Appendix B

Some Notes on the Statistical Comparison of Two Samples

The following presents methods for testing the spatial differences between two distributions. At this point, *CrimeStat* does not include routines for testing the differences between two or more samples. The following is provided for the reader's information. Chapter 4 discussed the calculation of these statistics as a single distribution.

Differences in the Mean Center of Two Samples

For differences between two samples in the mean center, it is necessary to test both differences in the X coordinate and differences in the Y coordinates. Since *CrimeStat* outputs both the mean X, mean Y, standard deviation of X, and standard deviation of Y, a simple t-test can be set up. The null hypothesis is that the mean centers are equal

$$H_0: \quad \mu_{XA} = \mu_{XB}$$
$$\mu_{YA} = \mu_{YB}$$

and the alternative hypothesis is that the mean centers are not equal

$$H_1: \quad \mu_{XA} \neq \mu_{XB}$$
$$\mu_{YA} \neq \mu_{YB}$$

Because the true standard deviations of sample A, σ_{XA} and σ_{YA}, and sample B, σ_{XB} and σ_{YB}, are not known, the sample standard deviations are taken, S_{XA}, S_{YA}, S_{XB} and S_{YB}. However, since there are two different variables being tested (mean of X and mean of Y for groups 1 and 2), the alternative hypothesis has two fundamentally different interpretations:

Comparison I: That EITHER $\mu_{XA} \neq \mu_{XB}$ OR $\mu_{YA} \neq \mu_{YB}$ is true

Comparison II: That BOTH $\mu_{XA} \neq \mu_{XB}$ AND $\mu_{YA} \neq \mu_{YB}$ are true

In the first case, the mean centers will be considered not being equal if either the mean of X *or* the mean of Y are significantly different. In the second case, both the mean of *and* the mean of Y must be significantly different for the mean centers to be considered not equal. The first case is clearly easier to fulfill than the second.

Significance levels

By tradition, significance tests for comparisons between two means are made at the $\alpha \leq .05$ or $\alpha \leq .01$ levels, though there is nothing absolute about those levels. The significance levels are selected to minimize *Type 1 Errors*, inadvertently declaring a difference in the means when, in reality, there is not a difference. Thus, a test establishes that the likelihood of falsely rejecting the null hypothesis be less than one-in-twenty (less strict) or one-in-one hundred (more strict).

However, with multiple comparisons, the chances increase for finding 'significance' due to the multiple tests. For example, with two tests - a difference in the means of the X coordinate and a difference in the means of the Y coordinate, the likelihood of rejecting the first null hypothesis ($\mu_{XA} \neq \mu_{XB}$) is one-in-twenty and the likelihood of rejecting the second null hypothesis ($\mu_{YA} \neq \mu_{YB}$) is also one-in-twenty, then the likelihood of rejecting either one null hypothesis or the other is actually one-in-ten.

To handle this situation, comparison I - the 'either/or' condition, a Bonferoni test is appropriate (Anselin, 1995; Systat, 1996). Because the likelihood of achieving a given significance level increases with multiple tests, a 'penalty' must be assigned in finding either the differences in means for the X coordinate or differences in means for the Y coordinates significant. The Bonferoni criteria divides the critical probability level by the number of tests. Thus, if the $\alpha \leq .05$ level is taken for rejecting the null hypothesis, the critical probability for each mean must be .025 (.05/2); that is, differences in either the mean of X or mean of Y between two groups must yield a significance level less than .025.

