therefore, the bigger the coefficients, the greater the deviation between the model and the grand mean.
- In experimental research the parameters ( $b$ ) represent the differences between group means. The bigger the differences between group means, the greater the difference between the model and the grand mean.
- If the differences between group means are large enough, then the resulting model will be a better fit of the data than the grand mean.



- If this is the case we can infer that our model (i.e., predicting scores from the group means) is better than not using a model (i.e., predicting scores from the grand mean). Put another way, our group means are significantly different.

Just as we have done before, we use the  $F$ -ratio to compare the improvement in fit due to using the model (rather than the grand mean) to the error that still remains. In other words, the  $F$ -ratio is the ratio of the explained to the unexplained variation.

## Total sum of squares ( $SS_T$ )

To find the total amount of variation within our data we calculate 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 ( $SS_T$ ):

$$SS_T = \sum_{i=1}^N (x_i - \bar{x}_{\text{grand}})^2 \quad \text{Eq. 2}$$

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 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:

$$\begin{aligned} SS_T &= s_{\text{grand}}^2(n - 1) \\ &= 3.124(15 - 1) \\ &= 3.124 \times 14 \\ &= 43.74 \end{aligned}$$

Before we move on, it is important to understand degrees of freedom. We saw before that when we estimate population values, the degrees of freedom are typically one less than the number of scores used to calculate the population value. This is because to get these estimates we have to hold something constant in the population (in this case the mean), which leaves all but one of the scores free to vary. For  $SS_T$ , we used the entire sample (i.e., 15 scores) to calculate the sums of squares and so the total degrees of freedom ( $df_T$ ) are one less than the total sample size ( $N - 1$ ). For the Viagra data, this value is 14.

## Model sum of squares ( $SS_M$ )

In your regression lecture you 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 (the coloured dashed horizontal lines in Figure 2). The bottom panel in Figure 2 shows the model sum of squared error: it is the sum of the squared distances between what the model predicts for each data point (i.e., the dotted horizontal line for the group to which the data point belongs) and the overall mean of the data (the solid horizontal line).

For each participant the value predicted by the model is the mean for the group to which the participant belongs. In the Viagra example, the predicted value for the five participants in the placebo group will be 2.2, for the five participants in the low-dose condition it will be 3.2, and for the five participants in the high-dose condition it will be 5. The model sum of squares requires us to calculate the differences between each participant'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 participants in a particular group is the mean of that group. Therefore, the easiest way to calculate  $SS_M$  is to:

- Calculate the difference between the mean of each group and the grand mean.
- Square each of these differences.
- Multiply each result by the number of participants within that group ( $n_k$ ).
- Add the values for each group together.

The mathematical expression of this process is:



$$SS_M = \sum_{k=1}^k n_k (\bar{x}_k - \bar{x}_{grand})^2 \quad \text{Eq. 3}$$

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 ( $df_M$ ) will always be one less than the number of parameters estimated. In short, this value will be the number of groups minus one (which you'll see denoted as  $k-1$ ). So, in the three-group case the degrees of freedom will always be 2 (because the calculation of the sums of squares is based on the group means, two of which will be free to vary in the population if the third is held constant).

## 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, telling you to do this provides little insight into what is being calculated and, of course, if you've messed up the calculations of either  $SS_M$  or  $SS_T$  (or both!) then  $SS_R$  will be incorrect also.

We saw in the regression lectures that the residual sum of squares is the difference between what the model predicts and what was actually observed. In ANOVA, the values predicted by the model are the group means (the coloured dashed horizontal lines in Figure 2). The top right panel shows the residual sum of squared error: it is the sum of the squared distances between each point and the dotted horizontal line for the group to which the data point belongs.

We already know that for a given participant, 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. In graphical terms the vertical lines in Figure 1 represent this sum of squares. These distances between each data point and the group mean are squared and then added together to give the residual sum of squares,  $SS_R$ :

$$SS_R = \sum (x_{ik} - \bar{x}_k)^2 \quad \text{Eq. 4}$$

The sum of squares for each group represents the sum of squared differences between each participant's score in that group and the group mean. Therefore, we can express  $SS_R$  as  $SS_R = SS_{group1} + SS_{group2} + SS_{group3} \dots$  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:

$$SS_R = \sum s_k^2 (n_k - 1) \quad \text{Eq. 5}$$

This just means take the variance from each group ( $s_k^2$ ) and multiply it by one less than the number of people in that group ( $n_k - 1$ ). When you've done this for each group, add them all up. For the Viagra data, this gives us:

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



The degrees of freedom for  $SS_R$  ( $df_R$ ) are the total degrees of freedom minus the degrees of freedom for the model ( $df_R = df_T - df_M = 14 - 2 = 12$ ). Put another way, it's  $N - k$ : the total sample size,  $N$ , minus the number of groups,  $k$ .

## Mean squares

$SS_M$  tells us the *total* variation that the regression model (e.g., the experimental manipulation) explains and  $SS_R$  tells us the *total* variation that is due to extraneous factors. However, because both of these values are summed values they will be 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 12 and 14 values respectively. 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. 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).

