APPENDIX D

Non-parametric significance tests

This appendix contains additional non-parametric tests of hypotheses to augment those described in Chapter 16 (see Exhibit 16.7).

One-sample case

Kolmogorov–Smirnov test

This test is appropriate when the data are at least ordinal and the research situation calls for a comparison of an observed sample distribution with a theoretical distribution. Under these conditions, the Kolmogorov–Smirnov (KS) one-sample test is more powerful than the X^2 test and can be used for small samples when the X^2 test cannot. The KS is a test of goodness of fit in which we specify the *cumulative* frequency distribution that would occur under the theoretical distribution and compare it with the observed cumulative frequency distribution. The theoretical distribution represents our expectations under H_0. We determine the point of greatest divergence between the observed and theoretical distributions and identify this value as D (maximum deviation). From a table of critical values for D, we determine whether such a large divergence is likely on the basis of random sampling variations from the theoretical distribution. The value for D is calculated as follows:

$$D = \text{maximum} /F_O(X) - F_T(X)/$$

in which

$F_O(X)$ = The observed cumulative frequency distribution of a random sample of n observations. Where X is any possible score, $F_O(X) = k/n$, where k = the number of observations equal to or less than X.

$F_T(X)$ = The theoretical frequency distribution under H_0.

We illustrate the KS test with an analysis of the results of the dining club study, in terms of various class levels. Take an equal number of interviews from each class, but secure unequal numbers of people interested in joining. Assume class levels are ordinal measurements. The testing process is as follows (see also Exhibit D.1).

Exhibit D.1 ***Testing process of Kolmogorov–Smirnov test.***

	First-year student	Second-year student	Junior	Senior	Graduate
Number in each class	5	9	11	16	19
$F_0(X)$	5/60	14/60	25/60	41/60	60/60
$F_T(X)$	12/60	24/60	36/60	48/60	60/60
$\lvert F_0(X) - F_T(X) \rvert$	7/60	10/60	11/60	7/60	0
$D = 11/60 = .183$					
$n = 60$					

1 NULL HYPOTHESIS

H_0: There is no difference among student classes as to their intention of joining the dining club.
H_A: There is a difference among students in various classes as to their intention of joining the dining club.

2 STATISTICAL TEST

Choose the KS one-sample test because the data are ordinal measures and we are interested in comparing an observed distribution with a theoretical one.

3 SIGNIFICANCE LEVEL

$\alpha = .05,\ n = 60.$

4 CALCULATED VALUE

$D = \text{Maximum}\ |F_O(X) - F_T(X)|.$

5 CRITICAL TEST VALUE

We enter the table of critical values of D in the KS one-sample test (see Appendix E, Exhibit E.5) and learn that with $\alpha = .05$ the critical value for D is

$$D = \frac{1.36}{60} = .175$$

6 INTERPRET

The calculated value is greater than the critical value, indicating that we should reject the null hypothesis.

Two-samples case

Sign test

The sign test is used with matched pairs when the only information is the identification of the pair member that is larger or smaller, or has more or less of some characteristic. Under H_0, one would expect the number of cases in which $X_A > X_B$ to equal the number of pairs in which $X_B > X_A$. All ties are dropped from the analysis, and n is adjusted to allow for these eliminated pairs. This test is based on the binomial expansion and has a good power efficiency for small samples.

Wilcoxon matched-pairs test

When you can determine both *direction* and *magnitude* of difference between carefully matched pairs, use the Wilcoxon matched-pairs test. This test has excellent efficiency and can be more powerful than the t-test in cases where the latter is not particularly appropriate. The mechanics of calculation are also quite simple. Find the difference score (d_i) between each pair of values, and rank-order the differences from smallest to largest without regard to sign. The actual signs of each difference are then added to the rank values and the test statistic T is calculated. T is the sum of the ranks with the less frequent sign. Typical of such research situations might be a study where husband and wife are matched, where twins are used, where a given subject is used in a before/after study, or where the outputs of two similar machines are compared.

