

## Topic 14: Nonparametric Methods (ST & D Chapter 24)

### Introduction

All of the statistical tests discussed up until now have been based on the assumption that the data are normally distributed. Implicitly we are estimating the parameters of this distribution, the mean and variance. These are *sufficient statistics* for this distribution, that is, specifying the mean and variance of a normal distribution specifies it completely. The central limit theorem provides a justification for the normality assumption in many cases, and in still other cases the robustness of the tests with respect to normality provides a justification. *Parametric statistics* deal with the estimation of parameters (e.g., means, variances) and testing hypotheses for continuous normally distributed variables.

In cases where the assumption of normality cannot be employed, however, *nonparametric*, or *distribution-free* methods may be appropriate. These methods lack the underlying theory of the parametric methods and we will simply discuss them as a collection of tests. *Nonparametric statistics* do not relate to specific parameters (the broad definition). They maintain their distributional properties irrespective of the underlying distribution of the data and for this reason they are called distribution-free methods. *Nonparametric statistics* compare distributions rather than parameters. Therefore, nonparametric statistics are less restrictive in terms of the assumptions compared to parametric techniques. Although some assumptions, for example, **samples are random and independent**, are still required. In cases involving ranked data, i.e. data that can be put in the order, and/or categorical data *nonparametric statistics* are necessary. *Nonparametric statistics* are not generally as powerful (sensitive) as parametric statistics if the assumptions regarding the distribution are valid for the parametric test. That is, type II errors (false null hypothesis is accepted) are more likely.

### 14.1. Advantages of using nonparametric techniques are the following.

1. They are appropriate when only weak assumptions can be made about the distribution.
2. They can be used with categorical data when no adequate scale of measurement is available.
3. For data that can be ranked, nonparametric test using ranked data may be the best option.
4. They are relatively quick and easy to apply and to learn since they involve counts, ranks and signs.

## 14.2 The $\chi^2$ test of goodness of fit (ST&D Chapter 20, 21)

The goodness of fit test involves a comparison of the observed frequency of occurrence of classes with that predicted by a theoretical model. Suppose there are  $n$  classes with *observed frequencies*  $O_1, O_2, \dots, O_n$ , and corresponding *expected frequencies*  $E_1, E_2, \dots, E_n$ . The expected frequency is the average number or expected value when the hypothesis is true and is simply calculated as  $n$  multiplied by the hypothesized population proportion. The statistics

$$X^2 = \sum \frac{(O - E)^2}{E}$$

has a distribution that is distributed *approximately* as  $\chi^2$  with  $n - 1$  degrees of freedom. This approximation becomes better as  $n$  increases. If parameters from the data are used to calculate the expected distributions the degrees of freedom of the  $\chi^2$  will be  $n - 1 - p$ ; where  $p$  is the number of parameteres estimated. For example, if we want to test that a distribution is normal and we estimate the mean and the variance from the data to calculate the expected frequencies, the df will be  $n - 1 - 2$  (ST&D482). If the hypothesis is extrinsic to the data, like in a genetic proportion, then  $p = 0$  and  $df = n - 1$ .

There are some restrictions to the utilization of  $\chi^2$  tests. The approximation is good for  **$n > 50$** . There should be **no 0** expected frequencies and expected **frequencies < 5 should not be present in more than 20% of the classes**. If this is not possible, an alternative is to use **Fischer's Exact Test** (ST&D 512, provided by **SAS PROC FREQ**).

We can formulate the hypothesis test as follows. Let  $H_0$  be  $O_1 = E_1, \dots, O_n = E_n$ , and let  $H_1$  be that  $O_i \neq E_i$  for at least one  $i$ . Then  $H_0$  is rejected at the  $\alpha$  level of significance if

$$X^2 = \sum \frac{(O - E)^2}{E} \geq \chi^2_{1-\alpha, n-1}$$

An adjusted  $\chi^2$  can be used when the criterion has a single degree of freedom in order to make the distribution of  $X^2$  more close to a  $\chi^2$  distribution (Yate's correction for continuity). This adjustment produces a lower  $\chi^2$  and a more conservative test.

$$adjusted X^2 = \sum \frac{(|O - E| - .5)^2}{E}$$

### 14.2.1. One way classification

#### Tests of hypotheses

Test of hypothesis using the  $\chi^2$  criterion can be exemplified by tests of 1:1 sex ratio or 3:1 segregation test of dominance in  $F_2$  generation.

