

# An Introduction to Biostatistics

Third Edition

Additional Appendices

Thomas Glover

*Hobart and William Smith Colleges*

Kevin Mitchell

*Hobart and William Smith Colleges*

© 1999–2015 by Thomas Glover and Kevin Mitchell.

This material may be reproduced for classroom use with the text *An Introduction to Biostatistics*, Third Edition, published by Waveland Press. Please do not post this material on the internet. For all other uses, please obtain written permission of the authors.

C O N T E N T S

<b>D Tests of Hypothesis Involving Two Samples</b>	<b>1</b>
D.1 Confidence Intervals for $M_X - M_Y$ . . . . .	1
D.2 Problems . . . . .	4
D.3 Answers . . . . .	6
<b>E <math>k</math>-Sample Tests of Hypothesis: The Analysis of Variance</b>	<b>7</b>
E.1 The Bonferroni $t$ Test and the Student-Newman-Keuls Test . . . . .	7
E.2 Duncan's Multiple Range Test . . . . .	12
E.3 Problems . . . . .	16
E.4 Answers . . . . .	19
<b>F Two-Factor Analysis</b>	<b>26</b>
F.1 The Durbin Test: Balanced Incomplete Blocks . . . . .	26
F.2 Problems . . . . .	31
F.3 Answers . . . . .	35
<b>G Sampling Distributions</b>	<b>40</b>
G.1 Confidence Intervals for a Population Proportion . . . . .	40
G.2 Problems . . . . .	45
G.3 Answers . . . . .	47
<b>H Field Methods</b>	<b>51</b>
H.1 Estimating the Size of a Population with the Petersen Method . . . . .	51
H.2 The Point-Centered Quarter Method . . . . .	55
H.3 Diversity Indices . . . . .	68
H.4 Problems . . . . .	76
H.5 Answers . . . . .	83
<b>I Proofs of Selected Results</b>	<b>86</b>
I.1 The Method of Least Squares for Linear Regression . . . . .	86

<b>J</b>	<b>Tables of Distributions and Critical Values</b>	<b>90</b>
J.1	Table of Critical Values for the Wilcoxon Rank-Sum Test . . . . .	91
J.2	Cumulative Standard Normal Distribution . . . . .	95
J.3	Student's $t$ Distribution . . . . .	97
J.4	Cumulative Chi-Square Distribution . . . . .	100
J.5	Critical Values of the $q$ Statistic for the Student-Newman-Keuls Test . . . . .	102
J.6	Critical Values for Duncan's Multiple Range Test . . . . .	102

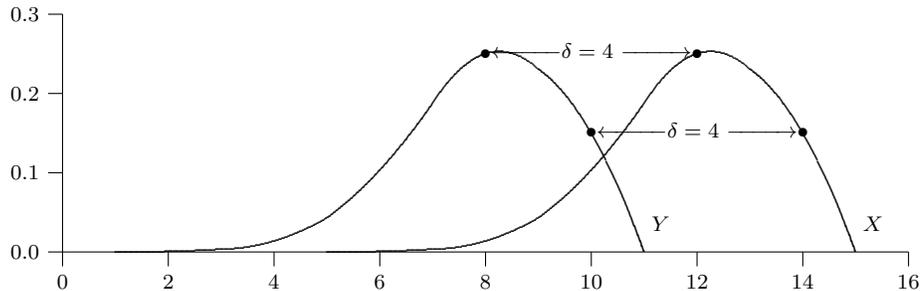
# Tests of Hypothesis Involving Two Samples

This material could be inserted between Sections 7.5 and 7.6 in the text *An Introduction to Biostatistics* by Glover and Mitchell (Waveland, 2015). It cross-references that text and follows the same format.

## D.1 Confidence Intervals for $M_X - M_Y$

Oftentimes we want to know not only whether two populations differ in some measurement of interest but by how much they differ. This is generally expressed as a confidence interval for the difference in their means or medians. The Wilcoxon rank-sum test can be adapted to calculate confidence intervals for the difference in the medians of two populations  $M_X - M_Y$ . The difference in the medians  $M_X - M_Y$  is not the same as the median difference  $M_{X-Y}$  in paired sample tests. The median difference makes sense only when observations are paired and one is trying to estimate  $X_i - Y_i$ . The difference of the medians makes sense when the samples are independent of each other where one is trying to estimate the difference  $X_i - Y_j$  for any  $i$  and  $j$ .

The basic assumption is that  $X = Y + \delta$ , that is, the shape of the distributions for  $X$  and  $Y$  are the same, but that the  $X$ 's are larger (or smaller) than the  $Y$ 's by some amount  $\delta$ . Hence the medians differ by the same amount,  $M_X - M_Y = \delta$ .



**FIGURE D.1.** These distributions have the same shape, but the distribution for  $X$  is shifted  $\delta = 4$  units to the right of the distribution for  $Y$ .

It is then natural to use the differences  $X_i - Y_j$  as the set of **elementary estimates** of  $\delta$ . We continue to use the conventions of the Wilcoxon rank-sum test with  $m$  denoting the  $X$  sample size  $n$  the  $Y$  sample size, where  $m \leq n$ . Since there are  $m$  choices for  $X_i$  and  $n$  choices for  $Y_j$ , there are  $mn$  differences in all for the elementary estimates.

**EXAMPLE D.1.** In a project for a botany class, 11 sunflower seeds were randomly assigned to and planted in pots. One group of pots served as a control and contained commercial potting soil. The other group of pots contained potting soil plus a fertilizer treatment. The height of each plant (in cm) two weeks after germination is recorded in the table below. List the elementary estimates of  $\delta$  for these data.

Fertilizer	29	32	35	36	37	
Control	22	25	26	27	28	32

**SOLUTION.** This is most easily done in tabular form with the  $X$ 's along the left edge and the  $Y$ 's along the top. It is convenient to order the observations in each sample. The difference  $X_i - Y_j$  is then put in row  $i$ , column  $j$  of the table.

$X_i - Y_j$	22	25	26	27	28	32
29	7	4	3	2	1	-3
32	10	7	6	5	4	0
35	13	10	9	8	7	3
36	14	11	10	9	8	4
37	15	12	11	10	9	5

In the example there are  $mn = 5 \cdot 6 = 30$  elementary estimates. As with confidence intervals based on the sign test, the next step is to eliminate a certain number of the most extreme estimates with the remaining estimates providing the range for the confidence interval. The endpoint depths for the confidence interval are listed in Table J.1. Calculating 95% confidence intervals is equivalent to two-tailed hypothesis testing at the  $\alpha = 0.05$  level. So locate the columns for a two-tailed test with  $\alpha = 0.05$ . Find the row with  $m = 5$  and  $n = 6$ . The table entry in the depth  $d$ -column is 4. Therefore, for a 95% confidence interval, the endpoints are the elementary estimates with a depth of  $d = 4$ . Removing the 3 smallest and 3 largest elementary estimates from the table, we find that a 95% confidence interval for the difference in median heights of the sunflowers is  $[2, 12]$ . In other words, we are 95% confident that  $X = Y + \delta$  where  $2 < \delta < 12$ . (How would you interpret the fact that 0 is not in this interval?)

## Theory

We now show how the endpoint depths in Table J.1 were derived from the values for  $W_X$ . Ninety-five percent confidence interval calculations and two-tailed hypothesis testing are complementary processes. To find a 95% confidence interval for the difference  $\delta = X - Y$  we must find the largest and smallest values of  $D$  such that when a two-sided hypothesis test  $H_0: X = Y + D$  versus  $H_a: X \neq Y + D$  is done, then  $H_0$  is *accepted*. For this example, with  $m = 5$  and  $n = 6$ , Table J.1 lists the pair (18,42) at the  $\alpha = 0.05$  level to *reject*  $H_0$ . Thus, to *accept*  $H_0: X = Y + D$ , we must find the minimum and maximum values of  $D$  so that  $W_X$  is *between* 18 and 42.

An alternative form of Wilcoxon rank-sum test, called the **Mann-Whitney  $U$ -test**, makes this easy to do. The elementary estimates correspond to all possible ways that the  $X$  and  $Y$  observations can be compared. The test statistic  $U$  is defined as the number of comparisons in which  $X_i$  is larger than  $Y_j$ . In other words,  $U$  is the number of positive entries in the table of differences  $X_i - Y_j$ . (Ties between  $X_i$  and  $Y_j$ , which appear as 0's in the difference table, are counted as 0.5.) For instance, in Example D.1,  $U = 28.5$  because there were 28 positive differences in the table and a 0 (indicating a tie).

The test statistic  $U$  is related to  $W_X$  in a simple way. Recall that  $W_X$  is the sum of the ranks of the  $X$  observations in the *combined*  $X$  and  $Y$  data set. To determine the *combined rank* of each  $X_i$ , we need to know the number of  $X$  and  $Y$  observations that are smaller than  $X_i$ . That is, we need the rank of  $X_i$  *within* the  $X$  sample and the number of  $Y$  observations smaller than  $X_i$ . In Example D.1,  $X_1 = 29$  is the smallest  $X$  observation, so its  $X$  rank is 1.

But it is larger than five  $Y$  observations. Thus,  $X_1 = 29$  has a rank of  $1 + 5 = 6$  in the combined set of  $X$  and  $Y$  observations. The table below shows how this works for the entire  $X$  sample.

$X_i$	29	32	35	36	37	Sum	
$X$ rank	1	2	3	4	5	15	$\leftarrow 1 + 2 + \dots + 5$
Number of $Y$ 's smaller than $X_i$	5	5.5	6	6	6	28.5	$\leftarrow U$
Combined rank	6	7.5	9	10	11	43.5	$\leftarrow W_X$

More generally, since there are  $m$  observations in the  $X$  sample, the  $X$  ranks will always sum to  $1 + 2 + \dots + m = \frac{m(m+1)}{2}$ . So the sum of the combined ranks is  $W_X = \frac{m(m+1)}{2} + U$  or

$$U = W_X - \frac{m(m+1)}{2},$$

where  $m$  is the smaller sample size. For the data in Example D.1,

$$U = W_X - \frac{m(m+1)}{2} = 43.5 - \frac{5(6)}{2} = 28.5,$$

the same as the number of positive entries in the difference table.

Earlier we said that to find the confidence interval, we need the minimum and maximum values of  $D$  so that  $H_0: X = Y + D$  is accepted, that is, so that  $W_X$  is *between* 18 and 42. Converting the  $W_X$  values to  $U$  values, since  $m = 5$ , then  $U$  is *between*  $18 - \frac{5(5+1)}{2} = 3$  and  $42 - \frac{5(5+1)}{2} = 27$ . But  $U$  is just the number of positive entries in the table of differences. So we must find the minimum value of  $D$  so that when  $D$  is added to each  $Y_j$ , the (new) table of differences contains more than 3 positive entries, i.e., at least 4 nonnegative entries. And we must find the maximum value for  $D$  so that when  $D$  is added to each  $Y_j$ , the table of differences contains fewer than 27 positive entries. Since there are  $mn = 30$  differences, fewer than 27 positive differences is equivalent to 4 or more nonpositive differences. That is why we focused on a depth of 4 in the original table to find the confidence interval endpoints.

Look back at the table of differences in Example D.1. The depth 4 values were 2 and 12. If we add  $D = 2$  to each  $Y$  value, then  $X_i - (Y_j + 2) = (X_i - Y_j) - 2$ . That is, each difference decreases by 2 (see the table below). Since 2 had depth 4 in the original table, now  $2 - D = 2 - 2 = 0$  has depth 4 in the new table. The table now contains 4 nonpositive entries. The corresponding  $U$  value is 26.5, for the 26 positive values and one 0 (reflecting the tie between the  $X$  and  $Y + D$  values) in the new table.

$X_i - (Y_j + 2)$	24	27	28	29	30	34
29	5	2	1	0	-1	-5
32	8	5	4	3	2	-2
35	11	8	7	6	5	1
36	12	9	8	7	6	2
37	13	10	9	8	7	3

Similarly, if we add  $D = 12$  to each  $Y$  value, then each difference from the original table is decreased by 12. Now the table contains 3 positive entries and a 0. The corresponding  $U$  value is 3.5.

$X_i - (Y_j + 12)$	34	37	38	39	40	44
29	-5	-8	-9	-10	-11	-15
32	-2	-5	-6	-7	-8	-12
35	1	-2	-3	-4	-5	-9
36	2	-1	-2	-3	-4	-8
37	3	0	-1	-2	-3	-7

Notice that the  $U$  values are just greater than 3 and just less than 27. Converting to  $W_X$  values by adding  $\frac{m(m+1)}{2} = 15$  to each, we see that  $18.5 \leq W_X \leq 41.5$ , just inside the range to accept  $H_0: X = Y + D$ .

Thus, we see that the depth values  $d$  listed in Table J.1 were derived by taking the lower critical value for  $W_X$ , converting to a lower critical value for  $U$  by subtracting  $\frac{m(m+1)}{2}$  and then adding 1 to move out of the “reject  $H_0$ ” range and into the “accept  $H_0$ ” range.

## D.2 Problems

- Return to Example 7.10 in the main text where we considered two color morphs (green and red) of the same species of sea star were found at Polka Point, North Stradbroke Island.

Red	108	64	80	92	40	
Green	102	116	98	132	104	124

- List the elementary estimates for these data.
  - find a 95% confidence interval for the difference  $\delta = X - Y$ .
- An undergraduate ecology student doing research on niche dimensions decided to repeat part of R. H. MacArthur’s famous study of foraging behavior of warblers in northeastern coniferous forests. She marked the heights of various branches in several conifers with colored tape and observed two similar species of warbler with binoculars and recorded the average foraging height for each bird. The heights in feet for the individuals observed are recorded below. Find a 95% confidence interval for the difference in *median* foraging heights.

Bay-breasted warbler	17	10	13	12			
Blackburnian warbler	15	16	18	19	23	24	26

- Interpret your confidence interval as a test of hypothesis about the difference in median foraging heights for the two species.
- The Southern cavefish, *Typhlichthys subterraneus*, is found mostly in the dolomite aquifers of the Salem Plateau in the South Central Ozarks. Populations in two different caves in central Missouri were sampled and the length of each individual was measured (in cm) and is recorded below. Find a 95% confidence interval for the difference in *median* lengths.

Cave X	3.8	6.6	3.2	3.9	4.5	6.8	2.8
Cave Y	3.0	3.5	3.8	3.9	4.3	4.6	6.0

- Interpret your confidence interval as a test of hypothesis about the difference in median lengths for the two populations.

4. In order for *Banksia serrata* to germinate, fire is required to open its seed cones. In order to understand the effect of fire on these trees, two neighboring sites were surveyed. Site A has been burned frequently during the last 80 years, while fire at Site B has been infrequent during this period. It is hypothesized that the different frequency of fire will result in larger median DBH for *B. serrata* (in cm) at location B than at location A. Find a 95% confidence interval for the difference in median DBH for the two locations. Interpret your result.

Site A	12	16	20	26	30	32	36	46
Site B	28	34	44	48	48	50		

### D.3 Answers

1. (a) Order the observations in each sample. The difference  $X_i - Y_j$  is then put in row  $i$ , column  $j$  of the table.

$X_i - Y_j$	98	102	104	116	124	132
40	-58	-62	-64	-76	-84	-92
64	-34	-38	-40	-52	-60	-68
80	-18	-22	-24	-36	-44	-52
92	-6	-10	-12	-24	-32	-40
108	10	6	4	-8	-16	-24

- (b) Locate the columns in Table J.1 for a two-tailed test with  $\alpha = 0.05$ . Find the row with  $m = 5$  and  $n = 6$ . The table entry under the depth  $d$ -column reads 4. Therefore, for a 95% confidence interval, the endpoints are the elementary estimates with a depth of  $d = 4$ . Thus, a 95% confidence interval for the difference in median lengths of the sea stars is  $[-68, -6]$ .
2. (a) The table of differences is

$X_i - Y_j$	15	16	18	19	23	24	26
10	5	6	8	9	13	14	16
12	3	4	6	7	11	12	14
13	2	3	5	6	10	11	13
17	-2	-1	1	2	6	7	9

From Table J.1 with  $m = 4$  and  $n = 7$ , the endpoint depth is 4. Therefore, the confidence interval is  $[2, 13]$ .

- (b) Since 0 is not in the interval, there is a significant difference in foraging heights.
3. (a) The table of differences is

$X_i - Y_j$	3.0	3.5	3.8	3.9	4.3	4.6	6.0
2.8	-0.2	-0.7	-1.0	-1.1	-1.5	-1.8	-3.2
3.2	0.2	-0.3	-0.6	-0.7	-1.1	-1.4	-2.8
3.8	0.8	0.3	0.0	-0.1	-0.5	-0.8	-2.2
3.9	0.8	0.3	0.0	-0.1	-0.5	-0.8	-2.2
4.5	1.5	1.0	0.7	0.6	0.2	-0.1	-1.5
6.6	3.6	3.1	2.8	2.7	2.3	2.0	0.6
6.8	3.8	3.3	3.0	2.9	2.5	2.2	0.8

From Table J.1 with  $m = 7$  and  $n = 7$ , the endpoint depth is 9. Therefore, the confidence interval is  $[-1.1, 2.5]$ .

- (b) Since 0 is in the interval, there is no evidence for a significant difference in lengths of cavefish at the two locations.
4. For a 95% confidence interval for differences in median DBH for the two locations use  $d = 8$  as the endpoint depth. From the table of differences below, the 95% confidence interval for  $B - A$  is  $[2, 30]$ . It does not contain 0 so there is evidence of a significant difference in DBH for the two locations.

$B - A$	12	16	20	26	30	32	36	46
28	16	12	8	2	-2	-4	-8	-18
34	22	18	14	8	4	2	-2	-12
44	32	28	24	18	14	12	8	-2
48	36	32	28	22	18	16	12	2
48	36	32	28	22	18	16	12	2
50	38	34	30	24	20	18	14	4

# *k*-Sample Tests of Hypothesis: The Analysis of Variance

This material follows naturally as third and fourth subsections of Section 8.2 (Mean Separation Techniques for Model I ANOVAs) in the text *An Introduction to Biostatistics*, Third Edition, by Glover and Mitchell (Waveland, 2015). It cross-references that text and follows the same format. The material is repetitive because each section is meant to stand alone. One would not ordinarily cover five different mean separation techniques.

## E.1 The Bonferroni $t$ Test and the Student-Newman-Keuls Test

When  $H_0$  is rejected as in Examples 8.1 and 8.2, the conclusion at that point is *at least one pair of means is not equal*. The task at hand is to discover exactly which means are different. The techniques to do this are called **mean separation techniques** or multiple comparisons. There are many of these techniques available, each with its own advantages and disadvantages. In Chapter 8 we presented two of the more common approaches: the **Bonferroni-Holm  $t$  test** and the Tukey test. We now consider two additional approaches: **Bonferroni  $t$  test** and the **Student-Newman-Keuls test**. There are some similarities to the earlier tests and a researcher would choose one of these techniques to employ at the time the experiment is planned.

In Example 8.1 there are three sample means, and so there are

$$\binom{3}{2} = \frac{3!}{2!1!} = 3$$

possible pairwise mean comparisons:

$$\begin{array}{lll} H_0: \mu_U = \mu_{90\%} & H_0: \mu_U = \mu_{80\%} & H_0: \mu_{90\%} = \mu_{80\%} \\ H_a: \mu_U \neq \mu_{90\%} & H_a: \mu_U \neq \mu_{80\%} & H_a: \mu_{90\%} \neq \mu_{80\%}. \end{array}$$

### The Bonferroni $t$ Test

The two-sample  $t$  test presented in Chapter 7 with the assumption of equal variances had test statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{with } \nu = n_1 + n_2 - 2.$$

We used  $s_p^2$ , a pooled variance, because we assumed equal population variances based on the preliminary  $F$  test. The assumptions about one-way ANOVAs include equal variances for *all* populations under investigation. The estimate of this variance is  $MS_{\text{Error}}$  in the ANOVA table. The Bonferroni  $t$  test uses this fact to generate the following test statistic.

**FORMULA E.1.** The Bonferroni  $t$  test statistic for Model I ANOVAs is

$$\text{Bonferroni } t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

where  $i$  and  $j$  represent any two treatment means to be compared and the degrees of freedom  $\nu$  is  $N - k$ , where  $N$  is the total number of observations in the experiment and  $k$  is the number of treatments. That is,  $\nu$  is the degrees of freedom of  $\text{MS}_{\text{Error}}$ .

**Note:** The Bonferroni  $t$  test statistic is the same as the Bonferroni-Holm  $t$  test statistic (compare to Formula 8.6 in the main text).

Bonferroni  $t$  tests are usually run at a smaller  $\alpha$  level than the global  $F$  test of the ANOVA table. This is necessary because we need to manage the probability of Type I error. If all Bonferroni  $t$  tests are performed at  $\alpha$ , then the overall probability of at least one Type I error, denoted by  $\alpha'$ , is larger than  $\alpha$  and its value is usually unknown. It can be shown that with three tests conducted at the  $\alpha$  level,  $\alpha'$  is at most  $1 - (1 - \alpha)^3$ . For example, if  $\alpha = 0.05$  and we do three Bonferroni  $t$  tests, the probability of making at least one Type I error becomes at most  $1 - (1 - .05)^3 = 0.143$ . Using  $\alpha = 0.05$  with six comparisons,  $\alpha' = 0.265$  and with 10 comparisons,  $\alpha' = 0.401$ . So as the number of means to be separated increases, the overall probability of a Type I error can quickly become unacceptably high.

Most Bonferroni  $t$  tests are, therefore, conducted at an  $\alpha$  level lower than the global  $F$  test. A general rule is to determine an *experimentwise* acceptable upper boundary for the probability of Type I error, say,  $b$ , and divide this probability by  $c$ , the actual number of comparisons run, to determine the  $\alpha$  level for each comparison. For Example 8.1 if we wished the experimentwise  $\alpha'$  to be 0.05, we would use  $\frac{0.05}{3} = 0.01\bar{6}$  as the  $\alpha$  level for each Bonferroni  $t$  test.

**SOLUTION TO EXAMPLE 8.1: Bonferroni Mean Separation.** Let's now complete the analysis for Example 8.1 via Bonferroni  $t$  tests with experimentwise  $\alpha' = 0.05$ , so that  $\frac{0.05}{3} = 0.01\bar{6}$  is the  $\alpha$  level for each test. Recall from earlier calculations that the ordered means are

$\bar{X}_U = 2.44$	$\bar{X}_{90\%} = 3.18$	$\bar{X}_{80\%} = 3.54$
$n_U = 5$	$n_{90\%} = 5$	$n_{80\%} = 5$
$\text{MS}_{\text{Error}} = 0.203$	$\nu = N - k = 12$	

Since all the sample sizes are equal, the denominator in the expression for the Bonferroni  $t$  statistic is the same for all tests,

$$\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} = \sqrt{0.203 \left( \frac{1}{5} + \frac{1}{5} \right)} = 0.285.$$

From Table C.4 the critical  $t$  value is  $t_{12, 0.01\bar{6}} = 2.779$ .

For the first comparison:  $H_0: \mu_{80\%} - \mu_U$ . From Formula E.1, the test statistic is

$$\text{Bonferroni } t = \frac{\bar{X}_{80\%} - \bar{X}_U}{\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_{80\%}} + \frac{1}{n_U} \right)}} = \frac{3.54 - 2.44}{0.285} = 3.86 > 2.779.$$

Reject  $H_0$ . The mean for 80% diet is *significantly different* (higher) than the unlimited diet.

For the second comparison:  $H_0: \mu_{80\%} - \mu_{90\%}$ .

$$\text{Bonferroni } t = \frac{3.54 - 3.18}{0.285} = 1.263 < 2.779.$$

Accept  $H_0$ . The mean for 80% diet is *not significantly different* than the 90% diet.

For the third comparison:  $H_0: \mu_{90\%} - \mu_U$ .

$$\text{Bonferroni } t = \frac{3.18 - 2.44}{0.285} = 2.596 < 2.779.$$

Accept  $H_0$ . The mean for 90% diet is *not significantly different* than the unlimited diet.

In summary, the Bonferroni  $t$  tests indicate that the mean for unlimited diet is *significantly* different from the 80% diet. The unlimited diet is not significantly different from the 90% diet and the 90% and 80% diets are not significantly different from each other. While the difference between the unlimited and 90% diets is not significant statistically, it is large enough that further study with larger samples might produce data that will allow rejection of  $H_0$ . We summarize these results with the superscripts  $a$  and  $b$ :

U	90%	80%
2.44 <sup>a</sup>	3.18 <sup>a,b</sup>	3.54 <sup>b</sup>

## The Student-Newman-Keuls Test

We now present a second method called the **Student-Newman-Keuls test (SNK)** that is used to separate means in a fixed treatment or Model I ANOVA. This test uses the rank orders of the sample means to determine the sequence in which different means are compared. This test is only slightly more complicated to apply than the Tukey test discussed in Section 8.2. It involves the following protocol.

**Step 1.** Linearly order the  $k$  sample means from smallest to largest.

**Step 2.** The Student-Newman-Keuls test uses the so-called  $q$ -statistic to compare means, that is described in the next step. Use Table J.5 to determine the  $q$  critical value at the selected  $\alpha$ -level for the the number of means in the range being investigated and for the degrees of freedom of  $MS_{\text{Error}}$  in the ANOVA.

**Step 3.** A convenient way to proceed through the comparisons is to compare the largest mean with the smallest, then the largest with the next smallest, and so on until the largest mean has been compared to all means. Then compare the second largest mean to the smallest, and then the second largest with the next smallest, and so on. For example, we would compare four means (ranked 1 through 4) as follows: 4 vs. 1, 4 vs. 2, 4 vs. 3, 3 vs. 1, 3 vs. 2, and 2 vs. 1. To compare means  $i$  and  $j$  in this rank order ( $H_0: \mu_i = \mu_j$ ) calculate the  $q$ -statistic

**FORMULA E.2.** The  $q$ -statistic for use in the Student-Newman-Keuls mean separation tests is defined as

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{Error}}}{2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

where  $MS_{\text{Error}}$  comes from the ANOVA table and  $n_i$  and  $n_j$  are the treatment sample sizes. When all the sample sizes are equal to  $n$ , this simplifies to

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{Error}}}{n}}}$$

Always subtract the larger mean from the smaller mean so that the  $q$ -statistic is positive. If  $q$  is larger than the critical value from Step 2, reject  $H_0$ : The means are different. Otherwise accept  $H_0$ . **Note:** The  $q$ -statistic is calculated exactly as for the Tukey test (see Formula 8.7 in the main text), however, the  $q$  critical values change depending on the number of means in the range being compared. (See the example below.)

We now analyze the means in Example 8.1 with the Student-Newman-Keuls test (SNK). In real life we would choose either the Bonferroni's  $t$  test or the SNK test, but not both. Each test is attempting the same thing, to separate means that are significantly different.

**SOLUTION TO EXAMPLE 8.1: SNK Mean Separation.** We use the procedure outlined above.

**Step 1.** For Example 8.1 the ordered means are

$\bar{X}_U = 2.44$	$\bar{X}_{90\%} = 3.18$	$\bar{X}_{80\%} = 3.54$
$n_U = 5$	$n_{90\%} = 5$	$n_{80\%} = 5$
$MS_{\text{Error}} = 0.203$	$\nu = N - k = 12$	

**Step 2.** Let  $\alpha = 0.05$ . Using Table J.5 with  $\nu = 12$  (the degrees of freedom of  $MS_{\text{Error}}$ ) and  $k = 3$  means to compare, the  $q$  critical value for comparing a range of three means (the smallest to the largest) is  $q_{(0.05,12,3)} = 3.773$  and the  $q$  critical value for comparing a range of two means (adjacent means) is  $q_{(0.05,12,2)} = 3.083$ .

**Step 3.** In this example  $n = 5$  is the common treatment sample size so the denominator of all the  $q$ -statistics will be

$$\sqrt{\frac{MS_{\text{Error}}}{n}} = \sqrt{\frac{0.203}{5}} = 0.201.$$

Now use the ordering from Step 1 and then proceed through the comparisons as outlined in the general description of Step 3 on page 9.

(a) Compare the 80% diet to unlimited diet,  $H_0: \mu_{80\%} = \mu_U$ , a range of three means.

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{Error}}}{n}}} = \frac{3.54 - 2.44}{0.201} = 5.473 > 3.773.$$

Reject  $H_0$ . The means the for 80% and unlimited diets are significantly different.

(b) Compare the 80% diet to the 90% diet,  $H_0: \mu_{80\%} = \mu_{90\%}$ , a range of two means.

$$q = \frac{3.54 - 3.18}{0.201} = 1.791 < 3.083.$$

Accept  $H_0$ . The means the for 80% and 90% diets are not significantly different.

(c) Compare the 90% diet to the unlimited diet,  $H_0: \mu_{90\%} = \mu_U$ .

$$q = \frac{3.18 - 2.44}{0.201} = 3.682 > 3.083.$$

Reject  $H_0$ . The means the for 90% and unlimited diets are significantly different. Overall, only the 80% and 90% diets differ from the unlimited diet. This result is

<b>U</b>	<b>90%</b>	<b>80%</b>
2.44 <sup>a</sup>	3.18 <sup>b</sup>	3.54 <sup>b</sup>

That the unlimited diet is significantly different from both of the restricted diets (90% and 80%) is indicated by the different superscripts used. That the 90% and 80% diets are not significantly different is reflected in the use of the same superscript.

The outcome with the SNK test is slightly different from the Bonferroni  $t$  test because we were quite conservative in our choice of  $\alpha$  for the Bonferroni statistics. *Both* analyses indicate that the “lean and mean” rats live longer. In general the Bonferroni  $t$  tests will be more conservative than the SNK test if both are applied at the same experimentwise alpha level. By more conservative we mean it will be more difficult to reject  $H_0$  and will produce more Type II errors.

**SOLUTION TO EXAMPLE 8.2: SNK Mean Separation.** We now complete the analysis of the field study on the cephalon lengths of trilobites with a Student-Newman-Keuls test.

**Step 1.** The ordered means are

Site	IV	I	III	II
$\bar{X}$	5.0	7.0	8.0	9.0
$n$	15	8	11	5
	$MS_{\text{Error}} = 4.71$			

**Step 2.** Let  $\alpha = 0.05$ . Use Table J.5 with  $\nu = 35$  (the degrees of freedom of  $MS_{\text{Error}}$ ) to obtain the  $q$  critical values: Since 35 degrees of freedom isn't in the table, use the nearest *smaller* value, 30, to be conservative. Check that for a range of four means  $q_{(0.05,30,4)} = 3.845$ , for a range of three means  $q_{(0.05,30,3)} = 3.486$ , and for a range of two means  $q_{(0.05,30,2)} = 2.888$ .

**Step 3.** As with many field studies the sample sizes are unequal simply because all the fossils found were measured and different numbers were found at different sites. This requires care in computing the  $q$  statistic. Use the ordering from Step 1 and then proceed through the comparisons as outlined in the general description of Step 3 on page 9.

(a) Compare site II to site IV,  $H_0: \mu_{\text{II}} = \mu_{\text{IV}}$ , a range of four means.

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{Error}}}{2} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} = \frac{9.0 - 5.0}{\sqrt{\frac{4.71}{2} \left( \frac{1}{5} + \frac{1}{15} \right)}} = 5.048 > 3.845.$$

Reject  $H_0$ . Mean cephalon lengths at sites II and IV are significantly different.

(b) Compare site II to site I,  $H_0: \mu_{\text{II}} = \mu_{\text{I}}$ , a range of three means.

$$q = \frac{9.0 - 7.0}{\sqrt{\frac{4.71}{2} \left( \frac{1}{5} + \frac{1}{8} \right)}} = 2.286 < 3.486.$$

Accept  $H_0$ . The mean cephalon lengths at sites II and I are not significantly different.

(c) Compare site II to site III,  $H_0: \mu_{\text{II}} = \mu_{\text{III}}$ , a range of two means.

$$q = \frac{9.0 - 8.0}{\sqrt{\frac{4.71}{2} \left( \frac{1}{5} + \frac{1}{11} \right)}} = 1.208 < 2.888.$$

Accept  $H_0$ . The mean cephalon lengths at sites II and III are not significantly different.

(d) Compare site III to site IV,  $H_0: \mu_{III} = \mu_{IV}$ , a range of three means

$$q = \frac{8.0 - 5.0}{\sqrt{\frac{4.71}{2} \left( \frac{1}{11} + \frac{1}{15} \right)}} = 4.925 > 3.486.$$

Reject  $H_0$ . Mean cephalon lengths at these sites are significantly different.

(e) Compare site III to site I,  $H_0: \mu_{III} = \mu_I$ , a range of two means.

$$q = \frac{8.0 - 7.0}{\sqrt{\frac{4.71}{2} \left( \frac{1}{11} + \frac{1}{8} \right)}} = 1.402 < 2.888.$$

Therefore, mean cephalon lengths at sites III and I are not significantly different, either.

(f) Compare site I to site IV,  $H_0: \mu_I = \mu_{IV}$ , a range of two means.

$$q = \frac{7.0 - 5.0}{\sqrt{\frac{4.71}{2} \left( \frac{1}{8} + \frac{1}{15} \right)}} = 2.979 > 2.888.$$

Reject  $H_0$ . Mean cephalon lengths at these sites I and IV significantly different.

From these comparisons we find that site IV has a *significantly smaller mean* than the other sites, while sites I, II, and III are *not significantly different*, as is summarized below.

<b>IV</b>	<b>I</b>	<b>III</b>	<b>II</b>
5.0 <sup>a</sup>	7.0 <sup>b</sup>	8.0 <sup>b</sup>	9.0 <sup>b</sup>

This may mean that site IV represents a different species of trilobite. The biological significance must be added after determination of the statistical significance. Note also the difference of 2 mm between means IV and I is significant, but not between means II and I. The larger sample size at site IV is principally the cause for this discrepancy.

In both Examples 8.1 and 8.2, the global  $F$  tests indicated some differences in means. The mean separation techniques of Bonferroni  $t$  tests or Student-Newman-Keuls tests allowed us to pinpoint the differences.