Two types of tie may occur with this test. When two observations are equal, the d score becomes zero, and we drop this pair of observations from the calculation. When two or more pairs have the same d value, we average their rank positions. For example, if two pairs have a rank score of 1, we assign the rank of 1.5 to each and rank the next largest difference as third. When $n = 25$, use the table of critical values (see Appendix E, Exhibit E.4). When $n > 25$, the sampling distribution of T is approximately normal with

$$\text{Mean} = \mu_T = \frac{n(n+1)}{4}$$

Standard deviation $= \sigma_T \sqrt{\frac{n(n+1)}{(2n+1)}}$

The formula for the test is: $z = \frac{T - \mu_T}{\sigma_T}$

Suppose you conduct an experiment on the effect of brand name on quality perception. Ten subjects are recruited and asked to taste and compare two samples of a product, one identified as a well-known drink and the other as a new product being tested. In truth, however, the samples are identical. The subjects are then asked to rate the two samples on a set of scale items judged to be ordinal. Test these results for significance by the usual procedure.

1 NULL HYPOTHESIS

H_0: There is no difference between the perceived qualities of the two samples.
H_A: There is a difference in the perceived quality of the two samples.

2 STATISTICAL TEST

The Wilcoxo]n matched-pairs test is used because the study is of related samples in which the differences can be ranked in magnitude.

3 SIGNIFICANCE LEVEL

$\alpha = .05$, with $n = 10$ pairs of comparisons minus any pairs with a d of zero.

4 CALCULATED VALUE

T equals the sum of the ranks with the less frequent sign. Assume we secure the following results:

Exhibit D.2 ***Results of Wilcoxon-matched pairs test.***

Pair	Branded	Unbranded	d_i	Rank of d_i	Rank with less frequent sign
1	52	48	4	4	
2	37	32	5	5.5*	
3	50	52	–1	–2	2
4	45	32	13	9	
5	56	59	–3	–3	3
6	51	50	1	1	
7	40	29	11	8	
8	59	54	5	5.5*	
9	38	38	0	*	
10	40	32	8	7	$T = 5$

Note: *There are two types of tie situation. We drop out the pair with the type of tie shown by pair 9. Pairs 2 and 8 have a tie in rank of difference. In this case, we average the ranks and assign the average value to each.

5 CRITICAL TEST VALUE

Enter the table of critical values of T with $n = 9$ (see Appendix E, Exhibit E.4) and find that the critical value with $\alpha = .05$ is 6. Note that with this test, the calculated value must be smaller than the critical value to reject the null hypothesis.

6 INTERPRET

Since the calculated value is less than the critical value, reject the null hypothesis.

Kolmogorov–Smirnov (KS) two-samples test

When a researcher has two independent samples of ordinal data, the (KS) two-samples test is useful. Like the one-sample test, this two-samples test is concerned with the agreement between two cumulative distributions, but both represent sample values. If the two samples have been drawn from the same population, the cumulative distributions of the samples should be fairly close to each other, showing only random deviations from the population distribution. If the cumulative distributions show a large enough maximum deviation D (defined by the formula below), it is evidence for rejecting the H_0. To secure the maximum deviation, one should use as many intervals as are available so as not to obscure the maximum cumulative difference.

$$D = \text{Maximum } |FN_1(X) - FN_2(X)| \text{ (two-tailed test)}$$

$$D = \text{Maximum } |FN_1(X) - FN_2(X)| \text{ (one-tailed test)}$$

D is calculated in the same manner as before, but the table for critical values for the numerator of D, K_D (two-samples case) is presented in Appendix E, Exhibit E.6 when $n_1 = n_2$ and is less than 40 observations. When n_1 and/or n_2 are larger than 40, D from Appendix E, Exhibit E.7 should be used. With this larger sample, it is not necessary that $n_1 = n_2$.