## 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. In other words, it is the ratio of how good the model is against how bad it is (how much error there is). It can be calculated by dividing the model mean squares by the residual mean squares.

$$F = \frac{MS_M}{MS_R} \tag{Eq. 6}$$

The  $F$ -ratio is a measure of the ratio of systematic variation to unsystematic variation. In experimental research, 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 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. 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. We can compare the obtained value of  $F$  against the maximum value we would expect to get by chance if the group means were equal in an  $F$ -distribution with the same degrees of freedom (these values can be found in **Error! Reference source not found.**); if the value we obtain exceeds this critical value we can be confident that this reflects an effect of our independent variable. In this case, with 2 and 12 degrees of freedom the critical values are 3.89 ( $p = .05$ ) and 6.93 ( $p = .01$ ). The observed value, 5.12, is, therefore, significant at a .05 level of significance but not significant at a .01 level. The exact significance produced by SPSS should, therefore, fall somewhere between .05 and .01 (which, incidentally, it does).

## References

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage.

Field, A. P. (2016). *An adventure in statistics: the reality enigma*. London: Sage.

Hawton, K. (1989). Sexual dysfunctions. In K. Hawton, P. M. Salkovskis, J. Kirk, & D. M. Clark (Eds.), *Cognitive behaviour therapy for psychiatric problems: a practical guide*. (pp. 370-405). Oxford: Oxford University Press.



## Terms of Use

This handout contains material from:

Field, A. P. (2013). *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll (4th Edition)*. London: Sage.

This material is copyright Andy Field (2000-2016).

This document is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), basically you can use it for teaching and non-profit activities but not meddle with it without permission from the author.

## Critical values of the F-distribution

The values below are taken from Field (2016)

$p = 0.05$

df <sub>n</sub>	df <sub>m</sub>													
	1	2	3	4	5	6	7	8	9	10	15	25	50	100
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	249.26	251.77	253.04
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.46	19.48	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.63	8.58	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.77	5.70	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.52	4.44	4.41
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.83	3.75	3.71
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.40	3.32	3.27
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.11	3.02	2.97
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.89	2.80	2.76
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.73	2.64	2.59
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.50	2.40	2.35
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.34	2.24	2.19
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.23	2.12	2.07
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.14	2.04	1.98
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.07	1.97	1.91
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.02	1.91	1.85
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	1.97	1.86	1.80
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.94	1.82	1.76
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.91	1.79	1.73
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.88	1.76	1.70
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18	2.13	1.98	1.84	1.72	1.66
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	1.96	1.82	1.70	1.63
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.78	1.66	1.59
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	1.89	1.75	1.63	1.55
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.73	1.60	1.52
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06	2.01	1.85	1.71	1.58	1.50
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.69	1.56	1.48
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03	1.98	1.82	1.68	1.54	1.46
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.81	1.66	1.53	1.45
75	3.97	3.12	2.73	2.49	2.34	2.22	2.13	2.06	2.01	1.96	1.80	1.65	1.52	1.44
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.79	1.64	1.51	1.43
85	3.95	3.10	2.71	2.48	2.32	2.21	2.12	2.05	1.99	1.94	1.79	1.64	1.50	1.42
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.78	1.63	1.49	1.41
95	3.94	3.09	2.70	2.47	2.31	2.20	2.11	2.04	1.98	1.93	1.77	1.62	1.48	1.40
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.77	1.62	1.48	1.39

$p = 0.01$

df <sub>n</sub>	df <sub>m</sub>													
	1	2	3	4	5	6	7	8	9	10	15	25	50	100
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6157.28	6239.83	6302.52	6334.11
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.43	99.46	99.48	99.49
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	26.87	26.58	26.35	26.24
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.20	13.91	13.69	13.58
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.72	9.45	9.24	9.13
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.30	7.09	6.99
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.06	5.86	5.75
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.26	5.07	4.96
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.71	4.52	4.41
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.31	4.12	4.01
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.76	3.57	3.47



# DISCOVERING STATISTICS



14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.66	3.41	3.22	3.11
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.41	3.16	2.97	2.86
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.23	2.98	2.78	2.68
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.84	2.64	2.54
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.98	2.73	2.53	2.42
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.89	2.64	2.44	2.33
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.81	2.57	2.36	2.25
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.75	2.51	2.30	2.19
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.45	2.25	2.13
33	7.47	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00	2.91	2.63	2.39	2.18	2.06
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.60	2.35	2.14	2.02
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.27	2.06	1.94
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.46	2.21	2.00	1.88
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.42	2.17	1.95	1.82
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.38	2.13	1.91	1.78
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.35	2.10	1.88	1.75
65	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69	2.61	2.33	2.07	1.85	1.72
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.31	2.05	1.83	1.70
75	6.99	4.90	4.05	3.58	3.27	3.05	2.89	2.76	2.65	2.57	2.29	2.03	1.81	1.67
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.27	2.01	1.79	1.65
85	6.94	4.86	4.02	3.55	3.24	3.02	2.86	2.73	2.62	2.54	2.26	2.00	1.77	1.64
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.24	1.99	1.76	1.62
95	6.91	4.84	3.99	3.52	3.22	3.00	2.83	2.70	2.60	2.51	2.23	1.98	1.75	1.61
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.22	1.97	1.74	1.60

Values computed by Andy Field using R