## E.2 Duncan's Multiple Range Test

When  $H_0$  is rejected as in Examples 8.1 and 8.2, the conclusion at that point is *at least one pair of means are not equal*. The task at hand is to discover exactly which means are different. The techniques to do this are called mean separation techniques or multiple comparisons. There are many of these techniques available, each with its own advantages and disadvantages. In Section 8.2 of the text we described two of these: the **Bonferroni-Holm  $t$  test** and the **Tukey test**. In Section G.1 we described two additional mean separation methods: the **Bonferroni  $t$  test** and the **Student-Newman-Keuls test**. In this section we present another such test, the **Duncan's multiple range test**. In Example 8.1 there are three sample means, and so there are

$$\binom{3}{2} = \frac{3!}{2!1!} = 3$$

possible pairwise comparisons.

$$\begin{array}{lll} H_0: \mu_U = \mu_{90\%} & H_0: \mu_U = \mu_{80\%} & H_0: \mu_{90\%} = \mu_{80\%} \\ H_a: \mu_U \neq \mu_{90\%} & H_a: \mu_U \neq \mu_{80\%} & H_a: \mu_{90\%} \neq \mu_{80\%} \end{array}$$

The Duncan's multiple range test uses the rank orders of the sample means to determine the *shortest significant range* or  $SSR_p$ . Any two means that differ by more than this value are considered significantly different. This test is rather easy to apply and is among the oldest mean separation techniques.

**Step 1.** Linearly order the  $k$  sample means from smallest to largest.

**Step 2.** Calculate the shortest significant range ( $SSR_p$ ) utilizing

**FORMULA E.3.**

$$SSR_p = r_p \sqrt{\frac{(n_i + n_j)MS_{\text{Error}}}{2n_i n_j}},$$

where

- $r_p$  is a table value from Table J.6,
- $MS_{\text{Error}}$  is the error mean square from the ANOVA Table,
- $n_i$  and  $n_j$  are the sample sizes of the treatments being compared,
- $\nu = N - k$  is the degrees of freedom for the  $MS_{\text{Error}}$  used in the ANOVA table.

Note: When all of the treatment sample sizes are equal to  $n$ , the formula simplifies to

$$SSR_p = r_p \sqrt{\frac{MS_{\text{Error}}}{n}}.$$

**Step 3.** For any subset of  $p$  sample means  $2 \leq p \leq k$ , compare the absolute value of the difference of their means (range) to the appropriate  $SSR_p$  in Table J.6. If the range of the means under consideration is greater than the  $SSR_p$ , the population means are considered significantly different. A convenient way to proceed through the comparisons is to compare the largest mean with the smallest, then the largest with the next smallest, and so on until the largest mean has been compared to all means. Then compare the second largest mean to the smallest, and then the second largest with the next smallest, and so on. For example, we would compare four means (ranked 1 through 4) as follows: 4 vs. 1, 4 vs. 2, 4 vs. 3, 3 vs. 1, 3 vs. 2, and 2 vs. 1. This ensures that the differences are always positive.

**EXAMPLE E.1.** Analyze the means in Example 8.1 with Duncan's multiple range test (DMRT). In real life one would choose in advance either the Bonferroni-Holm  $t$  test, the Tukey test, Bonferroni's  $t$  test, the Student-Newman-Keuls test, or DMRT, but not more than one of these. Each test is attempting to separate means.

**SOLUTION.** Use the process outlined above.

**Step 1.** For Example 8.1 the means are already rank-ordered.

$\bar{X}_U = 2.44$	$\bar{X}_{90\%} = 3.18$	$\bar{X}_{80\%} = 3.54$
$n_U = 5$	$n_{90\%} = 5$	$n_{80\%} = 5$
$MS_{\text{Error}} = 0.203$	$\nu = N - k = 12$	

**Step 2.** Here  $\alpha = 0.05$ , and the common sample size is  $n = 5$ . So

$$\sqrt{\frac{MS_{\text{Error}}}{n}} = \sqrt{\frac{0.203}{5}} = 0.201$$

with  $\nu = 12$ . Use Formula E.3 and Table J.6 to calculate  $SSR_p$  in the table below.

	Number of means	
	2	3
$r_p$	3.082	3.225
$SSR_p = r_p \sqrt{\frac{MS_{\text{Error}}}{n}}$	0.619	0.648

**Step 3.** Start by comparing the largest to the smallest mean.

- (a) Compare the 80% diet to unlimited diet,  $H_0: \mu_{80\%} = \mu_U$ , a range of three means.

$$|3.54 - 2.44| = 1.10.$$

Here  $SSR_p = 0.648$  for a range of three means. Since  $1.10 > 0.648$ , the two means are considered *significantly different*. Reject  $H_0$ .

- (b) Compare the 80% diet to the 90% diet,  $H_0: \mu_{80\%} = \mu_{90\%}$ , a range of two means.

$$|3.54 - 3.18| = 0.36.$$

The two means are adjacent in the ordering so  $SSR_p = 0.619$ . Since  $0.36 < 0.619$ , the two means are *not* considered significantly different. Accept  $H_0$ .

- (c) Compare the 90% diet to the unlimited diet,  $H_0: \mu_{90\%} = \mu_U$ , a range of two means.

$$|3.18 - 2.44| = 0.74 > 0.619 = SSR_p.$$

Reject  $H_0$ . The means for 90% and unlimited diets are significantly different. We summarize these results with the superscripts  $a$  and  $b$ .

$$2.44^a \quad 3.18^b \quad 3.54^b.$$

That the unlimited diet is significantly different from both of the restricted diets (90% and 80%) is indicated by the different superscripts used. That the 90% and 80% diets are not significantly different is reflected in the use of the same superscript. The outcome with the DMRT is slightly different from the Bonferroni  $t$  test because we were quite conservative in our choice of  $\alpha$  for the Bonferroni statistics. *Both* analyses indicate that the “lean and mean” rats live longer.

**EXAMPLE E.2.** Complete the analysis of Example 8.2, a field study on the cephalon lengths of trilobites using DMRT.

**SOLUTION.** As with many field studies the sample sizes are unequal simply because all the fossils found were measured and different numbers were found at different sites. The overall ANOVA indicated that some of the means are significantly different. We now look for specific differences with a DMRT.

**Step 1.** For Example 8.2 the ordered means are

Site	IV	I	III	II
$\bar{X}$	5.0	7.0	8.0	9.0
$n$	15	8	11	5
	$MS_{\text{Error}} = 4.71$			

**Step 2.** Let  $\alpha = 0.05$ . Use Table J.6 with  $\nu = 35$  to obtain the values of  $r_p$  below. Since 35 degrees of freedom isn't in Table J.6, use the next *lower* value, in this case 30. All four sample sizes are different requiring a new calculation for each  $SSR_p$ .

	Number of means		
	2	3	4
$r_p$	2.888	3.035	3.131

**Step 3.** Start with the largest and smallest means. We briefly indicate the calculations involved.

- (a) To compare II to IV, a range of four means, first calculate  $SSR_p$  using Formula E.3.

$$SSR_p = r_p \sqrt{\frac{(n_i + n_j)MS_{\text{Error}}}{2n_i n_j}} = 3.131 \sqrt{\frac{(5 + 15)4.71}{2(5)(15)}} = 2.481.$$

Since

$$|\bar{X}_i - \bar{X}_j| = |9.0 - 5.0| = 4.0 > 2.481 = SSR_p,$$

the means are significantly different.

- (b) Next compare II to I, a range of three means.

$$SSR_p = 3.035 \sqrt{\frac{(5 + 8)4.71}{2(5)(8)}} = 2.655.$$

This time

$$|\bar{X}_i - \bar{X}_j| = |9.0 - 7.0| = 2.0 < 2.655 = SSR_p,$$

so the means are not significantly different.

- (c) To compare II to III, a range of two means,

$$SSR_p = 2.888 \sqrt{\frac{(5 + 11)4.71}{2(5)(11)}} = 2.390.$$

Here

$$|\bar{X}_i - \bar{X}_j| = |9.0 - 8.0| = 1.0 < 2.390 = SSR_p,$$

so the means are not significantly different.

- (d) To compare III to IV, a range of three means.

$$SSR_p = 3.035 \sqrt{\frac{(11 + 15)4.71}{2(11)(15)}} = 1.849.$$

Since

$$|\bar{X}_i - \bar{X}_j| = |8.0 - 5.0| = 3.0 > 1.849 = SSR_p,$$

the means are significantly different.

- (e) To compare III to I, a range of two means,

$$SSR_p = 2.888 \sqrt{\frac{(11 + 8)4.71}{2(11)(8)}} = 2.059.$$

Here

$$|\bar{X}_i - \bar{X}_j| = |8.0 - 7.0| = 1.0 < 2.059 = SSR_p,$$

so the means are not significantly different.

(f) Finally, to compare I to IV, a range of two means,

$$SSR_p = 2.888 \sqrt{\frac{(8 + 15)4.71}{2(8)(15)}} = 1.940.$$

Since

$$|\bar{X}_i - \bar{X}_j| = |7.0 - 5.0| = 2.0 > 1.940 = SSR_p,$$

the means are significantly different.

From these comparisons we find that site IV has a *significantly smaller mean* than the other sites, while sites I, II, and III are *not significantly different*, as is summarized below.

Site			
IV	I	III	II
5.0 <sup>a</sup>	7.0 <sup>b</sup>	8.0 <sup>b</sup>	9.0 <sup>b</sup>

This result may mean that site IV represents a different species. The biological significance must be layered upon the statistical significance. Note also the difference of 2 mm between means IV and I is significant, but not between means I and II. The larger sample size in site IV is principally the cause for this discrepancy.

### E.3 Problems

- To test the effectiveness of various denture adhesives, an instrument called the TA.XT2i Texture Analyzer made by Texture Technologies Corp. was used to measure the amount of force required to separate dentures from a mouth and gum cast. The force required for separation was recorded in decigrams. The adhesives were: (A) karaya, (B) karaya with sodium borate, (C) carboxymethylcellulose sodium (32%) and ethylene oxide homopolymer, and (D) carboxymethylcellulose sodium (49%) and ethylene oxide homopolymer.

A	B	C	D	
71	76	75	80	
79	70	81	82	
80	90	60	91	
72	80	66	95	
88	76	74	84	
66	82	58	72	
$T_i$	456	474	414	504
$\bar{X}_i$	76	79	69	84
$s_i^2$	62	46	83.2	66.8
CSS <sub>i</sub>	310	230	416	334

- Are there any significant differences among these denture adhesives holding abilities? Analyze with ANOVA.
- If mean separation is need use DMRT.
- Alternatively, carry out mean comparisons with a Student-Newman-Keuls  $t$  test, if necessary.

2. Workers at a tree farm decided to test the efficacy of three fertilizer mixtures on the growth of Norway maple seedlings, *Acer platanoides*. The table below contains the heights of seedlings (in feet) for the three fertilizer treatments.

	Fertilizer mixture		
	A	B	C
	2.0	2.3	3.1
	2.1	2.9	1.5
	2.4	1.5	2.2
	2.8	1.2	2.9
	2.9	1.9	1.7
	3.1	1.9	2.1
	3.2	3.4	2.8
	3.7	2.1	1.5
	3.8	2.6	2.8
	4.1	2.4	2.2
$T_i$	30.1	22.2	22.8
CSS $_i$	4.61	3.82	3.20
$s_i$	0.72	0.61	0.60

$$\sum_i \sum_j X_{ij}^2 = 203.49, \quad T_{..} = 75.1, \quad SS_{\text{Total}} = 15.49$$

- (a) Determine if there are significant difference in the heights among the three treatments.
- (b) If mean separation is required use either DMRT or SNK.
3. Current fitness industry trends include a growing interest in non-competitive boxing programs. A study was conducted to examine the physiological responses of punching at various tempos. Using commercially available boxing equipment, twelve male boxing-trained subjects performed straight left and right punches at four different tempos (60, 80, 100, and 120 punches per minute). Each trial lasted two minutes, similar to that of traditional boxing-round training. Oxygen consumption (ml/kg/min), and heart rate (bpm) were monitored continuously during each round using open circuit spirometry and telemetry. Summary data are recorded below. (Based on concepts in L. Kravitz, et al. [www.unm.edu/~lkravitz/Research%20Abstrs/punchingtempo.html](http://www.unm.edu/~lkravitz/Research%20Abstrs/punchingtempo.html).)

Punches/min	60	80	100	120	Total
$n_i$	12	12	12	12	48
$\bar{X}_i$	25.2	28.0	28.3	29.9	
$s_i$	3.2	3.4	3.0	3.1	
$T_i$	302.4	336	339.6	358.8	1336.8

- (a) Use ANOVA to determine whether there is a difference in oxygen consumption among the various tempos.
- (b) If so, which pairs are different? Use either Bonferroni  $t$  tests or DMRT.
4. An instructor in a second-term calculus course wishes to determine whether the year in college has any effect on the performance of his students on their final exam. The table below lists the exam grade (out of 150) for students categorized by year. Are there significant differences in performance among years? Analyze with ANOVA and Bonferroni  $t$  tests, if necessary.

First year	Sophomore	Junior
122	117	72
111	97	91
104	113	71
118	123	72
113	130	96
98		121
129		
111		
127		

5. For many years practitioners of traditional Chinese medicine have advocated oolong tea consumption for weight control. Researchers worked with a strain of adult female rats that spontaneously become obese on a normal diet. For 10 weeks, the researchers let the animals eat all they wanted but laced the chow of some with a dried extract of brewed oolong tea. All the animals ate about the same amount of food. Group I got 2% of their food as tea extract by weight. Group II got 4% of their food as tea extract by weight. Group III didn't have any tea extract in their chow. Data are weight gains in grams with 15 rats per group.

Group	I	II	III
$\bar{X}_i$	40	20	120
CSS <sub><i>i</i></sub>	1400	1120	1890

Analyze the data with ANOVA techniques. If mean comparisons are required, use SNK.

6. In a study on brain development in the Norway rat various diets were compared for their effects on cortex weight. Diet one consisted of commercial lab rat chow. Diet two was the commercial rat chow augmented with vitamins and minerals. Diet three consisted of fresh grains and vegetables. Diet four was similar to three except a 2% alcohol solution replaced the normal drinking water. Twenty two-week old rats were randomly assigned to one of the diets; all other conditions remained the same. After four months the rats were sacrificed and their cortex weights recorded in milligrams.

Cortex weight (mg)			
Diet I	Diet II	Diet III	Diet IV
650	660	700	675
670	700	710	700
665	680	690	660
680	690	725	685
685	695	715	675

Did diet affect the cortex weights? Analyze with one-way ANOVA. Use Bonferroni  $t$  tests if mean comparisons are appropriate.

## E.4 Answers

1. (a) The hypotheses are  $H_0 : \mu_A = \mu_B = \mu_C = \mu_D$  versus  $H_a$  : at least one pair of  $\mu_i$ 's are different. Test with  $\alpha = 0.05$ .

$$SS_{\text{Total}} = 144294 - \frac{(1848)^2}{24} = 1998$$

$$SS_{\text{Error}} = \sum_i CSS_i = 1290$$

$$SS_{\text{Treat}} = 1998 - 1290 = 708$$

Source of variation	Sum of squares	df	MS	F	c.v.
Adhesives	708	3	236.0	3.66	3.10
Error	1290	20	64.5		
Total	1998	23			

Since  $3.66 > 3.10$ , reject  $H_0$ . At least some of the denture adhesives are significantly different.

- (b) Use a Duncan's multiple range test to locate these differences.

$$\sqrt{\frac{MS_{\text{Error}}}{n}} = \sqrt{\frac{64.5}{6}} = 3.28.$$

	Number of means		
	2	3	4
$r_p$	2.950	3.097	3.190
$SSR_p = r_p \sqrt{\frac{MS_{\text{Error}}}{n}}$	9.676	10.158	10.463

Look at the differences in the means and compare to the appropriate  $SSR_p$ . Adhesives C, A, and B are not significantly different. A, B, and D are not significantly different. Only C and D are significantly different.

C	A	B	D
69 <sup>a</sup>	76 <sup>ab</sup>	79 <sup>ab</sup>	84 <sup>b</sup>

- (c) Use a Student-Newman-Keuls test to locate these differences. From Table C.9 with  $\alpha = 0.05$  and  $\nu = 20$ , we obtain the critical values for the test.

k	2	3	4
q c.v.	2.950	3.578	3.958

To calculate the  $q$ -statistic use

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{Error}}}{n}}} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{64.5}{6}}} = \frac{\bar{X}_i - \bar{X}_j}{3.28}.$$

To compare the largest and smallest means,

$$q = \frac{84 - 69}{3.28} = 4.573.$$

The results of the calculations are summarized below.

Comparison	$q$ statistic	$k$	c.v.	Conclusion
D to C	4.573	4	3.958	Reject $H_0$
D to A	2.439	3	3.578	Accept $H_0$
D to B	1.524	2	2.950	Accept $H_0$
B to C	3.049	3	3.578	Accept $H_0$
B to A	0.915	2	2.950	Accept $H_0$
A to C	2.134	2	2.950	Accept $H_0$

Adhesives C, A, and B are not significantly different. A, B, and D are not significantly different. Only C and D are significantly different.

C	A	B	D
69 <sup>a</sup>	76 <sup>ab</sup>	79 <sup>ab</sup>	84 <sup>b</sup>

2. The hypotheses are  $H_0 : \mu_A = \mu_B = \mu_C$  versus  $H_a$  : at least one pair of  $\mu_i$ 's are different.

$$SS_{\text{Error}} = \sum_i CSS_i = 11.63$$

$$SS_{\text{Treat}} = 15.49 - 11.63 = 3.86$$

Source of variation	Sum of squares	df	MS	$F$	c.v.
Locations	3.86	2	1.93	4.49	3.35
Error	11.63	27	0.43		
Total	15.49	29			

Since  $4.49 > 3.35$ , reject  $H_0$ . At least one treatment mean is significantly different from the others.

Method 1: Use DMRT to locate the differences. The ordered means are

	B	C	A
$\bar{X}_i$	2.22	2.28	3.01

To calculate the  $SSR_p$  use

$$\sqrt{\frac{MS_{\text{Error}}}{n}} = \sqrt{\frac{0.43}{10}} = 0.21.$$

	Number of means	
	2	3
$r_p$	2.919	3.066
$SSR_p = r_p \sqrt{\frac{MS_{\text{Error}}}{n}}$	0.613	0.644

Look at the differences in the means and compare to the appropriate  $SSR_p$ . Treatment A is significantly different from treatments B and C.

B	C	A
2.22 <sup>a</sup>	2.28 <sup>a</sup>	3.01 <sup>b</sup>

Method 2: Use SNK to locate the differences. The ordered means are

	<b>B</b>	<b>C</b>	<b>A</b>
$\bar{X}_i$	2.22	2.28	3.01

From Table C.9 with  $\alpha = 0.05$  and  $\nu = 27$ , we obtain the critical values for the test.

<b>k</b>	<b>2</b>	<b>3</b>
$q$ c.v.	2.919	3.532

To calculate the  $q$ -statistic use

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{Error}}}{n}}} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{0.43}{10}}} = \frac{\bar{X}_i - \bar{X}_j}{0.207}.$$

To compare the largest and smallest means,

$$q = \frac{3.01 - 2.22}{0.207} = 3.816.$$

The results of the calculations are summarized below.

Comparison	$q$ statistic	$k$	c.v.	Conclusion
A to B	3.816	3	3.532	Reject $H_0$
A to C	3.527	2	2.919	Reject $H_0$
C to B	0.290	2	2.919	Accept $H_0$

Treatment A is significantly different from treatments B and C.

<b>B</b>	<b>C</b>	<b>A</b>
2.22 <sup>a</sup>	2.28 <sup>a</sup>	3.01 <sup>b</sup>

3. (a) The hypotheses are  $H_0$ : Oxygen consumption is the same at all tempos versus  $H_a$ : At least one pair of  $\mu_i$ 's are different. Test with  $\alpha = 0.05$ . Using Formula 8.5,

$$SS_{\text{Error}} = \sum_{i=1}^4 (n_i - 1)s_i^2 = 11 \cdot (3.2)^2 + 11 \cdot (3.4)^2 + 11 \cdot (3.0)^2 + 11 \cdot (3.1)^2 = 444.51.$$

Using Formula 8.4,

$$SS_{\text{Tempos}} = \sum \frac{T_{i.}^2}{n_i} - \frac{T_{..}^2}{N} = 137.40.$$

Source of variation	Sum of squares	df	MS	F	c.v.
Tempos	137.40	3	45.800	4.53	2.79
Error	444.51	44	10.103		
Total	581.91	47			

Since  $4.53 > 2.79$ , reject  $H_0$ . At least some of the tempos have different oxygen consumptions.

- (b) **Method 1:** Use Bonferroni  $t$  tests with  $\alpha = 0.05$  to compare the means. The ordered means are

Punches/min	<b>60</b>	<b>80</b>	<b>100</b>	<b>120</b>
$\bar{X}_i$	25.2	28.0	28.3	29.9

Since all the sample sizes are equal, the denominator in the expression for the Bonferroni  $t$  is the same for all tests,

$$\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} = \sqrt{10.103 \left( \frac{1}{12} + \frac{1}{12} \right)} = 1.298.$$

There are  $\binom{4}{2} = 6$  comparisons to make. From Table C.4 the critical  $t$  value for  $\frac{\alpha}{6} = 0.008\bar{3}$  with  $\nu = N - k = 48 - 4 = 44$  degrees of freedom is 2.763.

First compare the 120 and 60 tempos. From Formula E.1, the test statistic is

$$\text{Bonferroni } t = \frac{\bar{X}_{120} - \bar{X}_{60}}{\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_{120}} + \frac{1}{n_{60}} \right)}} = \frac{29.9 - 25.2}{1.298} = 3.621 > 2.763.$$

Reject  $H_0$ , the mean oxygen consumptions are different for the 120 and 60 punch tempos.

The other comparisons are carried out in a similar fashion leading to the following results.

Comparison	$t$ statistic	c.v.	Conclusion
120 to 60	3.621	2.763	Reject $H_0$
120 to 80	1.464	2.763	Accept $H_0$
120 to 100	1.233	2.763	Accept $H_0$
100 to 60	2.388	2.763	Accept $H_0$
100 to 80	0.231	2.763	Accept $H_0$
80 to 60	2.157	2.763	Accept $H_0$

The 60 punch tempo is significantly lower in oxygen consumption compared to the other tempos. There are no other differences at  $\alpha = 0.05$  significance level.

$$25.2^a \quad 28.0^{ab} \quad 28.3^{ab} \quad 29.9^b.$$

**Method 2:** Use DMRT to compare the (ordered) means.

Punches/min	60	80	100	120
$\bar{X}_i$	25.2	28.0	28.3	29.9

Since all the sample sizes are equal, to calculate  $\text{SSR}_p$ , use

$$\sqrt{\frac{\text{MS}_{\text{Error}}}{n}} = \sqrt{\frac{10.103}{12}} = 0.918.$$

Use Formula E.3 with  $\alpha = 0.05$  and  $\nu = N - k = 44$  to calculate  $\text{SSR}_p$  in the table below.

	Number of means		
	2	3	4
$r_p$	2.858	3.006	3.102
$\text{SSR}_p = r_p \sqrt{\frac{\text{MS}_{\text{Error}}}{n}}$	2.624	2.760	2.848

Look at the differences in the means and compare to the appropriate  $\text{SSR}_p$ . The 60 punch tempo is significantly lower in oxygen consumption compared to the other tempos. There are no other differences at  $\alpha = 0.05$  significance level.

$$25.2^a \quad 28.0^b \quad 28.3^b \quad 29.9^b.$$

4. The hypotheses are  $H_0: \mu_1 = \mu_2 = \mu_3$  versus  $H_a$ : At least one pair of  $\mu_i$ 's is different. The summary data and ANOVA table are given below.

	First year	Sophomore	Junior
$\bar{X}_i$	114.8	116.0	87.2
$n_i$	9	5	6

$$N = 20, \quad \sum_i \sum_j X_{ij} = 2136, \quad \sum_i \sum_j X_{ij}^2 = 234,852$$

Source of variation	Sum of squares	df	MS	$F$	c.v.
Years	3308.8	2	1654.41	8.23	3.59
Error	3418.4	17	201.08		
Total	6727.2	19			

Since  $8.23 > 3.59$ , reject  $H_0$ . There are significant differences in the final exam grades by year. Use Bonferroni  $t$  tests to separate means.

There are  $\binom{3}{2} = 3$  comparisons to make and  $\frac{\alpha}{3} = 0.01\bar{6}$ . First compare sophomores and juniors.

$$\text{Bonferroni } t = \frac{\bar{X}_S - \bar{X}_J}{\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_S} + \frac{1}{n_J} \right)}} = \frac{116.0 - 87.2}{\sqrt{201.08 \left( \frac{1}{5} + \frac{1}{6} \right)}} = 3.354.$$

The other test statistic calculations are similar. The following table summarizes the results. Using Table C.4 with  $\nu = 17$  we find:

	Sophomore – Junior	Sophomore – First Year	First year – Junior
Bonferroni $t$	3.354	0.152	3.692
$P$ value	< 0.005	> 0.20	< 0.005

The first year and sophomore means are not different, but both of these means differ significantly from the juniors in final exam means.

<b>Jr</b>	<b>FY</b>	<b>So</b>
87.2 <sup>a</sup>	114.8 <sup>b</sup>	116.0 <sup>b</sup>

5. Begin by calculating the ANOVA table and global  $F$  test to see if there are any significant differences among the groups. The hypotheses are

$$H_0: \mu_I = \mu_{II} = \mu_{III}$$

$$H_a: \text{At least one pair of the } \mu_i \text{'s is not equal.}$$

Test with  $\alpha = 0.05$ . Since  $n_i \bar{X}_i = T_i$ , we have the additional summary statistics:

	I	II	III
$n_i$	15	15	15
$T_i$	600	300	1800

Further,  $\sum_i n_i = 45$  and  $\sum_i T_i = T_{..} = 2700$ . Next we need the three sums of squares:  $SS_{\text{Total}}$ ,  $SS_{\text{Treat}}$ , and  $SS_{\text{Error}}$ . From Formula 8.4,

$$\begin{aligned} SS_{\text{Treat}} = SS_{\text{Groups}} &= \sum_i \left[ \frac{T_i^2}{n_i} \right] - \frac{T_{..}^2}{N} \\ &= \left[ \frac{(600)^2}{15} + \frac{(300)^2}{15} + \frac{(1800)^2}{15} \right] - \frac{(2700)^2}{45} \\ &= 246,000 - 162,000 \\ &= 84,000. \end{aligned}$$

From Formula 8.5, we can calculate  $SS_{\text{Error}}$ :

$$SS_{\text{Error}} = \sum_i (n_i - 1)s_i^2 = \sum_i CSS_i = 1400 + 1120 + 1890 = 4410.$$

Source of variation	Sum of squares	df	MS	F	c.v.
Groups	84,000	2	42,000	400	2.92
Error	4,410	42	105		
Total	88,410	44			

Using Table C.7 since  $400 \gg 3.23$ , we reject  $H_0$  and conclude that the groups are significantly different. We use SNK mean separation. The means are rank-ordered from smallest to largest:

<b>II</b>	<b>I</b>	<b>III</b>
20	40	120

From Table C.9 with  $\alpha = 0.05$  and  $\nu = N - k = 42$ , we obtain

<b>k</b>	<b>2</b>	<b>3</b>
q c.v.	2.858	3.442

Comparison	q statistic	k	c.v.	Conclusion
II to III	37.796	3	3.442	Reject $H_0$
II to I	30.237	2	2.858	Reject $H_0$
I to III	7.559	2	2.858	Reject $H_0$

All treatments produce differing weight gains.

<b>II</b>	<b>I</b>	<b>III</b>
20 <sup>a</sup>	40 <sup>b</sup>	120 <sup>c</sup>

6. The hypotheses are  $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$  There are no differences in mean cortex weights in rats on the various diets versus  $H_a$ : At least one pair of  $\mu_i$ 's (mean cortex weights) is different. The summary data and ANOVA table are given below.

Groups	Count	Sum	Mean	Variance
Diet I	5	3350	670.0	187.50
Diet II	5	3425	685.0	250.00
Diet III	5	3540	708.0	182.50
Diet IV	5	3395	679.0	217.50
Total	20	13,710		

Source of variation	Sum of squares	df	MS	F	c.v.
Diets	3945	3	1315.000	6.28	3.24
Error	3350	16	209.375		
Total	7295	19			

Since  $6.28 > 3.24$ , reject  $H_0$ . There are significant differences in the mean cortex weights of rats on at least two of the diets. Use Bonferroni  $t$  tests to separate means with  $\nu = N - k = 20 - 4 = 16$  and  $\alpha = \frac{0.05}{6} = 0.008\bar{3}$ .

First compare diets III and I. From Formula E.1, the test statistic is

$$\text{Bonferroni } t = \frac{\bar{X}_{\text{III}} - \bar{X}_{\text{I}}}{\sqrt{\text{MS}_{\text{Error}} \left( \frac{1}{n_{\text{III}}} + \frac{1}{n_{\text{I}}} \right)}} = \frac{708.0 - 670.0}{\sqrt{209.375 \left( \frac{1}{5} + \frac{1}{5} \right)}} = 4.152.$$

The other calculations are similar and are listed below.

	III - I	III - IV	III - II	II - I	II - IV	IV - I
Bonferroni $t$	4.152	3.169	2.513	1.639	0.656	0.983
$P$ value	$< 0.005$	$< 0.008\bar{3}$	$> 0.02$	$> 0.10$	$> 0.20$	$> 0.20$
Conclusion	Reject $H_0$	Reject $H_0$	Accept $H_0$	Accept $H_0$	Accept $H_0$	Accept $H_0$

The differences in the means are summarized below.

<b>I</b>	<b>IV</b>	<b>II</b>	<b>III</b>
670.0 <sup>a</sup>	679.0 <sup>a</sup>	685.0 <sup>a,b</sup>	708.0 <sup>b</sup>

# Two-Factor Analysis

This material follows naturally as Section 9.5 in the text *An Introduction to Biostatistics* by Glover and Mitchell (Waveland, 2015). It cross-references that text and follows the same format.

## F.1 The Durbin Test: Balanced Incomplete Blocks

### Introduction

When the number  $k$  of treatments is large, it can be impractical to apply all treatments to each block. For example, if 15 soft drinks were going to be compared, any judge would find it very difficult to rank all 15 accurately. However, if each judge were to taste only 2 or 3 soft drinks, reliable rankings could be obtained. Those experimental designs in which not all treatments are applied to each block are called **incomplete block designs**.

**EXAMPLE F.1.** Apple scab disease, caused by the fungus *Venturia inaequalis*, is an important pathogen of commercial apples. The fungus causes lesions on both the fruit and the leaves of apple trees. The lesions, which are usually circular, rough-surfaced, olive-green spots, can greatly reduce the marketability of fresh apples. Apple scab is typically prevented by applying fungicides at regular intervals throughout the Spring. To test the efficacy of various formulations of a relatively new class of fungicides called strobilurins, 6 different concentrations were tested on 10 plots, with 3 different concentrations applied to different sections of each plot. The number of apple scab lesions per 1000 apples in each section of the plot were recorded and are presented in the following table.

Block	Fungicide					
	A	B	C	D	E	F
1	3	17	12			
2	15	7		18		
3	6		27		14	
4	11			24		18
5	5				12	20
6		19	23			13
7		15		21	6	
8		23			7	12
9			24	14	8	
10			3	26		21

The design of the experiment above was arranged so that every block (plot) received three levels of treatment. Every treatment appeared in 5 different blocks and every treatment appeared with every other treatment twice.

An experimental design with the general features of the one above is called a **balanced incomplete block design**. Analysis is done using the **Durbin test**, which is very similar to the Friedman test.

## Notation

The following notation and properties summarize the Durbin test design.

- $t$  denotes the total number of treatments to be examined.
- $k$  denotes the block size ( $k < t$ ).
- $b$  denotes the total number of blocks.
- $r$  denotes the number of times each treatment appears ( $r < b$ ).
- Each pair of treatments occurs in the same number of blocks ensuring the same number of “head-to-head” comparisons.

The data are arrayed in a balanced incomplete design as in Example F.1.  $X_{ij}$  represents the result of treatment  $j$  in the  $i$ -th block (if it appears there). Rank the  $X_{ij}$  within each block from smallest to largest (from 1 to  $k$ ). Let  $R_j$  denote the sum of the  $r$  ranks associated with the  $j$ -th treatment.

## Assumptions

There are two assumptions for the Durbin test.

1. The blocks are independent of each other.
2. Within each block, the observations may be arranged in increasing order. (With the use of midranks, a small number of ties is tolerable.)

## Hypotheses

The hypotheses are identical to those for the Friedman test.

- $H_0$ : “The treatments have identical effects” versus  $H_a$ : “At least one of the treatments tends to yield larger effects than at least one other.”

## Test Statistic and Decision Rule

If the null hypothesis is true, then all treatments should produce nearly equal rank sums  $R_j$ . Since there are  $k$  ranks per block, the mean rank is  $\frac{(k+1)}{2}$ . Since each treatment is ranked  $r$  times, the mean rank sum is  $\frac{r(k+1)}{2}$ . So if the null hypothesis holds, then all the rank sums  $R_j$  should be close to this mean sum. The test statistic is based on the difference between this mean and the various  $R_j$ . The Durbin test statistic is defined as

$$T = \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2.$$

This may be written in the following form for easier hand computation.

$$T = \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t R_j^2 - \frac{3r(t+1)(k+1)}{k-1}.$$

Since the test statistic will be close to 0 if all  $R_j$  are close to the mean rank sum,  $H_0$  is accepted when  $T$  is small and rejected when  $T$  is sufficiently large. The distribution of  $T$  is reasonably well-approximated by the chi-square distribution with  $t - 1$  degrees of freedom. So the critical values for the test can be found in Table C.5. If the test is done at the  $\alpha$  level of significance, then  $H_0$  should be rejected when  $T$  exceeds the value in the Table C.5 for the  $1 - \alpha$  quantile.