Here we use a different sample from the smoking-accident study. (To make $n_1 = n_2$, we increased the sample size of no accidents to 34. Non-smokers with no accidents is 24.) Suppose the smoking classifications represent an ordinal scale, and you test these data with the KS two-samples test. Proceed as follows.

1 NULL HYPOTHESIS

H_0: There is no difference in on-the-job accident occurrences between smokers and nonsmokers.
H_A: The more a person smokes, the more likely that person is to have an on-the-job accident.

2 STATISTICAL TEST

The KS two-samples test is used because it is assumed the data are ordinal.

3 SIGNIFICANCE LEVEL

$\alpha = .05.\ n_1 = n_2 = 34.$

4 CALCULATED VALUE

See the one-sample calculation (KS test) and compare with Exhibit D.3 below.

Exhibit D.3 ***Results of Kolmogorov–Smirnov two-samples test.***

	Heavy smoker	Moderate smoker	Non-smoker
$F_{n_1}(X)$	12/34	21/34	34/34
$F_{n_2}(X)$	4/34	10/34	34/34
$D_i = K_{D/n}$	8/34	11/34	0

5 CRITICAL TEST VALUE

We enter Appendix E, Exhibit E.6 with $n = 34$ to find that $K_D = 11$ when $p = \leq .05$ for a one-tailed distribution.

6 INTERPRET

Since the critical value equals the largest calculated value, we reject the null hypothesis.

Mann–Whitney U test

This test is also used with two independent samples if the data are at least ordinal; it is an alternative to the t-test without the latter's limiting assumptions. When the larger of the two samples is 20 or less, there are special tables for interpreting U; when the larger sample exceeds 20, a normal curve approximation is used.

In calculating the U test, treat all observations in a combined fashion and rank them, algebraically, from smallest to largest. The largest negative score receives the lowest rank. In case of ties, assign the average rank as in other tests. With this test, you can also test samples that are unequal. After the ranking, the rank values for each sample are totalled. Compute the U statistic as follows:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \text{ or } U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

in which

n_1 = Number in sample 1
n_2 = Number in sample 2
R_1 = Sum of ranks in sample 1

With this equation, you can secure two U values, one using R_1 and the second using R_2. For testing purposes, use the smaller U.

Exhibit D.4 ***Process of Mann–Whitney U test.***

Sales per week per salesperson			
Training method A	Rank	Training method B	Rank
1500	15	1340	10
1540	16	1300	8.5
1860	22	1620	18
1230	6	1070	3
1370	12	1210	5
1550	17	1170	4
1840	21	1770	20
1250	7	950	1
1300	8.5	1380	13
1350	11	1460	14
1710	19	1030	2
	$R_1 = 154.5$		$R_2 = 98.5$

$$U = (11)(11) + \frac{11(11+1)}{2} - 154.5 = 32.5 \qquad U = (11)(11) + \frac{11(11+1)}{2} - 98.5 = 88.5$$

An example may help to clarify the U statistic calculation procedure. Let's consider the sales training example with the t distribution discussion. Recall that salespeople with training method A averaged higher sales than salespeople with training method B. While these data are ratio measures, one still might not want to accept the other assumptions that underlie the t-test. What kind of a result could be secured with the U test? While the U test is designed for ordinal data, it can be used with interval and ratio measurements.

1 NULL HYPOTHESIS

H_0: There is no difference in sales results produced by the two training methods.
H_A: Training method A produces sales results superior to the results of method B.

2 STATISTICAL TEST

The Mann–Whitney U test is chosen because the measurement is at least ordinal, and the assumptions under the parametric t-test are rejected.

3 SIGNIFICANCE LEVEL

We calculate Mann–Whitney U values as shown in Exhibit D.4.

$\alpha = .05$ (one-tailed test).

5 CRITICAL TEST VALUE

Enter Appendix E, Exhibit E.8 with $n_1 = n_2 = 11$, and find a critical value of 34 for $\alpha = 0.5$, one-tailed test. Note that with this test, the calculated value must be smaller than the critical value to reject the null hypothesis.