For comparison II - the 'both/and' condition, on the other hand, the test is more stringent since the differences between the means of X and the means of Y must both be significant. Following the logic of the Bonferoni criteria, the critical probability level is multiplied by the number of tests. Thus, if the $\alpha = .05$ level is taken for rejecting the null hypothesis, then *both* tests must be significant at the $\alpha \leq .10$ level (i.e., .05*2).[1]

Tests

The statistics used are for the t-test of the difference between means (Kanji, 1993).

a. First, test for equality of variances by taking the ratio of the variances (squared sample standard deviations) of both the X and Y coordinates:

$$F_X = \frac{S_{XA}^2}{S_{XB}^2} \tag{B.1}$$

$$F_Y = \frac{S_{YA}^2}{S_{YB}^2} \tag{B.2}$$

with (N_A - 1) and (N_B - 1) degrees of freedom for groups *A* and *B* respectively. This test is usually done with the larger of the variances in the numerator. Since there are two variances being compared (for X and Y, respectively), the logic should follow either *I* or *II* above (i.e., if either are to be true, then the critical α will be actually $\alpha/2$ for each; if both must be true, then the critical α

will be actually $2*\alpha$ for each).

b. Second, if the variances are considered equal, then a t-test for two group means with unknown, but equal, variances can be used (Kanji, 1993; 28). Let

$$S_{XAB} = \text{SQRT}\left[\frac{\sum_{i=1}^{N(A)} (X_{Ai} - \bar{X}_A)^2 + \sum_{i=1}^{N(B)} (X_{Bi} - \bar{X}_B)^2}{(N_A + N_B - 2)}\right] \quad \text{(B.3)}$$

$$S_{YAB} = \text{SQRT}\left[\frac{\sum_{i=1}^{N(A)} (Y_{Ai} - \bar{Y}_A)^2 + \sum_{i=1}^{N(B)} (Y_{Bi} - \bar{Y}_B)^2}{(N_A + N_B - 2)}\right] \quad \text{(B.4)}$$

where the summations are for i=1 to N within each group separately. Then the test becomes

$$t_X = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{XA} - \mu_{XB})}{S_{XAB} * \text{SQRT}\left[\frac{1}{N_A} + \frac{1}{N_B}\right]} \quad \text{(B.5)}$$

$$t_Y = \frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_{YA} - \mu_{YB})}{S_{YAB} * \text{SQRT}\left[\frac{1}{N_A} + \frac{1}{N_B}\right]} \quad \text{(B.6)}$$

with $(N_A + N_B - 2)$ degrees of freedom for each test.

c. Third, if the variances are not equal, then a t-test for two group means with unknown and unequal variances should be used (Kanji, 1993; 29).

$$t_X = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{XA} - \mu_{XB})}{SQRT\left\{\left[\frac{S_{XA}^2}{N_A} + \frac{S_{XB}^2}{N_B}\right]\right\}} \tag{B.7}$$

$$t_Y = \frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_{YA} - \mu_{YB})}{SQRT\left\{\left[\frac{S_{YA}^2}{N_A} + \frac{S_{YB}^2}{N_B}\right]\right\}} \tag{B.8}$$

with degrees of freedom

$$v = \left\{\frac{\left[\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}\right]}{\left[\frac{S_A^4}{N_A^2(N_A + 1)} + \frac{S_B^4}{N_B^2(N_B + 1)}\right]}\right\} - 2 \tag{B.9}$$

for both the X and Y test. Even though this latter formula is cumbersome, in practice, if the sample size of each group is greater than 100, then the t-values for infinity can be taken as a reasonable approximation and the above degrees of freedom need not be tested ($t=1.645$ for $\alpha=.05$; $t=1.960$ for $\alpha=.01$).

d. The significance levels are those selected above. For comparison I - that either differences in the means of X or differences in the means of Y are significant, the critical probability level is $\alpha/2$ (e.g., .05/2 = .025; .01/2 = .005). For comparison II - that both differences in the means of X and differences in the means of Y are significant, the critical probability level is $\alpha*2$ (e.g., .05*2 = .10; .01*2 - .02).

e. Reject the null hypothesis if:

Comparison I: Either tested t-value (t_x or t_y) is greater than the Critical t for $\alpha/2$

Comparison II: Both tested t-values (t_x and t_y) are greater than the critical t for $\alpha*2$

Example 1: Burglaries and Robberies in Baltimore County

To illustrate, compare the distribution of burglaries in Baltimore County with those of robberies, both for 1996. Figure B.1 shows the mean center of all robberies (blue square) and all residential burglaries (red triangle). As can be seen, the mean centers are located within Baltimore City, a property of the unusual shape of the county (which surrounds the city on three sides). Thus, these mean centers cannot be considered an unbiased estimate of the metropolitan area, but unbiased estimates for the County only. When the relative positions of the two mean centers are compared (figure 4.12 in chapter 4), the center of robberies is south and west of the center for burglaries. Is this difference significant or not?

