Introduction to Nonparametric Methods

UCR GradQuant Workshop

4/30/2015

UCR GradSuccess
graduate.ucr.edu/success

What we will do today…

- Give a brief overview of nonparametric statistical methods
- Introduce the following methods and their implements in R…
  - Wilcoxon ranked sum and signed rank (alternatives to t-tests)
  - Wilcoxon Rank-Sum (Mann-Whitney U) Test (alternatives to two- sample t-test)
  - Kruskal-Wallis (alternative to one-way ANOVA)
  - Friedman test (alternative to Within-subjects ANOVA)
  - Spearman correlation (alternative to Pearson correlation)

Overview of Distribution-Free Tests

- In statistics, parametric usually means some specific probability distribution is assumed
  - Normal (regression, ANOVA, etc.)
  - Exponential (survival)
- May also involve assumptions about parameters (variance)

Overview of Distribution-Free Tests

- Most inferential statistics assume normal distributions. Although these statistical tests work well even if the assumption of normality is violated, extreme deviations from normality can distort the results.
- The effect of violating the assumption of normality often causes a substantial decrease in power.
- Alternative means of dealing with skewed distributions such as taking logarithms or square roots of the data are available.

Overview of Distribution-Free Tests

- There is a collection of tests called distribution-free tests that do not make any assumptions about the distribution from which the numbers were sampled; thus the name “distribution-free”.
- The main advantage of distribution-free tests is that they provide more power than traditional tests when the samples are from highly-skewed distributions.

Overview of Distribution-Free Tests

- Distribution-free tests are sometimes referred to as non-parametric tests because, strictly speaking, they do not test hypotheses about population parameters. Nonetheless, they do provide a basis for making inferences about populations and are therefore classified as inferential statistics.

Overview of Distribution-Free Tests

Assumptions nonparametric methods *do* make

- Randomness
- Independence
- In multi-sample tests, distributions are of the same shape

Nonparametric Methods

- Advantages
  - Nonparametric methods require no or very limited assumptions to be made about the format of the data, and they may therefore be preferable when the assumptions required for parametric methods are not valid.
  - Nonparametric methods can be useful for dealing with unexpected, outlying observations that might be problematic with a parametric approach.
  - Nonparametric methods are intuitive and are simple to carry out by hand, for small samples at least.
  - Nonparametric methods are often useful in the analysis of ordered categorical data in which assignation of scores to individual categories may be inappropriate.

Nonparametric Methods

- Disadvantages
  - Loss of power when data *does* follow usual assumptions and larger differences are needed before the null hypothesis can be rejected compared to parametric counterparts.
  - Reduces the data's information
  - Larger sample sizes needed due to less efficiency.
- Use caution in selecting nonparametric method.
  - If the parametric assumptions can be met, the parametric methods are preferred.
  - When parametric assumptions cannot be met, the nonparametric methods are a valuable tool for analyzing the data.

Rank Test

- Rank tests are a simple group of nonparametric tests
- Instead of using the actual numerical value of something, we use its *rank,* or relative position on the number line to the other observations

Rank Test

- As an example
- What about ties?
  - Their ascending values are added together and divided by the total number of ties

Y Value	Ascending Value	Rank
32	1	1
45	2	2
54	3	3
64	4	4
311	5	5
2000	6	6

Y Value	Ascending Value	Rank
32	1	1
64	2	(2+3+4)/3=3
64	3	(2+3+4)/3=3
64	4	(2+3+4)/3=3
311	5	5
2000	6	6

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

- Alternative to the independent sample t-test (recall that the independent sample t-test compares two independent samples, testing if the means are equal)
- The t-test assumes normality of the data, and equal variances (though adjustments can be made for unequal variances)
- The Wilcoxon rank-sum test assumes that the two samples follow continuous distributions, as well as the usual randomness and independence

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

- Assume two independent samples of sizes $n1$ and $n2$.
- The test is very simple and consists of combining the two samples into one sample of size $n1 + n2$, sorting the result, assigning ranks to the sorted values (giving the average rank to any 'tied' observations), and then letting T be the sum of the ranks for the observations in the first sample.
- If the two populations have the same distribution, then the sum of the ranks of the first sample and those in the second sample should be close to the same value.

