The statistical tests we have performed, *z, t,* and *F,* are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ tests. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ tests are statistical tests for population parameters such as means, variances, and proportions that involve \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ about the populations from which the samples were selected. One of these \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is that the populations are normally distributed.

However, suppose the population for a particular hypothesis testing situation is not \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ distributed. Statisticians developed \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ or distribution-free statistics which are used when the population from which samples are selected is not normally distributed.

Nonparametric statistics can be used to test hypotheses that do not involve specific population \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, such as *μ*, *σ*, or *p*.

**Nonparametric statistical tests** are used to test hypotheses about population parameters when the assumption about \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ cannot be met.

# 13 - 1. Advantages and Disadvantages of Nonparametric Methods

## Objective 1. State the Advantages and Disadvantages of Nonparametric Methods.

Nonparametric tests and statistics can be used instead of their parametric counterparts when the assumption of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ cannot be met. However, one should not assume that these statistics are a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ alternative than parametric methods.

### Advantages of Nonparametric Methods Over Parametric Methods

1. Nonparametric methods can be used to test population parameters when the variable is \_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ distributed.
2. Nonparametric methods can be used when data are \_\_\_\_\_\_\_\_\_\_\_\_ or \_\_\_\_\_\_\_\_\_\_\_\_\_.
3. Nonparametric methods can be used to test hypotheses that do not involve population \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
4. In some cases, the\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for nonparametric methods are easier than those for the parametric counterparts.
5. Nonparametric methods are \_\_\_\_\_\_\_\_\_\_\_\_\_ to understand.
6. There are fewer \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ that have to be met, and the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ are easier to verify, for nonparametric methods.

### Disadvantages of Nonparametric Statistics

1. Nonparametric statistics are \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ than their parametric counterparts when the assumptions of the parametric methods are met. Therefore, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ differences are needed before the null hypothesis can be rejected.
2. Nonparametric statistics tend to use \_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ than the parametric tests. For example, the sign test only requires determining \_\_\_\_\_\_\_\_\_\_\_\_\_\_ data values are above or below the median, without needing to know \_\_\_\_\_\_ \_\_\_\_\_\_\_\_ above or below the median each value is.
3. Nonparametric statistics are \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ than their parametric counterparts when the assumptions of the parametric methods are met. Thus, \_\_\_\_\_\_\_\_\_\_\_\_\_ sample sizes are needed to overcome the loss of information.

If parametric assumptions can be met, the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ methods are preferred. When the parametric assumptions cannot be met, the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ methods provide a valuable tool for analyzing data.

### Assumptions for Nonparametric Statistics

1. The sample or samples are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ selected.
2. If two or more samples are used, they must be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of each other unless otherwise stated.

### Ranking

Many nonparametric tests involve the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ of data. Position a data value in a data array according to a rating scale.

For example, suppose teacher rates students’ speeches on a scale from 1 to 10, with 1 being the best and 10 being the worst. Two students could earn the same rating. The ratings given for two different groups of students are shown:

Students from Class 1:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Speaker** | A | B | C | D | E |
| **Rating** | 8 | 6 | 10 | 3 | 1 |

Students from Class 2:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Speaker** | G | H | J | K | L |
| **Rating** | 8 | 6 | 9 | 6 | 5 |

Each of the speakers are then \_\_\_\_\_\_\_\_\_\_\_ so that the student with the lowest score is ranked with 1 point, the student with the next highest score is ranked with 2 points. The student of the five with the highest score is ranked with 5 points.

The \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for the two classes follow:

**Students from Class 1:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Speaker** | E | D | B | A | C |
| **Rating** | 1 | 3 | 6 | 8 | 10 |
| ***Ranking*** |  |  |  |  |  |

When two or more students have the same rating, they are tied and the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of the ranks they are tied for are awarded.