**SOLUTION TO EXAMPLE F.1.** Carry out the test at the  $\alpha = 0.05$  level. In the example,  $t = 6$  is the total number of treatments,  $k = 3$  is the block size,  $b = 10$  is the total number of blocks, and  $r = 5$  is the number of times each treatment appears. Notice that every treatment is compared with every other variety exactly twice. Thus, this is a balanced incomplete block design and the Durbin test may be used. First the data are ranked and the treatment rank sums are calculated.

Block	Fungicide					
	A	B	C	D	E	F
1	1	3	2			
2	1	3		2		
3	1		3		2	
4	1			3		2
5	1				2	3
6		2	3			1
7		2		3	1	
8		3			1	2
9			3	2	1	
10			3	2		1
$R_j$	5	13	14	12	7	9

The mean rank is  $\frac{r(k+1)}{2} = \frac{5(3+1)}{2} = 10$ . So the test statistic is

$$\begin{aligned}
 T &= \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2 \\
 &= \frac{12(6-1)}{5(6)(3^2-1)} [(5-10)^2 + (13-10)^2 + (14-10)^2 + (12-10)^2 + (7-10)^2 + (9-10)^2] \\
 &= 16.0.
 \end{aligned}$$

From Table C.5, the critical value for a test at the  $\alpha = 0.05$  level with  $df = t - 1 = 6 - 1 = 5$  is 11.1. Since  $T > 11.1$ ,  $H_0$  is rejected. At least one fungicide has a different effect on apple scab disease.

## Paired Comparisons

Paired comparisons for the Durbin test can be carried out when  $H_0$  has been rejected. The set up is exactly the same as for the Friedman test. (You may wish to review the material in Chapter 9.3 on Friedman paired comparisons and selection of significance levels  $\alpha'$  and  $\alpha$ .) The hypotheses are:

- $H_0$ : “The effects of the  $i$ -th and  $j$ -th treatments are the same” versus  $H_a$ : “The effects of the  $i$ -th and  $j$ -th treatments are different.”

The comparisons are carried out using the rank sum differences,  $|R_i - R_j|$ . The test statistic for a two-tailed comparison of the  $i$ -th and  $j$ -th treatments is

$$z_{ij} = \frac{|R_i - R_j|}{\sqrt{\frac{rt(k^2-1)}{6(t-1)}}}.$$

$H_0$  is rejected if  $z_{ij} \geq z_{1-\alpha}$ , or more simply if

$$|R_i - R_j| \geq z_{1-\alpha} \sqrt{\frac{rt(k^2-1)}{6(t-1)}}.$$

If there are  $t$  treatments, then there will be a total of  $\binom{t}{2}$  paired comparisons.

**EXAMPLE F.2.** Carry out a set of paired comparisons for the data in Example F.1.

**SOLUTION.** The hypotheses are  $H_0$ : “The  $i$ -th and  $j$ -th fungicides have the same effect on apple scab disease” versus  $H_a$ : “The  $i$ -th and  $j$ -th fungicides differ in their effects on apple scab disease.” If the test is done at the  $\alpha' = 0.05$  level, then the individual comparisons are made at the  $\alpha = \frac{0.05}{\binom{6}{2}} = \frac{0.05}{15} = 0.0033$  level. Thus,  $1 - \alpha = 0.9967$  and  $z_{0.9967} = 2.72$ .  $H_0$  is rejected only if

$$|R_i - R_j| \geq z_{1-\alpha} \sqrt{\frac{rt(k^2-1)}{6(t-1)}} = 2.72 \sqrt{\frac{5 \cdot 6 \cdot 8}{6 \cdot 5}} = 7.69.$$

Fungicide A differs from (is more effective) than fungicides B and C since only these rank sums differ by at least 7.69.

## Incomplete Block Design Project: Ice Cream Taste Test

As a fitting conclusion to this section, apply your knowledge of statistics to plan, carry out, and analyze the results of an ice cream or soft drink tasting for your class using an incomplete block design.

### Design Restrictions

Incomplete block designs are a bit complicated to create.

- The number of blocks,  $b$ , is the number of people participating in the tasting.
- The total number of flavors to be compared in the tasting is  $t$ .
- Every person should taste the same number of flavors,  $k$ .
- Every flavor should be tasted by the same number of people,  $r$ .
- Each pair of flavors should appear in the same number of blocks. This ensures that every flavor appears in the same number of “head-to-head” tastings with every other flavor.

These conditions constrain the values of  $b$ ,  $k$ ,  $t$ , and  $r$ . Since there are  $b$  people each tasting  $k$  flavors, there are a total of  $bk$  individual tastings. On the other hand, since there are  $t$  flavors and each is tasted  $r$  times, then there are  $tr$  tastings. So  $bk$  must equal  $tr$  in the design layout that you propose.

Next we show that the final condition in the list above implies that the quantity  $r(k - 1)$  must be divisible by  $t - 1$ . There are  $\binom{t}{2}$  different head-to-head pairs. Let  $s$  denote the number of times each pair occurs in the design, making a total of  $s\binom{t}{2}$  pairs. But there are  $\binom{k}{2}$  such pairs in each of the  $b$  blocks. So

$$s\binom{t}{2} = b\binom{k}{2} \Rightarrow \frac{st(t-1)}{2} = \frac{bk(k-1)}{2}.$$

Multiplying by 2 and using the fact that  $bk = rt$ , it follows that

$$st(t-1) = rt(k-1) \Rightarrow s(t-1) = r(k-1).$$

Therefore  $t - 1$  divides  $r(k - 1)$ .

**EXAMPLE F.3.** Suppose that there were  $b = 30$  tasters and  $t = 10$  flavors to taste. Find values of  $k$  and  $r$  that satisfy the conditions of a balanced incomplete block design.

**SOLUTION.** We must have  $30k = 10r$  or  $3k = r$ . Thus, a possible solution would be  $k = 4$  and  $r = 12$  since  $30 \times 4 = 10 \times 12$  and  $12 \times (4 - 1) = 36$  is divisible by  $10 - 1 = 9$ . You would then have to produce a layout that fit the design.

**EXAMPLE F.4.** Suppose that there were  $b = 15$  tasters and  $t = 6$  flavors to taste. Find values of  $k$  and  $r$  that satisfy the conditions of a balanced incomplete block design.

**SOLUTION.** This time  $15k = 6r$  or  $5k = 2r$ . Thus, a possible solution is  $k = 2$  and  $r = 5$  because  $15 \times 2 = 6 \times 5$  and  $5 \times (2 - 1) = 5$  is divisible by  $6 - 1 = 5$ . Another possible solution might be  $k = 4$  and  $r = 10$  because  $15 \times 4 = 6 \times 10$  and  $10 \times (4 - 1) = 30$  is divisible by  $6 - 1 = 5$ . In each case you'd then have to find a layout to fit the design parameters.

## The Assignment

Your task is to provide the entire design layout for the tasting following these guidelines.

1. You must specify the values of  $b$ ,  $k$ ,  $t$ , and  $r$
2. You must provide the pattern for the layout for the test. Make sure that your design fulfills the requirements of a *balanced incomplete block design*. Note that a basic layout can be “repeated.” For example, if your design calls for 10 blocks and there are 30 people, each block would be repeated three times.
3. You must specify the *type* or *flavor*, that is, only chocolate (use various brands), only “lite,” or only flavors of a specific brand. The more specific you are, the better.
4. Carefully state the null and alternative hypotheses of your proposed test.
5. Select a significance level for a Durbin test and determine the corresponding critical value for your incomplete block design.
6. Similarly, select a significance level  $\alpha'$  for the paired comparisons that may be required for your design. Determine the corresponding significance level  $\alpha$  and find the difference in ranks necessary for there to be a significant difference between two flavors.

An additional factor that you should take into account is that it is probably not possible for one to accurately rank very many flavors. So make  $k$  relatively small. This also makes producing the layout easier.

## F.2 Problems

1. Seven different brands of breakfast bars were taste-tested by 7 subjects. Each person was asked to taste and rank 4 of the 7 bars (with a rank of 1 assigned to the preferred brand). A balanced incomplete block design was created and the following data were collected. Do the results indicate a significant difference in taste?

Person	A	B	C	D	E	F	G
1			3		1	4	2
2	1			3		4	2
3	2	1			3		4
4	1	2	4			3	
5		1	4	3			2
6	2		4	1	3		
7		1		2	3	4	

2. (a) Forty consumers were asked to rate five brands of coffee. Each consumer tasted exactly two brands (with a rank of 1 assigned to the preferred brand). A balanced incomplete block design was arranged where  $t = 5$ ,  $k = 2$ ,  $b = 40$ , and  $r = 16$ . The resulting rank sums were  $R_1 = 25$ ,  $R_2 = 18$ ,  $R_3 = 21$ ,  $R_4 = 29$ , and  $R_5 = 27$ . Was there a significant difference in the preferences of the consumers?
- (b) If appropriate, carry out a set of paired comparisons for these data. Interpret the results.
3. A physician is interested in comparing the effects of six different allergens in persons extremely sensitive to a ragweed skin allergy test. Ten patients received an injection of an antihistamine. Subsequently, the allergens were applied topically in different sections of the left arm. The areas (in  $\text{mm}^2$ ) of redness, measured after two days, are reported in the table below.

Person	A	B	C	D	E	F
1	41	25				40
2	42	46			37	
3		37	45	33		
4	46			35	34	
5		31		34		42
6			56		36	65
7	42		67	33		
8				37	39	49
9	40		59			55
10		36	57		34	

- (a) Verify that this is an incomplete block design. Determine the values of  $t$ ,  $k$ ,  $b$ , and  $r$ .
- (b) Is there a significant difference among the patients' responses to the allergens? Carry out a Durbin test at the  $\alpha = 0.05$  level.
- (c) Carry out the full set of paired comparisons if appropriate.
4. Students on a term abroad program in Australia designed and carried out and analyzed the results of a beer tasting using an incomplete block design. Five Australian bitter beers were compared. The raw data from their test appears on the next page. Do an analysis using *all appropriate* tests. Make sure to state all hypotheses and significance levels. Interpret the results.

Block	Castlemaine	Victoria	Foster's Special	Powers	Toohey's Red
1	2	3	1		
2	1	3		2	
3	2	3			1
4	3		2	1	
5	3		2		1
6	3			1	2
7		1	2	3	
8		1	3		2
9		1		3	2
10			3	1	2
11	1	3	2		
12	3	2		1	
13	1	3			2
14	2		3	1	
15	1		2.5		2.5
16	1			2	3
17		1	2	3	
18		1	3		2
19		3		1	2
20			1	2	3
21	3	2	1		
22	3	1		2	
23	2	3			1
24	3		1	2	
25	1		3		2
26	3			1	2
27		1	3	2	
28		2.5	1		2.5
29		2		1	3
30			2	1	3
31	2	1	3		
32	3	2		1	
33	3	2			1
34	2		3	1	
35	1		3		2
36	3			1	2
37		1	2	3	
38		1	2		3
39		1		3	2
40			3	1	2

5. Provide the layout for an incomplete block design with  $b = 15$  blocks,  $t = 6$  treatments,  $k = 2$ , and  $r = 5$ .
6. (a) Is the following an incomplete block design layout? If so, give the values of the parameters  $b$ ,  $t$ ,  $r$ , and  $k$  and also the number of times each comparison occurs. Otherwise, explain why it is not.

	A	B	C	D	E	F	G	H	I	J	K
1	x		x	x	x				x		
2		x		x	x	x				x	
3			x		x	x	x				x
4	x			x		x	x	x			
5		x			x		x	x	x		
6			x			x		x	x	x	
7				x			x		x	x	x
8	x				x			x		x	x
9	x	x				x			x		x
10	x	x						x	x		x

- (b) Is the following an incomplete block design layout? If so, give the values of the parameters  $b$ ,  $t$ ,  $r$ , and  $k$  and also the number of times each comparison occurs. Otherwise, explain why it is not.

	A	B	C	D	E	F
1	x	x	x			
2		x	x	x		
3			x	x	x	
4				x	x	x
5	x				x	x
6	x	x				x

- (c) Fill in entries in the two rows to make this an incomplete block design. Hint: Begin by determining  $r$ .

	A	B	C	D	E	F
1				x	x	x
2			x		x	x
3		x		x		x
4		x	x		x	
5		x	x	x		
6						
7						
8	x		x	x		
9	x	x				x
10	x	x			x	

7. (a) Show that  $kb = rt$  in any balanced incomplete block design. Hint: Count the number of observations in two different ways.
- (b) If we allowed  $r = b$  and  $k = t$ , show that the Durbin test statistic would be exactly the same as the Friedman test statistic.
- (c) Suppose that 16 consumers are to rate five brands of ice cream. Each consumer is given three flavors to rate. Is a balanced incomplete block design possible? Explain.

8. To evaluate the antibacterial benefits of a multicare dentifrice relative to other selected oral hygiene regimens, the following balanced incomplete block design study was conducted. Forty healthy subjects each tested 3 of 5 possible regimens (brushing with (A) distilled water or brushing with (B) regular, (C) baking soda, (D) tartar control, or (E) multicare toothpastes. Treatments were administered with a one week washout period between them. Subjects brushed for 60 seconds with a dentifrice dose sufficient to cover the head of a standard toothbrush. After brushing, subjects expectorated the dentifrice and rinsed for 10 seconds with 10 ml of water. Microbial samples were taken from the gingival margin of both the facial and lingual surfaces of the lower dentition using a swab prior to and immediately following brushing. Swab samples were evaluated for total facultative anaerobic bacteria. The results reported below give the *decrease*, (before – after) in log cfu/ml (log colony forming units per ml) for each regimen. Analyze these data and determine whether there are any significant differences among the treatments and, if so, determine which treatments are different.

Person	A	B	C	D	E
1	67	49	65		
2	51	41	79		
3	74	78	63		
4	86	73	69		
5	31	40		54	
6	72	46		20	
7	69	64		78	
8	56	76		55	
9	69	56			88
10	53	68			87
11	33	82			88
12	63	79			122
13	75		61	58	
14	49		66	85	
15	74		51	49	
16	50		80	82	
17	48		47		106
18	59		68		99
19	38		40		105
20	61		69		86
21	56			68	87
22	41			69	105
23	44			57	54
24	66			84	69
25		79	68	54	
26		57	76	37	
27		73	58	49	
28		64	35	55	
29		78	61		67
30		61	58		59
31		68	67		70
32		54	59		108
33		50		53	113
34		80		48	86
35		59		79	91
36		67		58	73
37			68	72	69
38			70	58	111
39			77	41	65
40			38	63	145

### F.3 Answers

1. Here  $t = 7$ ,  $k = 4$ ,  $b = 7$ , and  $r = 4$ .  $H_0$ : “All seven bars taste the same” versus  $H_a$ : “At least one of the bars tastes different than the others.” Let  $\alpha = 0.05$ . Since  $df = t - 1 = 6$ , the c.v. = 12.6. The ranks sums are:

A	B	C	D	E	F	G
6	5	15	9	10	15	10

The expected rank sum is  $\frac{r(k+1)}{2} = \frac{4(4+1)}{2} = 10$ ,

$$\begin{aligned} T &= \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2 \\ &= \frac{12(7-1)}{4(7)(4^2-1)} [(6-10)^2 + (5-10)^2 + (15-10)^2 + (10-10)^2 + (15-10)^2 + (10-10)^2] \\ &= 15.77. \end{aligned}$$

Since  $T > 12.6$ , reject  $H_0$ . There is evidence of a significant taste difference among the bars.

Carry out paired comparisons. The hypotheses are  $H_0$ : “The  $i$ th and  $j$ th brands taste the same” versus  $H_a$ : “The  $i$ th and  $j$ th brands taste different.” If the test is done at the  $\alpha' = 0.05$  level, then the individual comparisons are made at the  $\alpha = \frac{0.05}{\binom{7}{2}} = \frac{0.05}{21} = 0.0024$  level. Thus,  $1 - \alpha = 0.9976$  and  $z_{0.9976} = 2.82$ .  $H_0$  is rejected only if

$$|R_i - R_j| \geq z_{1-\alpha} \sqrt{\frac{rt(k^2-1)}{6(t-1)}} = 2.82 \sqrt{\frac{4 \cdot 7 \cdot 15}{6 \cdot 6}} = 9.63.$$

Brand B was judged significantly different (better) in taste than either brands C or F.

If the test were done at the  $\alpha' = 0.10$  level, then the individual comparisons are made at the  $\alpha = \frac{0.05}{\binom{7}{2}} = \frac{0.10}{21} = 0.0048$  level and  $z_{0.9952} = 2.60$ .  $H_0$  is rejected only if

$$|R_i - R_j| \geq 2.60 \sqrt{\frac{4 \cdot 7 \cdot 15}{6 \cdot 6}} = 8.88.$$

Under these less restrictive conditions, Brands A and B would be judged significantly different (better) in taste from either brands C or F.

2. (a) The Durbin test statistic is

$$\begin{aligned} T &= \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2 \\ &= \frac{12(5-1)}{16(5)(2^2-1)} [(25-24)^2 + (18-24)^2 + (21-24)^2 + (29-24)^2 + (27-24)^2] \\ &= 14.2. \end{aligned}$$

From Table C.5, the critical value for a test at the  $\alpha = 0.05$  level with  $df = t - 1 = 5 - 1 = 4$  is 9.49. Since  $T > 9.49$ ,  $H_0$  is rejected. At least one brand of coffee tastes different from that of another.

- (b) Carry out paired comparisons. The hypotheses are  $H_0$ : “The  $i$ th and  $j$ th brands taste the same” versus  $H_a$ : “The  $i$ th and  $j$ th brands taste different.” If the test is done at the  $\alpha' = 0.05$

level, then the individual comparisons are made at the  $\alpha = \frac{0.05}{\binom{5}{2}} = \frac{0.05}{10} = 0.005$  level. Thus,  $1 - \alpha = 0.995$  and  $z_{0.995} = 2.58$ .  $H_0$  is rejected only if

$$|R_i - R_j| \geq z_{1-\alpha} \sqrt{\frac{rt(k^2 - 1)}{6(t-1)}} = 2.58 \sqrt{\frac{16 \cdot 5 \cdot 3}{6 \cdot 4}} = 8.16.$$

Brand 2 was judged significantly different (better) in taste than either brands 4 or 5.

3. (a) Here  $t = 6$ ,  $k = 3$ ,  $b = 10$ , and  $r = 5$ .  
 (b) The ranks and their sums are given in the table that follows.

Person	A	B	C	D	E	F
1	3	1				2
2	2	3			1	
3		2	3	1		
4	3			2	1	
5		1		2		3
6			2		1	3
7	2		3	1		
8				1	2	3
9	1		3			2
10		2	3		1	
Sum	11	9	14	7	6	13

Test  $H_0$ : “All six allergens have the same effect” versus  $H_a$ : “At least one of the allergens elicits a different response.” Let  $\alpha = 0.05$ . Since  $t = 6$ ,  $df = 5$ , and the c.v. = 11.1.

$$\begin{aligned} T &= \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2 \\ &= \frac{12(6-1)}{5(6)(3^2-1)} [(11-10)^2 + (9-10)^2 + (14-10)^2 + (7-10)^2 + (6-10)^2 + (13-10)^2] \\ &= 13.0. \end{aligned}$$

Since  $T > 11.1$ ,  $H_0$  is rejected. The response associated with at least one allergen differs from that of another.

- (c) Carry out paired comparisons. The hypotheses are  $H_0$ : “The response associated with the  $i$ th and  $j$ th allergens is the same” versus  $H_a$ : “The response associated with the  $i$ th and  $j$ th allergens is different.” If the test is done at the  $\alpha' = 0.05$  level, then the individual comparisons are made at the  $\alpha = \frac{0.05}{\binom{6}{2}} = \frac{0.05}{15} = 0.0033$  level. Thus,  $1 - \alpha = 0.9967$  and  $z_{0.9967} = 2.72$ .  $H_0$  is rejected only if

$$|R_i - R_j| \geq z_{1-\alpha} \sqrt{\frac{rt(k^2 - 1)}{6(t-1)}} = 2.72 \sqrt{\frac{5 \cdot 6 \cdot 8}{6 \cdot 5}} = 7.69.$$

Only the responses to allergens C and E can be judged significantly different.

4. Here  $t = 5$ ,  $k = 3$ ,  $b = 40$ , and  $r = 24$ .  $H_0$ : “All five beers taste the same” versus  $H_a$ : “At least one of the beers tastes different than the others.” Let  $\alpha = 0.05$ . Since  $t = 5$ ,  $df = 4$ , and the c.v. = 9.49. The ranks sums are:

Castlemaine	Victoria	Foster's Special	Powers	Toohey's Red
52	44.5	53.5	40	50

The expected rank sum is  $\frac{r(k+1)}{2} = \frac{24(3+1)}{2} = 48$ ,

$$\begin{aligned} T &= \frac{12(5-1)}{24(5)(3^2-1)} [(52-48)^2 + (44.5-48)^2 + (53.5-48)^2 + (40-48)^2 + (50-48)^2] \\ &= \frac{16 + 12.25 + 30.25 + 64 + 4}{20} = 6.325. \end{aligned}$$

Since  $T < 9.49$ , accept  $H_0$ . There is no evidence of a taste difference among the beers.

5. Each possible pair occurs once.

Block	A	B	C	D	E	F
1	x	x				
2	x		x			
3	x			x		
4	x				x	
5	x					x
6		x	x			
7		x		x		
8		x			x	
9		x				x
10			x	x		
11			x		x	
12			x			x
13				x	x	
14				x		x
15					x	x

6. (a) Yes,  $b = t = 11$  and  $k = r = 5$ . Each comparison occurs twice.  
 (b) No, because the comparison (A,B) occurs twice, but the comparison (A,C) occurs only once.  
 (c) Since  $k = 3$  and two rows are missing, we must add a total of 6 x's to the layout, 3 in each row. Notice that one column (A) has 3 entries, while four columns (C, D, E, and F) have 4 entries, and one column (B) has 5. Since all columns must have the same number of entries  $r$ , we see that  $r$  must be at least 5 (from B) and at most 5 (because at most 2 x's can be added to column A). So  $r = 5$ . Add the 2 x's to column A. Notice that the pair (A,C) occurs only in row 7, while the pair (A,B) occurs in rows 9 and 10. So every pair must occur at least twice. If we put an x in row 6, column C, then we must also put an x in row 6, column F to obtain the second (C,F)-pair. Now row 6 has its 3 entries. Row 7 is completed by noting that we still need entries in columns D and E. So the final two rows (in either order) are:

	A	B	C	D	E	F
6	x		x			x
7	x			x	x	

7. (a) Since there are  $b$  people each tasting  $k$  flavors, there are a total of  $bk$  individual tastings. On the other hand, since there are  $t$  flavors and each is tasted  $r$  times, then there are  $tr$  tastings. So  $bk$  must equal  $tr$  in a balanced incomplete block design.

(b) If  $r = b$  and  $k = t$ , then

$$\begin{aligned}
 T_{\text{Durbin}} &= \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2 \\
 &= \frac{12(k-1)}{bk(k^2-1)} \sum_{j=1}^k \left( R_j - \frac{b(k+1)}{2} \right)^2 \\
 &= \frac{12(k-1)}{bk(k-1)(k+1)} \sum_{j=1}^k \left( R_j - \frac{b(k+1)}{2} \right)^2 \\
 &= \frac{12}{bk(k+1)} \sum_{j=1}^k \left[ R_j - \frac{b(k+1)}{2} \right]^2 \\
 &= T_{\text{Friedman}}.
 \end{aligned}$$

(c) No. Here  $t = 5$ ,  $k = 3$ ,  $b = 16$ , and  $r$  is unknown. But  $rt = bk$ , so  $5r = 16 \cdot 3 = 48$  which implies that  $r = \frac{48}{5} = 9.6$ . But  $r$  is supposed to be an integer.

8. Here  $t = 5$ ,  $k = 3$ ,  $b = 40$ , and  $r = 24$ .  $H_0$ : “All treatments have the same effect” versus  $H_a$ : “At least one treatment is different than the others.” Let  $\alpha = 0.05$ . Since  $\text{df} = t - 1 = 4$ , the c.v. = 9.49. (The ranks and their sums are given in the table at the end of this answer.)

The expected rank sum is  $\frac{r(k+1)}{2} = \frac{24(3+1)}{2} = 48$ ,

$$\begin{aligned}
 T &= \frac{12(t-1)}{rt(k^2-1)} \sum_{j=1}^t \left( R_j - \frac{r(k+1)}{2} \right)^2 \\
 &= \frac{12(5-1)}{24(5)(3^2-1)} [(41-48)^2 + (46-48)^2 + (43-48)^2 + (44-48)^2 + (66-48)^2] \\
 &= 20.9.
 \end{aligned}$$

Since  $T > 9.49$ , reject  $H_0$ . There is evidence of a significant differences among the treatments.

Carry out paired comparisons. The hypotheses are  $H_0$ : “The  $i$ th and  $j$ th treatments have the same effect” versus  $H_a$ : “The  $i$ th and  $j$ th treatments have different effects.” If the test is done at the  $\alpha' = 0.05$  level, then the individual comparisons are made at the  $\alpha = \frac{0.05}{\binom{5}{2}} = \frac{0.05}{10} = 0.005$  level. Thus,  $1 - \alpha = 0.995$  and  $z_{0.995} = 2.58$ .  $H_0$  is rejected only if

$$|R_i - R_j| \geq z_{1-\alpha} \sqrt{\frac{rt(k^2-1)}{6(t-1)}} = 2.58 \sqrt{\frac{24 \cdot 5 \cdot 8}{6 \cdot 4}} = 16.3.$$

Toothpaste E was judged significantly different (better) than all others.

Person	A	B	C	D	E
1	3	1	2		
2	2	1	3		
3	3	1	2		
4	3	2	1		
5	1	2		3	
6	3	2		1	
7	2	1		3	
8	2	3		1	
9	2	1			3
10	1	2			3
11	1	2			3
12	1	2			3
13	3		2	1	
14	1		2	3	
15	3		2	1	
16	1		2	3	
17	2		1		3
18	1		2		3
19	1		2		3
20	1		2		3
21	1			2	3
22	1			2	3
23	1			3	2
24	1			3	2
25		3	2	1	
26		2	3	1	
27		3	2	1	
28		3	1	2	
29		3	1		2
30		3	1		2
31		2	1		3
32		1	2		3
33		1		2	3
34		2		1	3
35		1		2	3
36		2		1	3
37			1	3	2
38			2	1	3
39			3	1	2
40			1	2	3
Sum	41	46	43	44	66

# Sampling Distributions

## G.1 Confidence Intervals for a Population Proportion

Confidence intervals for many other parametric statistics can be developed if the sampling distribution is known or can be reasonably approximated. Let's consider now the confidence interval for  $p$ , the population proportion in a binomial distribution.

As we saw in Chapter 1 many statistical studies involve counts (categorical data) rather than measurement data. Among the most common are opinion polls and surveys.

**EXAMPLE G.1.** A local epidemiologist wishes to determine the rate of breast cancer in women under age 35 in a rural county in Ireland. She surveys a random sample of 1200 women in this age group and determines that exactly 6 have had this form of cancer sometime during their lifetime. She wishes to use this information to estimate the population rate  $p$  of breast cancer and determine a confidence interval for this estimate.

We will present two solutions to this question. The first solution makes use of material we developed regarding the binomial distribution in Chapter 3. You will find this method has been used in many studies in the scientific literature.

### Method 1

If  $X$  is the number of “successes” in a sample of size  $n$ , then a simple way to estimate the population proportion is to use the **sample proportion**

$$\hat{p} = \frac{X}{n}.$$

This symbol is read as “ $p$  hat.” In the epidemiologist's case

$$\hat{p} = \frac{X}{n} = \frac{6}{1200} = 0.005.$$

She now has an estimate of the population rate of breast cancer and needs a way of expressing her confidence in this value.

From Section 3.2, a binomial distribution has an approximately normal distribution with mean  $p$  and population variance  $\sigma^2 = p(1 - p)$  if  $np > 5$  and  $n(1 - p) > 5$ . The sample proportion,  $\hat{p}$ , is an unbiased estimator of  $p$ , the population parameter.

Since  $\hat{p}$  is a sample mean, then by the Central Limit Theorem (Theorem 4.1) the variance of  $\hat{p}$  is

$$\sigma_{\hat{p}}^2 = \frac{\sigma^2}{n} = \frac{p(1 - p)}{n}.$$

and its standard error is

$$SE_{\hat{p}} = \sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}}. \quad (\text{G.1})$$

We have a point estimate  $\hat{p}$  for  $p$  and its standard error and we know that the sampling distribution is approximately normal (at least for large  $n$ ). So using (4.3), the end points of a  $(1 - \alpha)100\%$  confidence interval for  $p$  should be

$$L_1 = \hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

and

$$L_2 = \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}},$$

where  $z_{1-\frac{\alpha}{2}}$  is the confidence factor ( $z$  value) corresponding to  $1 - \frac{\alpha}{2}$ . For example, for a  $(1 - \alpha)100\% = 95\%$  confidence interval,  $\alpha = 0.05$  and the corresponding  $z$  value is  $z_{1-0.025} = z_{0.975} = 1.960$ .

Notice that there is a problem here. These limits include the population parameter  $p$ , which we are trying to determine. If we actually knew  $p$ , then we wouldn't need any confidence limits! To overcome this paradox, we estimate  $p$  by using  $\hat{p}$  in the formula for SE. Then

$$L_1 = \hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

and

$$L_2 = \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Because the standard error is estimated from the sample, we should also change the  $z$  values ( $\pm z_{1-\frac{\alpha}{2}}$ ) to appropriate  $t$  values. However, earlier we assumed that sample size  $n$  was large. If we assume that  $n > 30$ , then the  $t$  distribution is very nearly equal to the standard normal distribution, so  $t$  and  $z$  values will be very close. (As  $df$  goes to infinity, the  $t$  distribution approaches the  $z$  distribution.) Thus,

**FORMULA G.1.** A good approximation for the  $(1-\alpha)100\%$  confidence limits of a population proportion when  $n\hat{p} > 5$  and  $n(1-\hat{p}) > 5$  is given by

$$L_1 = \hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad \text{and} \quad L_2 = \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

**SOLUTION TO EXAMPLE G.1.** Since  $n = 1200$  and  $\hat{p} = 0.005$ , then

$$n\hat{p} = 1200 \times 0.005 = 6 > 5$$

and

$$n(1-\hat{p}) = 1200 \times 0.995 = 1194 > 5.$$

The approximation in Formula G.1 will be appropriate here. The 95% confidence limits are

$$L_1 = 0.005 - 1.960 \sqrt{\frac{(0.005)(0.995)}{1200}} = 0.005 - 0.00399 = 0.00101$$

and

$$L_2 = 0.005 + 1.960 \sqrt{\frac{(0.005)(0.995)}{1200}} = 0.005 + 0.00399 = 0.00899.$$

The 95% confidence interval for  $p$ , the proportion of Irish women under age 35 who have experienced breast cancer, is 0.00101 to 0.00899.

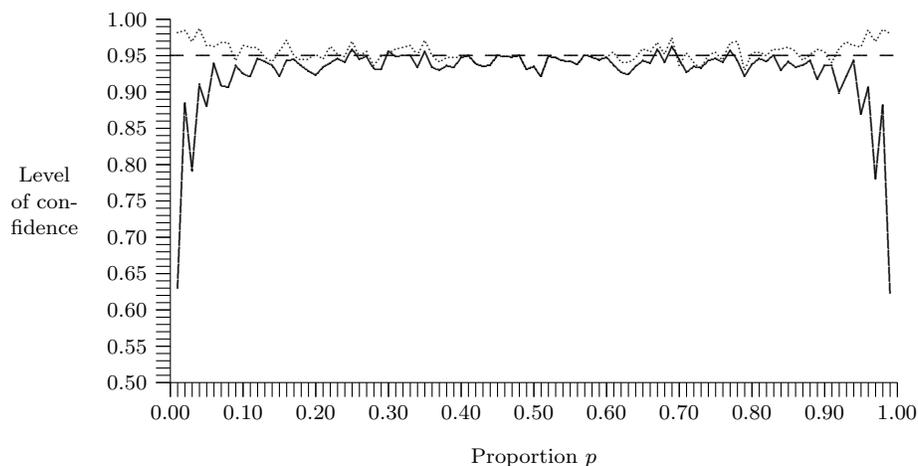
We can say that we are 95% confident that the true proportion  $p$  is between 0.00101 and 0.00899 or, as these statistics are often stated, between 101 and 899 per 100,000. Remember that we interpret this as meaning that with repeated samples of size 1200 about 95% of the intervals constructed as above would include the true population proportion  $p$ .

## Method 2 (Optional)

The method used to construct the confidence interval for Example G.1 required that the sample size be relatively large. What happens when  $np(1-p) \leq 5$ , as in the case when  $n$  is small and  $p$  is either very close to 0 as in Example G.1 or very large (close to 1)?

In the 1990's with the use of computer simulations, it was shown that confidence intervals based on the method above can be inaccurate, even for large samples. Figure G.1 illustrates how problematic Method 1 can be. For comparison purposes, the horizontal dashed line (- -) represents the desired 0.95 confidence level. For each proportion  $p$  between 0.01 to 0.99 one-thousand samples of size 100 were drawn. Next, 95% confidence intervals were computed in two ways. The solid graph (—) represents the proportion of time the true value of  $p$  was contained in the confidence interval using Method 1 with  $\hat{p}$ . In general the true level of confidence is somewhat less than 95% for most values of  $p$ . In particular, when  $p \leq 0.05$  or  $p \geq 0.95$ , the confidence interval computed from the sample contained the true population proportion  $p$  less than 85% of the time.

On the other hand, dotted graph (·····) represents the proportion of time the true value of  $p$  was contained in the confidence interval using Method 2 which will be described below. For nearly all values of  $p$ , especially large and small, the confidence intervals contained the true population proportion  $p$  *more than* 95% of the time. That is, the method is conservative and slightly understates the true level of confidence.



