Introduction to the Independent-Measures Design

Until this point, all of the inferential statistics we have considered involve using one sample as the basis for drawing conclusions about one population. Although these *single-sample* techniques are used occasionally in real research, most research studies require the comparison of two (or more) sets of data.

For example, a social psychologist may want to compare men and women in terms of their political attitudes, an educational psychologist may want to compare two methods for teaching mathematics, or a clinical psychologist may want to evaluate a therapy technique by comparing depression scores for patients before therapy with their scores after therapy. In each case, the research question concerns a mean difference between two sets of data.

There are two general research designs that can be used to obtain the two sets of data to be compared:

1. The two sets of data could come from two completely separate groups of participants. For example, the study could involve a sample of men compared with a sample of women. Or the study could compare grades for one group of freshmen who are given laptop computers with grades for a second group who are not given computers.

2. The two sets of data could come from the same group of participants. For example, the researcher could obtain one set of scores by measuring depression for a sample of patients before they begin therapy and then obtain a second set of data by measuring the same individuals after 6 weeks of therapy.

The first research strategy, using completely separate groups, is called an *independentmeasures* research design or a *between-subjects* design. These terms emphasize the fact that the design involves separate and independent samples and makes a comparison between two groups of individuals. The structure of an independent-measures research design is shown in Figure 10.1. Notice that the research study uses two separate samples to represent the two different populations (or two different treatments) being compared.

Definition: A research design that uses a separate group of participants for each treatment condition (or for each population) is called an **independent-measures research design** or a **between-subjects research design.**

The *t* Statistic for an Independent-Measures Research Design

i. State hypothesis

As always, the null hypothesis states that there is no change, no effect, or, in this case, no difference. Thus, in symbols, the null hypothesis for the independent-measures test is

H_0: $\mu_1 - \mu_2 = 0$ (No difference between the population means)

The alternative hypothesis states that there is a mean difference between the two populations,

H_1: $\mu_1 - \mu_2 \neq 0$ (There is a mean difference.)

ii. Set the Critical Regian

Usual α-levels are 0.05, 0.01 and 0.001.

$$df_{total} = df_1 + df_2$$

Where $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$

iii. Compute test statistics

The independent-measures t uses the difference between two sample means to evaluate a hypothesis about the difference between two population means. Thus, the independent-measures t formula is

$$t = \frac{sample\ mean\ difference - population\ mean\ difference}{estimated\ standard\ error}$$

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}}$$

<u>The estimated standard error</u>: For the independent-measures *t* statistic, we want to know the total amount of error involved in using *two* sample means to approximate *two* population means. To do this, we find the error from each sample separately and then add the two errors together.

$$s_{(M_1 - M_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad \text{(Equation 1)}$$

<u>Pooled Variance:</u> Although (Equation 1) accurately presents the concept of standard error for the independent-measures *t* statistic, this formula is limited to situations in which the two samples are exactly the same size (that is $n_1 = n_2$) . For situations in which the two sample sizes are different, the formula is *biased* and, therefore, inappropriate.