**Students from Class 2:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Speaker** | L | H | K | G | J |
| **Rating** | 5 | 6 | 6 | 8 | 9 |
| ***Ranking*** |  |  |  |  |  |

### Example 13-1. Rank data

Rank the following data: 25, 68, 36, 63, 36, 74, 39

*Solution:*

Data: 25 68 39 63 36 74 39

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ordered: |  |  |  |  |  |  |  |
| Rank: |  |  |  |  |  |  |  |

# 13 – 2. The Sign Test

## Objective 2. Test Hypotheses, Using the Sign Test.

### Single-Sample Sign Test

The sign test for a single sample is a nonparametric test used to test the value of a population median.

### The hypothesis is that the \_\_\_\_\_\_\_\_\_\_\_\_\_ is a specific value of a population. For a random set of data, a plus sign is assigned to a value above the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_, a minus sign is assigned to a value below the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_, and 0 is assigned to a value equal to the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_. The numbers of plus and minus signs are compared to determine if those numbers are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ different. The null hypothesis is rejected when there is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ number of either plus or minus signs.

### Test Value for the Sign Test

If , the test value is the smaller number of plus or minus signs.

When , the test value is

where *X* is the smaller number of plus signs and *n* is the total number of plus or minus signs.

When , use Table J in Appendix A to find the critical value.

When , the normal approximation with Table E can be used for the critical values.

### Performing the Sign Test

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value. Use Table J for and Table E for .

**Step 3** Compute the test value.

**Step 4** Make the decision.

**Step 5** Summarize the result.

### Example 13-2. Death Due to Severe Weather

A meteorologist suggests that the median number of deaths per year from tornadoes in the United States is 60. The number of deaths for a randomly selected sample of 11 years is shown.

53 39 39 67 69 40 25 33 30 130 94

At , is there enough evidence to reject the claim? If you took proper safety precautions during a tornado, would you feel relatively safe?

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0*: The median number of deaths each year is 60.

*H1*: The median number of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

**Step 2** Find the critical value. Use Table J for and Table E for .

For , and , from Table J, the critical value is \_\_\_\_\_.

**Step 3** Compute the test value.

List a minus sign for each value below the median and a plus sign for each value above the median. Count the minus signs and the plus signs.

There are \_\_\_\_\_ minus signs and there are \_\_\_\_\_ plus signs.

The test value is \_\_\_\_\_ which is \_\_\_\_\_\_\_\_\_\_ than the critical value.

**Step 4** Make the decision.

The decision is to \_\_\_\_\_\_\_\_\_\_\_\_\_\_ the null hypothesis.

**Step 5** Summarize the result.

There (is / is not) enough evidence to conclude that the median number of deaths each year is not 60. There is enough evidence to \_\_\_\_\_\_\_\_\_\_ the meteorologist’s claim.

### Example 13-3. Wave Heights

An oceanographer wishes to test the claim that the median height of waves in a resort town on the Atlantic Ocean is 2.4 feet. A random sample of 50 days shows the heights of the waves on 20 days were at least 2.4 feet. At , test the claim that the median height of the waves is at least 2.4 feet.

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0*:

*H1*:

**Step 2** Find the critical value. Use Table J for and Table E for .

**Step 3** Compute the test value.

The number of plus signs is 20 and the number of minus signs is 30. .

**Step 4** Make the decision.

The test value is \_\_\_\_\_\_\_\_\_\_\_\_\_ than the critical value, so the decision is to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ the null hypothesis.

**Step 5** Summarize the result.

### Paired-Sample Sign Test

The paired-sample sign test is a nonparametric test that is used to test the difference between two population medians when the samples are dependent.

The test can be left-tailed, right-tailed, or two-tailed. The paired data values are each subtracted, , and a plus or minus sign given to each answer. Zeros are ignored.

If the number of plus signs is approximately equal to the number of minus signs, then the null hypothesis is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. If the difference in the number of plus and minus signs is significant, then the null hypothesis is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

### Two Assumptions for the Paired-Sign Test

1. The sample is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
2. The variables are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ or \_\_\_\_\_\_\_\_\_\_\_\_\_\_.