## &lt;Example 14.1&gt; (ST &amp; D p 488)

Suppose a certain  $F_1$  generation of a *Drosophila* species with 35 males and 46 female. Test the hypothesis of a 1:1 sex ratio. ( $H_0: p = q$  (with  $q = 1-p$ ))

| Sex    | Observed<br>( <i>O</i> ) | Expected<br>( $E = p*n$ ) | Deviation<br>( $O-E$ ) | $(O-E)^2$ | $(O-E)^2/E$  |
|--------|--------------------------|---------------------------|------------------------|-----------|--------------|
| Male   | 35                       | 40.5                      | -5.5                   | 30.25     | 0.747        |
| Female | 46                       | 40.5                      | 5.5                    | 30.25     | 0.747        |
| Sum    | 81                       | 81                        | 0                      | 60.5      | <b>1.494</b> |

The  $\chi^2$  value is 1.494 with 1 df (= no. of classes -1). From table A.5, the probability for 1.49 with 1 df is between 0.1 and 0.25. Therefore, we fail to reject the null hypothesis that the sex ratio is 1:1.

<Example 14.2> Using SAS, test the hypothesis of a 9:3:3:1 ratio, normal dihybrid segregation for the data of  $F_2$  progeny of a barley cross (ST&D p500). The observed characters are non-two-row versus two-row, and green versus chlorina plant color. The data were 1178: 291: 273: 156. (1=green, non-two-row; 2=green, two-row; 3=chlorina, non-two-row; 4= chlorina, two-row).

```
data f2;
  input pheno count @@;
cards;
  1 1178  2  291  3  273  4  156

proc freq;
  weight count;
  tables pheno / testp = (0.5625, 0.1875, 0.1875, 0.0625); /* 9:3:3:1 */
run;
```

The **FREQ procedure** produces one-way to n-way frequency and crosstabulation (contingency) tables. For one-way frequency tables, PROC FREQ can compute statistics to test for equal proportions, specified proportions, or the binomial proportion. For contingency tables, PROC FREQ can compute various statistics to examine the relationships between two classification variables adjusting for any stratification variables.

Since the input data are in cell count form, the **WEIGHT** statement is required. The WEIGHT statement names the variable Count, which provides the frequency of each combination of data values. In the **TABLES** statement, Pheno specifies a table where the rows are pheno (in two way TABLES AA\*BB specifies a table where the rows are AA and the columns are BB).

## OUTPUT

| PHENO | Frequency | Percent | Test<br>Percent | Cumulative<br>Frequency | Cumulative<br>Percent |
|-------|-----------|---------|-----------------|-------------------------|-----------------------|
| 1     | 1178      | 62.1    | 56.3            | 1178                    | 62.1                  |
| 2     | 291       | 15.3    | 18.8            | 1469                    | 77.4                  |
| 3     | 273       | 14.4    | 18.8            | 1742                    | 91.8                  |
| 4     | 156       | 8.2     | 6.3             | 1898                    | 100.0                 |

Chi-Square Test for Specified Proportions  
-----  
**Statistic = 54.313      DF = 3      Prob = 0.001**

The number of degrees of freedom is one less than the number of classes. We conclude that the data don't follow the ratio of 9:3:3:1 with a probability of 0.001.

### 14.2.2. Contingency tables

For more than one variable, data can be conveniently represented by two-way tables called *contingency* table. These tables are useful to test if two classification criteria are independent (test of independence) and if two samples belong to the same population in relation to one-classification criteria (test of homogeneity). These tests are based on the principle that if two events are independent, the probability of their occurring together can be computed as the product of their separate probabilities.