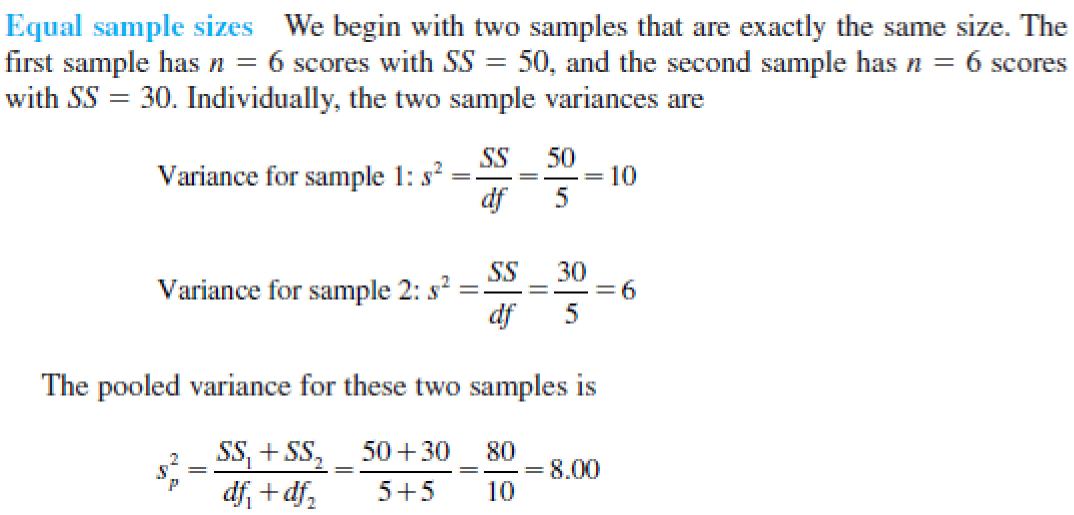
The bias comes from the fact that (Equation 1) treats the two sample variances equally. However, when the sample sizes are different, the two sample variances are not equally good and should not be treated equally,

The law of large numbers, which states that statistics obtained from large samples tend to be better (more accurate) estimates of population parameters than statistics obtained from small samples.

This same fact holds for sample variances: The variance obtained from a large sample is a more accurate estimate of σ^2 than the variance obtained from a small sample.

One method for correcting the bias in the standard error is to combine the two sample variances into a single value called **the pooled variance**. The pooled variance is obtained by averaging, or "pooling," the two sample variances using a procedure that allows the bigger sample to carry more weight in determining the final value.

$$pooled\ variance = s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$



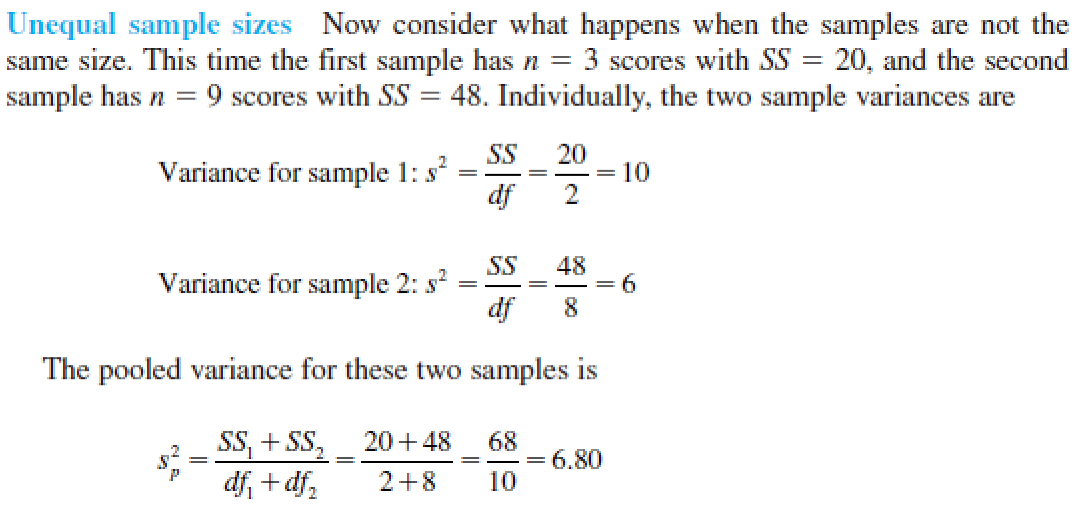
Equal sample sizes We begin with two samples that are exactly the same size. The first sample has $n = 6$ scores with $SS = 50$, and the second sample has $n = 6$ scores with $SS = 30$. Individually, the two sample variances are

$$\text{Variance for sample 1: } s^2 = \frac{SS}{df} = \frac{50}{5} = 10$$

$$\text{Variance for sample 2: } s^2 = \frac{SS}{df} = \frac{30}{5} = 6$$

The pooled variance for these two samples is

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{50+30}{5+5} = \frac{80}{10} = 8.00$$