### Example 13-4. Exam Scores

A statistics professor wants to investigate the relationship between a student’s midterm examination score and the score on the final. Eight students were randomly selected, and their scores on the two examinations are noted.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| **Midterm** | 75 | 92 | 68 | 85 | 65 | 80 | 75 | 80 |
| **Final** | 82 | 90 | 79 | 95 | 70 | 83 | 72 | 79 |

At the 0.10 level of significance is there sufficient evidence to conclude that there is a difference in scores?

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*Ho*: The scores on the Midterm and Final will be the same.

*H1*: The scores on the Midterm and Final will be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

**Step 2** Find the critical value. Use Table J for and Table E for .

, , the critical value is \_\_\_\_\_\_\_.

**Step 3** Compute the test value.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| **Midterm** | 75 | 92 | 68 | 85 | 65 | 80 | 75 | 80 |
| **Final** | 82 | 90 | 79 | 95 | 70 | 83 | 72 | 79 |
| **Sign of difference** |  |  |  |  |  |  |  |  |

, there are 5 minus signs and 3 plus signs. The test value is 3.

**Step 4** Make the decision.

Compare the test value \_\_\_\_ with the critical value \_\_\_\_\_. The test value is \_\_\_\_\_\_\_\_\_\_\_\_\_\_ than the critical value. The decision is to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ the null hypothesis.

*Note: The null hypothesis is rejected when the test value is less than or equal to the critical value and it is not rejected when the test value is greater than the critical value.*

**Step 5** Summarize the result.

There (is / is not) enough evidence to conclude that the scores on the Midterm and Final will be different.

# 13 – 3. The Wilcoxon Rank Sum Test

## Objective 3. Test Hypotheses, Using the Wilcoxon Rank Sum Test.

The Wilcoxon tests consider differences in \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ that the sign test do not consider. The Wilcoxon rank sum test is used for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ samples and the Wilcoxon signed-rank test is used for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ samples. The parametric equivalents, the *z* and *t* tests, assume the samples are from populations that are approximately normally distributed, whereas the Wilcoxon tests do not make these assumptions.

### Wilcoxon Rank Sum Test

The Wilcoxon rank sum test is a nonparametric test that uses ranks to determine if two \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ samples were selected from populations that have the same distributions.

In the Wilcoxon rank sum test, the values of the data for both samples are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and then \_\_\_\_\_\_\_\_\_\_\_\_\_. If the values in each sample should be ranked approximately the same, meaning there is no difference in the population distributions, then the null hypothesis is not rejected. That is, the difference in the sums of the ranks are approximately equal. If there is a large difference in the sums of the ranks, then it is concluded that the distributions are not the same and the null hypothesis is rejected.

### Assumptions for the Wilcoxon Rank Sum Test

1. The samples are random and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of one another.
2. The size of each sample must be greater than or equal to \_\_\_\_\_\_\_\_.

### Formula for the Wilcoxon Rank Sum Test When Samples Are Independent

where

*R* = sum of \_\_\_\_\_\_\_\_\_\_ for smaller sample size ()

= \_\_\_\_\_\_\_\_\_\_\_\_\_\_ of the sample sizes

= \_\_\_\_\_\_\_\_\_\_\_\_\_\_ of the sample sizes

and

Note that if both samples are the same size, either size can be used as .

Table E is used for critical values.

### Wilcoxon Rank Sum Test

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value(s) using Table E.

**Step 3** Compute the test value.

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ the data from the two samples, arrange the combined data in \_\_\_\_\_\_\_\_\_\_\_\_\_, and \_\_\_\_\_\_\_\_\_\_\_ each value.
2. \_\_\_\_\_\_\_\_\_\_ the ranks of the group with the\_\_\_\_\_\_\_\_\_\_\_\_\_ sample size. (Note: If both samples are the same size, either size can be used as .)
3. Use the formulas to find the test value.

Where *R* is the sum of the \_\_\_\_\_\_\_\_\_ of the data in the \_\_\_\_\_\_\_\_\_\_\_ sample and and are each greater than or equal to \_\_\_\_\_\_\_\_.