**FIGURE G.1.** For each proportion  $p$  between 0.01 to 0.99 one-thousand samples of size 100 were drawn and 95% confidence intervals were computed in two ways. The horizontal dashed line (- -) represents the desired 95% confidence level. The solid graph (—) represents the proportion of time the true value of  $p$  was contained in the confidence interval using  $\hat{p}$  while the dotted graph (·····) represents the proportion of time the true value of  $p$  was contained in the confidence interval using  $\tilde{p}$ .

To begin Method 2, we adjust the estimate for the population proportion. The **Wilson estimate** of the population proportion is

$$\tilde{p} = \frac{X + 2}{n + 4},$$

where  $X$  is the number of “successes” in a sample of size  $n$ . This is read as “ $p$  tilde.” It is

equivalent to adding two successes and two failures to the sample. The consequence of this adjustment is that  $\tilde{p}$  is a bit larger than  $\hat{p}$  if  $\hat{p} < 0.5$  and slightly smaller than  $\hat{p}$  if  $\hat{p} > 0.5$ . The difference between  $\tilde{p}$  and  $\hat{p}$  is more pronounced when  $\hat{p}$  is near 0 and 1 or when  $n$  is small.

In a similar fashion, we adjust the standard error. The standard error of  $\tilde{p}$  is

$$\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}. \quad (\text{G.2})$$

Using (4.3) with the new point estimate and standard error, and continuing to use the normal distribution, we obtain the following formula for the confidence interval limits.

**FORMULA G.2.** The  $(1 - \alpha)100\%$  confidence limits of a population proportion when  $n > 5$  are given by

$$L_1 = \tilde{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \quad \text{and} \quad L_2 = \tilde{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}.$$

Notice that these confidence limits may be used for all but the smallest sample sizes. As noted earlier, the confidence limits from Formula G.2 are generally conservative, even for small values of  $p$ , unlike Method 1. Since the calculations are no harder than in Method 1, Formula G.2 is preferred.

One might ask why we made the particular adjustment of adding two successes and two failures to the sample to produce the new estimate. The complete answer is beyond the scope of this text. However, it can be shown that this simple adjustment leads to a very, very good approximation of a significantly more complicated method of computing exact confidence intervals for  $p$ .

**SOLUTION TO EXAMPLE G.1.** In the epidemiologist's case, since  $n = 1200$  and  $X = 6$ , then

$$\tilde{p} = \frac{X+2}{n+4} = \frac{8}{1204} = 0.00664,$$

which is nearly one-third larger than  $\hat{p}$ . This is to be expected since we added 2 successes to  $X = 6$  while only adding 4 to  $n = 1200$ .

Using Formula G.2, the 95% confidence limits are

$$L_1 = 0.00664 - 1.960 \sqrt{\frac{(0.01039)(0.98961)}{1204}} = 0.00664 - 0.00573 = 0.00091$$

and

$$L_2 = 0.00664 + 1.960 \sqrt{\frac{(0.01039)(0.98961)}{1204}} = 0.00664 + 0.00573 = 0.01237.$$

This time the 95% confidence interval for  $p$ , the proportion of Irish women under age 35 who have experienced breast cancer, is 0.00091 to 0.01237 or between 91 and 1237 per 100,000. The confidence interval is wider (more conservative) and different in its center than the interval calculated with Method 1.

The upper half of Table G.1 illustrates the effect of sample size on the calculation of 95% confidence intervals for the population proportion using the two different methods. For very large samples, both methods will produce very similar confidence intervals. The lower half of the table shows that the closer  $p$  is to 0.5, the less pronounced the difference between the two intervals will be.

**TABLE G.1.** The effect of sample size on 95% confidence intervals using the sample proportion  $\hat{p}$  (Method 1) and the Wilson estimate  $\tilde{p}$  (Method 2).

		Standard: Method 1				Wilson: Method 2			
$n$	$X$	$\hat{p}$	SE	$L_1$	$L_2$	$\tilde{p}$	SE	$L_1$	$L_2$
200	1	0.005	0.00499	-0.00478	0.01478	0.01471	0.00843	-0.00181	0.03123
600	3	0.005	0.00288	-0.00064	0.01064	0.00828	0.00369	0.00105	0.01551
1000	5	0.005	0.00223	0.00063	0.00937	0.00697	0.00263	0.00182	0.01212
5000	25	0.005	0.00100	0.00304	0.00696	0.00540	0.00104	0.00337	0.00743
10000	50	0.005	0.00071	0.00361	0.00639	0.00520	0.00072	0.00379	0.00661
100	10	0.1	0.03000	0.04120	0.15880	0.11538	0.03133	0.05398	0.17678
100	20	0.2	0.04000	0.12160	0.27840	0.21154	0.04005	0.13305	0.29003
100	30	0.3	0.04583	0.21017	0.38983	0.30769	0.04526	0.21899	0.39639
100	40	0.4	0.04899	0.30398	0.49602	0.40385	0.04811	0.30955	0.49815
100	50	0.5	0.05000	0.40200	0.59800	0.50000	0.04903	0.40390	0.59610

## Choosing Sample Sizes

In some situations a researcher may wish to design a study that will produce a confidence interval for the population proportion  $p$  of a certain pre-specified width. This can be accomplished by using a sufficiently large sample size,  $n$ .

**EXAMPLE G.2.** Using county medical records, the epidemiologist now wishes to determine the five-year survival rate,  $p$ , of all women diagnosed with breast cancer in the 1970's. She would like to determine  $p$  to within 2%. More precisely, she wishes to construct a 95% confidence interval whose endpoints are within 2% of the five-year survival rate. How large of a sample will she need to do this?

We will carry out the analysis for this problem using the Wilson estimate method; you are asked to carry out the analysis for Method 1 in the exercises at the end of the chapter.

The epidemiologist wants to determine  $p \pm 0.02$ . Using Formula G.2 the endpoints of a  $(1 - \alpha)100\%$  confidence interval are

$$\tilde{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}.$$

We define the **margin of error** for a  $(1 - \alpha)100\%$  confidence interval for a population proportion to be

$$m = z_{1-\frac{\alpha}{2}} \text{SE}_{\tilde{p}} = z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}. \quad (\text{G.3})$$

The researcher specifies  $m$  and the confidence level. But having chosen the confidence level, the value of  $z_{1-\frac{\alpha}{2}}$  is determined. So to solve for the sample size  $n$  we simply need  $\tilde{p}$ . However, since the researcher has not yet done the study,  $\tilde{p}$  is unknown. But the researcher can't do the study until the sample size is known. We seem to be stuck.

There are two ways around this problem. If resources allow, one could do a pilot study to estimate  $\tilde{p}$ . The other solution is to observe that in (G.3), for any fixed value of  $n$ , the margin of error is largest when  $\tilde{p} = 0.5$ .<sup>1</sup> Using  $\tilde{p} = 0.5$  is the *conservative* approach as it will produce an overestimate of  $n$ .

<sup>1</sup>You may recall that for a concave down quadratic equation such as  $\tilde{p}(1-\tilde{p}) = \tilde{p} - \tilde{p}^2$ , the maximum value occurs half way between the two roots, which, in this case, is half way between 0 and 1.

With  $\tilde{p} = 0.5$  in (G.3), we have

$$m \leq z_{1-\frac{\alpha}{2}} \sqrt{\frac{0.5(1-0.5)}{n+4}}$$

or, squaring both sides of the equation,

$$m^2 \leq z_{1-\frac{\alpha}{2}}^2 \left( \frac{0.25}{n+4} \right).$$

Thus,

$$n+4 \leq 0.25 \left( \frac{z_{1-\frac{\alpha}{2}}}{m} \right)^2 = \left( \frac{z_{1-\frac{\alpha}{2}}}{2m} \right)^2.$$

Solving for  $n$ , we obtain

**FORMULA G.3.** A  $(1-\alpha)100\%$  confidence interval for a population proportion  $p$  will have a margin of error no greater than the pre-specified value  $m$  if the sample size is

$$n = \left( \frac{z_{1-\frac{\alpha}{2}}}{2m} \right)^2 - 4.$$

As mentioned, this value of  $n$  is conservative (larger than necessary). For values of  $\tilde{p}$  near 0.5 Formula G.3 will give very good estimates of the sample size required. However, if  $\tilde{p} < 0.15$  or  $\tilde{p} > 0.85$ , then the sample size given by Formula G.3 will be at least twice as large as necessary.

**SOLUTION TO EXAMPLE G.2.** We may now apply Formula G.3 to determine the sample size for the breast cancer survival rate study. Since the epidemiologist wants a 95% confidence interval with a 2% margin of error,  $\alpha = 0.05$  and  $m = 0.02$ . Thus,

$$n = \left( \frac{z_{1-\frac{\alpha}{2}}}{2m} \right)^2 - 4 = \left( \frac{1.960}{2(0.02)} \right)^2 - 4 = 2397.$$

The researcher should use a sample size of roughly 2400.

## G.2 Problems

1. To determine the frequency of type O blood (the universal donor) in a population, a random sample of 100 people were blood typed for the ABO group. Of this 100, 42 were found to be type O. Calculate the 95% confidence limits for the proportion of the population that has type O blood.
2. Dabbling ducks such as black ducks and mallards often have very skewed sex ratios. In the sample above, 10 were found to be females and the rest males. Use this information to generate a 95% confidence interval for the proportion of females in the flock.
3. In a study of the effectiveness of acupuncture to relieve the symptoms of shoulder impingement, 53 of 80 patients claimed to have complete recovery within one year of the end of a series of acupuncture treatments. Is this number significantly higher than the 50% that claim the same result with surgery? Find the 95% confidence interval for the proportion  $p$  of those obtaining complete recovery from acupuncture and use this interval to answer the question.
4. In a population where the proportion of people who have the ability to taste umami is unknown, a random sample of 100 people were tested and 74 could clearly distinguish umami from saltiness. Calculate the 95% and 99% population proportion confidence intervals for this finding.

5. Derive the analogue of Formula 4.4 for sample size for use with  $\hat{p}$  and Method 1 of computing confidence intervals.
6. (a) A pharmaceutical company wants to obtain a 95% confidence interval for the population proportion  $p$  of people who obtain relief from pain of headache in under 30 minutes using their product with a margin of error of no more than 3%. What size should their survey sample be?  
(b) How would the sample size need to change if a 99% confidence interval were required?
7. In a pilot study, a public health survey of 2000 people in the 40–49 age group in Boston reveals that 29 have AIDS. Estimate the population proportion and find a 98% confidence interval for the estimate. Express your answer as a rate per 100,000.
8. (a) A researcher is interested in determining the incidence rate of Ehlers-Danlos syndrome, a rare autosomal dominant disorder characterized by a collagen defect affecting joint hypermobility and skin hyperextensibility. Suppose that an initial pilot study is conducted to find a confidence interval for the incidence rate  $p$ . The survey uses a random sample of 500 people and locates no one with the condition. Determine a 95% confidence interval first using Method 1 and then using Method 2. Comment on the result.  
(b) Suppose instead that the survey used 1000 people and found none with the condition. Again determine 95% confidence intervals using both methods. Comment on the result.  
(c) Suppose that the true incidence rate were  $p = 0.001$ . How likely was it that a survey with 500 people locates no one with the condition? What if the survey had a sample size of 1000? Comment on this in relation to the earlier parts.
9. A 2007 telephone survey of 1,213 people in Sydney, Australia revealed that 40% of the respondents thought that global warming was a greater threat to security than terrorism. The poll indicated that there was a margin of error of no more than  $m = 3\%$ . However the confidence level  $1 - \alpha$  was not given. Determine this level. (Source: the *New York Times*, [www.nytimes.com/2007/10/04/world/asia/04australia.html](http://www.nytimes.com/2007/10/04/world/asia/04australia.html))

### G.3 Answers

1. Using Method 1,  $\hat{p} = 0.42$ .

$$L_1 = \hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.42 - 1.960\sqrt{\frac{(0.42)(0.58)}{100}} = 0.42 - 0.097 = 0.323.$$

$$L_2 = \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.42 + 1.960\sqrt{\frac{(0.42)(0.58)}{100}} = 0.42 + 0.097 = 0.517.$$

Using Method 2,  $\tilde{p} = \frac{44}{104} = 0.4231$ .

$$L_1 = \tilde{p} - z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = 0.4231 - 1.960\sqrt{\frac{(0.4231)(0.5769)}{104}} = 0.4231 - 0.0950 = 0.3281.$$

$$L_2 = \tilde{p} + z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = 0.4231 + 1.960\sqrt{\frac{(0.4231)(0.5769)}{104}} = 0.4231 + 0.0950 = 0.5181.$$

2. Using Method 1 with  $\hat{p} = 0.4$

$$L_1 = \hat{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.4 - 1.960\sqrt{\frac{0.4(0.6)}{25}} = 0.208$$

and

$$L_2 = \hat{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.592.$$

Using Method 2 with  $\hat{p} = \frac{12}{29} = 0.4138$

$$L_1 = \tilde{p} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.4138 - 1.960\sqrt{\frac{0.4138(0.5862)}{29}} = 0.2345$$

and

$$L_2 = \tilde{p} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.5931.$$

3. Using Method 1,  $\hat{p} = \frac{53}{80} = 0.6625$ . From Formula 4.3 the endpoints for a 95% confidence interval for  $p$  are

$$L_1 = \hat{p} - 1.960\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.6625 - 1.960\sqrt{\frac{0.6625(1-0.6625)}{80}} = 0.5589$$

and

$$L_2 = \hat{p} + 1.960\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.6625 + 1.960\sqrt{\frac{0.6625(1-0.6625)}{80}} = 0.7661.$$

Since 50% lies outside (below) this interval, the recovery rate  $p$  for acupuncture is significantly different (higher) than for surgery.

Using Method 2, the results are the same. This time  $\tilde{p} = \frac{55}{82} = 0.6548$ , so

$$L_1 = \tilde{p} - z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = 0.6548 - 1.960\sqrt{\frac{(0.6548)(0.3452)}{82}} = 0.6548 - 0.1017 = 0.5531;$$

$$L_2 = \tilde{p} + z\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n}} = 0.6548 + 1.960\sqrt{\frac{(0.6548)(0.3452)}{82}} = 0.6548 + 0.1017 = 0.7565.$$

4. For Method 1 use Formula 4.3 with  $\hat{p} = 0.74$ . Note that  $n\hat{p} = 100 \times 0.74 = 74 > 5$  and  $n(1 - \hat{p}) = 100 \times 0.26 = 26 > 5$ . For a 95% confidence interval, the endpoints are

$$L_1 = \hat{p} - 1.960\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad \text{and} \quad L_2 = \hat{p} + 1.960\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

Thus,

$$L_1 = 0.74 - 1.960\sqrt{\frac{(0.74)(0.26)}{100}} = 0.74 - 0.086 = 0.654$$

and  $L_2 = 0.74 + 0.086 = 0.826$ .

For a 99% confidence interval, the endpoints are

$$L_1 = 0.74 - 2.58\sqrt{\frac{(0.74)(0.26)}{100}} = 0.74 - 0.113 = 0.627$$

and  $L_2 = 0.74 + 0.113 = 0.853$ .

Using Method 2,  $\tilde{p} = \frac{74}{104} = 0.7308$ , so

$$L_1 = \tilde{p} - z\sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n}} = 0.7308 - 1.960\sqrt{\frac{(0.7308)(0.2692)}{104}} = 0.7308 - 0.0852 = 0.6456.$$

$$L_2 = \tilde{p} + z\sqrt{\frac{\tilde{p}(1 - \tilde{p})}{n}} = 0.7308 + 1.960\sqrt{\frac{(0.7308)(0.2692)}{104}} = 0.7308 + 0.0852 = 0.8160.$$

For a 99% confidence interval, the endpoints are

$$L_1 = 0.7308 - 2.58\sqrt{\frac{(0.7308)(0.2692)}{100104}} = 0.7308 - 0.1122 = 0.6186$$

and  $L_2 = 0.7308 + 0.1122 = 0.8430$ .

5. This time

$$m = z_{1-\frac{\alpha}{2}}\text{SE}_{\hat{p}} = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

Again using the conservative approach with  $p = 0.5$ ,

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{2m}\right)^2.$$

There is only a difference of 4 between the estimated sample sizes for the two methods.

6. (a) Using Formula 4.4 (for Method 2), Thus,

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{2m}\right)^2 - 4 = \left(\frac{1.960}{2(0.03)}\right)^2 - 4 = 1063.$$

Using the analogous formula for Method 1 yields

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{2m}\right)^2 = \left(\frac{1.960}{2(0.03)}\right)^2 = 1067.$$

- (b) This time using Formula 4.4

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{2m}\right)^2 - 4 = \left(\frac{2.58}{2(0.03)}\right)^2 - 4 = 1845$$

while for Method 1 we will obtain  $n = 1849$ .

7. Using Method 1 with  $\hat{p} = \frac{29}{2000} = 0.0145$

$$L_1 = \hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.0145 - 2.33 \sqrt{\frac{0.0145(0.9855)}{2000}} = 0.00827$$

and

$$L_2 = \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.02072.$$

This is the same as an incident rate of 827 to 2072 per 100,000 people in the 40–49 age group.

Using Method 2, the Wilson estimate is

$$\tilde{p} = \frac{X + 2}{n + 4} = \frac{31}{2004} = 0.01547.$$

The endpoints of a 98% confidence interval are

$$L_1 = \tilde{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.01547 - 2.33 \sqrt{\frac{0.01547(1-0.01547)}{2004}} = 0.00905$$

and

$$L_2 = 0.01547 + 2.33 \sqrt{\frac{0.01547(1-0.01547)}{2004}} = 0.02189.$$

This is the same as an incident rate of 905 to 2189 per 100,000 people in the 40–49 age group.

8. (a) Using Method 1 with  $\hat{p} = \frac{0}{500} = 0$  yields

$$L_1 = 0 - 1.96 \sqrt{\frac{0(1)}{500}} = 0 \quad \text{and} \quad L_2 = 0 + 1.96 \sqrt{\frac{0(1)}{500}} = 0.$$

The 95% confidence interval consists of the single point 0.

Using Method 2 with  $\tilde{p} = \frac{2}{504} = 0.00397$  yields

$$L_1 = 0.00397 - 1.96 \sqrt{\frac{0.00397(1-0.00397)}{504}} = -0.00152$$

and

$$L_2 = 0.00397 + 1.96 \sqrt{\frac{0.00397(1-0.00397)}{504}} = 0.00946.$$

Since a proportion cannot be negative,  $L_1 = 0$ . So the 95% confidence interval using Method 2 is  $[0, 0.00946]$ . The result from Method 1 cannot be correct. It suggests that we are certain that the incidence rate is 0. Method 2 gives a more reasonable estimate. Notice that even though the sample size is “large”, given that  $\hat{p} = 0$ , we should be suspicious of using Method 1 which requires  $np(1-p) > 5$ .

- (b) Using Method 1 with  $\hat{p} = 0$  again yields  $L_1 = L_2 = 0$ . Using Method 2 with  $\tilde{p} = \frac{2}{1004} = 0.00199$  yields

$$L_1 = 0.00199 - 1.96 \sqrt{\frac{0.00199(1-0.00199)}{1004}} = -0.00077$$

and

$$L_2 = 0.00199 + 1.96 \sqrt{\frac{0.00199(1-0.00199)}{1004}} = 0.00475.$$

The 95% confidence interval using Method 2 is  $[0, 0.00475]$ . Method 1 is still inadequate. Doubling the sample size decreased the size of the confidence interval for Method 2.

(c) Use a binomial distribution with  $p = 0.001$ . If  $n = 500$ , then

$$P(X = 0) = \binom{500}{0} (0.001)^0 (0.999)^{500} = 0.60638.$$

If  $n = 1000$ , then

$$P(X = 0) = \binom{1000}{0} (0.001)^0 (0.999)^{1000} = 0.36770.$$

In the first case, there is a 60% chance that no one with the condition will be found, yet the Method 1 interval does not contain the correct proportion (and consists of a single point). In the second case, there is nearly a 37% chance that no one with the condition will be found and Method 1 again fails to capture the true proportion. In other words, in two situations where the outcome of the survey is (very) likely, Method 1 fails to capture the true proportion. Notice that Method 2 does.

9. Using Formula 4.4 (for Method 2),

$$\begin{aligned} n &= \left( \frac{z_{1-\frac{\alpha}{2}}}{2m} \right)^2 - 4 \implies 1213 = \left( \frac{z_{1-\frac{\alpha}{2}}}{2(0.03)} \right)^2 - 4 \\ &\implies 1217 = \left( \frac{z_{1-\frac{\alpha}{2}}}{0.06} \right)^2 \\ &\implies \sqrt{1217} = \frac{z_{1-\frac{\alpha}{2}}}{0.06} \\ &\implies z_{1-\frac{\alpha}{2}} = 0.06\sqrt{1217} \approx 2.09. \end{aligned}$$

From Table C.3,  $1 - \frac{\alpha}{2} = 0.9817$  so  $\frac{\alpha}{2} = 0.183$  or  $1 - \alpha = 0.9634$ .

Using the analogous formula for Method 1 yields  $z_{1-\frac{\alpha}{2}} = 0.06\sqrt{1214} \approx 2.09$  and again  $1 - \alpha = 0.9634$ .

# Field Methods

## Concepts in Appendix D:

- The Petersen Method
- The Point-Centered Quarter Method
- Diversity Indices

### H.1 Estimating the Size of a Population with the Petersen Method

One of the most important and difficult to measure characteristics of a natural population is its size. Animals that are highly mobile or that have ephemeral life histories are particularly hard to census. Various capture-mark-recapture methods have been employed to solve this problem. One of the first ecological uses of mark and recapture was developed by C. G. J. Petersen in 1896 to study fish movements and subsequently to estimate fish population sizes. His method is relatively straight forward and utilizes concepts developed earlier in this text.

#### Assumptions

The assumptions of the Petersen Method are:

- The population is closed, that is, the population does not change size during the study period because of births, deaths, or movements.
- All animals in the population have the same chance of getting caught in a sample.
- Marking individuals does not affect their catchability.
- Animals do not lose their marks between sampling periods and all marked animals in the second sample are identified.

#### The Petersen Method

The basic procedure is to capture a number of animals in a short period of time, mark and release them, then to capture a second random sample and check these individuals for marks. The following data are obtained:

- $M$  is the number of individuals marked and released in the first sample ( $M$  = marked).
- $C$  is the total number of individuals captured in the second sample ( $C$  = captured).
- $R$  is the number of individuals in the second sample that are marked ( $R$  = recaptured).

Using these variables we can estimate  $N$ , the size of the population at the time of marking. Utilizing a proportionality argument

$$\frac{N}{M} = \frac{C}{R}.$$

That is, the number marked ( $M$ ) of the total population ( $N$ ) should be proportional to the number recaptured ( $R$ ) out of the second random sample ( $C$ ). So  $\hat{N}$  is the estimated population size

$$\hat{N} = \frac{CM}{R}. \quad (\text{H.1})$$

This estimate has been widely used because it is intuitively clear but, unfortunately, it produces a biased estimator of the population size, tending to overestimate the actual population. This bias can be large for small sample sizes and Seber has suggested the following modification

$$\hat{N} = \frac{(C+1)(M+1)}{R+1} - 1 \quad (\text{H.2})$$

which is unbiased if  $(C+M) > N$  and is nearly unbiased if  $R > 7$ , that is, if there are at least seven recaptures.

Once the population size,  $\hat{N}$ , has been estimated with one of these methods, it is usually appropriate to generate some measure of the confidence in the estimate. Confidence intervals may be used to provide a clearer understanding of the accuracy of the population size estimate. Unfortunately there are three forms of confidence intervals. *Each is appropriate for different types of experimental outcomes.*

- a) If the ratio  $\frac{R}{C} > 0.10$ , use binomial confidence intervals to generate confidence limits for  $N$ . This method involves finding the confidence limits for the proportion  $\frac{R}{C}$  and then using the lower and upper values of  $\frac{R}{C}$  in the  $\hat{N} = \frac{CM}{R}$  equation to find the confidence limits for  $\hat{N}$ .

The lower limit for  $\frac{R}{C}$  is

$$L_1 = \frac{R}{R + (C - R + 1)F_{\frac{\alpha}{2}}(\nu_1, \nu_2)}$$

where  $\nu_1 = 2(C - R + 1)$  and  $\nu_2 = 2R$ .

The upper limit for  $\frac{R}{C}$  is

$$L_2 = \frac{(R+1)F_{\frac{\alpha}{2}}(\nu'_1, \nu'_2)}{C - R + (R+1)F_{\frac{\alpha}{2}}(\nu'_1, \nu'_2)}$$

where  $\nu'_1 = 2(R+1) = \nu_2 + 2$  and  $\nu'_2 = 2(C - R) = \nu_1 - 2$ .

**EXAMPLE H.1.** Suppose in a large study of mammal populations, the first night 250 ( $M$ ) white-footed mice, *Peromyscus leucopus*, were caught in live traps and marked with fluorescent paint then released. The second night 200 ( $C$ ) mice were caught of which 50 ( $R$ ) were marked. Estimate the mouse population size and find the appropriate 95% confidence interval for this estimate.

**SOLUTION.** Using Petersen's original equation (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{200(250)}{50} = 1000.$$

Using Seber's modified equation (H.2)

$$\hat{N} = \frac{(C+1)(M+1)}{R+1} - 1 = \frac{(201)(251)}{51} - 1 = 988.$$

Next

$$\frac{R}{C} = \frac{50}{200} = 0.25.$$

Using  $\nu_1 = 2(200 - 50 + 1) = 302$  and  $\nu_2 = 2(50) = 100$ , The lower limit for  $\frac{R}{C}$  is

$$L_1 = \frac{50}{50 + (200 - 50 + 1)F_{0.025}(302,100)} = 0.187.$$

Using  $\nu'_1 = 100 + 2 = 102$  and  $\nu'_2 = 302 - 2 = 300$ , the upper limit for  $\frac{R}{C}$  is

$$L_2 = \frac{(50+1)F_{0.025}(102,300)}{200 - 50 + (51)F_{0.025}(102,300)} = 0.329$$

Using the lower and upper confidence limits for  $\frac{R}{C}$  in the formula for  $\hat{N}$ , we find that:

$$\text{the lower 95\% confidence limit for } N \text{ is: } \frac{1}{0.329}(250) = 769$$

$$\text{the upper 95\% confidence limit for } N \text{ is: } \frac{1}{0.187}(250) = 1337.$$

We are 95% confident that  $N$  is between 769 and 1337 mice.

- b) If the ratio  $\frac{R}{C} < 0.10$  and  $R > 50$ , then use a normal approximation to develop the confidence limits for  $N$ . Again, first find confidence limits for  $\frac{R}{C}$  and then use these in the Petersen formula to generate the limits for  $N$ . Here the lower and upper limits for the  $\frac{R}{C}$  ratio are

$$L_1 = \frac{R}{C} - z_{\frac{\alpha}{2}} \sqrt{\frac{(\frac{R}{C})(1 - \frac{R}{C})}{C - 1}}$$

and

$$L_2 = \frac{R}{C} + z_{\frac{\alpha}{2}} \sqrt{\frac{(\frac{R}{C})(1 - \frac{R}{C})}{C - 1}}.$$

**EXAMPLE H.2.** If 1200 mice were caught on day one and 800 on day two with 60 of them being recaptures, find the estimated population size and its 95% confidence limits.

**SOLUTION.** In this case, using (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{800(1200)}{60} = 16,000.$$

Since  $\frac{R}{C} = \frac{60}{800} = 0.075 < 0.10$  and  $R = 60 > 50$ , use the normal approximation for the confidence interval. The confidence limits for  $\frac{R}{C}$  are

$$L_1 = \frac{60}{800} - 1.960 \sqrt{\frac{(\frac{60}{800})(1 - \frac{60}{800})}{800 - 1}} = 0.075 - 0.0183 = 0.0567$$

and

$$L_2 = \frac{60}{800} + 1.960 \sqrt{\frac{(\frac{60}{800})(1 - \frac{60}{800})}{800 - 1}} = 0.075 + 0.0183 = 0.0933.$$

We find that:

$$\begin{aligned} \text{the lower 95\% confidence limit for } N \text{ is: } & \frac{1}{0.0933}(1200) = 12,861 \\ \text{the upper 95\% confidence limit for } N \text{ is: } & \frac{1}{0.0567}(1200) = 21,315. \end{aligned}$$

So we are 95% confident that  $N$  is between 12,861 and 21,315 mice.

- c) Finally, if the  $\frac{R}{C} < 0.10$  and  $R < 50$ , use Poisson confidence intervals for  $R$  to generate confidence limits for  $N$  as follows. The lower confidence limit for  $R$  is

$$L_1 = \frac{\chi_{\frac{\alpha}{2}}^2}{2},$$

where  $\nu = 2R$ . The upper limit for  $R$  is

$$L_2 = \frac{\chi_{1-\frac{\alpha}{2}}^2}{2},$$

where  $\nu = 2(R + 1)$ .

**EXAMPLE H.3.** If 200 mice were caught in day one and 150 on day two with 12 being recaptures, find the estimated population size and its 95% confidence limits.

**SOLUTION.** This time using (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{150(200)}{12} = 3125.$$

Since  $\frac{R}{C} = \frac{12}{150} = 0.08 < 0.10$  and  $R = 12 < 50$ , use the Poisson technique to generate confidence intervals. Confidence limits for  $R$  are

$$L_1 = \frac{\chi_{0.025(24)}^2}{2} = \frac{12.4}{2} = 6.2,$$

where  $\nu = 2(12) = 24$ . The upper limit for  $R$  is

$$L_2 = \frac{\chi_{0.975(26)}^2}{2} = \frac{41.9}{2} = 20.95,$$

where  $\nu = 2(12 + 1) = 26$ . We find that:

$$\begin{aligned} \text{the lower 95\% confidence limit for } N \text{ is: } & \frac{150(200)}{20.95} = 1431 \\ \text{the upper 95\% confidence limit for } N \text{ is: } & \frac{150(200)}{6.2} = 4839 \end{aligned}$$

So we estimate this population size  $N$  to be between 1431 and 4839 mice.

Although the confidence intervals are calculated somewhat differently in each of the three cases above, their interpretation remains the same as with any confidence interval discussed earlier (see Chapter 4 and Chapter 10). For further discussion of Petersen's Method and other more complex mark and recapture techniques see Krebs (1999).

## H.2 The Point-Centered Quarter Method

### Introduction and History

A wide variety of methods have been used to study forest structure parameters such as population density, basal area, and biomass. While these are sometimes estimated using aerial surveys or photographs, most studies involve measurement of these characteristics for individual trees using a number of different sampling methods. These methods fall into two broad categories: plot-based and plot-less. Plot-based methods begin with one or more plots (quadrats, belts) of known area in which the characteristics of interest are measured for each plant. In contrast, plot-less methods involve measuring distances for a random sample of trees, typically along a transect, and recording the characteristics of interest for this sample. The point-centered quarter method is one such plot-less method.

The advantage to using plot-less methods rather than standard plot-based techniques is that they tend to be more efficient. Plot-less methods are faster, require less equipment, and may require fewer workers. However, the main advantage is speed. The question, then, is whether accuracy is sacrificed in the process.

Stearns (1949) indicated that the point-centered quarter method dates back a least 150 years and was used by surveyors in the mid-nineteenth century making the first surveys of government land. In the late 1940s and early 1950s, several articles appeared that described a variety of plot-less methods and compared them to sampling by quadrats. In particular, Cottam, Curtis, and Hale (1953) compared the point-centered quarter method to quadrat sampling and derived empirically a formula that could be used to estimate population density from the distance data collected. Since the current paper is intended as an introduction to these methods, it is worth reminding ourselves what the goal of these methods is by recalling part of the introduction to their paper:

As our knowledge of plant communities increases, greater emphasis is being placed on the methods used to measure the characteristics of these communities. Succeeding decades have shown a trend toward the use of quantitative methods, with purely descriptive methods becoming less common. One reason for the use of quantitative techniques is that the resulting data are not tinged by the subjective bias of the investigator. The results are presumed to represent the vegetation as it actually exists; any other investigator should be able to employ the same methods in the same communities and secure approximately the same data.

Under the assumption that trees are distributed randomly throughout the survey site, Morisita (1954) provided a mathematical proof for the formula that Cottam, Curtis, and Hale (1953) had derived empirically for the estimation of population density using the point-centered quarter method. In other words, the point-centered quarter method could, in fact, be used to obtain accurate estimates of population densities with the advantage that the point-centered quarter method data could be collected more quickly than quadrat data. Subsequently, Cottam and Curtis (1956) provided a more detailed comparison of the point-centered quarter method and three other plot-less methods (the closest individual, the nearest neighbor, and the random pairs methods). Their conclusion was:

The quarter method gives the least variable results for distance determinations, provides more data per sampling point, and is the least susceptible to subjective bias...

It is the opinion of the authors that the quarter method is, in most respects, superior to the other distance methods studied, and its use is recommended.

Beasom and Haucke (1975) compared the same four plotless methods and also concluded that point-centered quarter method provides the most accurate estimate of density. In a comparison of a more diverse set of methods (Engeman, et al. 1994) have a more nuanced opinion of whether the point-centered quarter method is more efficient in the field and more accurate in its density estimates, especially in situations where individuals are not distributed randomly.

In recent years, as the point-centered quarter method has been used more widely, variations have been proposed by Dahdouh-Guebas and Koedam (2006) to address a number of practical problems that arise in the field (multi-stem trees, quarters where no trees are immediately present).

One use of the point-centered quarter method is to determine the **relative importance** of the various tree species in a community. The term “importance” can mean many things depending on the context. An obvious factor influencing the importance of a species to a community is the number of trees present of that species. However, the importance of some number of small trees is not the same as the importance of the same number of large trees. So the size of the trees also plays a role. Further, how the trees are distributed throughout the community also has an effect. A number of trees of the same species clumped together should have a different importance value than the same number of trees distributed more evenly throughout the community.

Measuring importance can aid understanding the successional stages of a forest habitat. At different stages, different species of trees will dominate. Importance values are one objective way of measuring this dominance.