Unequal sample sizes Now consider what happens when the samples are not the same size. This time the first sample has $n = 3$ scores with $SS = 20$, and the second sample has $n = 9$ scores with $SS = 48$. Individually, the two sample variances are

$$\text{Variance for sample 1: } s^2 = \frac{SS}{df} = \frac{20}{2} = 10$$

$$\text{Variance for sample 2: } s^2 = \frac{SS}{df} = \frac{48}{8} = 6$$

The pooled variance for these two samples is

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{20+48}{2+8} = \frac{68}{10} = 6.80$$

Finally,

$$estimated\ standar\ error\ for\ M_1 - M_2 = \ s_{(M_1 - M_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Assumptions Underlying the Independent- Measures *t* Formula

There are three assumptions that should be satisfied before you use the independent measures *t* formula for hypothesis testing:

1. The observations within each sample must be independent .

2. The two populations from which the samples are selected must be normal.

3. To justify using the pooled variance, the two populations from which the samples are selected must have equal variances.

The third assumption is referred to as homogeneity of variance and states that the two populations being compared must have the same variance.

Homogeneity of variance is most important when there is a large discrepancy between the sample sizes. With equal (or nearly equal) sample sizes, this assumption is less critical, but still important.

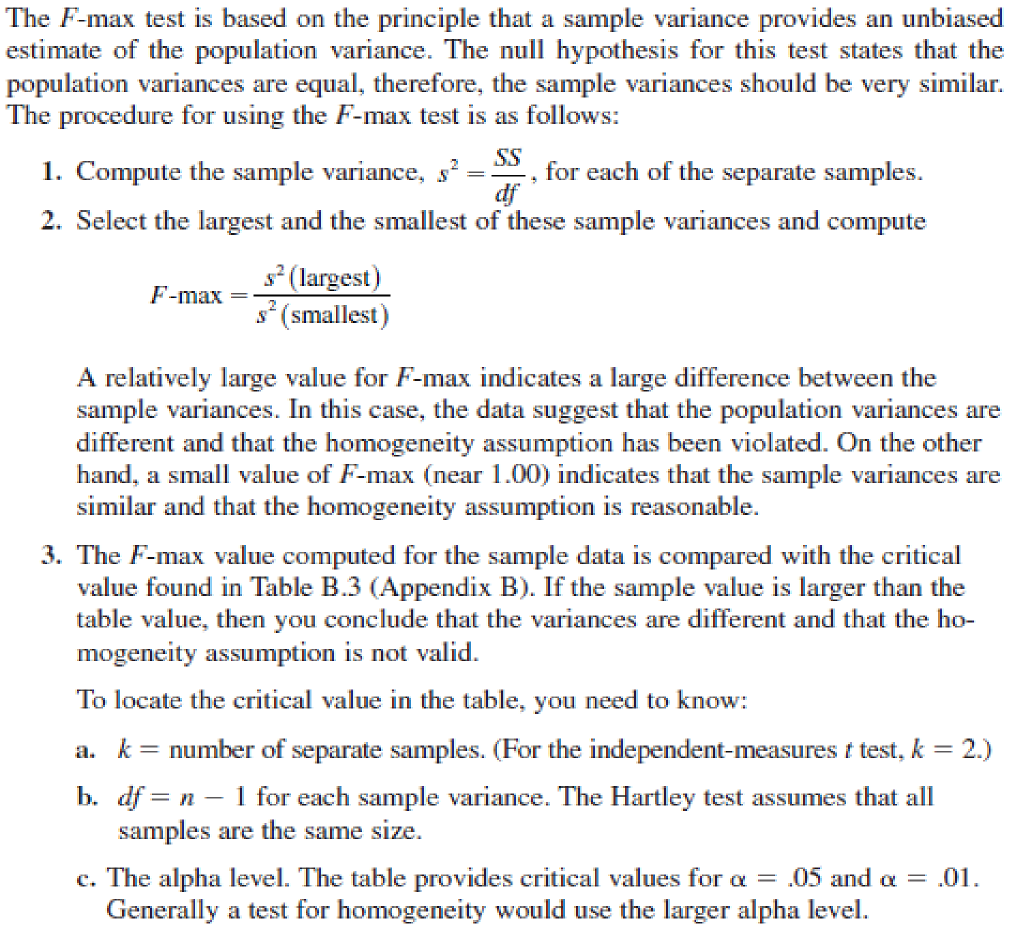
Violating the homogeneity of variance assumption can prevent any meaningful interpretation of the data from an independent-measures experiment.