Example 14.2 can be represented using a two-way table if a different question is asked. The new question is if the color and number of rows are independent. In genetic terms this is a test for linkage between the genes affecting the two traits. Example 14.3 demonstrates how to generate a contingency table using SAS and test the hypothesis of independence between two characters. The test of independence tests the goodness of fit of the observed cell frequencies to their expected frequencies. In this case, the test criterion will be the same as one way except

**degrees of freedom** =(row-1)\*(column-1)

and the expected frequency can be obtained by:

$$E = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

<**Example 14.3**> Perform the test of Independence with a 2 by 2 contingency table using the data from example 14.2.

We test the hypothesis ( $H_0: p_{ij} = p_{i.} * p_{.j}$ ) regardless of the true ratio. As a result, we reject the null hypothesis that two characters are independent.

```
data f2;
```

```

input c1 $ c2 $ pheno count;
cards;
1green non2row 1 1178
1green two_row 2 291
2chlor non2row 3 273
2chlor two_row 4 156

proc freq;
weight count;
tables c1*c2 / chisq nopercnt nocum norow nocol;
run;

```

**CHISQ** option requests chi-square statistics for assessing association.

To simplify the output:

**NOPERCENT** suppresses display of the percentage in crosstabulation tables

**NOCUM** suppresses display of cumulative frequencies and cumulative percentages in one-way frequency tables and in list format

**NOROW** suppresses display of the row percentage for each cell

**NOCOL** suppresses display of the column percentage for each cell

```

                TABLE OF C1 BY C2
          C1          C2
Frequency,non2row ,two_row , Total
1green      , 1178 , 291 , 1469
2chlor      , 273 , 156 , 429
Total              1451      447      1898

          STATISTICS FOR TABLE OF C1 BY C2
Statistic                DF      Value      Prob
Chi-Square                1    50.538    0.001
Likelihood Ratio Chi-Square      1    47.258    0.001
Continuity Adj. Chi-Square      1    49.623    0.001
Mantel-Haenszel Chi-Square      1    50.511    0.001
Fisher's Exact Test (Left)
                                (Right)
                                (2-Tail)
                                1.000
                                4.60E-12
                                7.11E-12

Phi Coefficient                0.163
Contingency Coefficient        0.161
Cramer's V                    0.163
Sample Size = 1898

```

<**Example 14.4**> One example in ecology is found in Sokal & Rohlf p 731. A plant ecologist samples 100 trees of a rare species from a 400 square mile area. He records for each tree whether or not it is rooted in serpentine soils and whether its leaves are pubescent or smooth.

| <i>Soil</i>    | <i>Pubescent</i> | <i>Smooth</i> |
|----------------|------------------|---------------|
| Serpentine     | 12               | 22            |
| Non serpentine | 16               | 50            |

&lt;Output&gt;

| STATISTICS FOR TABLE OF SOIL BY LEAF |    |        |       |
|--------------------------------------|----|--------|-------|
| Statistic                            | DF | Value  | Prob  |
| Chi-Square                           | 1  | 1.360  | 0.244 |
| Likelihood Ratio Chi-Square          | 1  | 1.332  | 0.248 |
| Continuity Adj. Chi-Square           | 1  | 0.867  | 0.352 |
| Mantel-Haenszel Chi-Square           | 1  | 1.346  | 0.246 |
| Fisher's Exact Test (Left)           |    |        | 0.176 |
| (Right)                              |    |        | 0.918 |
| (2-Tail)                             |    |        | 0.251 |
| Phi Coefficient                      |    | -0.117 |       |
| Contingency Coefficient              |    | 0.116  |       |
| Cramer's V                           |    | -0.117 |       |

Sample Size = 100

We accept the null hypothesis that the leaf type is independent of the soil type in which the tree is rooted.

## 14.4. One-sample tests

### 14.4.1 The Kolmogorov-Smirnov test and the normal probability plot (ST&D p 564)

The  $\chi^2$  test is useful to test hypothesis about the distribution of data that fall in **categories**. For a single sample of data, the Kolmogorov-Smirnov test is used to test whether or not the sample of data is consistent with a specified **continuous distribution** function. This is a useful nonparametric test for goodness of fit applicable to continuous distributions. The Kolmogorov-Smirnov test does not require the assumption that the population is normally distributed.