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

- Remember, for the t-test, we are testing

$$H_{0:}\ \mu_1 = \mu_2$$

vs.

$$H_{a:}\ \mu_1 \neq \mu_2$$

- Where μ_1 is the mean of group 1 and μ_2 is the mean of group 2

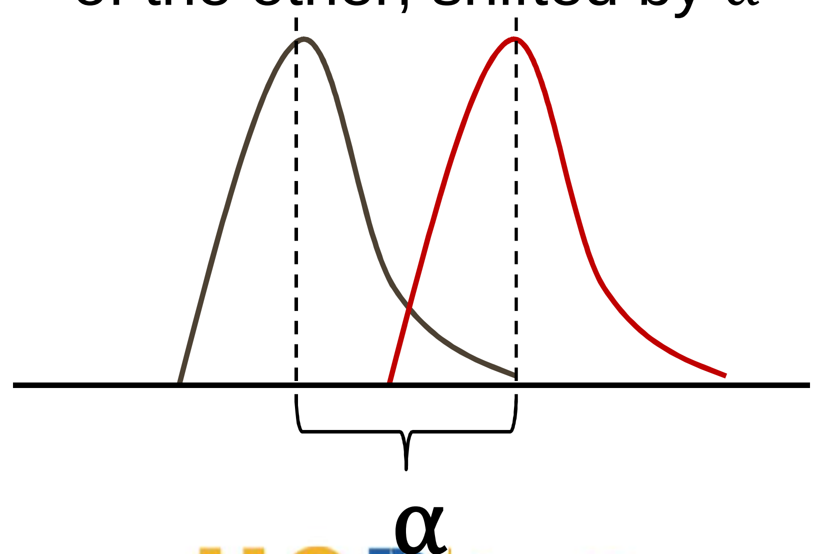
Rank-Sum (Mann-Whitney) Test for Two Independent Samples

- In the Wilcoxon rank-sum test, we are testing

$$H_{0:}\ F(G_1)=F(G_2)$$
vs.
$$H_a\colon F(G_1)=F(G_2-\alpha)$$

- Where $F(G_1)$ is the distribution of group 1 and $F(G_2)$ is the distribution of group 2, and α is the "location shift"

- What this really means is that $F(G_1)$ and $F(G_2)$ are basically the same, except one is to the right of the other, shifted by α



UCR GradSuccess
graduate.ucr.edu/success

Wilcoxon Rank-Sum (Mann-Whitney) Test for Two Independent Samples

Example

To compare the running speed of the first grade boys and girls, the following information is collected.

- **Rank** 1 2 3 4 5 6 7 8 9 10 11
- **Sex** G B G G G G B B B G B

Is there a difference between the running speed of boys and girls?

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

Solution:

Analyze the data according to sex:

Boys:

5 observations: 2, 7, 8, 9, 11

Rank Sum: 37

Girls:

6 observations: 1, 3, 4, 5, 6, 10

Rank Sum: 29

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

To test:

H0: There is no difference in running speed of boys and girls

vs *H1*: There is difference in running speed of boys and girls (2-tailed test)

at $\alpha = 5\%$ level

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

R = sum of ranks in boy group

= 37

From Wilcoxon Rank Sum table with $n1 = 5$ and $n2 = 6$, we see that the critical value is (18, 42).

So, we do not reject *H0* and conclude that there is no difference in the speed of boys and girls.

Rank-Sum (Mann-Whitney) Test for Two Independent Samples

Normal approximation for larger samples (just for illustration)

$\mu = ½ \times 5 \times (5+6+1) = 30$

$\sigma^2 = [\, 5 \times 6 \times (5+6+1) \,] / 12 = 30$

$z_{\mathrm{cal}} = (37 - 30) / \sqrt{30} = 1.278 < 1.96$

So, at 5% level, we do not reject *H0* and conclude that there is no difference in the speed of boys and girls.

Note:

Approximation should be applied only when both sample sizes are more than 10.

The Sign Test for Paired Replicates

Moving to a paired situation:

- Suppose we have two samples, but they are paired in some way
- Suppose pairs of people are couples, or plants are paired off into different plots, dogs are paired by breeds, or whatever

The Sign Test for Paired Replicates

- Given *n* pairs of data, the sign test tests the hypothesis that the *median of the differences* in the pairs is zero.
- The test statistic is the number of positive differences. If the null hypothesis is true, then the numbers of positive and negative differences should be approximately the same.
- Let *X* is the number of positive differences.