To test this, the standard deviations of the two distributions are first compared and the F-test of the larger to the smaller variance is used (equations B.1 and B.2). *CrimeStat* provides the standard deviation of both the X and Y coordinates; the variance is the square of the standard deviation. In this case, the variance for burglaries is slightly larger than for robberies for both the X and Y coordinates.

$$F_X = \frac{S_{XA}^2}{S_{XB}^2} = \frac{0.0154}{0.0145} = 1.058$$

$$F_Y = \frac{S_{YA}^2}{S_{YB}^2} = \frac{0.0058}{0.0029} = 2.007$$

Because both samples are fairly large (1180 robberies and 6051 burglaries), the degrees of freedom are also very large. The F-tables are a little indeterminate with large samples, but the variance ratio approaches 1.00 as the sample reaches infinity. An approximate critical F-ratio can be obtained by the next largest pair of values in the table (1.22 for $p \leq .05$ and 1.32 for $p \leq .01$). Using this criteria, differences in the variances for the X coordinate are probably not significant while that for the Y coordinates definitely are significant. Consequently, the test for a difference in means with unequal variances is used (equations B.7, B.8 and B.9).

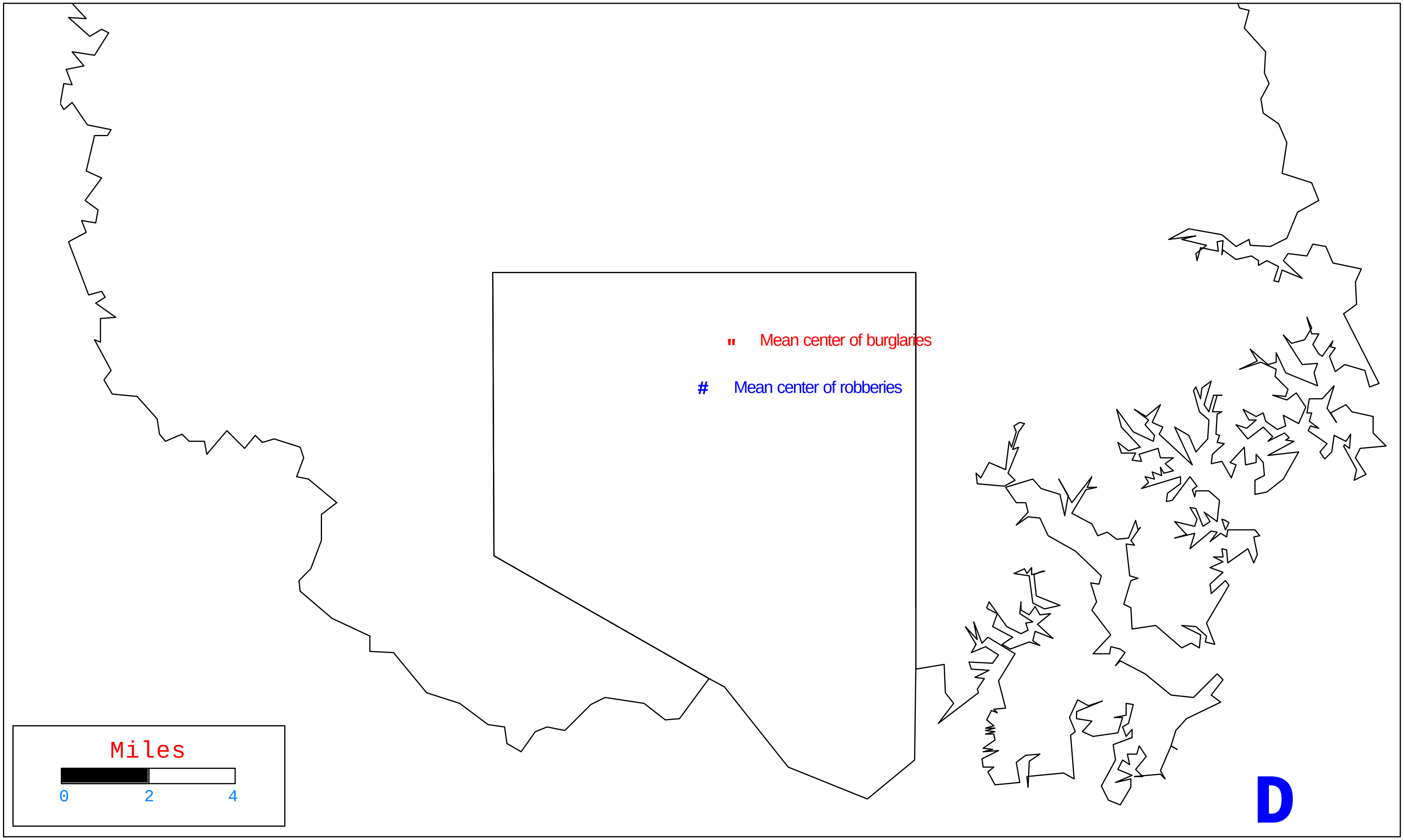
$$t_X = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{XA} - \mu_{XB})}{\text{SQRT}\left\{\left[\frac{S_{XA}^2}{N_A} + \frac{S_{XB}^2}{N_B}\right]\right\}} = \frac{-76.608482 - (-76.620838)}{\text{SQRT}\left\{\left[\frac{0.0154}{6051} + \frac{0.0145}{1180}\right]\right\}}$$