As we already saw in the previous chapters, a common graphical way for assessing normality of the data is the *normal probability plot*. If the data has a normal distribution, the plotted points will be close to a straight line with intercept ( $\bar{Y}$ ) and slope (s). The linearity of this plot can be measured by the correlation r between the data and their normal scores. A statistic (W) by Shapiro-Wilk's test is used to test for normality for small to medium number of samples and Kormogorov-Smirnov statistic (D) is used for large samples. The test for normality of a single sample of data can be obtained using SAS proc univariate.

## 14.5. Two sample tests

### 14.5.1 The sign test for two paired samples (also one-sample test for median)

The sign test is designed to test a hypothesis about the location of a population distribution. It is most often used **to test the hypothesis about a population median**,

and often involves the use of matched pairs, for example, before and after data, in which case it tests for a median difference of zero. That is, the signed numbers serve to test the null hypothesis that each difference has a median of zero (pluses and minuses occur with the same chances). The sign test does not require the assumption that the population is normally distributed.

$$X^2 = \frac{(n_1 - n_2)^2}{n_1 + n_2}, \text{ with } 1 \text{ df}$$

The values,  $n_1$  and  $n_2$  are the numbers of pluses and minuses. In many applications, this test is used in place of the one sample t-test when the normality assumption is questionable. It is a less powerful alternative to the Wilcoxon signed ranks test, but does not assume that the population probability distribution is symmetric.

For paired observations we can also ask the question if treatment A gives a response that is  $C$  units better than that of B. To do this test record the signs of the differences  $Y_{1i} - (Y_{2i} + C)$  and apply the sign test.

With this test is impossible to detect a departure from the null hypothesis with fewer than six pairs of observations. With **20 or more pairs** this test is more useful. This test does not require a symmetric distribution.

This test can also be used as a **one-sample test**, to test the null hypothesis that the median is a specified value.

- Record the number of observations above ( $n_1$ ) and below ( $n_2$ ) the specified value
- Use the previous equation.

### 14.5.2 Wilcoxon's signed-rank test for paired treatments (also one-sample test for median)

The Wilcoxon signed-rank test is an improvement on the sign test in terms of detecting real differences with paired treatments. The improvement is attributable to the use of the magnitude of the differences. The steps of Wilcoxon's signed-rank test are;

1. Rank the differences between paired values from smallest to largest without regard to sign.
2. Assign the signs (tied ranks including both signs are given averages)

|             |    |    |    |    |     |      |      |     |
|-------------|----|----|----|----|-----|------|------|-----|
| Difference  | +2 | -4 | -6 | +8 | +10 | -12  | +12  | +15 |
| Signed rank | +1 | -2 | -3 | +4 | +5  | -6.5 | +6.5 | +8  |

3. Obtain  $T_+$  and  $T_-$  (sum of positive ranks and negative ranks respectively). Choose smaller one and call it  $T$ .

$T_+ = 24.5$

$T_- = 11.5$

4. Compare T with the critical value in Table A17. Note that small values of T are the significant ones.

**Critical T=4 11.5>4 ⇒ Not significant**

**The Wilcoxon signed rank test can be also used as a one-sample test to analyze if the median is a certain value.**

**The Wilcoxon signed rank test requires that the distribution be symmetric; the previous sign test (14.5.1) does not require this assumption.**

### 14.5.3 The Kolmogorov-Smirnov test for two independent samples

The null hypothesis is that the two independent samples come from an identical distribution.

This test is sensitive to differences in means and/or variances, since is a **test of the equality of distributions** rather than of specific parameters.

The algorithm for the test is:

- 1) Rank all observations in ascending order.
- 2) Determine the sample cumulative distribution functions  $F_n(Y_1)$  and  $F_n(Y_2)$ .
- 3) Compute  $|F_n(Y_1) - F_n(Y_2)|$  at each Y value.
- 4) Find the maximum difference D over all values of Y. Compare it with a critical value in Tables A.22A (balanced design) and A.22B (unbalanced design).