  Under *H0*, X ~ Bin(*n*,½).

The Sign Test for Paired Replicates

Example

In a study, the average number of seeds in two pods was recorded at both the top and the bottom of 10 plants. The objective of study was to determine whether the position on the plant affected the number of seeds in the pods.

Location	1	2	3	4	5	6	7	8	9	10
Top	4.0	5.2	5.7	4.2	4.8	3.9	4.1	3.0	4.6	6.8
Bottom	4.4	3.7	4.7	2.8	4.2	4.3	3.5	3.7	3.1	1.9

The Sign Test for Paired Replicates

Solution:

Location	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Top	4.0	5.2	5.7	4.2	4.8	3.9	4.1	3.0	4.6	6.8
Bottom	4.4	3.7	4.7	2.8	4.2	4.3	3.5	3.7	3.1	1.9
diff	-	+	+	+	+	-	+	-	+	+

Number of “+”, $X = 7$.

The Sign Test for Paired Replicates

To test:

H0: the position on the plant does NOT affect the number of seeds in the pods

vs *H1*: the position on the plant does affect the number of seeds in the pods

at α = 5% level.

The Sign Test for Paired Replicates

Under *H0*,

X ~ Bin(10,½) ≈ N(5, 2.5)

$$Z_{cal} = \frac{7-5}{\sqrt{2.5}} = 1.2649 < 1.645$$

Therefore, we do not reject *H0* and conclude that there is insufficient evidence to suggest that the position on the plant affect the number of seeds in the pods.

The Wilcoxon Signed Rank Test for Paired Replicates

- This test is similar to the sign test in that it tests for the median difference in paired data to be zero.
- The test consists of sorting the absolute values of the differences from smallest to largest, assigning ranks to the absolute values (rank 1 to the smallest, rank 2 to the next smallest, and so on) and then finding the sum of the ranks of the positive differences.
- If the null hypothesis is true, the sum of the ranks of the positive differences should be about the same as the sum of the ranks of the negative differences.

The Wilcoxon Signed Rank Test for Paired Replicates

Example

Consider the data in the previous example.

Location	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Top	4.0	5.2	5.7	4.2	4.8	3.9	4.1	3.0	4.6	6.8
Bottom	4.4	3.7	4.7	2.8	4.2	4.3	3.5	3.7	3.1	1.9
Difference, d_i	-.4	1.5	1.0	1.4	0.6	-.4	0.6	-.7	1.5	4.9
Rank of $\|d_i\|$	1.5	8.5	6	7	3.5	1.5	3.5	5	8.5	10

The sum of the positive ranks, $R+ = 47$, while

The sum of the negative ranks, $R– = 8$.

So, $T = \min(R+ , R–) = 8$.

The Wilcoxon Signed Rank Test for Paired Replicates

To test:

H0: the position on the plant does NOT affect the number of seeds in the pods

vs *H1*: the position on the plant does affect the number of seeds in the pods

at $\alpha = 5\%$ level.

The Wilcoxon Signed Rank Test for Paired Replicates

From the Wilcoxon Signed-Rank table:

The critical value for $N = 10$ at $\alpha = 5\%$ is (8,47).

Since $T = 8$, we don't reject *H0* and conclude that there is no sufficient evidence to suggest that the position on the plant affect the number of seeds in the pods.

Sign Test vs. Wilcoxon Signed-Ranks Test

The Sign Test

This is used when the direction of the difference between matched pairs of data can be determined, but the magnitude cannot. The test computes the difference between the two variables for all cases and classifies the differences as either positive, negative, or tied.

Sign Test vs. Wilcoxon Signed-Ranks Test

The Wilcoxon Signed-Ranks Test

This is used when both the direction of the difference between matched pairs of data and the magnitude can be determined. This test considers information about both the sign of the differences and the magnitude of the differences between pairs.

Conclusion:

If both the Sign test and the Wilcoxon test can be performed, the Wilcoxon test is the better choice, as it incorporates more information about the data.