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 13-5. Lifetimes of Handheld Video Games

To test the claim that there is no difference in the lifetimes of two brands of handheld video games, a researcher selects a random sample of 11 video games of each brand. The lifetimes (in months) of each brand are shown.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Brand A** | 42 | 34 | 39 | 42 | 22 | 47 | 51 | 34 | 41 | 39 | 28 |
| **Brand B** | 29 | 39 | 38 | 43 | 45 | 49 | 53 | 38 | 44 | 43 | 32 |

At , can the researcher conclude that there is a difference in the distributions of lifetimes for the two brands?

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0:* There is \_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in lifetimes of the two brands of handheld video games.

*H0:* There is \_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in lifetimes of the two brands of handheld video games.

**Step 2** Find the critical value. Since and this test is two-tailed, use Table E.

The critical value \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 3** Compute the test value.

1. Combine the data from the two samples, arrange the combined data in order, and rank each value.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Lifetime** | 22 | 28 | 29 | 32 | 34 | 34 | 38 | 38 | 39 | 39 | 39 |
| **Group** | A | A | B | B | A | A | B | B | A | A | B |
| **Rank** |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Lifetime** | 41 | 42 | 42 | 43 | 43 | 44 | 45 | 47 | 49 | 51 | 53 |
| **Group** | A | A | A | B | B | B | B | A | B | A | B |
| **Rank** |  |  |  |  |  |  |  |  |  |  |  |

1. Sum the ranks of the group with the smaller sample size. (Note: If both samples are the same size, either size can be used as .)  
   Group A and Group B are the same size. Use Group A.
2. Use the formulas to find the test value.

**Step 4** Make the decision.

\_\_\_\_\_\_\_\_\_ \_\_\_\_\_ \_\_\_\_\_\_\_\_\_, so the decision is to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ the null hypothesis.

**Step 5** Summarize the results.

# 13 – 4. The Wilcoxon Signed-Rank Test

## Objective 4. Test Hypotheses, Using the Wilcoxon Signed-Rank Test.

### Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a nonparametric test used to test whether two dependent samples have been selected from two populations having the same distributions.

### Assumptions

1. The \_\_\_\_\_\_\_\_\_\_\_\_ data have been obtained from a \_\_\_\_\_\_\_\_\_\_sample.
2. The population of differences has a distribution that is approximately \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

### Wilcoxon Signed-Rank Test Steps

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical value from Table K when and from Table E when .

**Step 3** Compute the test value.

When :

1. Make a table as shown:

| **Before**  ***XB*** | **After**  ***XA*** | **Difference**  ***= XB - XA*** | **Absolute**  **Value** | **Rank** | **Signed**  **Rank** |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |

1. Find the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (before – after), denoted by , and place the values in the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ column.
2. Find the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of each difference, and place the results in the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ column.
3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_ each absolute value from lowest to highest, and place the rankings in the \_\_\_\_\_\_\_\_\_\_\_ column.
4. Give each rank a positive or negative \_\_\_\_\_\_\_\_\_\_\_\_\_, according to the sign in the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ column.
5. Find the \_\_\_\_\_\_\_\_ of the positive ranks and the \_\_\_\_\_\_\_\_ of the negative ranks separately.
6. Select the \_\_\_\_\_\_\_\_\_\_\_\_\_ of the absolute values of the sums, and use this absolute value as the test value .  
   When , use Table E and the test value   
   where  
   *n* = number of pairs where difference is not 0  
   = \_\_\_\_\_\_\_\_\_\_\_\_\_\_ sum in absolute value of signed ranks

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 13-6. Bowling Scores

Eight randomly selected volunteers at a bowling alley were asked to bowl three games and pick their best score. They were then given a bowling ball made of a new composite material and were allowed to practice with the ball as much as they wanted. The next day they each bowled three games with the new ball and picked their best score.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Bowler** | A | B | C | D | E | F | G | H |
| **Day 1** | 141 | 176 | 178 | 174 | 135 | 190 | 182 | 141 |
| **Day 2** | 158 | 144 | 135 | 153 | 195 | 151 | 151 | 183 |

At the 0.05 level of significance, did the scores improve?

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0:* There is

*H1:* There is

**Step 2** Find the critical value. (Use Table K because .)

This is a one tailed test with , and , so the critical value is \_\_\_\_\_\_.