#### **Example from ST&D 571**

| $Y_1$ | $F_1(Y_1)$ | $Y_2$ | $F_2(Y_2)$ | $ F_1(Y_1) - F_2(Y_2) $    |
|-------|------------|-------|------------|----------------------------|
| 53.2  | 1/7        |       |            | $ 1/7 - 0  = 1/7$          |
| 53.6  | 2/7        |       |            | $ 2/7 - 0  = 2/7$          |
| 54.4  | 3/7        |       |            | $ 3/7 - 0  = 3/7$          |
| 56.2  | 4/7        |       |            | $ 4/7 - 0  = 4/7$          |
| 56.4  | 5/7        |       |            | $ 5/7 - 0  = 5/7$          |
| 57.8  | 6/7        |       |            | $ 6/7 - 0  = 6/7$ <b>D</b> |
|       |            | 58.7  | 1/6        | $ 6/7 - 1/6  = 29/42$      |
|       |            | 59.2  | 2/6        | $ 6/7 - 2/6  = 22/42$      |
|       |            | 59.8  | 3/6        | $ 6/7 - 3/6  = 15/42$      |
| 61.9  | 7/7        |       |            | $ 7/7 - 3/6  = 1/2$        |
|       |            | 62.5  | 4/6        | $ 7/7 - 4/6  = 1/3$        |
|       |            | 63.1  | 5/6        | $ 7/7 - 5/6  = 1/6$        |
|       |            | 64.2  | 6/6        | $ 7/7 - 6/6  = 0$          |

In this case the maximum difference **D** =  $6/7 = 0.857$

The critical value in Table A22B for  $\alpha = 0.01$  is  $5/6 = 0.83$

Since **D** > critical value, we reject  $H_0$

**We conclude that the samples belong to different populations**

### 14.5.4. The Wilcoxon-Mann-Whitney **location test** for two independent samples

This tests the hypothesis that two data sets have the same **location parameter** (the same median).

Assume the data sets have size  $n_1$  and  $n_2$ , where  $n_1 < n_2$ . The test procedure is this:

- 1) Rank all observations from both observations together from smallest to largest (Tied observations are given the mean rank).
- 2) Add the ranks for the smaller sample. Call this T.
- 3) Compute  $T' = n_1(n_1 + n_2 + 1) - T$ .  $T'$  is the value you would get by adding the ranks if the observations are ranked in the **opposite direction**, from the largest to the smallest.
- 4) Compare **the smaller** of T and  $T'$  with Table A.18 in ST&D. Note that values **smaller than the tabulated value lead to rejection** of the null hypothesis.

In the sheep - cow data (ST&D96) here are the steps:

1)

|       |      |      |      |      |      |      |      |      |      |      |      |      |      |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Rank  | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   |
| Set   | S    | S    | S    | S    | S    | S    | C    | C    | C    | S    | C    | C    | C    |
| Value | 53.2 | 53.6 | 54.4 | 56.2 | 56.4 | 57.8 | 58.7 | 59.2 | 59.8 | 61.9 | 62.5 | 63.1 | 64.2 |

- 2) C is the smaller set, with  $n_1 = 6$ .  $T = 7+8+9+11+12+13 = 60$ .
- 3)  $T' = 6(6 + 7 + 1) - 60 = 1+2+3+5+6+7 = 24$ .
- 4) From Table A.18,  $24 < 27$  so reject ( $P < 0.05$ ),  $24 = 24$  so don't reject ( $P = 0.01$ ).

Note that if the data of C8 and S5 were swapped, for example, the null hypothesis would not be rejected at the 0.05 level.

## 14.6 More than two sample tests (Non parametric CRD)

### 14.6.1 Kruskal-Wallis k-sample test

Kruskal-Wallis developed a nonparametric method to test the null hypothesis that all the populations have the same location (median).

The test can be used to compare three or more samples and is the *nonparametric analogue to the F-test* used in analysis of variance.

While analysis of variance tests depend on the assumption that all populations under comparison are normally distributed, the Kruskal-Wallis test places no such restriction. It is a logical extension of the Wilcoxon-Mann-Whitney Test. So, for  $k=2$  it is identical to Wilcoxon-Mann-Whitney test.