Kruskal-Wallis Test

- Suppose we are comparing more than two groups
- Normally we would use an analysis of variance, or ANOVA. But, of course, we need to assume normality for this as well
- The Kruskal-Wallis test is the non-parametric version of one way ANOVA and is a straightforward generalization of the Wilcoxon rank sum test for two independent samples.

Kruskal-Wallis Test

- Suppose that there are k independent samples of sizes $n1$, $n2$, … … , nk.
- Combine all the samples into one large sample of size n (= $n1 + n2 + \dots \dots + nk$), and sort the result from smallest to largest.
- Assign ranks 1, 2, … n, to the observations. For ties, assign the average rank in a group of tied observations.
- Then, find Ri, the sum of the ranks of the observations in the ith sample.

Kruskal-Wallis Test

The test statistic is

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \sim \chi^2(k-1)$$

We reject the null hypothesis that all *k* distributions are the same if $H_{\text{cal}} > \chi^2(k-1)$.

Kruskal-Wallis Test

Example

A study was conducted to compare the average starting salaries for 4 different majors: **F**inance, **A**ccounting, **M**arketing and **B**usiness **A**dministration.

The data obtained are:

Major	**Salary Rank** (1 = lowest, 22 = highest)
F	5 15 16 19 22
A	2 7 11 17 18 21
M	3 8 9 10 13 20
BA	1 4 6 12 14

Does the major affect starting salary?

Kruskal-Wallis Test

Solution:

To test:

H0: The average starting salaries are the same for different majors

vs *H1*: The average starting salaries are different for different majors

at $\alpha = 5\%$ level.

Under *H0*,

$$H \sim \chi^2_{4-1}$$

Kruskal-Wallis Test

From the data:

Major	Salary Rank	n	R
F	5 15 16 19 22	5	77
A	2 7 11 17 18 21	6	76
M	3 8 9 10 13 20	6	63
BA	1 4 6 12 14	5	37

$$H_{cal} = \frac{12}{22(22+1)}\left[\frac{77^2}{5} + \frac{76^2}{6} + \frac{63^2}{6} + \frac{37^2}{5}\right] - 3(22+1)$$

$= 12/506 \times 3083.767 - 3(23) = 4.132806$

Kruskal-Wallis Test

$$\chi^2_{3,0.05} = 7.81$$

Since $H_{cal} < 7.81$, we do not reject *H0* and conclude that there is insufficient evidence to suggest that the average starting salaries are different for different majors.

Friedman Test

Nonparametric equivalent of the repeated measures ANOVA

Example: Effects on worker mood of different types of music

Five workers. Each is tested three times, once under each of the following conditions:
condition 1: silence.
condition 2: "easy-listening” music.
condition 3: marching-band music.

DV: mood rating ("0" = unhappy, "10" = euphoric).
Ratings - so use a nonparametric test.

Friedman Test

	Silence (raw score)	Silence (ranked score)	Easy (raw score)	Easy (ranked score)	Band (raw score)	Band (ranked score)
Wkr 1:	4	1	5	2	6	3
Wkr 2:	2	1	7	2.5	7	2.5
Wkr 3:	6	1.5	6	1.5	8	3
Wkr 4:	3	1	7	3	5	2
Wkr 5:	3	1	8	2	9	3
	M = 3.60 SD = 1.52		M = 6.60 SD = 1.14		M = 7.00 SD = 1.58	

Step 1:
Rank *each subject's scores* individually.
Worker 1's scores are 4, 5, 6: these get ranks of 1, 2, 3.
Worker 4's scores are 3, 7, 5: these get ranks of 1, 3, 2 .

Friedman Test

	Silence (raw score)	Silence (ranked score)	Easy (raw score)	Easy (ranked score)	Band (raw score)	Band (ranked score)
Wkr 1:	4	1	5	2	6	3
Wkr 2:	2	1	7	2.5	7	2.5
Wkr 3:	6	1.5	6	1.5	8	3
Wkr 4:	3	1	7	3	5	2
Wkr 5:	3	1	8	2	9	3
rank total:		5.5		11		13.5

Step 2:

Find the rank total for each condition, using the ranks from all subjects within that condition.

Rank total for "Silence" condition: 1+1+1.5+1+1 = 5.5.

Rank total for "Easy Listening" condition = 11.

Rank total for "Marching Band" condition = 13.5.