Specifically, when you compute the *t* statistic in a hypothesis test, all of the numbers in the formula come from the data except for the population mean difference, which you get from H_0. Thus, you are sure of all of the numbers in the formula except one. If you obtain an extreme result for the *t* statistic (a value in the critical region), then you conclude that the hypothesized value was wrong.

But consider what happens when the homogeneity assumption is violated. In this case, you have two questionable values in the formula (the hypothesized population value and the meaningless average of the two variances). Now if you obtain an extreme *t* statistic, you do not know which of these two values is responsible. Specifically, you cannot reject the hypothesis because it may have been the pooled variance that produced the extreme *t* statistic. Without satisfying the homogeneity of variance requirement, you cannot accurately interpret a *t* statistic, and the hypothesis test becomes meaningless.

Hartley’s *F*-Max Test

One simple test involves just looking at the two sample variances. Logically, if the two population variances are equal, then the two sample variances should be very similar.



The *F*-max test is based on the principle that a sample variance provides an unbiased estimate of the population variance. The null hypothesis for this test states that the population variances are equal, therefore, the sample variances should be very similar. The procedure for using the *F*-max test is as follows:

1. Compute the sample variance, $s^2 = \frac{SS}{df}$, for each of the separate samples.
2. Select the largest and the smallest of these sample variances and compute

$$F\text{-max} = \frac{s^2(\text{largest})}{s^2(\text{smallest})}$$

   A relatively large value for *F*-max indicates a large difference between the sample variances. In this case, the data suggest that the population variances are different and that the homogeneity assumption has been violated. On the other hand, a small value of *F*-max (near 1.00) indicates that the sample variances are similar and that the homogeneity assumption is reasonable.

3. The *F*-max value computed for the sample data is compared with the critical value found in Table B.3 (Appendix B). If the sample value is larger than the table value, then you conclude that the variances are different and that the homogeneity assumption is not valid.

   To locate the critical value in the table, you need to know:

   a. k = number of separate samples. (For the independent-measures t test, $k = 2$.)

   b. $df = n - 1$ for each sample variance. The Hartley test assumes that all samples are the same size.

   c. The alpha level. The table provides critical values for $\alpha = .05$ and $\alpha = .01$. Generally a test for homogeneity would use the larger alpha level.

Levene's test

In statistics, **Levene's test** is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups. Some common statistical procedures assume that variances of the populations from which different samples are drawn are equal. Levene's test assesses this assumption. It tests the null hypothesis that the population variances are equal (called *homogeneity of variance* or *homoscedasticity*). If the resulting *P*-value of Levene's test is less than some significance level (typically 0.05), the obtained differences in sample variances are unlikely to have occurred based on random sampling from a population with equal variances. Thus, the null hypothesis of equal variances is rejected and it is concluded that there is a difference between the variances in the population.

Some of the procedures typically assuming homoscedasticity, for which one can use Levene's tests, include analysis of variance and t-tests.

Levene's test is often used before a comparison of means. When Levene's test shows significance, one should switch to generalized tests (non-parametric tests), free from homoscedasticity assumptions.