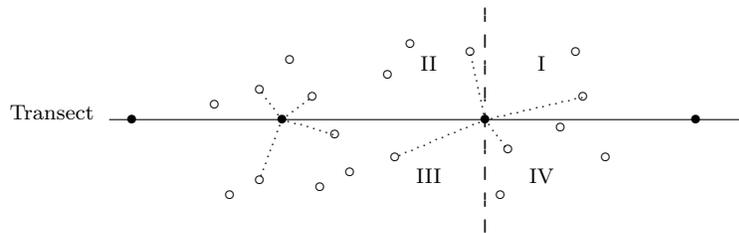
The three factors that we will use to determine the importance value of a species are the density, the size, and the frequency (distribution). Ideally, to estimate these factors, one would take a large sample, measuring, say, all the trees in a  $100 \times 100$  meter square (a hectare). This can be extraordinarily time consuming if the trees are very dense. The point-centered quarter method provides a quick way to make such estimates by using a series of measurements along a transect.

## Materials and Methods

The procedure outlined below describes how to carry out point-centered quarter method data collection along a 100 m transect. It can be scaled up or down, as appropriate, for longer or shorter transects. While this analysis can be carried out alone, groups of two or three can make for very efficient data collection. Material requirements include 50 or 100 meter tape, a shorter 5 or 10 meter tape, a notebook, a calculator, and a table of random numbers (Table H.13) if the calculator cannot generate them.

1. Generate a list of 15 to 20 random two-digit numbers. If the difference of any two is 4 or less, cross out the second listed number. There should be 10 or more two-digit numbers remaining; if not, generate additional ones. List the first 10 remaining numbers in increasing order. It is important to generate this list before doing any measurements.
2. Lay out a 100 m transect (or longer or shorter as required).

3. The random numbers represent the distances along the transect at which data will be collected. Random numbers are used to eliminate bias. Everyone always wants to measure that BIG tree along the transect, but such trees may not be representative of the community.<sup>1</sup> The reason for making sure that points are at least 5 meters apart is so that the same trees will not be measured repeatedly. Caution: If trees are particularly sparse, both the length of the transect and the minimum distance between points may need to be increased.
4. The smallest random number determines the first sampling point along the transect. At this (and every sampling) point, run an imaginary line perpendicular to the transect. This line and the transect divide the world into four quarters (hence the name, point-centered quarter method).



**FIGURE H.1.** Sample points along a transect with the nearest trees in each quarter indicated by ·····.

5. Select one of the quarters. In that quarter, locate the tree nearest to the sampling point. For the purposes of this exercise, to be counted as a “tree” it should have a minimum diameter of 4 cm or, equivalently, a minimum circumference of 12.5 cm. (Caution: In other situations, different minimum values may apply.)

For the each sampling point, record:

- (a) the quarter number (I, II, III, or IV);
- (b) the distance from the sampling point to the center of the trunk of the tree to the nearest 0.1 m (Caution: Review Chapter 1 on the 30–300 Rule.);
- (c) the species of the tree;
- (d) and the diameter at breast height (DBH) or circumference at chest height (CCH) to the nearest cm, but again observe the 30–300 Rule.

**Note:** Brokaw and Thompson (2000) have shown that it is important to use the same height to measure the diameter or circumference. They suggest using a standard height of 130 cm and employing the notation  $D_{130}$  rather than DBH to indicate this. Whatever height is used should be explicitly noted in the results.

**Note:** Tree calipers are an easy way to measure diameters, but are often unavailable. It may be more convenient to measure the girth (circumference) of each tree instead of the diameter.

---

<sup>1</sup>Even Cottam and Curtis (1956) warn us about this tendency: “Repeated sampling of the same stand with different investigators indicates that some individuals have a tendency to place the sampling points so that large or unusual trees occur more commonly than they occur in the stand.”

**Cautions:** If a tape is used to measure DBH, avoid protrusions on the trunk. If calipers are used, an average from three caliper readings is recorded. If girths are recorded, rather than convert each girth to a diameter, change the column heading from DBH to CCH. Make the appropriate scaling adjustment in later calculations whenever diameters are involved.

See Table H.1 for how this data should be organized. Repeat this for the other three quarters at this sampling point. If a tree species cannot be identified, simply record it as A, B, C, etc., and collect and label a sample leaf that for comparison purposes at other quarters and later taxonomic identification.

6. Repeat this process for the entire set of sampling points.
7. Carry out the data analysis as described below.

For trees with multiple trunks at breast height, record the diameter (circumference) of each trunk separately. What is the minimum allowed diameter of each trunk in a such multi-trunk tree? Such decisions should be spelled out in the methods section of the resulting report. At a minimum, one should ensure that the combined cross-sectional areas of all trunks meet the previously established minimum cross-sectional area for a single trunk tree. For example, with a 4 cm minimum diameter for a single trunk, the minimum cross-sectional area is

$$\pi r^2 = \pi(2)^2 = 4\pi \approx 12.6 \text{ cm}^2.$$

## Data Organization and Notation

### The Data Layout

Table H.1 illustrates how the data should be organized for the point-centered quarter method analysis. Note the multi-trunk Acacia (8 cm, 6 cm; D<sub>130</sub>) in the third quarter at the second sampling point. The only calculation required at this stage is to sum the distances from the sample points to each of the trees that was measured. **Note:** A sample of only five points as in Table H.1 is too few for most studies. These data are presented only to illustrate the method of analysis in a concise way.

### Notation

We will use the following notation throughout this section.

$n$	the number of sample points along the transect
$4n$	the number of samples or observations one for each quarter at each point
$i$	a particular transect point, where $i = 1, \dots, n$
$j$	a quarter at a transect point, where $j = 1, \dots, 4$
$R_{ij}$	the point-to-tree distance at point $i$ in quarter $j$

For example, the sum of the distances in the Table H.1 is

$$\sum_{i=1}^5 \sum_{j=1}^4 R_{ij} = 40.9.$$

**TABLE H.1.** Field data organized for point-centered quarter method analysis.

Sampling Point	Quarter No.	Species	Distance (m)	$D_{130}$ (cm)
1	1	Acacia	1.1	6
	2	Eucalyptus	1.6	48
	3	Casuarina	2.3	15
	4	Callitris	3.0	11
2	1	Eucalyptus	2.8	65
	2	Casuarina	3.7	16
	3	Acacia	0.9	8, 6
	4	Casuarina	2.2	9
3	1	Acacia	2.8	4
	2	Acacia	1.1	6
	3	Acacia	3.2	6
	4	Acacia	1.4	5
4	1	Callitris	1.3	19
	2	Casuarina	0.8	22
	3	Casuarina	0.7	12
	4	Callitris	3.1	7
5	1	Acacia	1.5	7
	2	Acacia	2.4	5
	3	Eucalyptus	3.3	27
	4	Eucalyptus	1.7	36
Total			40.9	

## Basic Analysis

The next three subsections outline the estimation of density, frequency, and cover. The most widely studied of the three is density. In Section H.2 we present a more robust way to determine the both a point estimate and a confidence interval for population density. In this section density, frequency, and cover are defined both in absolute and relative terms. The relative measures are then combined to create a measure of relative importance.

## Density

### Absolute Density

The **absolute density**  $\lambda$  of trees is defined as the number of trees per unit area. Since  $\lambda$  is most easily estimated per square meter and since a hectare is 10,000 m<sup>2</sup>,  $\lambda$  is often multiplied by 10,000 to express the number of tree per hectare. The distances measured using the point-centered quarter method may be used to estimate  $\lambda$  to avoid having to count every tree within such a large area.

Note that if  $\lambda$  is given as trees/m<sup>2</sup>, then its reciprocal  $1/\lambda$  is the mean area occupied by a single tree. This observation is the basis for the following estimate of  $\lambda$ . (Also see Section H.2.)

From the transect information, determine the **mean distance**  $\bar{r}$ , which is the sum of the

nearest neighbor distances in the quarters surveyed divided by the number of quarters,

$$\bar{r} = \frac{\sum_{i=1}^n \sum_{j=1}^4 R_{ij}}{4n}.$$

For the data in Table H.1,

$$\bar{r} = \frac{40.9}{20} = 2.05 \text{ m.}$$

Cottam, Curtis, and Hale (1953) showed empirically and Morisita (1954) demonstrated mathematically that  $\bar{r}$  is actually an estimate of  $\sqrt{1/\lambda}$ , the square root of the mean area occupied by a single tree. Consequently, an estimate of the density is given by

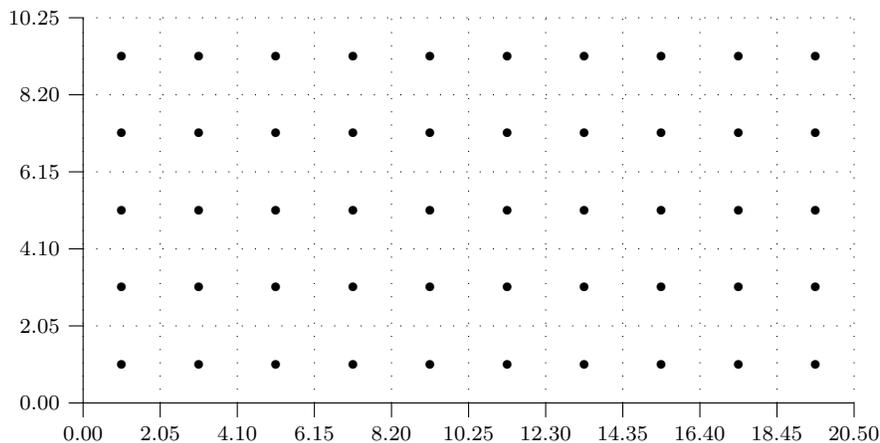
$$\text{Absolute density} = \tilde{\lambda} = \frac{1}{\bar{r}^2} = \frac{16n^2}{\left(\sum_{i=1}^n \sum_{j=1}^4 R_{ij}\right)^2}. \quad (\text{H.3})$$

For the data in Table H.1,

$$\tilde{\lambda} = \frac{1}{\bar{r}^2} = \frac{1}{2.05^2} = 0.2380 \text{ trees/m}^2,$$

or, equivalently, 2380 trees/ha.

One way to “see this” is to imagine a forest where the trees are uniformly distributed on a square grid whose sides are  $\bar{r} = 2.05$  m long. If a tree is located at the center of each square in this “forest,” then the mean distance  $\bar{r}$  between trees is 2.05 m. Such a forest is illustrated in Figure H.2. Each tree occupies a square side 2.05 m and so the density is  $1/2.05^2 = 0.2380$  trees/m<sup>2</sup>. Though such a uniform arrangement of trees violates the assumption of randomness, the figure does illustrate what is happening “on average” or in the mean.



**FIGURE H.2.** A grid-like forest with trees uniformly dispersed so that the nearest neighbor is 2.05 m.

### Absolute Density of Each Species

The absolute density of an individual species is the expected number of trees of that species per square meter (or hectare). The **absolute density  $\lambda_k$  of species  $k$**  is estimated as the proportion of quarters in which the species is found times the absolute density of all trees.

$$\hat{\lambda}_k = \frac{\text{Quarters with species } k}{4n} \times \hat{\lambda}. \quad (\text{H.4})$$

Table H.2 gives the absolute density for each species in Table H.1.

**TABLE H.2.** The absolute density of each species.

Species	Frequency/Quarter	Trees/ha
Acacia	$8/20 = 0.40$	$0.40 \times 2380 = 952$
Eucalyptus	$4/20 = 0.20$	$0.20 \times 2380 = 476$
Casuarina	$5/20 = 0.25$	$0.25 \times 2380 = 595$
Callitris	$3/20 = 0.15$	$0.15 \times 2380 = 357$
Total		2380

### Relative Density of a Species

The **relative density** of each species is the percentage of the total number observations of that species,

$$\text{Relative density (Species } k) = \frac{\hat{\lambda}_k}{\hat{\lambda}} \times 100.$$

Equivalently by making use of (H.4), we may define

$$\text{Relative density (Species } k) = \frac{\text{Quarters with species } k}{4n} \times 100. \quad (\text{H.5})$$

In the current example, using the first definition, the relative density of a species can be found by making use of the data in column 3 of Table H.2. For example,

$$\text{Relative density of Eucalyptus} = \frac{476}{2380} \times 100 = 20.0.$$

Using the alternative method in (H.5) as a check on earlier calculations we see that the relative density is just the proportion in column 2 of Table H.2 times 100. For example,

$$\text{Relative density of Eucalyptus} = \frac{4}{20} \times 100 = 20.0.$$

The relative densities should sum to 100 plus or minus a tiny round-off error.

**TABLE H.3.** The relative density of each species.

Species	Relative Density
Acacia	40.0
Eucalyptus	20.0
Casuarina	25.0
Callitris	15.0

Based on simulations, Cottam, Curtis, and Hale (1953) suggest that about 30 individuals of a particular species must be present in the total sample before confidence can be placed in any statements about relative frequency.

## Cover or Dominance of a Species

### Absolute Cover

The cover or dominance of an individual tree is measured by its **basal area** or cross-sectional area. Let  $d$ ,  $r$ ,  $c$ , and  $A$  denote the diameter, radius, circumference, and basal area of a tree, respectively. Since the area of a circle is  $A = \pi r^2$ , it is also  $A = \pi(d/2)^2 = \pi d^2/4$ . Since the circumference is  $c = 2\pi r$ , then the area is also  $A = c^2/4\pi$ . Either  $A = \pi d^2/4$  or  $A = c^2/4\pi$  can be used to determine basal area, depending on whether DBH or CCH was recorded in Table H.1.

The first step is to compute the basal area for each tree sampled, organizing the data by species. This is the most tedious part of the analysis. A calculator that can handle lists of data or a spreadsheet can be very handy at this stage. For the data in Table H.1, the basal area for each tree was obtained using the formula  $A = \pi d^2/4$ . For trees with multiple trunks, the basal area for each trunk was computed separately and the results summed. (See Acacia in Table H.4.)

TABLE H.4. The basal area of each tree.

Acacia		Eucalyptus		Casuarina		Callitris		Total
D <sub>130</sub> (cm)	Area (cm <sup>2</sup> )							
6	28.3	48	1809.6	15	176.7	11	95.0	
8, 6	78.5	65	3318.3	16	201.1	19	283.5	
4	12.6	27	572.6	9	63.6	7	38.5	
6	28.3	36	1017.9	22	380.1			
6	28.3			12	113.1			
5	19.6							
7	38.5							
5	19.6							
Total BA	253.7		6718.4		934.6		417.0	8323.7
Mean BA	31.71		1679.60		186.92		139.00	416.19

Next, determine the total cover or basal area of the trees in the sample by species, and then calculate the mean basal area for each species.<sup>2</sup> Be careful when computing the means as the number of trees for each species will differ. Remember that each multi-trunk tree counts as a single tree.

The **absolute cover** or **dominance** of each species is expressed as its basal area per hectare. This is obtained by taking the number of trees per species from Table H.2 and multiplying by the corresponding mean basal area in Table H.4. The units for cover are m<sup>2</sup>/ha (not cm<sup>2</sup>/ha), so a conversion factor is required. For Acacia,

$$\text{Absolute Cover (Acacia)} = 31.71 \text{ cm}^2 \times \frac{952}{\text{ha}} \times \frac{1 \text{ m}^2}{10,000 \text{ cm}^2} = 3.0 \frac{\text{m}^2}{\text{ha}}.$$

Finally, calculate the total cover per hectare by summing the per species covers.

<sup>2</sup>Note: Mean basal area **cannot** be calculated by finding the mean diameter for each species and then using the formula  $A = \pi d^2/4$ .

**TABLE H.5.** The total basal area of each species.

Species	Mean BA (cm <sup>2</sup> )	Number/ha	Total BA/ha (m <sup>2</sup> /ha)
Acacia	31.71	952	3.0
Eucalyptus	1679.60	476	79.9
Casuarina	186.92	595	11.1
Callitris	139.00	357	5.0
Total Cover/ha			99.0

### Relative Cover (Relative Dominance) of a Species

The **relative cover** or **relative dominance** [see Cottam and Curtis (1956)] for a particular species is defined to be the absolute cover for that species divided by the total cover times 100 to express the result as a percentage. For example, for Eucalyptus,

$$\text{Relative cover (Eucalyptus)} = \frac{79.9 \text{ m}^2/\text{ha}}{99.0 \text{ m}^2/\text{ha}} \times 100 = 80.7.$$

The relative covers should sum to 100% plus or minus a tiny round-off error. Note that the relative cover can also be calculated directly from the transect information in Table H.4.

$$\text{Relative cover (Species } k) = \frac{\text{Total BA of species } k \text{ along transect}}{\text{Total BA of all species along transect}} \times 100. \quad (\text{H.6})$$

For example,

$$\text{Relative cover (Eucalyptus)} = \frac{6718.4 \text{ cm}^2}{8323.7 \text{ cm}^2} \times 100 = 80.7.$$

**TABLE H.6.** The relative cover of each species.

Species	Relative Cover
Acacia	3.0
Eucalyptus	80.7
Casuarina	11.2
Callitris	5.1

### The Frequency of a Species

#### Absolute Frequency of a Species

The **absolute frequency** of a species is the percentage of sample points at which a species occurs. Higher absolute frequencies indicate a more uniform distribution of a species while lower values may indicate clustering or clumping. It is defined as

$$\text{Absolute frequency} = \frac{\text{No. of sample points with a species}}{\text{Total number of sample points}} \times 100. \quad (\text{H.7})$$

For example,

$$\text{Absolute frequency (Acacia)} = \frac{4}{5} \times 100 = 80\%.$$

Note that absolute frequency is based on the number of sample points, not the number of quarters!

**TABLE H.7.** The absolute cover of each species.

Species	Absolute Frequency
Acacia	$(4/5) \times 100 = 80$
Eucalyptus	$(3/5) \times 100 = 60$
Casuarina	$(3/5) \times 100 = 60$
Callitris	$(2/5) \times 100 = 40$
Total	240

Note that the total will sum to more than 100%.

### Relative Frequency of a Species

To normalize for the fact that the absolute frequencies sum to more than 100%, the **relative frequency** is computed. It is defined as

$$\text{Relative frequency} = \frac{\text{Absolute frequency of a species}}{\text{Total frequency of all species}} \times 100. \quad (\text{H.8})$$

For example,

$$\text{Relative frequency (Acacia)} = \frac{80}{240} \times 100 = 33.3.$$

The relative frequencies should sum to 100 plus or minus a tiny round-off error.

**TABLE H.8.** The relative frequency of each species.

Species	Relative Frequency
Acacia	33.3
Eucalyptus	25.0
Casuarina	25.0
Callitris	16.7

What is the difference between relative frequency and relative density? A high relative frequency indicates that the species occurs near relatively many different sampling points, in other words, the species is well-distributed along the transect. A high relative density indicates that the species appears in a relatively large number of quarters. Consequently, if the relative density is high and the relative frequency is low, then the species must appear in lots of quarters but only at a few points, that is, the species appears in clumps. If both are high, the distribution is relatively even and relatively common along the transect. If the relative density is low (appears in few quarters) and the relative frequency is high(er), then the species must be sparsely distributed (few plants, no clumping).

### The Importance Value of a Species

The **importance value** of a species is defined as the sum of the three relative measures:

$$\text{Importance value} = \text{Relative density} + \text{Relative cover} + \text{Relative frequency}. \quad (\text{H.9})$$

The importance value gives equal weight to the three factors of relative density, cover, and frequency. This means that small trees (that is, with small basal area) can be dominant only if there are enough of them widely distributed across the transect. The importance value can range from 0 to 300.

For the data in Table H.1, even though eucalypti are not very common, because of their size they turn out to be the most important species within the community.

**TABLE H.9.** The importance value of each species.

Species	Relative Density	Relative Cover	Relative Frequency	Importance
Acacia	40.0	3.0	33.3	76.3
Eucalyptus	20.0	80.7	25.0	125.7
Casuarina	25.0	11.2	25.0	61.2
Callitris	15.0	5.1	16.7	36.8

**Comment.** Each of the measures that make up relative importance may be calculated without knowing the absolute density of the trees at the site (review (H.5), (H.6), and (H.8).) In fact, any estimate for the absolute density of all species leads to the same relative densities for each species. Consequently, the actual value of density of the plot is not needed to determine relative importance. However, in most studies, absolute density is one the parameters of greatest interest. Because of this, there have been a number of different methods to estimate absolute density from point-centered quarter method data proposed in the literature. In the next section we explore one of these. Whichever method is used, relative importance is unaffected.

It has been shown by Pollard (1971) that the estimate of Cottam and Curtis (1956) of  $\lambda$  in (H.3) is biased.<sup>3</sup> Nonetheless, this estimate appears widely in the literature and, so, has been used here. Another drawback of the estimate in (H.3) is that no confidence limits are available for it. The next section addresses both of these issues.

## Population Density Reconsidered

Pollard (1971) and Seber (1982) derived an unbiased estimate of the absolute population density using point-centered quarter method data that we now present. It also has the advantage that it can be used to determine confidence intervals for the density estimate.

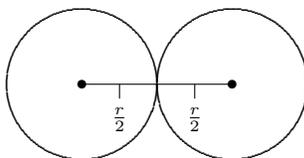
### Intuition

The discussion that follows is meant to inform our intuition and by no means constitutes a proof of any of the results, which requires a substantially more sophisticated argument. See (Mitchell, 2007) for further details).

The assumption of this model is that trees are randomly distributed in the survey area. Now think of the random points along the transect as representing “virtual trees”. The measured distance  $R_{ij}$  is a nearest neighbor distance from a virtual to a real tree. As such, it is an estimate of the actual mean nearest neighbor tree-to-tree distance.

<sup>3</sup>Pollard (1971) states that the reason for this is Cottam and Curtis (1956) chose to estimate the mean area  $A$  occupied by a tree as the reciprocal of  $\lambda$ . Rather than estimate  $A$  directly, as we saw in (H.3) they estimated  $\bar{r}$ , which is the reciprocal of the square root of  $A$ . Squaring and inverting leads to a biased estimate of  $A$ .

If an actual tree-to-tree distance were  $r$  meters, we could draw circles of radius  $r/2$  centered at each tree. See Figure H.3. Notice that the circles would not overlap and that only one tree would lie in each circle.



**FIGURE H.3.** When trees are  $r$  units apart, circles of radius  $r/2$  centered at each tree do not overlap and only one tree would lie in each circle.

The area of each circle is  $\pi(r/2)^2 = \pi r^2/4$  m<sup>2</sup>. Since there is exactly 1 tree per circle and since the circles don't overlap, the density is 1 tree per  $\pi r^2/4$  m<sup>2</sup>, or equivalently,

$$\frac{4}{\pi r^2} \text{ trees/m}^2.$$

The observed point-to-tree distances  $R_{ij}$  are the estimates of the actual distances. So  $\pi(R_{ij}/2)^2 = \pi R_{ij}^2/4$  m<sup>2</sup> is an estimate of the sample mean area of a circle occupied by a single organism. Using the  $4n$  area estimates along the transect, an unbiased estimate of the mean area occupied by an organism is

$$\frac{\sum_{i=1}^n \sum_{j=1}^4 \frac{\pi R_{ij}^2}{4}}{4n - 1} = \frac{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2}{4(4n - 1)}.$$

Note: For this estimate to be unbiased, the denominator is one less than the actual number of observations, that is,  $4n - 1$ . The density is the reciprocal of the mean circular area.

**FORMULA H.1.** An unbiased estimate of the **population density**  $\lambda$  is given by

$$\hat{\lambda} = \frac{4(4n - 1)}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2},$$

where the units are typically items/m<sup>2</sup>. Multiplying by 10,000 yields trees/ha. The variance is given by

$$\text{Var}(\hat{\lambda}) = \frac{\hat{\lambda}^2}{4n - 2}.$$

**EXAMPLE H.4.** Reanalyze the data in Table H.1 by calculating  $\lambda$  using Formula H.1.

**SOLUTION.** First we determine

$$\sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2 = (1.1)^2 + (1.6)^2 + \cdots + (1.7)^2 = 100.71.$$

Unlike in (H.3), remember to square the distances first, then sum. The density estimate is

$$10,000\hat{\lambda} = 10,000 \cdot \frac{4(4n - 1)}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2} = \frac{10,000(4(20 - 1))}{100.71\pi} = 2402 \text{ trees/ha}.$$

This estimate is about 1% higher than the earlier biased estimate of 2380.

### Confidence Intervals

Confidence interval estimates for  $\lambda$  may be calculated in the following way.

**FORMULA H.2.** For  $n > 7$ , the endpoints of a confidence interval at the  $(1 - \alpha)100\%$  level are determined by

$$\text{lower endpoint: } \lambda = \frac{(z_{\frac{\alpha}{2}} + \sqrt{16n - 1})^2}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2}$$

and

$$\text{upper endpoint: } \lambda = \frac{(z_{1-\frac{\alpha}{2}} + \sqrt{16n - 1})^2}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2},$$

where  $z_{\beta}$  is the standard normal  $z$ -value corresponding to probability  $\beta$ .

**EXAMPLE H.5.** The following data were collected at Lamington National Park in 1994. The data are the nearest point-to-tree distances for each of four quarters at 15 points along a 200 meter transect. The measurements are in meters. Estimate the tree density and find a 95% confidence interval for the mean.

Point	I	II	III	IV
1	1.5	1.2	2.3	1.9
2	3.3	0.7	2.5	2.0
3	3.3	2.3	2.3	2.4
4	1.8	3.4	1.0	4.3
5	0.9	0.9	2.9	1.4
6	2.0	1.3	1.0	0.7
7	0.7	2.0	2.7	2.5
8	2.6	4.8	1.1	1.2
9	1.0	2.5	1.9	1.1
10	1.6	0.7	3.4	3.2
11	1.8	1.0	1.4	3.6
12	4.2	0.6	3.2	2.6
13	4.1	3.9	0.2	2.0
14	1.7	4.2	4.0	1.1
15	1.8	2.2	1.2	2.8

**SOLUTION.** In this example, the number of points is  $n = 15$  and the number of samples is  $4n = 60$ . Therefore, the density estimate is

$$\hat{\lambda} = \frac{4(4n - 1)}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2} = \frac{4(59)}{347.63\pi} = 0.2161 \text{ trees/m}^2.$$

Since the number of points is greater than 7, confidence intervals may be calculated using Formula H.2. To find a  $1 - \alpha = 0.95$  confidence interval, we have  $\alpha = 0.05$  and so  $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$  and  $z_{0.025} = -z_{0.975} = -1.96$ . The lower endpoint of the confidence interval is

$$\frac{(z_{0.025} + \sqrt{16n - 1})^2}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2} = \frac{(-1.96 + \sqrt{16(15) - 1})^2}{347.63\pi} = 0.1669$$

and the upper endpoint is

$$\frac{(z_{0.975} + \sqrt{16n - 1})^2}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2} = \frac{(1.96 + \sqrt{16(15) - 1})^2}{347.63\pi} = 0.2778.$$

Therefore, the confidence interval for the density is

$$(0.1669, 0.2778) \text{ trees/m}^2.$$

Using Formula H.1, the point estimate for the density<sup>4</sup>

$$\hat{\lambda} = \frac{4(4n - 1)}{\pi \sum_{i=1}^n \sum_{j=1}^4 R_{ij}^2} = \frac{4(60 - 1)}{\sqrt{347.63}\pi} = 0.2161 \text{ trees/ha}$$

The units are changed to hectares by multiplying by 10,000. Thus,  $\hat{\lambda} = 2161$  trees/ha while the confidence interval is (1669, 2778) trees/ha.

## Cautions

The estimates and confidence intervals for density assume that the points along the transect are spread out sufficiently so that no organism is sampled in more than one quarter. Further, the density estimate assumes that the spatial distribution of the organisms is completely random. For example, it would be inappropriate to use these methods in an orchard or woodlot where the trees had been planted in rows.

## H.3 Diversity Indices

*There are three reasons why ecologists are interested in ecological diversity and its measurement. First, despite changing fashions and preoccupations, diversity has remained a central theme in ecology. The well documented patterns of spatial and temporal variation in diversity intrigued the early investigators of the natural world. . . . Second, measures of diversity are frequently seen as indicators of the wellbeing of ecological systems. Thirdly, considerable debate surrounds the measurement of diversity.*

Anne E. Magurran

The concept of species diversity and its measurement is a hotly contested one. There are many reasons for this, but a basic one is that diversity has two distinct components: (1) the total number of species in the community and (2) evenness—how the abundance data are distributed among the various species. Some diversity indices try to measure both components with a single number, which leads to confusion. Adding to this confusion, there are nearly as many diversity indices as there are studies of ecological diversity. Researchers adapt and modify existing measures to better capture the notion of diversity in their particular context. This makes comparisons between studies difficult. Despite these problems, diversity measures are widely used and these notes give a description of a couple of the more common ones. Do not take these measures as definitive. They are just a starting point! Be aware of their limitations.

## The Components of Diversity

As noted, species diversity is usually thought of as being comprised of two components. The first is the number of species in the ecological community. Ecologists often refer to this as **species richness**. The more species a community supports, the richer and more diverse it will be. But it is clear that richness, alone, does not capture the notion of diversity. A simple example shows why.

---

<sup>4</sup>Instead, if (H.3) were used, the density estimate would be quite similar, 2205 trees/ha.

**EXAMPLE H.6.** 100 individual trees were sampled at two locations with the results shown in the table below. What is the richness at each location? What about “diversity”?

Species	Location I	Location II
Acacia	97	25
Eucalyptus	1	25
Casuarina	1	25
Callitris	1	25
Total	100	100

**SOLUTION.** The richness is the same in both locations since four different species were observed. But ecologists would say that the diversity at the second location is much higher than in the first. Intuitively you can see why this is the case. If you walked around at the first location, you might not see anything other than Acacia. There is almost no diversity. While at the second location, all four species are likely to be seen.

The way in which the number of individuals is distributed among the various species in a community is the second major component of species diversity. This is often referred to as **species evenness** or **equitability**. The more equitable the distribution, the higher the diversity. Location II in Example H.6 has a much higher evenness than location I. Because location II has the same richness but a more equitable distribution than location I, it has a higher overall diversity. But how do you compare communities with different numbers of species and different relative abundances?

**EXAMPLE H.7.** 100 samples were taken at two more locations with results recorded below. At which location were trees more diverse?

Species	Location III	Location IV
Acacia	50	75
Eucalyptus	25	10
Casuarina	25	10
Callitris		5
Total	100	100

Do you see the difficulty? Location IV is “richer” since it has 4 species while location III has only 3. But location III has greater “evenness.” So which is more diverse?

There are richness and evenness indices and also diversity indices that attempt to capture both of these components. The major criticism of all diversity indices is that they combine (and hence confound) a number of variables that characterize an ecological community. Nonetheless, Example H.7 shows why ecologists have need for such indices.

## Richness

Species richness—what could be simpler? You might think that the index of species richness would be  $S$ , the total number of species in the community. However,  $S$  depends on the searching effort made, that is, the sample size, the area sampled, and the time spent. Several indices have been proposed to measure species richness that are independent of the sample size,  $n$ . Two well-known indices are **Margalef’s index**

$$R_1 = \frac{S - 1}{\ln n},$$

and **Menhinick's index**

$$R_2 = \frac{S}{\sqrt{n}}.$$

These indices are supposed to provide roughly constant values for a particular community no matter what sample size is used. While both are helpful when used properly, both also pose some difficulties. The main problem is that each presumes a particular sort of relationship between the number of species observed and the number of individuals sampled. This is easiest to see in the Menhinick index, which assumes that  $S = k\sqrt{n}$  for some constant  $k$ . This may be true in some communities, but if it is not, then  $R_2$  will vary with sample size.

An alternative to richness indices is to *use the direct count of species numbers in samples of equal size*. Not only is this simple, it avoids the problems alluded to in the previous paragraph. For those field exercises at sites where we are attempting to measure species diversity, try to use the same sample size. Corrections using a statistical technique called rarefaction may be used to adjust for different sample sizes, but let's avoid such complications!

## Diversity

The **Shannon Index**  $H$  has been one of the most widely used diversity indices in ecology. It is based on information theory.<sup>5</sup>  $H$  measures the average degree of “uncertainty” in predicting the species to which an individual chosen at random will belong.<sup>6</sup> Think about Example H.6 again. Notice that the uncertainty in guessing an individual's species increases as the distribution becomes more even. In the first community there is almost no uncertainty. You can “bet on” the individual being an Acacia. In the second community, the uncertainty is maximized. Each species is equally likely since each comprises 25% of the sample. Also notice that as the number of species in a community increases, the more uncertain one is when guessing the species of a particular individual, at least if the distributions are relatively even.

The formula for the Shannon index is based on the proportion of each species in the sample. Let  $n_i$  denote the number of individuals of the  $i$ -th species in the sample. Then  $p_i = n_i/n$  is the proportion of sample that is composed of the  $i$ -th species. The **Shannon index** is defined as

$$H = - \sum_{i=1}^S p_i \log_2 p_i.$$

Shannon used base 2 logarithms because the notion of uncertainty he was investigating was easier to understand when base 2 logs were used. However, some authors use natural logarithms or common logarithms because they are found on calculators. Such answers can be converted to base 2 logarithms. If natural logarithms are used, simply divide the final answer by  $\ln 2$  to obtain the base 2 result. If common (base 10) logarithms are used, divide the final answer by  $\log 2$  to obtain the base 2 result.

The Shannon index has two nice properties: (1)  $H = 0$  if and only if all individuals in the sample are of the same species and (2) for a given number  $S$  of species,  $H$  has a maximum value of  $\log_2 S$ . This maximum is achieved only when each species has the same number of individuals, that is, in a perfectly even distribution. The first property is “obvious” if you use

<sup>5</sup>Claude Shannon was a researcher for Bell labs and was interested in the amount of information that could be transmitted in various formats.

<sup>6</sup> $H$  is the number of “yes/no” questions required on average to guess the species to which the individual belongs, at least if base 2 logarithms are used in the definition of  $H$ .

$S = 1$  and  $p_1 = 1$  in the formula for  $H$ . The second property can be proven using the Mean Value Theorem from calculus (Mitchell and Kolmes, 1990).