- 1) Rank all observations TOGETHER from smallest to largest (from all treatments)
- 2) Sum the ranks for each sample
- 3) Compute the test criterion and compare with tabulated values Kruskal and Wallis (1952)

$$H = \frac{12}{n(n+1)} \sum_i \frac{R_i^2}{n_i} - 3(n+1)$$

#### 14.7 Use of SAS for non parametric statistics **PROC NPAR1WAY**

PROC NPAR1WAY is a nonparametric procedure for testing that the distribution of a variable has the same location parameter across different groups or, in the case of empirical distribution function tests, that the distribution is the same across the different groups.

PROC NPAR1WAY, computes simple linear rank statistics like Kruskal-Wallis test, or the W-M-W test for two samples, and also statistics like Kolmogorov-Smirnov that are based on the empirical distribution of the sample.

Here it is applied to the Coefficients of digestibility of dry matter, feed corn silage, in percentages. (ST&D p96, 579).

```
data digest;
  /* nonparametric tests for data from ST&D p96 */
  do animal = 'sheep', 'steers';
    do rep= 1 to 7;
      input dm @@;
      output;
    end;
  end;
cards;
57.8 56.2 61.9 54.4 53.6 56.4 53.2
64.2 58.7 63.1 62.5 59.8 59.2 .
proc npar1way anova edf wilcoxon data=digest;
  class animal;
  var dm;
run;
```

**ANOVA** requests a standard analysis of variance on the raw data.

**EDF** requests statistics based on the empirical distribution function including the Kolmogorov-Smirnov

**WILCOXON** requests an analysis using Wilcoxon scores. When there are two classification levels, or two samples, this option produces the Wilcoxon rank-sum test. For any number of classification levels, this option produces the Kruskal-Wallis test

N P A R 1 W A Y P R O C E D U R E  
Analysis of Variance for Variable DM Classified by Variable ANIMAL

### Parametric ANOVA

| ANIMAL | N | Mean       | Among MS   | Within MS  |
|--------|---|------------|------------|------------|
| sheep  | 7 | 56.2142857 | 81.9271978 | 7.32577922 |
| steer  | 6 | 61.2500000 |            |            |
|        |   |            | F Value    | Prob > F   |
|        |   |            | 11.183     | 0.0065     |

Output of an ordinary parametric ANOVA suggesting a rejection of the null hypothesis under the assumption of normality

### Wilcoxon 2-Sample Test

| Wilcoxon Scores (Rank Sums) for Variable DM<br>Classified by Variable ANIMAL |   |               |                   |                  |            |
|------------------------------------------------------------------------------|---|---------------|-------------------|------------------|------------|
| ANIMAL                                                                       | N | Sum of Scores | Expected Under H0 | Std Dev Under H0 | Mean Score |
| sheep                                                                        | 7 | 31.0          | 49.0              | 7.0              | 4.428      |
| steer                                                                        | 6 | 60.0          | 42.0              | 7.0              | 10.000     |

Wilcoxon 2-Sample Test (Normal Approximation)

(with Continuity Correction of .5)

S = 60.0000 Z = 2.50000 Prob > |Z| = **0.0124**

T-Test Approx. Significance = 0.0279

Sum of ranks within each group is produced

### Kruskal-Wallis Test

Kruskal-Wallis Test (Chi-Square Approximation)

CHISQ = 6.6122 DF = 1 Prob > CHISQ = **0.0101**

Output of Kruskal-Wallis test, which is appropriate when there are more than two samples

### Kolmogorov-Smirnov Test

Kolmogorov-Smirnov Test for Variable DM Classified by Variable ANIMAL

| ANIMAL | N  | EDF at Maximum | Deviation from Mean at Maximum |
|--------|----|----------------|--------------------------------|
| sheep  | 7  | 0.857142857    | 1.04667085                     |
| steer  | 6  | 0.000000000    | -1.13053373                    |
|        | 13 | 0.461538462    |                                |

Maximum Deviation Occurred at Observation 1

Value of DM at Maximum 57.8000000

Kolmogorov-Smirnov 2-Sample Test (Asymptotic)