THE INDEPENDENT-MEASURES *t* TEST

In a study of jury behavior, two samples of participants were provided details about a trial in which the defendant was obviously guilty. Although group 2 received the same details as group 1, the second group was also told that some evidence had been withheld from the jury by the judge. Later, the participants were asked to recommend a jail sentence. The length of term suggested by each participant is presented here. Is there a significant difference between the two groups in their responses?

Group 1	*Group 2*
4	3
4	7
3	8
2	5
5	4
1	7
1	6
4	8

For Group 1: $M = 3$ and $SS = 16$

For Group 2: $M = 6$ and $SS = 24$

There are two separate samples in this study. Therefore, the analysis uses the independent-measures *t* test.

State the hypothesis, and select an alpha level.

H_0: $\mu_1 - \mu_2 = 0$ (For the population, knowing that evidence has been withheld has no effect on the suggested sentence.)

H_1: $\mu_1 - \mu_2 \neq 0$ (For the population, knowing that evidence has been withheld has an effect on the jury's response.)

We set the level of significance to $\alpha = .05$, two tails.

Identify the critical region. For the independent-measures *t* statistic, degrees of freedom are determined by

$$\begin{aligned} df &= df_1 + df_2 \\ &= 7 + 7 \\ &= 14 \end{aligned}$$

The *t* distribution table is consulted, for a two-tailed test with $\alpha = .05$ and $df = 14$. The critical *t* values are +2.145 and –2.145.

Compute the test statistic. As usual, we recommend that the calculation of the *t* statistic be separated into three stages.

Pooled variance: For these data, the pooled variance equals

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{16 + 24}{7 + 7} = \frac{40}{14} = 2.86$$

Estimated standard error: Now we can calculate the estimated standard error for mean differences.

$$s_{(M_1-M_2)} = \sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}} = \sqrt{\frac{2.86}{8}+\frac{2.86}{8}} = \sqrt{0.358+0.358} = \sqrt{0.716} = 0.85$$

The t statistic: Finally, the *t* statistic can be computed.

$$t = \frac{(M_1-M_2)-(\mu_1-\mu_2)}{s_{(M_1-M_2)}} = \frac{(3-6)-0}{0.85} = \frac{-3}{0.85} = -3.53$$

Make a decision about H_0, and state a conclusion. The obtained *t* value of −3.53 falls n the critical region of the left tail (critical *t* = ±2.145). Therefore, the null hypothesis s rejected. The participants who were informed about the withheld evidence gave significantly longer sentences, *t*(14)= −3.53, $p < .05$, two tails.

0.2

EFFECT SIZE FOR THE INDEPENDENT-MEASURES *t*

We compute Cohen's *d* and r^2 for the jury decision data in Demonstration 10.1. For these data, the two sample means are $M_1 = 3$ and $M_2 = 6$, and the pooled variance is 2.86. Therefore, our estimate of Cohen's *d* is

$$\text{estimated } d = \frac{M_1-M_2}{\sqrt{s_p^2}} = \frac{3-6}{\sqrt{2.86}} = \frac{3}{1.69} = 1.78$$

With a *t* value of *t* = 3.53 and *df* = 14, the percentage of variance accounted for is

$$r^2 = \frac{t^2}{t^2+df} = \frac{(3.53)^2}{(3.53)^2+14} = \frac{12.46}{26.46} = 0.47 \text{ (or 47\%)}$$

Exercises

1-For each of the following, assume that the two samples are obtained from populations with the same mean, and calculate how much difference should be expected, on average, between the two sample means.

a. Each sample has $n = 4$ scores with $s^2 = 68$ for the first sample and $s^2 = 76$ for the second. (*Note:* Because the two samples are the same size, the pooled variance is equal to the average of the two sample variances.)

b. Each sample has $n = 16$ scores with $s^2 = 68$ for the first sample and $s^2 = 76$ for the second.

c. In part b, the two samples are bigger than in part a, but the variances are unchanged. How does sample size affect the size of the standard error for the sample mean difference?