6 INTERPRET

Since the calculated value is smaller than the critical value (34 > 32.5), reject the null hypothesis and conclude that training method A is probably superior.

Thus, one would reject the null hypothesis at $\alpha = .05$ in a one-tailed test using either the *t*-test or the *U* test. In this example, the *U* test has approximately the same power as the parametric test.

When $n > 20$ in one of the samples, the sampling distribution of *U* approaches the normal distribution with

$$\text{Mean} = \mu_U = \frac{n_1 n_2}{2}$$

$$\text{Standard deviation} = \sigma_U = \sqrt{\frac{(n_1)(n_2)(n_1 + n_2 + 1)}{12}}$$

$$\text{And } z = \frac{U - \mu_U}{\sigma_U}$$

Other non-parametric tests

Other tests are appropriate under certain conditions when testing two independent samples. When the measurement is only nominal, the Fisher exact probability test may be used. When the data are at least ordinal, use the median and Wald–Wolfowitz runs tests.

k-Samples case

You can use tests more powerful than X_2 with data that are at least ordinal in nature. One such test is an extension of the median test mentioned earlier. We illustrate here the application of a second ordinal measurement test known as the Kruskal–Wallis one-way analysis of variance.

Kruskal–Wallis test

This is a generalized version of the Mann–Whitney test. With it we rank all scores in the entire pool of observations from smallest to largest. The rank sum of each sample is then calculated, with ties being distributed as in other examples. We then compute the value of *H* as follows:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{T_i^2}{n_j} - 3(N+1)$$

where

T_j = Sum of ranks in column *j*
n_j = Number of cases in *j*th sample
$N = \alpha w_j$ = Total number of cases
k = Number of samples

When there are a number of ties, it is recommended that a correct factor (*C*) be calculated and used to correct the *H* value as follows:

$$C = 1 - \left\{ \sum_{j}^{G} (t_i^3 - t_j) \right\}$$

where

G = Number of sets of tied observations
t_i = Number tied in any set i
$H' = H/C$

To secure the critical value for H', use the table for the distribution of X_2 (see Appendix E, Exhibit E.3), and enter it with the value of H' and $d.f. = k - 1$.

To illustrate the application of this test, use the price discount experiment problem. The data and calculations are shown in Exhibit D.5 and indicate that, by the Kruskal–Wallis test, one again barely fails to reject the null hypothesis with $\alpha = .05$.

Exhibit D.5 ***Kruskal–Wallis one-way analysis of variance (price differentials).***

One eurocent		Three eurocents		Five eurocents	
X_A	Rank	X_B	Rank	X_C	Rank
6	1	8	5	9	8.5
7	2.5	9	8.5	9	8.5
8	5	8	5	11	14
7	2.5	10	11.5	10	11.5
9	8.5	11	14	14	18
11	14	13	16.5	13	16.5
	$T_j = 33.5$		60.5		77.0

$$T = 33.5 + 60.5 + 77 = 171$$

$$H = \frac{12}{18(18+1)}\left\{\frac{33.5^2 + 60.5^2 + 77^2}{6}\right\} - 3(18+1)$$

$$= \frac{12}{342}\left\{\frac{1122.25 + 3660.25 + 5929}{6}\right\} - 57$$

$$= 0.0351\left\{\frac{10711.5}{6}\right\} - 57$$

$$H = 5.66$$

$$C = 1 - \left\{\frac{[3(2)^3 - 2] + [2(3)^3 - 3] + [4(4)^3 - 4]}{18^3 - 18}\right\}$$

$$= 1 - \frac{18 + 48 + 60}{5814}$$

$$= 0.978$$

$$H' = \frac{H}{C} = \frac{5.66}{0.978} = 5.79$$

$$d.f. = k - 1 = 2$$

$$p = .05$$