Friedman Test

Step 3: Work out χ_r^2

$$\chi_r^2 = \left[\left(\frac{12}{N C (C+1)} \right) \times \sum Tc^2 \right] - 3 \times N \times (C+1)$$

C is the number of conditions.
N is the number of subjects.
ΣTc^2 is the sum of the squared rank totals for each condition.

Friedman Test

$$\chi_r^2 = \left[\frac{12}{N\,C\,(C+1)} \right] \cdot \sum Tc^2 - 3\,N\,(C+1)$$

To get ΣTc^2 :

(a) square each rank total:
$5.5^2 = 30.25$. $11^2 = 121$. $13.5^2 = 182.25$.

(b) Add together these squared totals.
$30.25 + 121 + 182.25 = 333.5$.

Friedman Test

In our example,

$$\chi_r^2 = \left[\left(\frac{12}{5 \times 3 \times 4}\right) \times 333.5\right] - 3 \times 5 \times 4$$

$\chi r^2 = 7.44$

Step 4:

Degrees of freedom = number of conditions minus one.

df = 3 - 1 = 2.

Friedman Test

Step 5:
Assessing the statistical significance of χr^2

Our obtained ¡ r^2 is 7.44.

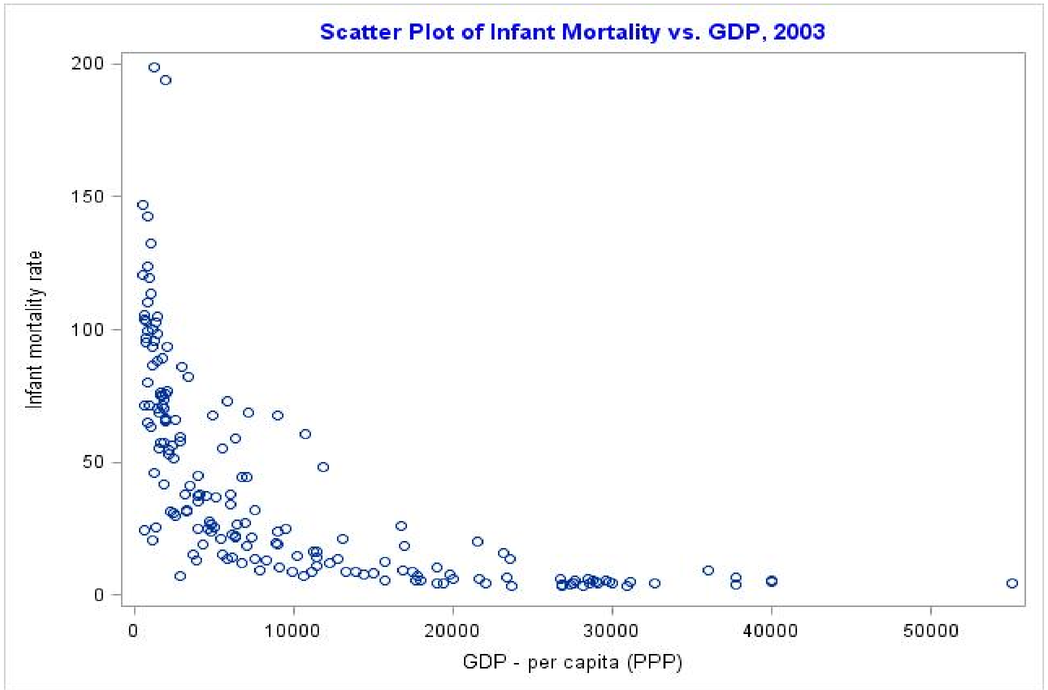
For 2 d.f., a ¡ ^{2} value of 5.99 would occur by chance with a probability of .05.
Our obtained value is *bigger* than 5.99.
Therefore our obtained ¡ r^2 is even *less* likely to occur by chance: $p < .05$.

Conclusion: the conditions are significantly different. Music *does* affect worker mood.