KS = 0.427302 D = 0.857143

KSa = 1.54066 Prob > KSa = **0.0174**

Output of Kolmogorov-Smirnov test, which looks for differences in any part of the distribution

### 14.8. Coefficient of rank correlation

The Pearson correlation coefficient  $r$  is applicable only to a bivariate normal distribution. Spearman's rank correlation coefficient  $r_s$  do not require this assumption. The Spearman's rank correlation coefficient measures correspondence between ranks, so it is not necessarily a measure of linear correlation. This coefficient is simply the ordinary correlation coefficient  $r$  between the ranked values of both variables.

The procedure is:

- 1) Rank the observations for each variable
- 2) Obtain the differences in ranks for the paired observations. Let  $d_i$  = the difference between the ranks of pair  $i$
- 3) Calculate  $r_s$  as follow

$$r_s = 1 - \frac{6 \sum_i d_i^2}{(n-1)(n+1)}$$

Where  $d_i$  is the difference for the  $i$ th pair, and  $n$  is the number of pairs.

The estimated  $r_s$  is compared with the critical value in the following table for samples of 10 or fewer pairs.

Table of two-tailed significance levels of the Spearman Rank Correlation  $r_s$  for  $n < 11$ . For  $n > 10$  use tables for regular Pearson correlation coefficient.

| Sample size n | Significance level |       |
|---------------|--------------------|-------|
|               | 5%                 | 1%    |
| 5             | 1.000              | none  |
| 6             | 0.886              | 1.00  |
| 7             | 0.786              | 0.929 |
| 8             | 0.738              | 0.857 |
| 9             | 0.683              | 0.817 |
| 10            | 0.648              | 0.781 |

If  $r_s$  is  $> 10$  then Student's distribution with  $n-2$  df is used to test the following statistics:

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

Spearman's rank correlation coefficient can be used to correlate order of emergence in a sample of insects with a ranking in size; or ranking in flower size of roses with order of branching; rating competence of taste panelist using known increase of certain flavoring.

If we know that the data do not have a bivariate normal distribution, Spearman's coefficient can be used to test for the significance of association between the two variables. This method uses a coefficient of rank correlation after ranking the variables.

<**Example 14.5**> The data of exercise 11.2.1 in ST&D p 290. Application of Spearman's rank correlation. The following values correspond to tube length (T) and limb length (L) of flowers of a Nicotiana cross.

```
data digest;
  /* Coefficient of rank correlation for the data from p290 ex11.2.1 */
  input t l @@;
  cards;
  49 27  44 24  32 12  42 22  32 13  53 29  36 14  39 20  37 16
  45 21  41 22  48 25  39 18  40 20  34 15  37 20  35 13

  proc corr;
  /* Pearson's correlation coefficients - ordinary corr. */
  var t l;

  proc corr Spearman;
  /* Spearman's correlation coefficients. */
  var t l;

  proc freq;
  table t*l / noprint measures cl;
run;
```

### **PROC FREQ**

- The option **MEASURES** in table statement enables SAS to print out correlations including **Spearman's** coefficient of rank correlation
- The **NOPRINT** option in the TABLES statement suppresses display of the crosstabulation tables but allows display of the requested statistics.
- The **CL** option in the TABLES statement, computes asymptotic confidence limits
- for all MEASURES statistics. The confidence coefficient is determined according to the value of the ALPHA=option, which by default equals 0.05 and produces 95 percent confidence limits.

### **PROC CORR**

- The option **SPEARMAN** after PROC CORR will also produce the calculation of the Spearman's rank correlation coefficient

| Statistic            | Value  | ASE    | Confidence Bounds |        |
|----------------------|--------|--------|-------------------|--------|
| Pearson Correlation  | 0.9538 | 0.0202 | 0.9143            | 0.9934 |
| Spearman Correlation | 0.9618 | 0.0203 | 0.9220            | 1.0000 |

Pearson correlation is the ordinary parametric correlation coefficient and Spearman correlation is the ranked correlation coefficient. Both indicate significant correlation.