2- For each of the following, calculate the pooled variance and the estimated standard error for the sample mean difference.

a. The first sample has $n = 4$ scores and a variance of $s^2 = 55$, and the second sample has $n = 6$ scores and a variance of $s^2 = 63$.

b. Now the sample variances are increased so that the first sample has $n = 4$ scores and a variance of $s^2 = 220$, and the second sample has $n = 6$ scores and a variance of $s^2 = 252$.

c. Comparing your answers for parts a and b, how does increased variance influence the size of the estimated standard error?

3- If other factors are held constant, explain how each of the following influences the value of the independent measures *t* statistic and the likelihood of rejecting the null hypothesis:

a. An increase in the mean difference between the samples.

b. An increase in the number of scores in each sample.

c. An increase in the variance for each sample.

4-Describe the homogeneity of variance assumption and explain why it is important for the independent measures *t* test.

5- A researcher would like to compare the political attitudes for college freshmen those for college seniors. Sample of n=10 freshmen and n=10 seniors are obtained, and each student is given a questionnaire measuring political attitudes on a scale from 0 to 100. The average score for the freshman is M=52 with SS=4800, and the seniors average M=39 with SS=4200. Do these data indicate significant differences in political attitude for freshman versus seniors? Test at the .05 level of significance.

6- A person's gender can have a tremendous influence on his or her personality and behavior. Psychologists classify individuals as masculine, feminine, or androgynous. Androgynous individuals possess both masculine and feminine traits. Among other things, androgynous individuals appear to cope with better with stress than do traditionally masculine or feminine people. In a typical study, depression scores are recorded for a sample of traditionally masculine or feminine participants and for a sample of androgynous participants, all of whom have recently experienced a series of strongly negative events. The average depression score for the sample of n=10 androgynous participants is M=63 with SS=700. The sample of n=10 traditional sex-typed participants averaged M=71 with SS=740. Do the data indicate that traditionally masculine and feminine people have significantly more depression than androgynous people?

7-A biopsychologist studies the role of the brain chemical serotonin in aggression. One sample of rats serves as a control group and receives a placebo. A second sample of rats receive a drug that lowers brain levels of serotonin. Then the researcher tests the animals by recording the number of aggressive responses each of the rats display. The data are presented below. Does the drug have a significant effect on aggression? Use an alpha level of .05, two tails.

Low Serotonin	Control
n = 6	n = 8
M = 22	M = 14
SS = 108	SS = 180

8-In a study examining the effects of environment on development, Krech and his colleagues (1962) divided a sample of infant rats into two groups One group was housed in a stimulus-rich environment containing ladders, platforms, tunnels and colorful decorations. The second group was housed in a stimulus-poor conditions consisting of plain gray cages. At maturity, maze-learning performance measured for all rats. The following hypothetical data stimulate Krechs's results. The mean for rich group was 26.0 with SS=214. For the poor group, M=34.2 and SS=313.61.

a. Does the data satisfy the homogeneity of variance assumption?

b. Do these data indicate a significant difference between the two groups? Test at the .01 level of significance

c. Estimate Cohen's d and compute r^2 for the test.

9-Friedman and Rosenman (1983) have classified people into two categories: Type A personalities and Type B personalities. Type As are hard-driving, competitive, and ambitious. Type Bs are more relaxed, easy-going people. One factor that differentiates these two groups is the chronically high level of frustration experience by Type As. To demonstrate this phenomenon, separate samples of Type As and Type Bs are obtained, with n=8 in each sample. The individual participants are all given a frustration inventory measuring level of frustration. The average score for Type As is M=84 with SS=740 and the Type Bs average M=71 with SS=660.

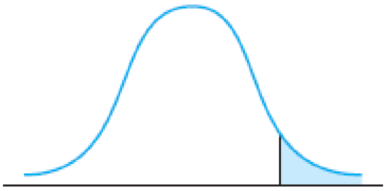
a. Does the data satisfy the homogeneity of variance assumption?

b. Do these data a significant difference between the two groups? Test at the .01 level of significance.