UCR GradSuccess
graduate.ucr.edu/success

Spearman Correlation

- Suppose you want to discover the association between infant mortality and GDP by country
- Here's a 2003 scatterplot of the situation



graduate.ucr.edu/success

Spearman Correlation

Pearson correlation

- This data comes from www.indexmundi.com
- In this example, the Pearson correlation is about -.63
- Still significant, but perhaps underestimates the monotone nature of the relationship between GDP and infant mortality rate
- In addition, the Pearson correlation assumes linearity, which is clearly not present here

Spearman Correlation

- This is a correlation coefficient based on ranks, which are computed in the y variable and x variable independently, with sample size n
- To calculate the coefficient, we do the following…
  - Take each x_i and each y_i, convert them into ranks (ranks of x and y are independent of each other
  - Subtract r_{xi} from r_{yi} to get d_i, the difference in ranks
  - The formula is

$$R_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Spearman Correlation

- In this case, the hypotheses are

$$H_{0:}\ Rs=0$$

vs.

$$H_{a:}\ Rs\neq 0$$

- Basically, we are attempting to see whether or not the two variables are independent, or if there is evidence of an association

Spearman Correlation

Let's return to the GDP data

- We will now plot the data in R, and see how to get the Spearman correlation
- R tests this by using an exact p-value for small sample sizes, and an approximate t-distribution for larger ones
- The test statistic in that case would follow a t-distribution with n-2 degrees of freedom

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

Thank you!

Appendix : Wilcoxon Rank-Sum Table

Probabilities relate to the distribution of W_A, the rank sum for group A when $H_0 : A = B$ is true. The tabulated value for the **lower tail** is the largest value of w_A for which $\text{pr}(W_A \leq w_A) \leq \textit{prob}$. The tabulated value for the **upper tail** is the smallest value of w_A for which $\text{pr}(W_A \geq w_A) \leq \textit{prob}$.

		Lower Tail						Upper Tail					
		prob						***prob***					
n_A	n_B	.005	.01	.025	.05	.10	.20	.20	.10	.05	.025	.01	.005
4	4			10	11	13	14	22	23	25	26		
	5		10	11	12	14	15	25	26	28	29	30	
	6	10	11	12	13	15	17	27	29	31	32	33	34
	7	10	11	13	14	16	18	30	32	34	35	37	38
	8	11	12	14	15	17	20	32	35	37	38	40	41
	9	11	13	14	16	19	21	35	37	40	42	43	45
	10	12	13	15	17	20	23	37	40	43	45	47	48
	11	12	14	16	18	21	24	40	43	46	48	50	52
	12	13	15	17	19	22	26	42	46	49	51	53	55
5	5	15	16	17	19	20	22	33	35	36	38	39	40
	6	16	17	18	20	22	24	36	38	40	42	43	44
	7	16	18	20	21	23	26	39	42	44	45	47	49
	8	17	19	21	23	25	28	42	45	47	49	51	53
	9	18	20	22	24	27	30	45	48	51	53	55	57
	10	19	21	23	26	28	32	48	52	54	57	59	61
	11	20	22	24	27	30	34	51	55	58	61	63	65
	12	21	23	26	28	32	36	54	58	62	64	67	69
6	6	23	24	26	28	30	33	45	48	50	52	54	55
	7	24	25	27	29	32	35	49	52	55	57	59	60
	8	25	27	29	31	34	37	53	56	59	61	63	65
	9	26	28	31	33	36	40	56	60	63	65	68	70
	10	27	29	32	35	38	42	60	64	67	70	73	75
	11	28	30	34	37	40	44	64	68	71	74	78	80
	12	30	32	35	38	42	47	67	72	76	79	82	84
7	7	32	34	36	39	41	45	60	64	66	69	71	73
	8	34	35	38	41	44	48	64	68	71	74	77	78
	9	35	37	40	43	46	50	69	73	76	79	82	84
	10	37	39	42	45	49	53	73	77	81	84	87	89
	11	38	40	44	47	51	56	77	82	86	89	93	95
	12	40	42	46	49	54	59	81	86	91	94	98	100
8	8	43	45	49	51	55	59	77	81	85	87	91	93
	9	45	47	51	54	58	62	82	86	90	93	97	99
	10	47	49	53	56	60	65	87	92	96	99	103	105
	11	49	51	55	59	63	69	91	97	101	105	109	111
	12	51	53	58	62	66	72	96	102	106	110	115	117
9	9	56	59	62	66	70	75	96	101	105	109	112	115
	10	58	61	65	69	73	78	102	107	111	115	119	122
	11	61	63	68	72	76	82	107	113	117	121	126	128
	12	63	66	71	75	80	86	112	118	123	127	132	135
10	10	71	74	78	82	87	93	117	123	128	132	136	139
	11	73	77	81	86	91	97	123	129	134	139	143	147
	12	76	79	84	89	94	101	129	136	141	146	151	154
11	11	87	91	96	100	106	112	141	147	153	157	162	166
	12	90	94	99	104	110	117	147	154	160	165	170	174
12	12	105	109	115	120	127	134	166	173	180	185	191	195