$$= \frac{0.0124}{0.0039} = 3.21 \ (p \leq .005)$$

Figure B.1:

1996 Burglaries and Robberies in Baltimore County

Comparison of Mean Centers



$$t_Y = \frac{(\bar{Y}_A - \bar{Y}_B) - (\mu_{YA} - \mu_{YB})}{SQRT\{[\frac{S_{YA}^2}{N_A} + \frac{S_{YB}^2}{N_B}]\}} = \frac{39.348368 - 39.334816}{SQRT\{[\frac{0.0058}{6051} + \frac{0.0029}{1180}]\}}$$

$$= \frac{0.0136}{0.0018} = 7.36 \quad (p \leq .005)$$

Therefore, whether we use the 'either/or' test (critical $\alpha \leq .025$) or the 'both/and' test (critical $\alpha \leq .1$), we find that the difference in the mean centers is highly significant. Burglaries have a different center of gravity than robberies in Baltimore County.

Differences in the Standard Distance Deviation of Two Samples

Since the standard distance deviation, S_{XY} (equation 4.6 in chapter 4) is a standard deviation, differences in the standard distances of two groups can be compared with an equality of variance test (Kanji, 1993, 37).

$$F = \frac{S_{XYA}^2}{S_{XYB}^2} \qquad \text{(B.10)}$$

with $(N_A - 1)$ and $(N_B - 1)$ degrees of freedom for groups A and B, respectively. This test is usually done with the larger of the variances in the numerator. Since there is only one variance being compared, the critical α are as listed in the tables.

From *CrimeStat*, we find that the standard distance deviation of burglaries is 8.44 miles while that for robberies is 7.42 miles. In chapter 4, figure 4.12 displayed these two standard distance deviations. As can be seen, the dispersion of incidents, as defined by the standard distance deviation, is greater for burglaries than for robberies. The F-test of the difference is calculated by

$$F = \frac{S_{XYA}^2}{S_{XYB}^2} = \frac{8.44^2}{7.42^2} = 1.29$$

with 6050 and 1180 degrees of freedom respectively. Again, the F-tables are slightly indeterminate with respect to large samples, but the next largest F beyond infinity is 1.25 for $p \leq .05$ and 1.38 for $p \leq .01$. Thus, it appears that burglaries have a significantly greater dispersion than robberies, at least at the $p \leq .05$ level.