**EXAMPLE H.8.** Compute the Shannon index for the two locations in Example H.6.

**SOLUTION.** For location I, the proportions are  $p_1 = 0.97$  and  $p_2 = p_3 = p_4 = 0.01$ . If we use natural logs to do the calculations, then

$$\begin{aligned} H &= -\frac{\sum_{i=1}^S p_i \ln p_i}{\ln 2} \\ &= -\frac{0.97 \ln 0.97 + 3(0.01 \ln 0.01)}{\ln 2} \\ &= -\frac{-0.02955 - 3(0.04605)}{0.69315} \\ &= 0.242. \end{aligned}$$

For location II, the proportions are  $p_1 = p_2 = p_3 = p_4 = 0.25$ . So

$$H = -\frac{\sum_{i=1}^S p_i \ln p_i}{\ln 2} = -\frac{4(0.25 \ln 0.25)}{\ln 2} = 2.000.$$

Notice that the uncertainty and hence the diversity of location II is much higher, which matches our intuition. We noted earlier that  $H$  achieves its maximum value of  $\log_2 S$  when the distribution is perfectly even, as it is here. Notice that  $\log_2 S = \log_2 4 = 2$ , which is the value of  $H$  for location II.

## Evenness

When all species in a sample are equally abundant, a measure of evenness should be at its maximum.  $H$  is maximized for a fixed value of  $S$  when the distribution of individuals is even across all species. This means that it can be used as a basis for an evenness index. The following index is among the most common used by ecologists,<sup>7</sup>

$$E = \frac{H}{\log_2 S}.$$

$E$  expresses  $H$  as a proportion of its maximum possible value,  $\log_2 S$ . Thus, the range of values for  $E$  is from 0, when all individuals are of one species, to 1, when all species are evenly represented.

When base 2 logs are unavailable,  $E$  can be “corrected” in the same way that  $H$  was. For example, when using natural logarithms

$$E = \frac{H}{\log_2 S} = \frac{H}{\frac{\ln S}{\ln 2}}.$$

But in this case, a correction already will have been required to compute  $H$ . The two corrections actually cancel each other and so  $E$  can be calculated more simply as

$$E = -\frac{\sum_{i=1}^S p_i \ln p_i}{\ln S}.$$

---

<sup>7</sup>This index is sometimes referred to as Pielou’s  $J'$  index.

**EXAMPLE H.9.** Calculate the evenness  $E$  for the two locations in Example H.6.

**SOLUTION.** For location I, we calculated that  $H = 0.242$  So

$$E = \frac{H}{\log_2 S} = \frac{H}{\frac{\ln S}{\ln 2}} = \frac{0.242}{\frac{\ln 4}{\ln 2}} = 0.121.$$

For location II,  $H = 2.000$ , so

$$E = \frac{H}{\log_2 S} = \frac{2}{\log_2 4} = \frac{2}{2} = 1.0,$$

which is the maximum possible evenness.

## Applications to Ecological Diversity

The most common biological application of information theory is in the quantification of ecological diversity. When an ecosystem possesses numerous plant and animal species, with many of them present in relatively high numbers, it will have a high  $H$  value and a high  $E$  value. An ecosystem with fewer types of organisms present, or with only a few common plant or animal species, will concomitantly have lower  $H$  or lower  $E$  values. We generally think of healthy biological communities in favorable habitats as being highly diverse. Decreased biological diversity might be due to harsher environmental conditions (desert versus a temperate zone forest) or to stresses upon a biological community (acid rain or pesticides), which eliminate susceptible species from the natural mix. Information theory measures allow us to quantify both stress-induced and natural differences between ecosystems in their biological diversity.

Even natural changes in the diversity of an ecosystem can be quite dramatic. Twice in the last twenty years there have been major infestations of the crown of thorns starfish, *Acanthaster planci*, on the Great Barrier Reef. The crown of thorns attacks certain species of corals that build and maintain the reef. In some areas as much as 98% of the coral is dead, though it is not clear whether the crown of thorns is responsible for all of this destruction. Initially marine ecologists in Australia were alarmed by these infestations. However, opinion is now beginning to shift.

Not only do some experts think the coral can recover from the crown of thorns. Some marine scientists also think it might be good for the reef to go through such growing pains. Their argument is essentially that the crown of thorns, like a renovator, may make a terrible mess on the way toward home improvements in the longer term. In particular, the crown of thorns cuts down the dominant species of coral and thus makes room for other species, currently crowded out, the upshot being a more diverse group of corals. Biologists have known for a long time that the starfish in temperate waters, by eating common things like mussels, clear out space for other species, and the same kind of benefits, they argue, may accrue from the “destruction” caused by the crown of thorns. [Ford 1988, 51]

Measuring changes in the diversity of an ecosystem are important, as the example above indicates. An  $H$  value can be computed for all of the organisms present in an environment, or for specific types of organisms such as trees or insects. Because the presence of uncommon species in nature can make information theory measures sensitive to the sizes of data sets (in general only very large sets will contain representatives of all the rare species), it is important to collect similar-sized data sets for comparisons of biological diversity.

### Tree Species Diversity

Habitats of the sandhill complex type have fairly simple mature tree communities ( $H = 0.97$ ) compared to an area of sand pine scrub ( $H = 1.55$ ). The most complex mature tree community is the climax southern mixed hardwoods ( $H = 2.56$ ) into which the other forest types listed change very slowly by a process known as **ecological succession**. Information theory measures do distinguish between these different natural communities. Table H.10 contains  $H$  values for mature trees found in different types of forests in Florida [Monk, 1967].

**TABLE H.10.**  $H$  values for trees in plant communities in Florida (from [Monk 1967, 175])

Community	$H$ of Mature Trees
Sandhill Complex	0.97
Cypress Heads	1.16
Sand Pine Scrub	1.55
Mixed Hardwood Swamps	2.28
Climax Southern Mixed Hardwoods	2.56

### Bird Species Diversity

Ornithologists have noted that more types of birds are present breeding in woodlands than in fields of similar sizes. MacArthur and MacArthur (1961) used first-order diversity measures to investigate the relationship between bird diversity and vegetation. They measured  $H$  values for the diversities of bird species breeding at 11 deciduous woodland locations in Pennsylvania, Vermont, and Maryland. In the same habitats they measured various aspects of the vegetation in order to look for any plant community characteristics that were strongly correlated with bird species diversities.

Plant species diversities were computed by using  $H$  values (see Table H.11). Foliage height diversities, which expressed the number of layers of leaves between the ground and the sky in different woodlands, were also measured. Zones of 0 to 2 feet, 2 to 25 feet, and greater than 25 feet above the ground were used as height categories. The number of leaves above points on the ground were estimated for each height zone, and  $H$  values for the foliage height diversity were then calculated. When the number of leaves above the ground in the three height zones are more nearly equal, the foliage height diversity measure increases to reflect the more complex physical environment.

The simplest model of how birds select a nesting habitat would be that as either foliage height diversity or plant species diversity increased, the attractiveness of the habitat would increase linearly. MacArthur and MacArthur (1961) looked to see whether one of these simple predictive relationships existed. Bird species diversity and foliage height diversity values were strongly correlated to one another. Figure H.4 shows this correlation as a linear relationship. The data closely approximate the line given by the equation:

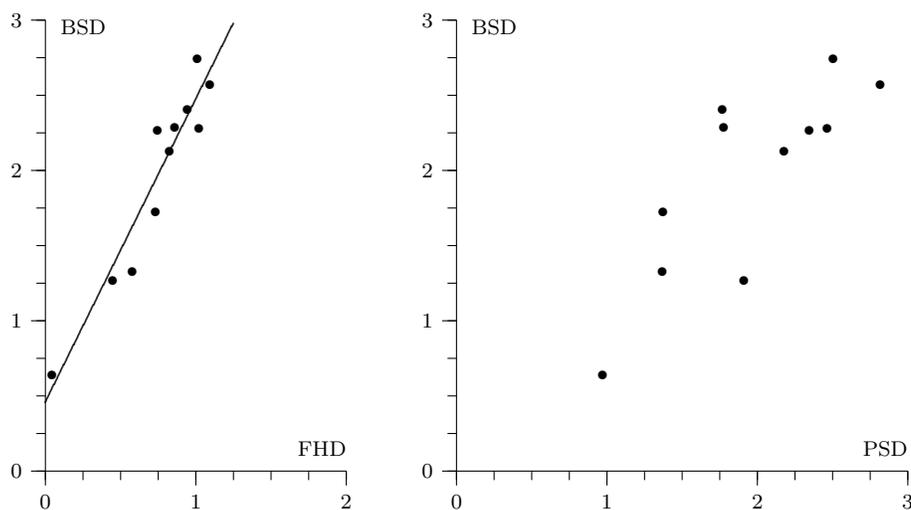
$$\text{bird species diversity} = 2.01 \times \text{foliage height diversity} + 0.46 .$$

On the basis of this linear relationship, birds appear to be selecting nesting habitats on the basis of foliage height diversities. Interestingly, there was a much weaker relationship between bird species diversity and plant species diversity measures [MacArthur and MacArthur 1961].

**TABLE H.11.** Bird species diversity (BSD), foliage height diversity (FHD), and plant species diversity (PSD) at 11 different locations. Adapted from [MacArthur and MacArthur 1961, 596]

Site	BSD	FHD	PSD
A	0.639	0.043	0.972
B	1.266	0.448	1.911
C	2.265	0.745	2.344
D	2.403	0.943	1.768
E	1.721	0.731	1.372
F	2.739	1.009	2.503
G	1.332	0.577	1.367
H	2.285	0.859	1.776
I	2.277	1.021	2.464
J	2.127	0.825	2.176
K	2.567	1.093	2.816

The plot of bird species diversity versus plant species diversity is considerably less linear than the plot using foliage height diversity. The physical structure of the woodlands, in terms of the leaves present in different height zones, seemed to matter more to the birds than the species of plants producing those physical structures.



**FIGURE H.4.** Bird species diversity (BSD) plotted against foliage height diversity (FHD) and against plant species diversity (PSD). Adapted from [MacArthur and MacArthur 1961, 596].

### Fire Ant Odor Trails

Fire ants of the species *Solenopsis saevissima* are social insects that live in underground nests containing many sterile workers and their queen. To obtain food, workers set forth from the nest and search the surrounding area. If a worker finds a food source large enough for a number of ants to harvest, a communication system based upon odor trails allows additional ants to be recruited [Wilson 1962].

To produce an odor trail, a worker returning to the nest periodically drags its sting along the ground while releasing a chemical produced by Dufour's gland through the extruded sting. The chemical released is attractive to other workers, and causes them to follow the odor trail towards the food. A truly abundant food source eventually produces a situation where many ants returning from the food to the nest have been excited into producing an odor trail, and the summed individual odor trails produces a virtual chemical "highway" leading to the food. The chemical secreted from Dufour's gland fades slowly over time, so that a depleted food source loses its attractiveness.

In the absence of any odor trail, a foraging worker leaving a fire ant nest might be expected to depart without bias towards any particular direction, that is, in any one of the  $360^\circ$  of directions that surround the nest. If we denote the diversity of directions that a departing group of ants might be expected to display in this uninformed initial circumstance by  $H_i$ , then

$$H_i = \log_2 360$$

where every 1 degree of direction is taken as a potential direction category.

If an ant odor trail to a food source exists, then the departing ants might be expected to leave from their nest in a smaller angular range of directions. If the smaller diversity of this range of directions is symbolized by  $H_s$ , then the transmission of information by the odor trail, denoted  $H_t$ , must be equal to the difference between the diversity of the array of departure directions displayed by the informed and the uninformed groups of ants. That is,

$$H_t = H_i - H_s.$$

Using a small drop of sugar solution placed on an index card as a food source, Wilson [1962] measured the direction indicated by the odor trail produced by a single fire ant and its influence upon the directions in which recruited foragers travelled from their nest. After this procedure was carried out a number of times to obtain replicate data sets, an estimate of the information transmitted by fire ant odor trails could be made.<sup>8</sup>

The results of the fire ant study showed a considerable amount of information transmission by the odor trails (see Table H.12). When food sources were placed at between 20mm and 100mm from an ant nest, the range of directional information transmitted by odor trails proved to be between 3 and 5 bits.

**TABLE H.12.** The amount of directional information transmitted to single workers by a single fire ant odor trail (adapted from [Wilson 1962, 152])

Target range (mm)	$H_t$
20	2.81
50	4.11
100	5.10

We can interpret one bit of information in this context in the following way. If a foraging ant could only inform another worker that a food source was either to the north or the south

<sup>8</sup> $H_s$  can either be measured by observing the distribution of directions by which ants depart from their nest and counting the number of ants in each degree-category, or by considering the data to be a normally distributed one-dimensional Gaussian distribution and applying the formula  $H_s = \log_2 \sqrt{2\pi e} \sigma$  where  $\sigma$  is the standard deviation. See Haldane and Spurway [1954] or Wilson [1962] for more details concerning the latter approach.

of the nest, there would be only two directional choices, so  $H_{1\max} = H_i = \log_2 2 = 1$ . If the communication that took place was perfect and the second worker always went in the correct direction, then  $H_s = 1\log_2 1 = 0$ . In this simple situation,

$$H_t = H_i - H_s = 1.$$

Now assume the foraging ant could perfectly transmit the information as to whether the food source was to the north, east, south, or west, then  $H_i = \log_2 4 = 2$  bits since there are now four directional categories. In the same manner, we can interpret the 3 to 5 bits of information conveyed by the foraging ants in this experiment. The 3 to 5 bits of information transmitted is equivalent to every departing forager being told what direction to walk and being equipped with a tiny compass upon which are marked between  $2^3 = 8$  and  $2^5 = 32$  directional points. Departing foragers given such equipment, and able to use it well, would be able to depart in a given direction as accurately as departing ants using an odor trail rather than a tiny compass as their guide [Wilson 1962, 154].

The main reason for converting the actual field data regarding trail communication into abstract “bits of information” is that it allows us to compare the very different communication systems employed by a wide range of insects and animals. For example, it allows us to compare the information conveyed by a honey bee’s waggle dance with the information conveyed by a fire ant’s odor trail.

A foraging honey bee that has located a rich source of nectar and pollen will perform a waggle dance upon its return to its hive. The dance allows the other bees to find the food source. The waggle dance looks rather like a “figure-8” performed upon a vertical surface inside the hive. Its interpretation involves both the orientation of the dance and its pace. The direction to a food source is encoded by the angle that the figure-8 deviates from the vertical. The distance to the food source is encoded by the number of turns per minute that a dancing bee performs. Haldane and Spurway [1954] analyzed the information transmitted by dancing bees to other foragers in a fashion similar to the one already described for fire ants. The honey bee data was quite similar to that gathered by Wilson [1962] for fire ants. Honey bees transmit between 2.5 and 4.0 bits of directional information in their waggle dance. Both types of social insects are quite good at directing nestmates to food sources, although the methods they use to communicate with each other differ dramatically.

## H.4 Problems

1. The following data were collected in interior Alaska. The data are the nearest point-to-tree distances in meters for each of four quarters at the first 25 points of 724 sample points. All trees were black spruce, *Picea mariana*. Estimate the tree density and find a 95% confidence interval for the mean. (Data reported in: Hollingsworth, T. 2005. Point-center quarter method black spruce heights, diameter, and location for 150 mature black spruce sites in interior Alaska. LTER Database:BNZD:138. [www.lter.uaf.edu/ascii/files138\\_pcqbs\\_ht\\_diam.txt](http://www.lter.uaf.edu/ascii/files138_pcqbs_ht_diam.txt). NSF awards DEB-0080609, 9810217, 9211769, 8702629 and USDA Forest Service, Pacific Northwest Research Station RJVA-PNW-01-JV-11261952-231.)
2. The following data were collected at Lamington National Park in 1994 by another group of students. The data are the nearest point-to-tree distances for each of four quarters at 14 points along a two-hundred meter transect. The measurements are in meters. Estimate the tree density and find a 95% confidence interval for the mean.

Point	I	II	III	IV	Point	I	II	III	IV
1	7.7	2.2	1.4	1.6	14	1.2	1.1	1.0	1.4
2	0.97	1.2	1.4	1.5	15	0.5	0.7	0.9	1.1
3	1.4	1.4	1.8	1.6	16	0.52	0.85	0.82	2.1
4	1.7	2.5	2.2	1.8	17	0.51	0.46	1.6	1.1
5	0.77	1.2	1.0	1.2	18	0.46	0.9	1.7	0.65
6	0.38	0.64	1.84	1.7	19	0.35	0.64	0.98	0.53
7	0.45	0.6	0.55	0.62	20	0.98	1.3	2.1	1.6
8	0.15	0.14	0.96	0.9	21	0.35	0.5	0.25	1.0
9	0.39	0.5	0.57	0.88	22	0.4	0.4	0.6	0.8
10	0.72	0.73	0.45	0.75	23	0.6	1.5	1.3	1.1
11	0.35	1.1	0.45	1.1	24	0.4	0.5	0.9	0.8
12	0.55	0.9	0.65	0.9	25	0.5	1.1	2.1	1.1
13	0.8	0.7	0.8	0.9					

I	II	III	IV	I	II	III	IV
0.6	1.4	3.6	2.0	3.4	3.4	2.9	2.6
0.6	0.9	3.2	1.8	1.7	3.2	2.7	4.2
2.0	3.9	1.8	2.2	3.8	4.2	3.2	4.4
4.1	7.0	1.6	4.0	1.8	1.1	4.3	3.4
3.2	2.0	1.0	3.8	2.8	0.9	2.7	2.3
2.8	3.3	1.3	0.8	1.4	5.0	4.5	2.7
3.1	1.9	2.9	3.4	2.0	0.2	3.0	4.0

3. Calculate  $H$  and  $E$  for the two locations in Example H.7.
4. Here are three data sets on the distribution of bird species in low and high altitude forests in the northeastern region of New South Wales. These data are for less common bird species. They were recorded in a 15 hour “area search” in each sampling area during the breeding season (W. S. Osborne, *The Rainforest Legacy*, p. 222). Birds were observed in four different forest habitats including the two recorded here: low altitude subtropical rain forest (LASRF) and high altitude subtropical rain forest (HASRF). The birds were classified into three general groups: canopy feeders, ground feeders, and others.
- (a) Calculate the species richness  $S$ , diversity  $H$ , and evenness  $E$  for ground feeding birds in high and low altitude subtropical rainforests and interpret the results.

Ground Feeders	LASRF	HASRF
Brush turkey	46	6
Wonga pigeon	1	7
Emerald fruit dove	6	3
Noisy pitta	4	6
Superb lyrebird	3	71
White’s thrush	9	17
Logrunner	9	26

- (b) Calculate the species richness  $S$ , diversity  $H$ , and evenness  $E$  for canopy feeding birds.

Canopy Feeders	LASRF	HASRF
Wampoo fruit-dove	48	1
Topknot pigeon	25	2
Rainbow lorikeet	1	7
Sulphur-crested cockatoo	8	
Satin bowerbird	30	17
Figbird	10	

(c) Carry out the same exercise for “other” birds.

Others	LASRF	HASRF
White-headed pigeon	2	1
Brown cuckoo dove	1	3
Yellow-tailed black cockatoo	1	
Australian king parrot	8	3
Crimson rosella	4	26
Laughing kookaburra	2	3
Green catbird	23	24
Paradise riflebird		4
Pied currawong	6	7

(d) Which forest/feeding combination exhibits the highest diversity? Explain.

5. Honey bees perform tasks inside the hive when they are newly-emerged workers, and switch to collecting nectar and pollen from flowers later in their lives. Numerous studies of honey bee behavior have revealed that the tasks they perform inside the hive are quite varied and complex. The data in below show all of the behavioral patterns recorded for three worker bees that represent a very small part of the group observed in an experiment [Kolmes 1985]. Although three bees is too small a group from which to draw any conclusions, we can still get a sense of the diversity of behavior patterns.

Behavior	Bee 1	Bee 2	Bee 3
Walk	9	7	8
Stand	7	7	2
Groom Self	4	7	5
Inspect/Feed	1	1	2
Into Empty Cell	2	5	1
Into Honey Cell	0	1	0
Into Pollen Cell	0	1	0
Build Comb	0	3	0
Groom Other Bee	0	0	4
Get Groomed	0	0	1
Get Fed	0	0	1
DVAV	1	0	0
Attend Dance	3	0	0
Antennate	2	2	0
Chew Hive	1	0	0
Fan	1	2	0

- (a) Determine the diversity of the behaviors,  $H$ , for each of the three bees.  
 (b) Determine the evenness measure,  $E$ , of the behaviors for each of the three bees.

6. Before settlement by Europeans, the Great Lakes Region of the United States was covered by a vast pine forest. Recent studies of these vanished forest areas have relied upon modern analysis of data collected during the General Land office survey of Michigan carried out in the 1800s [Whitney 1986]. The data in below show the percentage of trees reported as “bearing trees” by surveyors in two different habitats in Northern Michigan, sometime between 1836 and 1859 [Whitney 1986].

Species	Swamps	Uplands
Balsam fir	1.7	0.0
Red maple	1.6	2.6
Sugar maple	0.1	3.0
Yellow birch	0.5	0.0
White birch	2.9	0.4
Birch	2.5	0.8
Beech	0.2	17.2
White ash	0.7	0.0
Black ash	3.6	0.0
Tamarack	29.7	0.0
Spruce	10.6	0.0
Pitch pine	1.7	10.9
Norway pine	1.1	25.7
Pine	1.5	7.5
White pine	7.6	12.7
Poplar	3.8	2.2
White oak	0.0	5.7
Black oak	0.0	3.4
Cedar	25.2	0.0
Hemlock	3.2	7.3
Elm	1.0	0.1
Miscellaneous	0.8	0.5

- (a) Determine the diversity of the tree species,  $H$ , for each of the two habitats in Northern Michigan.
- (b) Determine the measure of evenness,  $E$ , of the tree species for each of the two areas.
7. The data below show how responsive different female white-crowned sparrows, *Zonotrichia leucophrys*, were to the recorded courtship songs of male sparrows of two species which had been raised in the laboratory. Responses to their own species are in the top row and responses to a related species, *Melospiza melodia* (song sparrow), are in the bottom row. Animal communication must follow predictable patterns for it to be effective, so we might expect a more consistent response by female sparrows to appropriate (same species) male songs than to inappropriate (different species) male courtship songs. Information theory measures are one way of addressing how consistently a group of experimental animals responds to different stimuli. Adapted from [Spitler-Nabors and Baker 1987, 383].

Female	a	b	c	d	e	f	g	h	i	j	k	l	m
W-CS	1	4	10	1	4	20	1	1	1	7	14	1	3
SS	2	0	10	0	0	14	27	0	0	0	10	0	4

- (a) Determine which of the laboratory-raised male sparrow’s (W-CS or SS) elicited a more diverse response from the thirteen female white-crowned sparrows.
- (b) From just looking at the data do you expect the evenness to be high or low for the song sparrows? Calculate the evenness for each data set. What might this mean biologically?

8. The Aldabra tortoise, *Geochelone gigantea*, is a large tortoise endemic to the Aldabra Atoll in the Seychelles in the Indian Ocean. Adult males have a carapace length of more than 100 cm and may weigh 250 kg while the females have a carapace length in excess of 90 cm carapace and weigh 159 kg. The thick carapace is brownish gray and highly domed. Among the tortoises, this species is second only in size to the slightly larger species found on the Galapagos Islands. Life span is believed to surpass 100 years.

A survey was conducted to estimate the population of Aldabra tortoises on Aldabra atoll. Initially 4214 tortoises were marked. In a second survey, 3701 were located with 221 being recaptures. Find the estimated population size and its 95% confidence limits. [Morgan, D. D. V. and Bourn, D. M. 1981. A comparison of two methods of estimating the size of a population of giant tortoises on Aldabra. *J. Applied Ecology*. 37–40.]

9. The striped bass, *Morone saxatilis*, is a highly prized sportfish and is widely stocked in the eastern United States. A mark and recapture study was conducted in the Delaware River to estimate population size of age-0 striped bass. Approximately 95,000 hatchery-reared fish were marked with wire tags and released at multiple locations within the nursery area along the river.
- (a) Recapture efforts yielded 53 tagged fish. If the total population was estimated to be  $\hat{N} = 975, 100$ , how many age-0 striped bass were caught ( $C$ ) during the recapture process?
- (b) Find 95% confidence limits for the population.
10. Using aerial photography, a population of humpback whales, *Megaptera novaeangliae*, in the Gulf of Maine off Nova Scotia were “marked.” A total of 27 were photographically recorded. A second survey in the same area captured 29 individuals, with 9 being recaptures. Estimate the population size and find 95% confidence limits.
11. Commercial sport fishing guides have expressed concern over the status of the rainbow trout, *Oncorhynchus mykiss*, population along the Kanektok River in Alaska. A study was conducted to estimate the trout population in this area. 400 trout were tagged and released. An estimated 7600 trout were subsequently caught by anglers later in the season with 110 of these being recaptures.
- (a) Estimate the population size and find 95% confidence limits.
- (b) Which assumptions of Petersen model were likely violated by this study?
12. The introduced feral goat, *Capra hircus*, occurs in all states and territories of Australia except the Northern Territory. The total has thought to exceed 2.5 million, with the greatest concentrations occurring in western New South Wales, southern Queensland, central eastern South Australia and Western Australia. Feral goats occur in a wide range of habitats, though they do not do well in deserts or wet tropical areas where dingoes are present.

Feral goats are a major environmental and agricultural pest. They compete with sheep for pasture and also cause land degradation through soil damage, overgrazing pasture plants and browsing established trees and shrubs preventing their regeneration.

In one study in a 234 square kilometer area of sheep range in South Australia, 96 goats were tagged and 1829 goats were later sighted. Estimate the population size and find 95% confidence limits. [Parkes, J., Henzell, R., and Pickles, G. Managing vertebrate pests: Feral goats. Australian Government Publishing Service. [www.affa.gov.au/corporate\\_docs/publications/word/rural\\_science/lms/ferals/goatguide.doc](http://www.affa.gov.au/corporate_docs/publications/word/rural_science/lms/ferals/goatguide.doc)]

## References

### For the point-centered quarter method:

- Beasom, S. and H. Haucke. 1975. A comparison of four distance sampling techniques in South Texas live oak mottes. *J. Range Management*. 28: 142–144.
- Cottam, Grant and J. T. Curtis. 1956. The use of distance measures in phytosociological sampling. *Ecology*. 37: 451–460.

- Cottam, Grant, J. T. Curtis, and B. W. Hale. 1953. Some sampling characteristics of a population of randomly dispersed individuals. *Ecology*. **34**: 741–757.
- Krebs, C. J. 1989. *Ecological Methodology*. New York, Harper & Row.
- Lamb, D. 1994. Lecture notes. University of Queensland, Australia.
- Mitchell, K. 2007. Quantitative analysis by the point-centered quarter method. <http://people.hws.edu/mitchell/PCQM.pdf>.
- Morisita, Masaaki. 1954. Estimation of population density by spacing method. *Memoirs Fac. Sci. Kyushu Univ. Ser. E*. **1**: 187–197.
- Pollard, J. H. 1971. On distance estimators of density in randomly distributed forests. *Biometrics*. **27**: 991–1002.
- Stearns, Forest. 1949. Ninety years change in a northern hardwood forest in Wisconsin. *Ecology*. **30**: 350–358.

### For diversity indices:

- Barrett, G. W. 1968. The effects of an acute insecticide stress on a semi-enclosed grassland ecosystem. *Ecology* **49**: 1019–1035.
- Ford, D. 1988. A reporter at large: crown of Thorns. *New Yorker* 25 July 1988: 34–63.
- Haldane, J. B. S. and H. Spurway. 1954. A statistical analysis of communication in *Apis mellifera* and a comparison with communication in other animals. *Insectes Sociaux* **1**: 247–283.
- Kolmes, S. A. 1985. An information-theory analysis of task specialization among worker honey bees performing hive duties. *Animal Behaviour* **33**: 181–187.
- Lloyd, M., J. H. Zar, and J. R. Karr. 1968. On the calculation of information theoretical measures of diversity. *American Midland Naturalist* **79**: 257–272.
- Ludwig, John A. and James F. Reynolds. 1988. *Statistical Ecology*. New York: Wiley.
- MacArthur, R. H., and J. W. MacArthur. 1961. On bird species diversity. *Ecology* **42**: 594–598.
- Magurran, A. E. 1988. *Ecological Diversity and its Measurement*, Princeton, NJ: Princeton University Press.
- Mitchell, Kevin and Steven Kolmes. 1990. Information theory and biological systems. *UMAP Journal* **11.1**: 25–62. Also published as: *UMAP Modules: Tools for Teaching 1990*. 43–78 and as UMAP Module 705.
- Monk, C. D. 1967. Tree species diversity in the eastern deciduous forest with particular reference to north central Florida. *American Naturalist* **101**: 173–187.
- Spitler-Nabors, and Baker. 1987. Sexual display response of female white-crowned sparrows to normal, isolate, and modified conspecific songs. *Animal Behaviour* **35**: 380–386.
- Whitney, G. G. 1986. Relation of Michigan's presettlement pine forests to substrate and disturbance history. *Ecology* **67**: 1548–1559.
- Wilson, E. O. 1962. Chemical Communication among workers of the fire ant *Solenopsis saevissima* (Fr. Smith). 2. An information analysis of the odor trail. *Animal Behaviour* **10**: 148–158.

## Appendix

**TABLE H.13.** Table of 2025 Random Digits.

<b>1</b>	60082	84894	87580	22864	25331	54562	44686	40649	51483
<b>2</b>	22224	12938	28165	75805	68172	80673	17717	53236	68851
<b>3</b>	60285	32511	72012	82652	34342	78292	76543	20885	73190
<b>4</b>	88812	28748	21729	61863	68489	21822	56358	52501	89453
<b>5</b>	44576	55744	46672	14593	64783	37256	93146	88197	76405
<b>6</b>	28890	23523	93040	14691	29545	74989	95987	28891	21203
<b>7</b>	33248	36833	92299	67498	42777	26268	17589	92760	46627
<b>8</b>	6486	93538	12667	83088	4615	65794	66354	60781	84674
<b>9</b>	17475	62049	17297	39937	65459	75082	78141	12139	89131
<b>10</b>	15274	37983	98317	94216	67221	93399	85141	77546	67711
<b>11</b>	68879	51475	98386	75048	29674	75489	12385	5994	63650
<b>12</b>	83496	72984	23660	95481	60220	39281	58264	52018	27812
<b>13</b>	26744	36792	72255	76361	19424	98679	36742	18622	19857
<b>14</b>	62711	87719	67049	44892	52839	15490	46973	74915	46364
<b>15</b>	31414	85454	16495	40617	2926	45817	96356	52240	47116
<b>16</b>	34554	98863	34967	85013	27775	14375	89156	21919	76635
<b>17</b>	95462	96714	49735	87824	97419	33554	17134	49235	97579
<b>18</b>	48093	46752	93317	37664	45035	72983	80716	30263	64913
<b>19</b>	60969	95257	40274	60833	74771	73456	27750	10135	49899
<b>20</b>	1096	16749	75350	87705	72326	68094	23155	91453	74633
<b>21</b>	39062	42448	18988	93694	57797	34517	10748	74680	21585
<b>22</b>	88966	87249	77126	1433	94406	15789	7692	17558	33372
<b>23</b>	55472	54559	42499	98779	34668	77150	4338	70459	31650
<b>24</b>	77115	91315	70052	14534	76386	18211	42522	31774	52762
<b>25</b>	68296	65967	27859	36237	3758	2576	31417	79768	23853
<b>26</b>	11891	92132	43614	25173	37475	92684	7525	12754	52073
<b>27</b>	67845	41815	87539	63773	33269	96363	83893	13684	54758
<b>28</b>	80715	3333	36746	42279	63932	91413	13015	45479	96152
<b>29</b>	93614	88328	22103	21134	73295	22175	46254	11747	36284
<b>30</b>	28017	18124	61320	52542	35362	27681	58562	53691	96599
<b>31</b>	95114	73345	78448	17128	94266	82197	10938	42871	39309
<b>32</b>	29631	61790	17394	87012	80142	12916	43588	88044	07429
<b>33</b>	72439	22965	22452	89352	84598	40162	51112	99370	58994
<b>34</b>	43206	76531	23736	90099	16631	62425	23619	94864	28797
<b>35</b>	19266	29669	79345	1827	15147	85505	58666	84693	65570
<b>36</b>	95222	14122	54382	71115	93771	35510	79567	96455	67252
<b>37</b>	17310	48813	33458	54178	34773	29541	75989	11419	81253
<b>38</b>	72494	45082	88616	80699	59886	36329	69658	71891	03236
<b>39</b>	89818	68866	13858	32642	41924	8469	14327	84551	47068
<b>40</b>	73182	66270	93939	45159	28426	43253	42189	61174	77953
<b>41</b>	41648	15786	24517	80227	79184	72866	96071	36856	92714
<b>42</b>	86633	67816	43550	765	88497	46434	10767	27709	14374
<b>43</b>	60762	91378	18649	96638	85675	33142	79869	18443	24879
<b>44</b>	29283	77878	61353	89214	72140	29236	11476	82552	47777
<b>45</b>	78114	48491	51326	68205	52576	54212	46363	61776	97791

## H.5 Answers

- 1382 trees/ha with a 95% confidence interval of (1057, 1792).
- For location III,

$$H = -(0.5\log_2 0.5 + 0.25\log_2 0.25 + 0.25\log_2 0.25) = 1.5 \text{ bits}$$

and (assuming four species were possible)  $E = \frac{1.5}{2.0} = 0.75$ . For location IV,

$$H = -(0.75\log_2 0.75 + 0.1\log_2 0.1 + 0.1\log_2 0.1 + 0.05\log_2 0.05) = 1.192 \text{ bits}$$

and  $E = \frac{1.192}{2.0} = 0.596$ . Though it has fewer species, location III has a higher diversity and a greater evenness.