TABLE 12.19

Lower and Upper Critical Values, W, of Wilcoxon Signed Ranks Test

ONE-TAIL	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
TWO-TAIL	$\alpha = .10$	$\alpha = .05$	$\alpha = .02$	$\alpha = .01$
n	*(Lower, Upper)*			
5	0,15	—,—	—,—	—,—
6	2,19	0,21	—,—	—,—
7	3,25	2,26	0,28	—,—
8	5,31	3,33	1,35	0,36
9	8,37	5,40	3,42	1,44
10	10,45	8,47	5,50	3,52
11	13,53	10,56	7,59	5,61
12	17,61	13,65	10,68	7,71
13	21,70	17,74	12,79	10,81
14	25,80	21,84	16,89	13,92
15	30,90	25,95	19,101	16,104
16	35,101	29,107	23,113	19,117
17	41,112	34,119	27,126	23,130
18	47,124	40,131	32,139	27,144
19	53,137	46,144	37,153	32,158
20	60,150	52,158	43,167	37,173

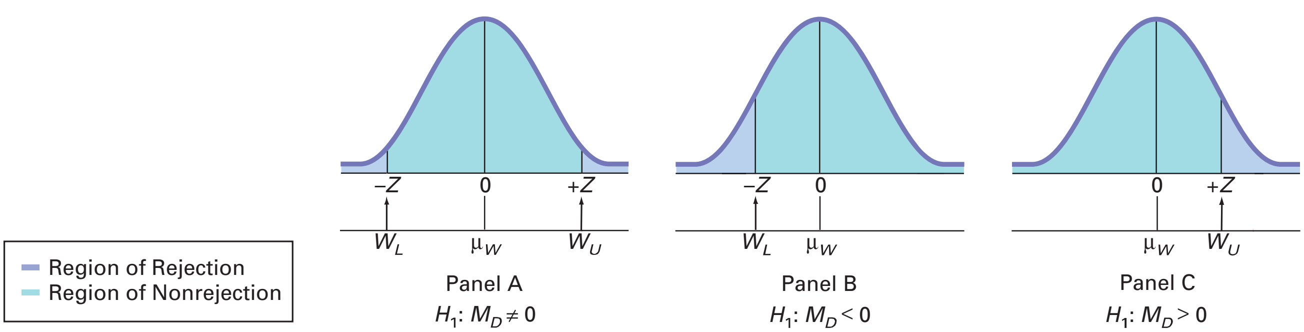
Source: Adapted from Table 2 of F. Wilcoxon and R. A. Wilcox, *Some Rapid Approximate Statistical Procedures* (Pearl River, NY: Lederle Laboratories, 1964), with permission of the American Cyanamid Company.

You use Table 12.19 to find the critical values of the test statistic W for both one- and two-tail tests for samples of $n' \leq 20$.

For a two-tail test, you reject the null hypothesis (Panel A of Figure 12.20) if the computed W test statistic equals or is greater than the upper critical value or is equal to or less than the lower critical value. For a one-tail test in the lower tail, you reject the null hypothesis if the computed W test statistic is less than or equal to the lower critical value (Panel B of Figure 12.20). For a one-tail test in the upper tail, the decision rule is to reject the null hypothesis if the computed W test statistic equals or is greater than the upper critical value (Panel C of Figure 12.20).

FIGURE 12.20

Regions of Rejection and Nonrejection Using the Wilcoxon Signed Ranks Test



For samples of $n' > 20$, the test statistic W is approximately normally distributed with mean μ_W and standard deviation σ_W. The mean of the test statistic W is

$$\mu_W = \frac{n'(n' + 1)}{4}$$

and the standard deviation of the test statistic W is

$$\sigma_W = \sqrt{\frac{n'(n' + 1)(2n' + 1)}{24}}$$