Differences in the Standard Deviational Ellipse of Two Samples

In a standard deviational ellipse, there are actually six variables being compared:

Mean of X
Mean of Y
Angle of rotation
Standard deviation along the transformed X axis
Standard deviation along the transformed Y axis
Area of the ellipse

Differences in the mean centers

Comparisons between the two mean centers can be tested with the above statistics.

Differences in the angle of rotation

Unfortunately, to our knowledge, there is not a formal test for the difference in the angle of rotation. Until this test is developed, we have to rely on subjective judgements.

Differences in the standard deviations along the transformed axes

The differences in the standard deviations along the transformed axes (X and Y) can be tested with an equality of variance test (Kanji, 1993, 37).

$$F_{Sx} = \frac{S_{X1}^2}{S_{X2}^2} \qquad \text{(B.11)}$$

$$F_{SY} = \frac{S_{Y1}^2}{S_{Y2}^2} \qquad \text{(B.12)}$$

with (N_A - 1) and (N_B - 1) degrees of freedom for groups A and B respectively. This test is usually done with the larger of the variances in the numerator. The example above for comparing the mean centers of Baltimore County burglaries and robberies illustrated the use of this test.

Differences in the areas of the two ellipses

Since an area is a variance, the differences in the areas of the two ellipses can be compared with an equality of variance test (Kanji, 1993, 37).

$$F = \frac{Area_A}{Area_B} \tag{B.13}$$

with (N_A - 1) and (N_B - 1) degrees of freedom for groups 1 and 2 respectively. This test is done with the larger of the variances in the numerator.

Significance levels

The testing of each of these parameters for the difference between two ellipses is even more complicated than the difference between two mean centers since there are up to six parameters which must be tested (differences in mean X, mean Y, angle of rotation, standard deviation along transformed X axis, standard deviation along transformed Y axis, and area of ellipse). However, as with differences in mean center of two groups, there are two different interpretations of differences.

Comparison I:	That the two ellipses differ on ANY of the parameters
Comparison II:	That the two ellipses differ on ALL parameters.

In the first case, the critical probability level, α, must be divided by the number of parameters being tested, α/p. In theory, this could involve up to six tests, though in practice some of these may not be tested (e.g., the angle of rotation). For example, if five of the parameters are being estimated, then the critical probability level at $\alpha \leq .05$ is actually $\alpha \leq .01$ (.05/5).

In the second case, the critical probability level, α, is multiplied by the number of parameters being tested, $\alpha * p$, since *all* tests must be significant for the two ellipses to be considered as different. For example, if five of the parameters are being estimated, then the critical probability level, say, at $\alpha \leq .05$ is actually $\alpha \leq .25$ (.05*5).

Differences in Mean Direction Between Two Groups

Statistical tests of different angular distributions can be made with the directional mean and variance statistics. To test the difference in the angle of rotation between two groups, a Watson-Williams test can be used (Kanji, 1993; 153-54). The steps in the test are as follows:

1. All angles, θ_i, are converted into radians

$$Radian_i = Angle_i * \pi/180 \tag{B.14}$$

2. For each sample separately, *A* and *B*, the following measures are calculated

$$C_j = \sum_{A=1}^{N_1} \cos\,\theta_j \qquad S_j = \sum_{A=1}^{N_1} \sin\,\theta_j \tag{B.15}$$

$$C_k = \sum_{B=1}^{N_2} \cos\,\theta_k \qquad S_k = \sum_{B=1}^{N_2} \sin\,\theta_k \tag{B.16}$$

where θ_j and θ_k are the individual angles for the respective groups, *A* and *B*.