- (a) For ground feeders:  $S_{\text{LASRF}} = 7$ ,  $H_{\text{LASRF}} = 1.934$  bits,  $E_{\text{LASRF}} = 0.748$ .  $S_{\text{HASRF}} = 7$ ,  $H_{\text{HASRF}} = 2.060$  bits,  $E_{\text{LASRF}} = 0.797$ .  
 (b) For canopy feeders:  $S_{\text{LASRF}} = 6$ ,  $H_{\text{LASRF}} = 2.106$  bits,  $E_{\text{LASRF}} = 0.815$ .  $S_{\text{HASRF}} = 5$ ,  $H_{\text{HASRF}} = 1.379$  bits,  $E_{\text{LASRF}} = 0.594$ .  
 (c) For other feeders:  $S_{\text{LASRF}} = 8$ ,  $H_{\text{LASRF}} = 2.245$  bits,  $E_{\text{LASRF}} = 0.746$ .  $S_{\text{HASRF}} = 8$ ,  $H_{\text{HASRF}} = 2.288$  bits,  $E_{\text{LASRF}} = 0.763$ .  
 (d) The other feeders in the low altitude subtropical rain forest were most diverse.
- (a) For Bee 1:  $H = 2.859$  bits. For Bee 2:  $H = 2.966$  bits. For Bee 3:  $H = 2.601$  bits.  
 (b) For Bee 1:  $E = 0.715$ . For Bee 2:  $E = 0.742$ . For Bee 3:  $H = 0.650$ .
- (a) For swamps:  $H = 3.132$  bits. For uplands:  $H = 3.170$  bits.  
 (b) For swamps:  $E = 0.702$ . For uplands:  $E = 0.711$ .
- (a) For white-crowned sparrows  $H = 2.950$  bits while for song sparrows  $H = 2.123$  bits.  
 (b) One should expect the evenness to be lower for song sparrows since the response pattern is so uneven. For white-crowned sparrows  $E = 0.797$  while for song sparrows  $E = 0.598$ . The data seem to indicate that most (but not all) female white-crowned sparrows can discriminate between calls of males of their own species and those of a related species.

- Using (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{3701(4214)}{221} = 70,570.$$

Since  $\frac{R}{C} = \frac{221}{3701} = 0.060 < 0.10$  and  $R = 221 > 50$ , use the normal approximation for the confidence interval. The confidence limits for  $\frac{R}{C}$  are

$$L_1 = \frac{221}{3701} - 1.960\sqrt{\frac{\left(\frac{221}{3701}\right)\left(1 - \frac{221}{3701}\right)}{3701 - 1}} = 0.0521$$

and

$$L_2 = \frac{221}{3701} + 1.960\sqrt{\frac{\left(\frac{221}{3701}\right)\left(1 - \frac{221}{3701}\right)}{3701 - 1}} = 0.0673.$$

We find that:

the lower 95% confidence limit for  $N$  is:  $\frac{1}{0.0673}(4214) = 62,615$

the upper 95% confidence limit for  $N$  is:  $\frac{1}{0.0521}(4214) = 80,883$ .

8. (a) Rewriting (H.1)

$$C = \frac{\hat{N}R}{M} = \frac{975,100(53)}{95,000} = 544.$$

(b) Since  $\frac{R}{C} = \frac{53}{95,000} < 0.10$  and  $R = 53 > 50$ , use the normal approximation for the confidence interval. The confidence limits for  $\frac{R}{C}$  are

$$L_1 = \frac{53}{544} - 1.960\sqrt{\frac{\left(\frac{53}{544}\right)\left(1 - \frac{53}{544}\right)}{544 - 1}} = 0.0725$$

and

$$L_2 = \frac{53}{544} + 1.960\sqrt{\frac{\left(\frac{53}{544}\right)\left(1 - \frac{53}{544}\right)}{544 - 1}} = 0.1224.$$

We find that:

$$\text{the lower 95\% confidence limit for } N \text{ is: } \frac{1}{0.1224}(95,000) = 776,144$$

$$\text{the upper 95\% confidence limit for } N \text{ is: } \frac{1}{0.0725}(95,000) = 1,310,345.$$

9. Using Petersen's original equation (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{29(27)}{9} = 87.$$

Using Seber's modified equation (H.2)

$$\hat{N} = \frac{(C+1)(M+1)}{R+1} - 1 = \frac{(30)(28)}{10} - 1 = 83.$$

Next

$$\frac{R}{C} = \frac{9}{29} = 0.31 > 0.10.$$

Using  $\nu_1 = 2(29 - 9 + 1) = 42$  and  $\nu_2 = 2(9) = 18$ , The lower limit for  $\frac{R}{C}$  is

$$L_1 = \frac{9}{59 + (29 - 9 + 1)F_{0.025(42,18)}} = 0.153.$$

Using  $\nu'_1 = 18 + 2 = 20$  and  $\nu'_2 = 42 - 2 = 40$ , the upper limit for  $\frac{R}{C}$  is

$$L_2 = \frac{(9+1)F_{0.025(20,40)}}{29 - 9 + (10)F_{0.025(20,40)}} = 0.508.$$

Using the lower and upper confidence limits for  $\frac{R}{C}$  in the formula for  $\hat{N}$ , we find that:

$$\text{the lower 95\% confidence limit for } N \text{ is: } \frac{1}{0.508}(27) = 53$$

$$\text{the upper 95\% confidence limit for } N \text{ is: } \frac{1}{0.153}(27) = 177.$$

10. Using (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{7600(400)}{110} = 27,636.$$

Since  $\frac{R}{C} = \frac{110}{7600} = 0.014 < 0.10$  and  $R = 110 > 50$ , use the normal approximation for the confidence interval. The confidence limits for  $\frac{R}{C}$  are

$$L_1 = \frac{110}{7600} - 1.960\sqrt{\frac{\left(\frac{110}{7600}\right)\left(1 - \frac{110}{7600}\right)}{7600 - 1}} = 0.0118$$

and

$$L_2 = \frac{110}{7600} + 1.960\sqrt{\frac{\left(\frac{110}{7600}\right)\left(1 - \frac{110}{7600}\right)}{7600 - 1}} = 0.0172.$$

We find that:

$$\text{the lower 95\% confidence limit for } N \text{ is: } \frac{1}{0.0172}(400) = 23,256$$

$$\text{the upper 95\% confidence limit for } N \text{ is: } \frac{1}{0.0118}(400) = 33,898.$$

11. Using Petersen's original equation (H.1)

$$\hat{N} = \frac{CM}{R} = \frac{366(96)}{34} = 1033.$$

Using Seber's modified equation (H.2)

$$\hat{N} = \frac{(C+1)(M+1)}{R+1} - 1 = \frac{(366)(97)}{35} - 1 = 1016.$$

Since

$$\frac{R}{C} = \frac{34}{366} < 0.10,$$

use the normal approximation for the confidence interval. The confidence limits for  $\frac{R}{C}$  are

$$L_1 = \frac{34}{366} - 1.960\sqrt{\frac{\left(\frac{34}{366}\right)\left(1 - \frac{34}{366}\right)}{366 - 1}} = 0.0631$$

and

$$L_2 = \frac{34}{366} + 1.960\sqrt{\frac{\left(\frac{34}{366}\right)\left(1 - \frac{34}{366}\right)}{366 - 1}} = 0.1227.$$

We find that:

$$\text{the lower 95\% confidence limit for } N \text{ is: } \frac{1}{0.1227}(96) = 782$$

$$\text{the upper 95\% confidence limit for } N \text{ is: } \frac{1}{0.0631}(96) = 1521.$$

# Proofs of Selected Results

## I.1 The Method of Least Squares for Linear Regression

This section presents another derivation of the least squares linear regression equation which, unlike the Section A.6 in the text, does not require calculus.

To predict  $Y$  for various values of  $X$  when the data indicate a straight-line relationship between the two variables, we need to find  $a'$  and  $b'$  in the standard equation for a straight line,  $Y = a' + b'X$ , with  $a'$  being the constant equal to the line's intercept on the  $Y$ -axis and  $b'$  being equal to the line's slope.

Each  $Y$  can be expressed as the regression equation value for  $X_i$  plus a deviation from that value,  $\varepsilon_i$ :

$$Y_i = a' + b'(X_i - \bar{X}) + \varepsilon_i.$$

Letting  $x_i = X_i - \bar{X}$ , the  $i$ th deviate of  $X$ , we can rewrite the last equation as

$$Y_i = a' + b'x_i + \varepsilon_i$$

or

$$\varepsilon_i = Y_i - a' - b'x_i.$$

So  $\varepsilon_i$  is the difference between the actual and predicted  $Y$  values for a particular  $X_i$ . The best-fit equation will minimize the deviations  $\varepsilon_i$ 's. More accurately, it will minimize the sum of the squares of the deviations

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - a' - b'x_i)^2 = \text{SS}_{\text{Error}}(a', b').$$

Notice that the sum of squares depends on the choice of  $a'$  and  $b'$ . We want to find values for the intercept  $a$  and slope  $b$  that *minimize* this number. In other words, for any choice of  $a'$  and  $b'$ , we should have

$$\text{SS}_{\text{Error}}(a', b') \geq \text{SS}_{\text{Error}}(a, b).$$

First we will determine  $a$ . Since  $\text{SS}_{\text{Error}}(a, b)$  is minimal, if we choose any  $a'$ , then

$$\text{SS}_{\text{Error}}(a', b) \geq \text{SS}_{\text{Error}}(a, b).$$

This is equivalent to

$$\sum_{i=1}^n (Y_i - a' - bx_i)^2 \geq \sum_{i=1}^n (Y_i - a - bx_i)^2$$

or

$$\sum_{i=1}^n \left[ (Y_i - a' - bx_i)^2 - (Y_i - a - bx_i)^2 \right] \geq 0. \quad (\text{I.1})$$

Now let  $\epsilon$  denote the error in the choice of  $a'$ , that is,  $\epsilon = a' - a$  or, equivalently,  $a' = a + \epsilon$ . Then (I.1) can be written as

$$\sum_{i=1}^n \left[ (Y_i - (a + \epsilon) - bx_i)^2 - (Y_i - a - bx_i)^2 \right] \geq 0. \quad (\text{I.2})$$

Each summand in (I.2) can be simplified as follows:

$$\begin{aligned} (Y_i - a - \epsilon - bx_i)^2 - (Y_i - a - bx_i)^2 &= [(Y_i - a - bx_i) - \epsilon]^2 - (Y_i - a - bx_i)^2 \\ &= -2(Y_i - a - bx_i)\epsilon + \epsilon^2. \end{aligned}$$

Consequently, (I.2) can be written as

$$\sum_{i=1}^n [-2(Y_i - a - bx_i)\epsilon + \epsilon^2] \geq 0$$

or, equivalently,

$$n\epsilon^2 - \left( 2 \sum_{i=1}^n (Y_i - a - bx_i) \right) \epsilon \geq 0. \quad (\text{I.3})$$

Now (I.3) is true for all choices of  $\epsilon$ . But as a function of  $\epsilon$ , the left-hand side of (I.3) is a quadratic. Since the coefficient of  $\epsilon^2$  is  $n$  which is greater than 0, the graph of the quadratic is a concave up parabola. If  $\sum (Y_i - a - bx_i) \neq 0$ , then this quadratic, which can be factored as

$$\epsilon \left[ n\epsilon - \left( 2 \sum_{i=1}^n (Y_i - a - bx_i) \right) \right],$$

has two distinct roots. For any value of  $\epsilon$  *between* the roots, the graph of the parabola lies *below* the horizontal axis, that is,

$$n\epsilon^2 - \left( 2 \sum_{i=1}^n (Y_i - a - bx_i) \right) \epsilon < 0.$$

But this contradicts (I.3). So it must be the case that

$$\sum_{i=1}^n (Y_i - a - bx_i) = 0 \quad (\text{I.4})$$

and then (I.3) simplifies to  $n\epsilon^2 \geq 0$ . (Of course, this quadratic has a single root at  $\epsilon = 0$  and is never negative, which checks.) Equation (I.4) can be written more simply as

$$na + \left( \sum_{i=1}^n x_i \right) \beta = \sum_{i=1}^n Y_i$$

Since the sum of the deviates is always 0 (see Fact 4), we obtain

$$na = \sum_{i=1}^n Y_i.$$

or

$$a = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}. \quad (\text{I.5})$$

We determine  $b$  in a similar fashion. Since  $\text{SS}_{\text{Error}}(a, b)$  is minimal, if we choose any  $b'$ , then

$$\text{SS}_{\text{Error}}(a, b') \geq \text{SS}_{\text{Error}}(a, b).$$

This is equivalent to

$$\sum_{i=1}^n (Y_i - a - b'x_i)^2 \geq \sum_{i=1}^n (Y_i - a - bx_i)^2.$$

or

$$\sum_{i=1}^n \left[ (Y_i - a - b'x_i)^2 - (Y_i - a - bx_i)^2 \right] \geq 0. \quad (\text{I.6})$$

Now choose  $\delta$  so that  $b' = b + \delta$ . Then (I.6) becomes

$$\sum_{i=1}^n \left[ (Y_i - a - bx_i - \delta x_i)^2 - (Y_i - a - bx_i)^2 \right] \geq 0. \quad (\text{I.7})$$

Simplify each summand in (I.7):

$$\begin{aligned} (Y_i - a - bx_i - \delta x_i)^2 - (Y_i - a - bx_i)^2 &= [(Y_i - a - bx_i) - \delta x_i]^2 - (Y_i - a - bx_i)^2 \\ &= -2(Y_i - a - bx_i)\delta x_i + \delta^2 x_i^2. \end{aligned}$$

Consequently, (I.7) can be written as

$$\sum_{i=1}^n [-2(Y_i - a - bx_i)\delta x_i + \delta^2 x_i^2] \geq 0$$

or, equivalently,

$$\left( \sum_{i=1}^n x_i^2 \right) \delta^2 - \left( 2 \sum_{i=1}^n (Y_i - a - bx_i)x_i \right) \delta \geq 0. \quad (\text{I.8})$$

Now repeat the argument we made earlier for  $\epsilon$ . As a function of  $\delta$ , the graph of the left-hand side of (I.8) is a concave up parabola. If  $\sum (Y_i - a - bx_i)x_i \neq 0$ , then this parabola has two distinct roots. For any value of  $\delta$  between the roots, the parabola lies below the horizontal axis so

$$\left( \sum_{i=1}^n x_i^2 \right) \delta^2 - \left( 2 \sum_{i=1}^n (Y_i - a - bx_i)x_i \right) \delta < 0.$$

But this contradicts (I.8). So it must be the case that

$$\sum_{i=1}^n (Y_i - a - bx_i)x_i = 0$$

or

$$\left( \sum_{i=1}^n x_i^2 \right) b + \left( \sum_{i=1}^n x_i \right) a = \sum_{i=1}^n Y_i x_i.$$

Again  $\sum x_i = 0$ , so we obtain

$$b = \frac{\sum_{i=1}^n Y_i x_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n Y_i (X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}. \quad (\text{I.9})$$

Use (A.4) to simplify the numerator for  $b$  in the last expression in (I.9).

$$\sum_{i=1}^n Y_i (X_i - \bar{X}) = \sum_{i=1}^n X_i Y_i - \bar{X} \sum_{i=1}^n Y_i = \sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}.$$

We can now rewrite (I.9) using the expression above as the numerator and the sum of squares formula (A.7) as the denominator to obtain the expression for the slope,

$$b = \frac{\sum_{i=1}^n X_i Y_i - \frac{(\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{n}}{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}}. \quad (\text{I.10})$$

Using (I.5) for the intercept and the expression for  $b$  in (I.10) for the slope makes

$$Y = \bar{Y} + b(X - \bar{X})$$

the best-fitting straight line according to the least squares criterion.

# Tables of Distributions and Critical Values

1. Table of Critical Values for the Wilcoxon Rank-Sum Test
2. Cumulative standard normal distribution
3. Student's  $t$  distribution
4. Cumulative chi-square distribution
5. Critical Values of the  $q$  Statistic for the Student-Newman-Keuls Test
6. Critical Values for Duncan's Multiple Range Test

### J.1 Table of Critical Values for the Wilcoxon Rank-Sum Test

The tables on the following pages provide critical values for the Wilcoxon rank-sum test for independent samples with sizes from 3 to 25. Column  $m$  is the sample size for the smaller sample and column  $n$  is the sample size for the larger sample. If the sample sizes are equal, either sample can be designated  $m$ . For each pair of sample sizes  $(m, n)$  there are two sets of critical values, one set for one-tail  $\alpha = 0.025$  and two-tail  $\alpha = 0.05$  and a second set for one-tail  $\alpha = 0.05$  and two-tail  $\alpha = 0.10$ . Suppose for a two-tailed test at  $\alpha = 0.05$  we have  $m = 8$  and  $n = 9$ . In the appropriate row and column we find the following numbers 51 93 16 0.0232. The 51 and 93 are the lower and upper critical values for  $W_X$ , the statistic testing  $H_0: M_X = M_Y$ . If  $W_X \leq 51$  or  $W_X \geq 93$ ,  $H_0$  would be rejected. The value 0.0232 is the exact  $P$  value for the critical values of 51 or 93. The 16 under the column heading  $d$  is called the depth. Basically  $d$  is the depth one must go into the rank-ordered elementary estimates from each end to find the confidence limit values. In this case, the 16th smallest elementary estimate and the 16th largest elementary estimate are the 95% confidence interval limits for  $M_x - M_y$ .

1-tail 2-tail		$\alpha = 0.025$ $\alpha = 0.05$			$\alpha = 0.05$ $\alpha = 0.10$			1-tail 2-tail		$\alpha = 0.025$ $\alpha = 0.05$			$\alpha = 0.05$ $\alpha = 0.10$						
<i>m</i>	<i>n</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>m</i>	<i>n</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>W</i>	<i>d</i>	<i>P</i>				
3	3				6	15	1	.0500	5	10	23	57	9	.0200	26	54	12	.0496	
3	4				6	18	1	.0286	5	11	24	61	10	.0190	27	58	13	.0449	
3	5	6	21	1	.0179	7	20	2	.0357	5	12	26	64	12	.0242	28	62	14	.0409
3	6	7	23	2	.0238	8	22	3	.0476	5	13	27	68	13	.0230	30	65	16	.0473
3	7	7	26	2	.0167	8	25	3	.0333	5	14	28	72	14	.0218	31	69	17	.0435
3	8	8	28	3	.0242	9	27	4	.0424	5	15	29	76	15	.0209	33	72	19	.0491
3	9	8	31	3	.0182	10	29	5	.0500	5	16	30	80	16	.0201	34	76	20	.0455
3	10	9	33	4	.0245	10	32	5	.0385	5	17	32	83	18	.0238	35	80	21	.0425
3	11	9	36	4	.0192	11	34	6	.0440	5	18	33	87	19	.0229	37	83	23	.0472
3	12	10	38	5	.0242	11	37	6	.0352	5	19	34	91	20	.0220	38	87	24	.0442
3	13	10	41	5	.0196	12	39	7	.0411	5	20	35	95	21	.0212	40	90	26	.0485
3	14	11	43	6	.0235	13	41	8	.0456	5	21	37	98	23	.0243	41	94	27	.0457
3	15	11	46	6	.0196	13	44	8	.0380	5	22	38	102	24	.0234	43	97	29	.0496
3	16	12	48	7	.0237	14	46	9	.0423	5	23	39	106	25	.0226	44	101	30	.0469
3	17	12	51	7	.0202	15	48	10	.0465	5	24	40	110	26	.0219	45	105	31	.0445
3	18	13	53	8	.0233	15	51	10	.0398	5	25	42	113	28	.0246	47	108	33	.0480
3	19	13	56	8	.0201	16	53	11	.0435	6	6	26	52	6	.0206	28	50	8	.0465
3	20	14	58	9	.0232	17	55	12	.0469	6	7	27	57	7	.0175	29	55	9	.0367
3	21	14	61	9	.0203	17	58	12	.0410	6	8	29	61	9	.0213	31	59	11	.0406
3	22	15	63	10	.0230	18	60	13	.0443	6	9	31	65	11	.0248	33	63	13	.0440
3	23	15	66	10	.0204	19	62	14	.0473	6	10	32	70	12	.0210	35	67	15	.0467
3	24	16	68	11	.0229	19	65	14	.0421	6	11	34	74	14	.0238	37	71	17	.0491
3	25	16	71	11	.0205	20	67	15	.0449	6	12	35	79	15	.0207	38	76	18	.0415
4	4	10	26	1	.0143	11	25	2	.0286	6	13	37	83	17	.0231	40	80	20	.0437
4	5	11	29	2	.0159	12	28	3	.0317	6	14	38	88	18	.0204	42	84	22	.0457
4	6	12	32	3	.0190	13	31	4	.0333	6	15	40	92	20	.0224	44	88	24	.0474
4	7	13	35	4	.0212	14	34	5	.0364	6	16	42	96	22	.0244	46	92	26	.0490
4	8	14	38	5	.0242	15	37	6	.0364	6	17	43	101	23	.0219	47	97	27	.0433
4	9	14	42	5	.0168	16	40	7	.0378	6	18	45	105	25	.0236	49	101	29	.0448
4	10	15	45	6	.0180	17	43	8	.0380	6	19	46	110	26	.0214	51	105	31	.0462
4	11	16	48	7	.0198	18	46	9	.0388	6	20	48	114	28	.0229	53	109	33	.0475
4	12	17	51	8	.0209	19	49	10	.0390	6	21	50	118	30	.0244	55	113	35	.0487
4	13	18	54	9	.0223	20	52	11	.0395	6	22	51	123	31	.0224	57	117	37	.0498
4	14	19	57	10	.0232	21	55	12	.0395	6	23	53	127	33	.0237	58	122	38	.0452
4	15	20	60	11	.0243	22	58	13	.0400	6	24	54	132	34	.0219	60	126	40	.0463
4	16	21	63	12	.0250	24	60	15	.0497	6	25	56	136	36	.0231	62	130	42	.0473
4	17	21	67	12	.0202	25	63	16	.0493	7	7	36	69	9	.0189	39	66	12	.0487
4	18	22	70	13	.0212	26	66	17	.0491	7	8	38	74	11	.0200	41	71	14	.0469
4	19	23	73	14	.0219	27	69	18	.0487	7	9	40	79	13	.0209	43	76	16	.0454
4	20	24	76	15	.0227	28	72	19	.0485	7	10	42	84	15	.0215	45	81	18	.0439
4	21	25	79	16	.0233	29	75	20	.0481	7	11	44	89	17	.0221	47	86	20	.0427
4	22	26	82	17	.0240	30	78	21	.0480	7	12	46	94	19	.0225	49	91	22	.0416
4	23	27	85	18	.0246	31	81	22	.0477	7	13	48	99	21	.0228	52	95	25	.0484
4	24	27	89	18	.0211	32	84	23	.0475	7	14	50	104	23	.0230	54	100	27	.0469
4	25	28	92	19	.0217	33	87	24	.0473	7	15	52	109	25	.0233	56	105	29	.0455
5	5	17	38	3	.0159	19	36	5	.0476	7	16	54	114	27	.0234	58	110	31	.0443
5	6	18	42	4	.0152	20	40	6	.0411	7	17	56	119	29	.0236	61	114	34	.0497
5	7	20	45	6	.0240	21	44	7	.0366	7	18	58	124	31	.0237	63	119	36	.0484
5	8	21	49	7	.0225	23	47	9	.0466	7	19	60	129	33	.0238	65	124	38	.0471
5	9	22	53	8	.0210	24	51	10	.0415	7	20	62	134	35	.0239	67	129	40	.0460

1-tail 2-tail		$\alpha = 0.025$ $\alpha = 0.05$			$\alpha = 0.05$ $\alpha = 0.10$			1-tail 2-tail		$\alpha = 0.025$ $\alpha = 0.05$			$\alpha = 0.05$ $\alpha = 0.10$						
<i>m</i>	<i>n</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>m</i>	<i>n</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>W</i>	<i>d</i>	<i>P</i>				
7	21	64	139	37	.0240	69	134	42	.0449	10	20	110	200	56	.0245	117	193	62	.0498
7	22	66	144	39	.0240	72	138	45	.0492	10	21	113	207	59	.0241	120	200	65	.0478
7	23	68	149	41	.0241	74	143	47	.0481	10	22	116	214	62	.0237	123	207	68	.0459
7	24	70	154	43	.0241	76	148	49	.0470	10	23	119	221	65	.0233	127	213	72	.0482
7	25	72	159	45	.0242	78	153	51	.0461	10	24	122	228	68	.0230	130	220	75	.0465
8	8	49	87	14	.0249	51	85	16	.0415	10	25	126	234	72	.0248	134	226	79	.0486
8	9	51	93	16	.0232	54	90	19	.0464	11	11	96	157	31	.0237	100	153	34	.0440
8	10	53	99	18	.0217	56	96	21	.0416	11	12	99	165	34	.0219	104	160	38	.0454
8	11	55	105	20	.0204	59	101	24	.0454	11	13	103	172	38	.0237	108	167	42	.0467
8	12	58	110	23	.0237	62	106	27	.0489	11	14	106	180	41	.0221	112	174	46	.0477
8	13	60	116	25	.0223	64	112	29	.0445	11	15	110	187	45	.0236	116	181	50	.0486
8	14	62	122	27	.0211	67	117	32	.0475	11	16	113	195	48	.0221	120	188	54	.0494
8	15	65	127	30	.0237	69	123	34	.0437	11	17	117	202	52	.0235	123	196	57	.0453
8	16	67	133	32	.0224	72	128	37	.0463	11	18	121	209	56	.0247	127	203	61	.0461
8	17	70	138	35	.0247	75	133	40	.0487	11	19	124	217	59	.0233	131	210	65	.0468
8	18	72	144	37	.0235	77	139	42	.0452	11	20	128	224	63	.0244	135	217	69	.0474
8	19	74	150	39	.0224	80	144	45	.0475	11	21	131	232	66	.0230	139	224	73	.0480
8	20	77	155	42	.0244	83	149	48	.0495	11	22	135	239	70	.0240	143	231	77	.0486
8	21	79	161	44	.0233	85	155	50	.0464	11	23	139	246	74	.0250	147	238	81	.0490
8	22	81	167	46	.0223	88	160	53	.0483	11	24	142	254	77	.0237	151	245	85	.0495
8	23	84	172	49	.0240	90	166	55	.0454	11	25	146	261	81	.0246	155	252	89	.0499
8	24	86	178	51	.0231	93	171	58	.0472	12	12	115	185	38	.0225	120	180	42	.0444
8	25	89	183	54	.0247	96	176	61	.0488	12	13	119	193	42	.0229	125	187	47	.0488
9	9	62	109	18	.0200	66	105	22	.0470	12	14	123	201	46	.0232	129	195	51	.0475
9	10	65	115	21	.0217	69	111	25	.0474	12	15	127	209	50	.0234	133	203	55	.0463
9	11	68	121	24	.0232	72	117	28	.0476	12	16	131	217	54	.0236	138	210	60	.0500
9	12	71	127	27	.0245	75	123	31	.0477	12	17	135	225	58	.0238	142	218	64	.0486
9	13	73	134	29	.0217	78	129	34	.0478	12	18	139	233	62	.0239	146	226	68	.0474
9	14	76	140	32	.0228	81	135	37	.0478	12	19	143	241	66	.0240	150	234	72	.0463
9	15	79	146	35	.0238	84	141	40	.0478	12	20	147	249	70	.0241	155	241	77	.0493
9	16	82	152	38	.0247	87	147	43	.0477	12	21	151	257	74	.0242	159	249	81	.0481
9	17	84	159	40	.0223	90	153	46	.0476	12	22	155	265	78	.0242	163	257	85	.0471
9	18	87	165	43	.0231	93	159	49	.0475	12	23	159	273	82	.0243	168	264	90	.0496
9	19	90	171	46	.0239	96	165	52	.0474	12	24	163	281	86	.0243	172	272	94	.0486
9	20	93	177	49	.0245	99	171	55	.0473	12	25	167	289	90	.0243	176	280	98	.0475
9	21	95	184	51	.0225	102	177	58	.0472	13	13	136	215	46	.0221	142	209	51	.0454
9	22	98	190	54	.0231	105	183	61	.0471	13	14	141	223	51	.0241	147	217	56	.0472
9	23	101	196	57	.0237	108	189	64	.0470	13	15	145	232	55	.0232	152	225	61	.0489
9	24	104	202	60	.0243	111	195	67	.0469	13	16	150	240	60	.0250	156	234	65	.0458
9	25	107	208	63	.0249	114	201	70	.0468	13	17	154	249	64	.0240	161	242	70	.0472
10	10	78	132	24	.0216	82	128	28	.0446	13	18	158	258	68	.0232	166	250	75	.0485
10	11	81	139	27	.0215	86	134	32	.0493	13	19	163	266	73	.0247	171	258	80	.0497
10	12	84	146	30	.0213	89	141	35	.0465	13	20	167	275	77	.0238	175	267	84	.0470
10	13	88	152	34	.0247	92	148	38	.0441	13	21	171	284	81	.0231	180	275	89	.0481
10	14	91	159	37	.0242	96	154	42	.0478	13	22	176	292	86	.0243	185	283	94	.0491
10	15	94	166	40	.0238	99	161	45	.0455	13	23	180	301	90	.0236	189	292	98	.0467
10	16	97	173	43	.0234	103	167	49	.0487	13	24	185	309	95	.0247	194	300	103	.0476
10	17	100	180	46	.0230	106	174	52	.0465	13	25	189	318	99	.0240	199	308	108	.0485
10	18	103	187	49	.0226	110	180	56	.0493	14	14	160	246	56	.0249	166	240	61	.0469
10	19	107	193	53	.0250	113	187	59	.0472	14	15	164	256	60	.0229	171	249	66	.0466

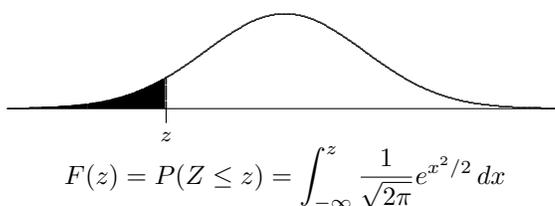
1-tail 2-tail		$\alpha = 0.025$ $\alpha = 0.05$			$\alpha = 0.05$ $\alpha = 0.10$			1-tail 2-tail		$\alpha = 0.025$ $\alpha = 0.05$			$\alpha = 0.05$ $\alpha = 0.10$						
<i>m</i>	<i>n</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>m</i>	<i>n</i>	<i>W</i>	<i>d</i>	<i>P</i>	<i>W</i>	<i>d</i>	<i>P</i>				
14	16	169	265	65	.0236	176	258	72	.0463	17	24	282	432	130	.0239	294	420	141	.0492
14	17	174	274	70	.0242	182	266	78	.0500	17	25	288	443	136	.0238	300	431	147	.0480
14	18	179	283	75	.0247	187	275	83	.0495	18	18	270	396	100	.0235	280	386	109	.0485
14	19	183	293	79	.0230	192	284	88	.0489	18	19	277	407	107	.0246	287	397	116	.0490
14	20	188	302	84	.0235	197	293	93	.0484	18	20	283	419	113	.0238	294	408	123	.0495
14	21	193	311	89	.0239	202	302	98	.0480	18	21	290	430	120	.0247	301	419	130	.0499
14	22	198	320	94	.0243	207	311	103	.0475	18	22	296	442	126	.0240	307	431	136	.0474
14	23	203	329	99	.0247	212	320	108	.0471	18	23	303	453	133	.0248	314	442	143	.0478
14	24	207	339	103	.0233	218	328	114	.0498	18	24	309	465	139	.0240	321	453	150	.0481
14	25	212	348	108	.0236	223	337	119	.0492	18	25	316	476	146	.0248	328	464	157	.0484
15	15	184	281	65	.0227	192	273	73	.0488	19	19	303	438	114	.0248	313	428	123	.0482
15	16	190	290	71	.0247	197	283	78	.0466	19	20	309	451	120	.0234	320	440	130	.0474
15	17	195	300	76	.0243	203	292	84	.0485	19	21	316	463	127	.0236	328	451	138	.0494
15	18	200	310	81	.0239	208	302	89	.0465	19	22	323	475	134	.0238	335	463	145	.0486
15	19	205	320	86	.0235	214	311	95	.0482	19	23	330	487	141	.0240	342	475	152	.0478
15	20	210	330	91	.0232	220	320	101	.0497	19	24	337	499	148	.0241	350	486	160	.0496
15	21	216	339	97	.0247	225	330	106	.0478	19	25	344	511	155	.0243	357	498	167	.0488
15	22	221	349	102	.0243	231	339	112	.0492	20	20	337	483	128	.0245	348	472	138	.0482
15	23	226	359	107	.0239	236	349	117	.0474	20	21	344	496	135	.0241	356	484	146	.0490
15	24	231	369	112	.0235	242	358	123	.0486	20	22	351	509	142	.0236	364	496	154	.0497
15	25	237	378	118	.0248	248	367	129	.0499	20	23	359	521	150	.0246	371	509	161	.0478
16	16	211	317	76	.0234	219	309	84	.0469	20	24	366	534	157	.0242	379	521	169	.0484
16	17	217	327	82	.0243	225	319	90	.0471	20	25	373	547	164	.0237	387	533	177	.0490
16	18	222	338	87	.0231	231	329	96	.0473	21	21	373	530	143	.0245	385	518	154	.0486
16	19	228	348	93	.0239	237	339	102	.0474	21	22	381	543	151	.0249	393	531	162	.0482
16	20	234	358	99	.0247	243	349	108	.0475	21	23	388	557	158	.0238	401	544	170	.0478
16	21	239	369	104	.0235	249	359	114	.0475	21	24	396	570	166	.0242	410	556	179	.0497
16	22	245	379	110	.0242	255	369	120	.0476	21	25	404	583	174	.0245	418	569	187	.0492
16	23	251	389	116	.0248	261	379	126	.0476	22	22	411	579	159	.0247	424	566	171	.0491
16	24	256	400	121	.0238	267	389	132	.0476	22	23	419	593	167	.0244	432	580	179	.0477
16	25	262	410	127	.0243	273	399	138	.0476	22	24	427	607	175	.0242	441	593	188	.0486
17	17	240	355	88	.0243	249	346	97	.0493	22	25	435	621	183	.0240	450	606	197	.0494
17	18	246	366	94	.0243	255	357	103	.0479	23	23	451	630	176	.0249	465	616	189	.0499
17	19	252	377	100	.0243	262	367	110	.0499	23	24	459	645	184	.0242	474	630	198	.0497
17	20	258	388	106	.0242	268	378	116	.0485	23	25	468	659	193	.0246	483	644	207	.0495
17	21	264	399	112	.0242	274	389	122	.0473	24	24	492	684	193	.0241	507	669	207	.0486
17	22	270	410	118	.0241	281	399	129	.0490	24	25	501	699	202	.0241	517	683	217	.0496
17	23	276	421	124	.0240	287	410	135	.0477	25	25	536	739	212	.0247	552	723	227	.0497

## J.2 Cumulative Standard Normal Distribution

Table C.3 is a tabulation of the cumulative standard normal distribution ( $Z$ ) for values from  $-3.99$  to  $3.99$ . This table can be used to calculate probabilities for various CDFs of normal distributions that have been standardized by the  $\frac{X-\mu}{\sigma}$  transformation. While normal distributions can have various values for the mean and standard deviation, the  $Z$  transformation has a mean of  $\mu = 0$  and a standard deviation of  $\sigma = 1$ . To use this table, transform the normal variate  $X$  to a  $Z$  value. For example, suppose  $X$  is distributed as normal with  $\mu = 100$  and  $\sigma = 6$ , then

$$P(X \leq 93) = P\left(Z = \frac{93 - 100}{6}\right) = P(Z \leq -1.17).$$

Next find the value at the intersection of the “ $-1.1$ ” row and the “ $0.07$ ” column,  $0.1210$ . This value is the probability of  $X$  less than or equal to  $93$  in the original distribution. For more examples, see Chapter 3.