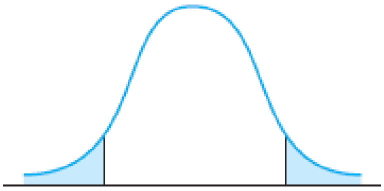
c. Estimate Cohen's d and compute r^2 for the test.

TABLE B.2 THE *t* DISTRIBUTION

Table entries are values of *t* corresponding to proportions in one tail or in two tails combined.



One tail
(either right or left)



Two tails
combined

	Proportion in One Tail					
	0.25	*0.10*	*0.05*	*0.025*	*0.01*	*0.005*
	Proportion in Two Tails Combined					
df	*0.50*	*0.20*	*0.10*	*0.05*	*0.02*	*0.01*
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
60	0.679	1.296	1.671	2.000	2.390	2.660
120	0.677	1.289	1.658	1.980	2.358	2.617
∞	0.674	1.282	1.645	1.960	2.326	2.576

Table III of Fisher, R. A., & Yates, F. (1974). *Statistical Tables for Biological, Agricultural and Medical Research* (6th ed.). London: Longman Group Ltd., 1974 (previously published by Oliver and Boyd Ltd., Edinburgh). Copyright ©1963 R. A. Fisher and F. Yates. Adapted and reprinted with permission of Pearson Education Limited.

TABLE B.3 CRITICAL VALUES FOR THE *F*-MAX STATISTIC*

*The critical values for α = .05 are in lightface type, and for α = .01, they are in boldface type.

	k = Number of Samples										
$n - 1$	2	3	4	5	6	7	8	9	10	11	12
4	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.4	44.6	48.0	51.4
	23.2	**37.**	**49.**	**59.**	**69.**	**79.**	**89.**	**97.**	**106.**	**113.**	**120.**
5	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
	14.9	**22.**	**28.**	**33.**	**38.**	**42.**	**46.**	**50.**	**54.**	**57.**	**60.**
6	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
	11.1	**15.5**	**19.1**	**22.**	**25.**	**27.**	**30.**	**32.**	**34.**	**36.**	**37.**
7	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
	8.89	**12.1**	**14.5**	**16.5**	**18.4**	**20.**	**22.**	**23.**	**24.**	**26.**	**27.**
8	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
	7.50	**9.9**	**11.7**	**13.2**	**14.5**	**15.8**	**16.9**	**17.9**	**18.9**	**19.8**	**21.**
9	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
	6.54	**8.5**	**9.9**	**11.1**	**12.1**	**13.1**	**13.9**	**14.7**	**15.3**	**16.0**	**16.6**
10	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
	5.85	**7.4**	**8.6**	**9.6**	**10.4**	**11.1**	**11.8**	**12.4**	**12.9**	**13.4**	**13.9**
12	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
	4.91	**6.1**	**6.9**	**7.6**	**8.2**	**8.7**	**9.1**	**9.5**	**9.9**	**10.2**	**10.6**
15	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
	4.07	**4.9**	**5.5**	**6.0**	**6.4**	**6.7**	**7.1**	**7.3**	**7.5**	**7.8**	**8.0**
20	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
	3.32	**3.8**	**4.3**	**4.6**	**4.9**	**5.1**	**5.3**	**5.5**	**5.6**	**5.8**	**5.9**
30	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
	2.63	**3.0**	**3.3**	**3.5**	**3.6**	**3.7**	**3.8**	**3.9**	**4.0**	**4.1**	**4.2**
60	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
	1.96	**2.2**	**2.3**	**2.4**	**2.4**	**2.5**	**2.5**	**2.6**	**2.6**	**2.7**	**2.7**

Table 31 of Pearson, E., and Hartley, H.O. (1958). *Biometrika Tables for Statisticians* (2nd ed.). New York: Cambridge University Press. Adapted and reprinted with permission of the Biometrika trustees.