3. Calculate the resultant lengths of each group

$$R_A = SQRT[\, C_A^2 + S_A^2 \,] \tag{B.17}$$

$$R_B = SQRT[\, C_B^2 + S_B^2 \,] \tag{B.18}$$

4. Resultant lengths for the combined sample are calculated as well as the length of the resultant vector.

$$C = C_A + C_B \tag{B.19}$$

$$S = S_A + S_B \tag{B.20}$$

$$R = SQRT[\, C^2 + S^2 \,] \tag{B.21}$$

$$N = N_A + N_B \tag{B.22}$$

$$R^* = \frac{(R_A + R_B)}{N} \tag{B.23}$$

5. An F-test of the two angular means is calculated with

$$F = g\,(N - 2) \frac{R_A + R_B - R}{N - (R_A + R_B)} \tag{B.24}$$

where

$$g = 1 - \frac{3}{8k} \tag{B.25}$$

with k being identified from a maximum likelihood Von Mises distribution by referencing R^* with 1 and N-2 degrees of freedom (Mardia, 1972; Gaile and Burt, 1980). Some of the reference k's are given in table B.1 (from Mardia, 1972; Kanji, 1993, table 38).

Table B.1

Maximum Likelihood Estimates for Given R* in the Von Mises Case
(from Mardia, 1972; Kanji, 1993, table 38)

R^*	k
0.00	0.00000
0.05	0.10013
0.10	0.20101
0.15	0.30344
0.20	0.40828
0.25	0.51649
0.30	0.62922
0.35	0.74783
0.40	0.87408
0.45	1.01022
0.50	1.15932
0.55	1.32570
0.60	1.51574
0.65	1.73945
0.70	2.01363
0.75	2.36930
0.80	2.87129
0.85	3.68041
0.90	5.3047
0.95	10.2716
1.00	infinity

Table B.2

Comparison of Two Groups for Angular Measurements
Angle of Deviation From Due North

Group A		Group B	
Incident	**Measured Angle**	**Incident**	**Measured Angle**
1	160	1	196
2	184	2	212
3	240	3	297
4	100	4	280
5	95	5	235
6	120	6	353
		7	190
		8	340

6. Reject the null hypothesis of no angular difference if the calculated F is greater than the critical value $F_{1, N-2}$.

Example 2: Angular comparisons between two groups

A fourth example is that of sets of angular measurements from two different groups, A and B. Table B.2 provides the data for the two sets. The angular mean for Group A is 144.83^0 with a directional variance of 0.35 while the angular mean for Group B is 258.95^0 with a directional variance of 0.47. The higher directional variance for Group B suggests that there is more angular variability than for Group A.

Using the Watson-Wheeler test, we compare these two distributions.

1. All angles are converted into radians (equation B.14).

2. The cosines and sines of each angle are taken and are summed within groups (equations B.15 and B.16).

   C_A = -3.1981 S_A = 2.2533
   C_B = -.8078 S_B = -4.1381

3. The resultants are calculated (equations B.17 and B.18).

   R_A = 3.9121
   R_B = 4.2162

4. Combined sample characteristics are defined (equations B.19 through B.23).

   C = -4.0059
   S = -1.8848
   R = 4.4271
   N = 14
   R^* = 0.5806

5. Once the parameter, k, is obtained (approximated from table 4.1 or obtained from Mardia, 1972 or Kanji, 1993), g is calculated, and an F-test is constructed (equations B.24 and B.25).

   k = 1.44
   g = 0.7396
   F = 5.59

6. The critical F for 1 and 12 degrees of freedom is 4.75 ($p \leq .05$) and 9.33 ($p \leq .01$). The test is significant at the $p \leq .05$ level and we reject the null hypothesis of no angular differences between the two groups. Group A has a different angular distribution than Group B.

Endnotes for Appendix B

1. There are limits to the Bonferoni logic. For example, if there were 10 tests, having a threshold significance level of .005 (.05 / 10) for the 'either/or' conditions and a threshold significance level of .50 (.05 * 10) for the 'both/and' would lead to an excessively difficult test in the first case and a much too easy test in the second. Thus, the Bonferoni logic should be applied to only a few tests (e.g., 5 or fewer).