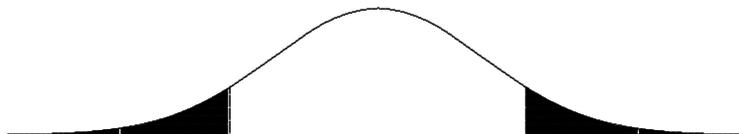


$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379



### J.3 Student's $t$ Distribution

Unlike Table C.3 that collapsed all normal distributions into a single standard normal distribution, Table C.4 represents a summary of partial CDFs for a number of  $t$  distributions. Each row of this table summarizes a different  $t$  distribution as indicated by its degrees of freedom ( $n - 1$ ). The body of the table presents  $t_0$  values that represent cutoff points for various probabilities ranging from 0.10 to 0.0025 for one tail and 0.20 to 0.005 for two tails and that are tabulated as various columns. For example, if  $n = 20$ , then  $df = 19$  and for  $P(-t_0 \leq t \leq t_0) = 0.95$ , the table value  $t_0$  would be 2.093. This value is found as the intersection of the row labeled “19” and the column headed “0.05” (2-tail). This means that  $-2.093$  and  $2.093$  cut off 0.025 in each tail of the distribution. For  $P(t \leq t_0) = 0.10$ , the table value  $t_0$  would be  $-1.328$ . Left tail values are negative. For  $P(t \geq t_0) = 0.10$ , the table value  $t_0$  would be  $+1.328$ . For  $P(t \leq t_0) = 0.90$ , the table value  $t_0$  would again be  $+1.328$ . For more examples see Chapters 4 and 6. Notice that the limit of the  $t$  distribution family (where  $df = \infty$ ) is the standard normal distribution ( $Z$ ). The values in the last row of Table C.4 can also be found using Table C.3.



	<b>1-tail</b>	<b>0.10</b>	<b>0.05</b>	<b>0.025</b>	0.0125	<b>0.01</b>	0.0083	0.00625	<b>0.005</b>	0.00416	<b>0.0025</b>	
	<b>2-tail</b>	<b>0.20</b>	<b>0.10</b>	<b>0.050</b>	0.0250	<b>0.02</b>	0.0166	0.01250	<b>0.010</b>	0.00833	<b>0.0050</b>	<b>2-tail</b>
<b>df: 1</b>		3.078	6.314	12.706	25.452	31.821	38.188	50.923	63.657	76.390	127.3	<b>df: 1</b>
<b>2</b>		1.886	2.920	4.303	6.205	6.965	7.649	8.860	9.925	10.886	14.09	<b>2</b>
<b>3</b>		1.638	2.353	3.182	4.177	4.541	4.857	5.392	5.841	6.232	7.453	<b>3</b>
<b>4</b>		1.533	2.132	2.776	3.495	3.747	3.961	4.315	4.604	4.851	5.598	<b>4</b>
<b>5</b>		1.476	2.015	2.571	3.163	3.365	3.534	3.810	4.032	4.219	4.773	<b>5</b>
<b>6</b>		1.440	1.943	2.447	2.969	3.143	3.287	3.521	3.707	3.863	4.317	<b>6</b>
<b>7</b>		1.415	1.895	2.365	2.841	2.998	3.128	3.335	3.499	3.636	4.029	<b>7</b>
<b>8</b>		1.397	1.860	2.306	2.752	2.896	3.016	3.206	3.355	3.479	3.833	<b>8</b>
<b>9</b>		1.383	1.833	2.262	2.685	2.821	2.933	3.111	3.250	3.364	3.690	<b>9</b>
<b>10</b>		1.372	1.812	2.228	2.634	2.764	2.870	3.038	3.169	3.277	3.581	<b>10</b>
<b>11</b>		1.363	1.796	2.201	2.593	2.718	2.820	2.981	3.106	3.208	3.497	<b>11</b>
<b>12</b>		1.356	1.782	2.179	2.560	2.681	2.779	2.934	3.055	3.153	3.428	<b>12</b>
<b>13</b>		1.350	1.771	2.160	2.533	2.650	2.746	2.896	3.012	3.107	3.372	<b>13</b>
<b>14</b>		1.345	1.761	2.145	2.510	2.624	2.718	2.864	2.977	3.069	3.326	<b>14</b>
<b>15</b>		1.341	1.753	2.131	2.490	2.602	2.694	2.837	2.947	3.036	3.286	<b>15</b>
<b>16</b>		1.337	1.746	2.120	2.473	2.583	2.673	2.813	2.921	3.008	3.252	<b>16</b>
<b>17</b>		1.333	1.740	2.110	2.458	2.567	2.655	2.793	2.898	2.984	3.222	<b>17</b>
<b>18</b>		1.330	1.734	2.101	2.445	2.552	2.639	2.775	2.878	2.963	3.197	<b>18</b>
<b>19</b>		1.328	1.729	2.093	2.433	2.539	2.625	2.759	2.861	2.944	3.174	<b>19</b>
<b>20</b>		1.325	1.725	2.086	2.423	2.528	2.613	2.744	2.845	2.927	3.153	<b>20</b>
<b>21</b>		1.323	1.721	2.080	2.414	2.518	2.601	2.732	2.831	2.912	3.135	<b>21</b>
<b>22</b>		1.321	1.717	2.074	2.405	2.508	2.591	2.720	2.819	2.899	3.119	<b>22</b>
<b>23</b>		1.319	1.714	2.069	2.398	2.500	2.582	2.710	2.807	2.886	3.104	<b>23</b>
<b>24</b>		1.318	1.711	2.064	2.391	2.492	2.574	2.700	2.797	2.875	3.091	<b>24</b>
<b>25</b>		1.316	1.708	2.060	2.385	2.485	2.566	2.692	2.787	2.865	3.078	<b>25</b>

	<b>1-tail</b>	<b>0.10</b>	<b>0.05</b>	<b>0.025</b>	0.0125	<b>0.01</b>	0.0083	0.00625	<b>0.005</b>	0.00416	<b>0.0025</b>	
	<b>2-tail</b>	<b>0.20</b>	<b>0.10</b>	<b>0.050</b>	0.0250	<b>0.02</b>	0.0166	0.01250	<b>0.010</b>	0.00833	<b>0.0050</b>	<b>1-tail</b>
												<b>2-tail</b>
<b>df: 26</b>		1.315	1.706	2.056	2.379	2.479	2.559	2.684	2.779	2.856	3.067	<b>df: 26</b>
<b>27</b>		1.314	1.703	2.052	2.373	2.473	2.552	2.676	2.771	2.847	3.057	<b>27</b>
<b>28</b>		1.313	1.701	2.048	2.368	2.467	2.546	2.669	2.763	2.839	3.047	<b>28</b>
<b>29</b>		1.311	1.699	2.045	2.364	2.462	2.541	2.663	2.756	2.832	3.038	<b>29</b>
<b>30</b>		1.310	1.697	2.042	2.360	2.457	2.536	2.657	2.750	2.825	3.030	<b>30</b>
<b>31</b>		1.309	1.696	2.040	2.356	2.453	2.531	2.652	2.744	2.818	3.022	<b>31</b>
<b>32</b>		1.309	1.694	2.037	2.352	2.449	2.526	2.647	2.738	2.812	3.015	<b>32</b>
<b>33</b>		1.308	1.692	2.035	2.348	2.445	2.522	2.642	2.733	2.807	3.008	<b>33</b>
<b>34</b>		1.307	1.691	2.032	2.345	2.441	2.518	2.638	2.728	2.802	3.002	<b>34</b>
<b>35</b>		1.306	1.690	2.030	2.342	2.438	2.515	2.633	2.724	2.797	2.996	<b>35</b>
<b>36</b>		1.306	1.688	2.028	2.339	2.434	2.511	2.629	2.719	2.792	2.990	<b>36</b>
<b>37</b>		1.305	1.687	2.026	2.336	2.431	2.508	2.626	2.715	2.788	2.985	<b>37</b>
<b>38</b>		1.304	1.686	2.024	2.334	2.429	2.505	2.622	2.712	2.783	2.980	<b>38</b>
<b>39</b>		1.304	1.685	2.023	2.331	2.426	2.502	2.619	2.708	2.780	2.976	<b>39</b>
<b>40</b>		1.303	1.684	2.021	2.329	2.423	2.499	2.616	2.704	2.776	2.971	<b>40</b>
<b>41</b>		1.303	1.683	2.020	2.327	2.421	2.496	2.613	2.701	2.772	2.967	<b>41</b>
<b>42</b>		1.302	1.682	2.018	2.325	2.418	2.494	2.610	2.698	2.769	2.963	<b>42</b>
<b>43</b>		1.302	1.681	2.017	2.323	2.416	2.491	2.607	2.695	2.766	2.959	<b>43</b>
<b>44</b>		1.301	1.680	2.015	2.321	2.414	2.489	2.605	2.692	2.763	2.956	<b>44</b>
<b>45</b>		1.301	1.679	2.014	2.319	2.412	2.487	2.602	2.690	2.760	2.952	<b>45</b>
<b>46</b>		1.300	1.679	2.013	2.317	2.410	2.485	2.600	2.687	2.757	2.949	<b>46</b>
<b>47</b>		1.300	1.678	2.012	2.315	2.408	2.483	2.597	2.685	2.755	2.946	<b>47</b>
<b>48</b>		1.299	1.677	2.011	2.314	2.407	2.481	2.595	2.682	2.752	2.943	<b>48</b>
<b>49</b>		1.299	1.677	2.010	2.312	2.405	2.479	2.593	2.680	2.750	2.940	<b>49</b>
<b>50</b>		1.299	1.676	2.009	2.311	2.403	2.477	2.591	2.678	2.747	2.937	<b>50</b>
<b>51</b>		1.298	1.675	2.008	2.310	2.402	2.476	2.589	2.676	2.745	2.934	<b>51</b>
<b>52</b>		1.298	1.675	2.007	2.308	2.400	2.474	2.588	2.674	2.743	2.932	<b>52</b>
<b>53</b>		1.298	1.674	2.006	2.307	2.399	2.472	2.586	2.672	2.741	2.929	<b>53</b>
<b>54</b>		1.297	1.674	2.005	2.306	2.397	2.471	2.584	2.670	2.739	2.927	<b>54</b>
<b>55</b>		1.297	1.673	2.004	2.304	2.396	2.469	2.583	2.668	2.737	2.925	<b>55</b>
<b>56</b>		1.297	1.673	2.003	2.303	2.395	2.468	2.581	2.667	2.735	2.923	<b>56</b>
<b>57</b>		1.297	1.672	2.002	2.302	2.394	2.467	2.579	2.665	2.733	2.920	<b>57</b>
<b>58</b>		1.296	1.672	2.002	2.301	2.392	2.465	2.578	2.663	2.732	2.918	<b>58</b>
<b>59</b>		1.296	1.671	2.001	2.300	2.391	2.464	2.577	2.662	2.730	2.916	<b>59</b>
<b>60</b>		1.296	1.671	2.000	2.299	2.390	2.463	2.575	2.660	2.729	2.915	<b>60</b>
<b>61</b>		1.296	1.670	2.000	2.298	2.389	2.462	2.574	2.659	2.727	2.913	<b>61</b>
<b>62</b>		1.295	1.670	1.999	2.297	2.388	2.461	2.573	2.657	2.726	2.911	<b>62</b>
<b>63</b>		1.295	1.669	1.998	2.296	2.387	2.460	2.571	2.656	2.724	2.909	<b>63</b>
<b>64</b>		1.295	1.669	1.998	2.295	2.386	2.459	2.570	2.655	2.723	2.908	<b>64</b>
<b>65</b>		1.295	1.669	1.997	2.295	2.385	2.458	2.569	2.654	2.721	2.906	<b>65</b>
<b>66</b>		1.295	1.668	1.997	2.294	2.384	2.457	2.568	2.652	2.720	2.904	<b>66</b>
<b>67</b>		1.294	1.668	1.996	2.293	2.383	2.456	2.567	2.651	2.719	2.903	<b>67</b>
<b>68</b>		1.294	1.668	1.995	2.292	2.382	2.455	2.566	2.650	2.718	2.902	<b>68</b>
<b>69</b>		1.294	1.667	1.995	2.291	2.382	2.454	2.565	2.649	2.716	2.900	<b>69</b>
<b>70</b>		1.294	1.667	1.994	2.291	2.381	2.453	2.564	2.648	2.715	2.899	<b>70</b>
<b>71</b>		1.294	1.667	1.994	2.290	2.380	2.452	2.563	2.647	2.714	2.897	<b>71</b>
<b>72</b>		1.293	1.666	1.993	2.289	2.379	2.451	2.562	2.646	2.713	2.896	<b>72</b>
<b>73</b>		1.293	1.666	1.993	2.289	2.379	2.450	2.561	2.645	2.712	2.895	<b>73</b>
<b>74</b>		1.293	1.666	1.993	2.288	2.378	2.450	2.560	2.644	2.711	2.894	<b>74</b>
<b>75</b>		1.293	1.665	1.992	2.287	2.377	2.449	2.559	2.643	2.710	2.892	<b>75</b>

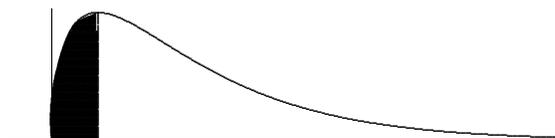
	<b>1-tail</b>	<b>0.10</b>	<b>0.05</b>	<b>0.025</b>	0.0125	<b>0.01</b>	0.0083 $\bar{3}$	0.00625	<b>0.005</b>	0.00416 $\bar{6}$	<b>0.0025</b>	
	<b>2-tail</b>	<b>0.20</b>	<b>0.10</b>	<b>0.050</b>	0.0250	<b>0.02</b>	0.0166 $\bar{6}$	0.01250	<b>0.010</b>	0.00833 $\bar{3}$	<b>0.0050</b>	<b>1-tail</b>
												<b>2-tail</b>
<b>df: 76</b>		1.293	1.665	1.992	2.287	2.376	2.448	2.559	2.642	2.709	2.891	<b>df: 76</b>
<b>77</b>		1.293	1.665	1.991	2.286	2.376	2.447	2.558	2.641	2.708	2.890	<b>77</b>
<b>78</b>		1.292	1.665	1.991	2.285	2.375	2.447	2.557	2.640	2.707	2.889	<b>78</b>
<b>79</b>		1.292	1.664	1.990	2.285	2.374	2.446	2.556	2.640	2.706	2.888	<b>79</b>
<b>80</b>		1.292	1.664	1.990	2.284	2.374	2.445	2.555	2.639	2.705	2.887	<b>80</b>
<b>81</b>		1.292	1.664	1.990	2.284	2.373	2.445	2.555	2.638	2.705	2.886	<b>81</b>
<b>82</b>		1.292	1.664	1.989	2.283	2.373	2.444	2.554	2.637	2.704	2.885	<b>82</b>
<b>83</b>		1.292	1.663	1.989	2.283	2.372	2.443	2.553	2.636	2.703	2.884	<b>83</b>
<b>84</b>		1.292	1.663	1.989	2.282	2.372	2.443	2.553	2.636	2.702	2.883	<b>84</b>
<b>85</b>		1.292	1.663	1.988	2.282	2.371	2.442	2.552	2.635	2.701	2.882	<b>85</b>
<b>86</b>		1.291	1.663	1.988	2.281	2.370	2.442	2.551	2.634	2.701	2.881	<b>86</b>
<b>87</b>		1.291	1.663	1.988	2.281	2.370	2.441	2.551	2.634	2.700	2.880	<b>87</b>
<b>88</b>		1.291	1.662	1.987	2.280	2.369	2.441	2.550	2.633	2.699	2.880	<b>88</b>
<b>89</b>		1.291	1.662	1.987	2.280	2.369	2.440	2.549	2.632	2.699	2.879	<b>89</b>
<b>90</b>		1.291	1.662	1.987	2.280	2.368	2.440	2.549	2.632	2.698	2.878	<b>90</b>
<b>91</b>		1.291	1.662	1.986	2.279	2.368	2.439	2.548	2.631	2.697	2.877	<b>91</b>
<b>92</b>		1.291	1.662	1.986	2.279	2.368	2.439	2.548	2.630	2.697	2.876	<b>92</b>
<b>93</b>		1.291	1.661	1.986	2.278	2.367	2.438	2.547	2.630	2.696	2.876	<b>93</b>
<b>94</b>		1.291	1.661	1.986	2.278	2.367	2.438	2.547	2.629	2.695	2.875	<b>94</b>
<b>95</b>		1.291	1.661	1.985	2.277	2.366	2.437	2.546	2.629	2.695	2.874	<b>95</b>
<b>96</b>		1.290	1.661	1.985	2.277	2.366	2.437	2.546	2.628	2.694	2.873	<b>96</b>
<b>97</b>		1.290	1.661	1.985	2.277	2.365	2.436	2.545	2.627	2.693	2.873	<b>97</b>
<b>98</b>		1.290	1.661	1.984	2.276	2.365	2.436	2.545	2.627	2.693	2.872	<b>98</b>
<b>99</b>		1.290	1.660	1.984	2.276	2.365	2.435	2.544	2.626	2.692	2.871	<b>99</b>
<b>100</b>		1.290	1.660	1.984	2.276	2.364	2.435	2.544	2.626	2.692	2.871	<b>100</b>
<b>110</b>		1.289	1.659	1.982	2.272	2.361	2.431	2.539	2.621	2.687	2.865	<b>110</b>
<b>120</b>		1.289	1.658	1.980	2.270	2.358	2.428	2.536	2.617	2.683	2.860	<b>120</b>
<b>130</b>		1.288	1.657	1.978	2.268	2.355	2.425	2.533	2.614	2.679	2.856	<b>130</b>
<b>140</b>		1.288	1.656	1.977	2.266	2.353	2.423	2.530	2.611	2.676	2.852	<b>140</b>
<b>150</b>		1.287	1.655	1.976	2.264	2.351	2.421	2.528	2.609	2.674	2.849	<b>150</b>
<b>160</b>		1.287	1.654	1.975	2.263	2.350	2.419	2.526	2.607	2.671	2.847	<b>160</b>
<b>170</b>		1.287	1.654	1.974	2.261	2.348	2.418	2.525	2.605	2.669	2.844	<b>170</b>
<b>180</b>		1.286	1.653	1.973	2.260	2.347	2.417	2.523	2.603	2.668	2.842	<b>180</b>
<b>190</b>		1.286	1.653	1.973	2.259	2.346	2.415	2.522	2.602	2.666	2.840	<b>190</b>
<b>200</b>		1.286	1.653	1.972	2.258	2.345	2.414	2.520	2.601	2.665	2.838	<b>200</b>
<b>250</b>		1.285	1.651	1.969	2.255	2.341	2.410	2.516	2.596	2.659	2.832	<b>250</b>
<b>300</b>		1.284	1.650	1.968	2.253	2.339	2.407	2.513	2.592	2.656	2.828	<b>300</b>
<b>350</b>		1.284	1.649	1.967	2.251	2.337	2.406	2.511	2.590	2.653	2.825	<b>350</b>
<b>400</b>		1.284	1.649	1.966	2.250	2.336	2.404	2.509	2.588	2.651	2.823	<b>400</b>
<b>450</b>		1.283	1.648	1.965	2.249	2.335	2.403	2.508	2.587	2.650	2.821	<b>450</b>
<b>500</b>		1.283	1.648	1.965	2.248	2.334	2.402	2.507	2.586	2.649	2.820	<b>500</b>
<b>550</b>		1.283	1.648	1.964	2.248	2.333	2.401	2.506	2.585	2.648	2.818	<b>550</b>
<b>600</b>		1.283	1.647	1.964	2.247	2.333	2.401	2.505	2.584	2.647	2.817	<b>600</b>
<b>650</b>		1.283	1.647	1.964	2.247	2.332	2.400	2.505	2.583	2.646	2.817	<b>650</b>
<b>700</b>		1.283	1.647	1.963	2.246	2.332	2.400	2.504	2.583	2.646	2.816	<b>700</b>
<b>750</b>		1.283	1.647	1.963	2.246	2.331	2.399	2.504	2.582	2.645	2.815	<b>750</b>
<b>800</b>		1.283	1.647	1.963	2.246	2.331	2.399	2.503	2.582	2.645	2.815	<b>800</b>
<b>900</b>		1.282	1.647	1.963	2.245	2.330	2.398	2.503	2.581	2.644	2.814	<b>900</b>
<b>1000</b>		1.282	1.646	1.962	2.245	2.330	2.398	2.502	2.581	2.644	2.813	<b>1000</b>
$\infty$		1.282	1.645	1.960	2.241	2.326	2.394	2.498	2.576	2.638	2.807	$\infty$

### J.4 Cumulative Chi-Square Distribution

Table J.4 summarizes partial CDFs for  $\chi^2$  distributions with degrees of freedom from 1 to 60, where

$$F(\chi^2) = P(X^2 \leq \chi^2).$$

Each row represents a different  $\chi^2$  distribution and each column a different CDF from 0.005 to 0.995. The body of the table presents  $\chi^2$  values associated with these various probabilities. For example, if  $n = 20$ , then  $df = 19$  and  $\chi_{0.025}^2 = 8.91$  is found as the intersection of row “19” and column “0.025.” For a  $\chi^2$  distribution with 19 degrees of freedom  $P(\chi^2 \leq 8.91) = 0.025$ .  $\chi_{0.90}^2 = 27.2$  found as the intersection of row “19” and column “0.90” implies  $P(\chi^2 \leq 27.2) = 0.90$ . For more examples see Chapters 4 and 6.



df	0.005	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99	0.995
1	0.0000393	0.000157	0.000982	0.00393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7

<b>df</b>	<b>0.005</b>	<b>0.01</b>	<b>0.025</b>	<b>0.05</b>	<b>0.10</b>	<b>0.90</b>	<b>0.95</b>	<b>0.975</b>	<b>0.99</b>	<b>0.995</b>
<b>31</b>	14.5	15.7	17.5	19.3	21.4	41.4	45.0	48.2	52.2	55.0
<b>32</b>	15.1	16.4	18.3	20.1	22.3	42.6	46.2	49.5	53.5	56.3
<b>33</b>	15.8	17.1	19.0	20.9	23.1	43.7	47.4	50.7	54.8	57.6
<b>34</b>	16.5	17.8	19.8	21.7	24.0	44.9	48.6	52.0	56.1	59.0
<b>35</b>	17.2	18.5	20.6	22.5	24.8	46.1	49.8	53.2	57.3	60.3
<b>36</b>	17.9	19.2	21.3	23.3	25.6	47.2	51.0	54.4	58.6	61.6
<b>37</b>	18.6	20.0	22.1	24.1	26.5	48.4	52.2	55.7	59.9	62.9
<b>38</b>	19.3	20.7	22.9	24.9	27.3	49.5	53.4	56.9	61.2	64.2
<b>39</b>	20.0	21.4	23.7	25.7	28.2	50.7	54.6	58.1	62.4	65.5
<b>40</b>	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
<b>41</b>	21.4	22.9	25.2	27.3	29.9	52.9	56.9	60.6	64.9	68.1
<b>42</b>	22.1	23.7	26.0	28.1	30.8	54.1	58.1	61.8	66.2	69.3
<b>43</b>	22.9	24.4	26.8	29.0	31.6	55.2	59.3	63.0	67.5	70.6
<b>44</b>	23.6	25.1	27.6	29.8	32.5	56.4	60.5	64.2	68.7	71.9
<b>45</b>	24.3	25.9	28.4	30.6	33.4	57.5	61.7	65.4	70.0	73.2
<b>46</b>	25.0	26.7	29.2	31.4	34.2	58.6	62.8	66.6	71.2	74.4
<b>47</b>	25.8	27.4	30.0	32.3	35.1	59.8	64.0	67.8	72.4	75.7
<b>48</b>	26.5	28.2	30.8	33.1	35.9	60.9	65.2	69.0	73.7	77.0
<b>49</b>	27.2	28.9	31.6	33.9	36.8	62.0	66.3	70.2	74.9	78.2
<b>50</b>	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5
<b>51</b>	28.7	30.5	33.2	35.6	38.6	64.3	68.7	72.6	77.4	80.7
<b>52</b>	29.5	31.2	34.0	36.4	39.4	65.4	69.8	73.8	78.6	82.0
<b>53</b>	30.2	32.0	34.8	37.3	40.3	66.5	71.0	75.0	79.8	83.3
<b>54</b>	31.0	32.8	35.6	38.1	41.2	67.7	72.2	76.2	81.1	84.5
<b>55</b>	31.7	33.6	36.4	39.0	42.1	68.8	73.3	77.4	82.3	85.7
<b>56</b>	32.5	34.3	37.2	39.8	42.9	69.9	74.5	78.6	83.5	87.0
<b>57</b>	33.2	35.1	38.0	40.6	43.8	71.0	75.6	79.8	84.7	88.2
<b>58</b>	34.0	35.9	38.8	41.5	44.7	72.2	76.8	80.9	86.0	89.5
<b>59</b>	34.8	36.7	39.7	42.3	45.6	73.3	77.9	82.1	87.2	90.7
<b>60</b>	35.5	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	92.0
<b>61</b>	36.3	38.3	41.3	44.0	47.3	75.5	80.2	84.5	89.6	93.2
<b>62</b>	37.1	39.1	42.1	44.9	48.2	76.6	81.4	85.7	90.8	94.4
<b>63</b>	37.8	39.9	43.0	45.7	49.1	77.7	82.5	86.8	92.0	95.6
<b>64</b>	38.6	40.6	43.8	46.6	50.0	78.9	83.7	88.0	93.2	96.9
<b>65</b>	39.4	41.4	44.6	47.4	50.9	80.0	84.8	89.2	94.4	98.1
<b>66</b>	40.2	42.2	45.4	48.3	51.8	81.1	86.0	90.3	95.6	99.3
<b>67</b>	40.9	43.0	46.3	49.2	52.7	82.2	87.1	91.5	96.8	100.6
<b>68</b>	41.7	43.8	47.1	50.0	53.5	83.3	88.3	92.7	98.0	101.8
<b>69</b>	42.5	44.6	47.9	50.9	54.4	84.4	89.4	93.9	99.2	103.0
<b>70</b>	43.3	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100.4	104.2
<b>71</b>	44.1	46.2	49.6	52.6	56.2	86.6	91.7	96.2	101.6	105.4
<b>72</b>	44.8	47.1	50.4	53.5	57.1	87.7	92.8	97.4	102.8	106.6
<b>73</b>	45.6	47.9	51.3	54.3	58.0	88.8	93.9	98.5	104.0	107.9
<b>74</b>	46.4	48.7	52.1	55.2	58.9	90.0	95.1	99.7	105.2	109.1
<b>75</b>	47.2	49.5	52.9	56.1	59.8	91.1	96.2	100.8	106.4	110.3
<b>76</b>	48.0	50.3	53.8	56.9	60.7	92.2	97.4	102.0	107.6	111.5
<b>77</b>	48.8	51.1	54.6	57.8	61.6	93.3	98.5	103.2	108.8	112.7
<b>78</b>	49.6	51.9	55.5	58.7	62.5	94.4	99.6	104.3	110.0	113.9
<b>79</b>	50.4	52.7	56.3	59.5	63.4	95.5	100.7	105.5	111.1	115.1
<b>80</b>	51.2	53.5	57.2	60.4	64.3	96.6	101.9	106.6	112.3	116.3

### J.5 Critical Values of the $q$ Statistic for the Student-Newman-Keuls Test

Table J.5 is a tabulation of critical values for the Student-Newman-Keuls test used to separate means in a Model I ANOVA with a significant global  $F$  test. To find the critical values for the SNK  $q$  statistics choose either  $\alpha = 0.05$  or  $\alpha = 0.01$ . Find the number of means ( $k$ ) in rank order to be compared at the top of the table and the degrees of freedom ( $\nu$ ) for  $MS_{\text{Error}}$  as a row value. As with earlier tables if the degrees of freedom are not found, be conservative, and use the closest lower value. For example, at  $\alpha = 0.05$  with 35 degrees of freedom in  $MS_{\text{Error}}$  and five means to separate, the critical values would be

$k$	2	3	4	5
$q$ c.v.	2.888	3.486	3.845	4.102

To compare the first and the last of 5 ordered means use 4.102 as the critical value for the  $q$  statistic. To compare any two adjacent means use 2.888 as the critical value for the  $q$  statistic. For means separated by one other mean use 3.486 as the critical value for the  $q$  statistic. In general,  $k$  is the number of means in the range of interest. For further examples, see Appendix E.1.

### J.6 Critical Values for Duncan's Multiple Range Test

The value of  $r_p$  for Duncan's Multiple Range test is the same as the  $q$  statistic for the Student-Newman-Keuls test. To find  $r_p$  use the degrees of freedom from  $MS_{\text{Error}}$  in the ANOVA to locate the row and the number of ordered means to locate the column. For comparison of four means in an ANOVA with 35 degrees of freedom in  $MS_{\text{Error}}$  use the row marked 30 (next lower value since 35 is not in the table). The  $r_p$  values would be 2.888, 3.035, and 3.131. Multiplied by the appropriate standard error these values generate the  $SSR_p$ 's that are the critical values for the DMRT. For further examples, see Appendix E.2.

$\alpha = 0.05$									
$k$									
$\nu$	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.085	8.331	9.798	10.88	11.74	12.44	13.03	13.54	13.99
3	4.501	5.910	6.825	7.502	8.037	8.478	8.853	9.177	9.462
4	3.927	5.040	5.757	6.287	6.707	7.053	7.347	7.602	7.826
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.802	6.995
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.998	6.158
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597	5.767	5.918
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432	5.595	5.739
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305	5.461	5.599
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.487
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119	5.265	5.395
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049	5.192	5.318
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990	5.131	5.254
15	3.014	3.674	4.076	4.367	4.595	4.782	4.940	5.077	5.198
16	2.998	3.649	4.046	4.333	4.557	4.741	4.897	5.031	5.150
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108
18	2.971	3.609	3.997	4.277	4.495	4.673	4.824	4.956	5.071
19	2.960	3.593	3.977	4.253	4.469	4.645	4.794	4.924	5.038
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.896	5.008
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915
30	2.888	3.486	3.845	4.102	4.302	4.464	4.602	4.720	4.824
40	2.858	3.442	3.791	4.039	4.232	4.389	4.521	4.635	4.735
60	2.829	3.399	3.737	3.977	4.163	4.314	4.441	4.550	4.646
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560
$\infty$	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474

$\alpha = 0.01$									
$k$									
$v$	2	3	4	5	6	7	8	9	10
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
2	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69
3	8.261	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69
4	6.512	8.120	9.173	9.958	10.58	11.10	11.55	11.93	12.27
5	5.702	6.976	7.804	8.421	8.913	9.321	9.669	9.972	10.24
6	5.243	6.331	7.033	7.556	7.973	8.318	8.613	8.869	9.097
7	4.949	5.919	6.543	7.005	7.373	7.679	7.939	8.166	8.368
8	4.746	5.635	6.204	6.625	6.960	7.237	7.474	7.681	7.863
9	4.596	5.428	5.957	6.348	6.658	6.915	7.134	7.325	7.495
10	4.482	5.270	5.769	6.136	6.428	6.669	6.875	7.055	7.213
11	4.392	5.146	5.621	5.970	6.247	6.476	6.672	6.842	6.992
12	4.320	5.046	5.502	5.836	6.101	6.321	6.507	6.670	6.814
13	4.260	4.964	5.404	5.727	5.981	6.192	6.372	6.528	6.667
14	4.210	4.895	5.322	5.634	5.881	6.085	6.258	6.409	6.543
15	4.168	4.836	5.252	5.556	5.796	5.994	6.162	6.309	6.439
16	4.131	4.786	5.192	5.489	5.722	5.915	6.079	6.222	6.349
17	4.099	4.742	5.140	5.430	5.659	5.847	6.007	6.147	6.270
18	4.071	4.703	5.094	5.379	5.603	5.788	5.944	6.081	6.201
19	4.046	4.670	5.054	5.334	5.554	5.735	5.889	6.022	6.141
20	4.024	4.639	5.018	5.294	5.510	5.688	5.839	5.970	6.087
24	3.956	4.546	4.907	5.168	5.374	5.542	5.685	5.809	5.919
30	3.889	4.455	4.799	5.048	5.242	5.401	5.536	5.653	5.756
40	3.825	4.367	4.696	4.931	5.114	5.265	5.392	5.502	5.559
60	3.762	4.282	4.595	4.818	4.991	5.133	5.253	5.356	5.447
120	3.702	4.200	4.497	4.709	4.872	5.005	5.118	5.214	5.299
$\infty$	3.643	4.120	4.403	4.603	4.757	4.882	4.987	5.078	5.157