**Step 3** Find the test value.

| **Before**  ***XB*** | **After**  ***XA*** | **Difference**  ***XB - XA*** | **Absolute**  **Value |D|** | **Rank** | **Signed**  **Rank** |
| --- | --- | --- | --- | --- | --- |
| 141  176  178  174  135  190  182  141 | 158  144  135  153  195  151  151  183 |  |  |  |  |

Positive rank sum:

Negative rank sum:

\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4** Make the decision.

The test value, \_\_\_\_\_\_\_, is not \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ to the critical value of \_\_\_\_. Therefore, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_the null hypothesis.

**Step 5** Summarize the result.

There (is / is not) enough evidence to support the claim that the bowling scores were higher when the bowlers used the new bowling ball.

# 13 – 5. The Kruskal-Wallis Test

## Objective 5. Test Hypotheses, Using the Kruskal-Wallis Test.

### The Kruskal-Wallis test

The Kruskal-Wallis test is a nonparametric test that is used to determine whether three or more samples came from populations with the same distributions.

### Assumptions for the Kruskal-Wallis Test

1. There are at least three \_\_\_\_\_\_\_\_\_\_\_\_\_\_ samples.
2. The size of each sample must be at least \_\_\_\_\_\_\_\_\_.

### Formula for the Kruskal-Wallis Test

where

sum of ranks of sample 1

size of sample 1

sum of ranks of sample 2

size of sample 2

sum of ranks of sample *k*

size of sample *k*

number of samples

### Kruskal-Wallis Test

**Step 1** State the hypothesis and identify the claim.

**Step 2** Find the critical value. Use the chi-square table, Table G,   
 with ( number of samples)

**Step 3** Compute the data value. (See the formula above)

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 13-7. Job Offers for Chemical Engineers

A recent study recorded the number of job offers received by randomly selected, newly graduated chemical engineers at three colleges. The data are shown here:

| **College A** | **College B** | **College C** |
| --- | --- | --- |
| 6 | 2 | 10 |
| 8 | 1 | 12 |
| 7 | 0 | 9 |
| 5 | 3 | 13 |
| 6 | 6 | 4 |

At , is there a difference in the average number of job offers received by the graduates at the three colleges?

*Solution:*

**Step 1** State the hypothesis and identify the claim.

*H0:*

*H1:*

**Step 2** Find the critical value. Use the chi-square table, Table G,   
 with ( number of samples)

With , and d.f. = \_\_\_\_\_\_\_\_\_ the critical value is \_\_\_\_\_\_\_\_\_.

**Step 3** Compute the data value. (See the formula above)

| **Amount** | **Group** | **Rank** |
| --- | --- | --- |
| 0  1  2  3  4  5  6  6  6  7  8  9  10  12  13 |  |  |

The sum of the ranks of the groups:

Group A: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Group B: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Group C: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

,

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4** Make the decision.

**Step 5** Summarize the results.

# 13 – 6. The Spearman Rank Correlation Coefficient and the Runs Test

## Objective 6. Compute the Spearman Rank Correlation Coefficient.

Correlation and regression methods assume that the population from which the samples are obtained are \_\_\_\_\_\_\_\_\_\_\_\_\_\_ distributed. However, this requirement \_\_\_\_\_\_\_\_\_\_\_\_ always be met. The nonparametric \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of using the Pearson product moment correlation coefficient is the Spearman rank correlation coefficient which can be used when data are ranked.

Each set of data are ranked. The difference in ranks is found and the coefficient is computed using the differences. If the both sets of data have the same ranks, the test statistic will be +1. If both sets of data are ranked in exactly the opposite way, the test statistic will be . If there is no relationship between the rankings, the test statistic will be near 0.

### Spearman Rank Correlation Coefficient

The Spearman rank correlation coefficient is a nonparametric statistic that uses ranks to determine if there is a relationship between two variables.

### Assumptions for Spearman’s Rank Correlation Coefficient

1. The sample is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_ sample.
2. The data consist of \_\_\_\_\_\_\_\_\_ measurements or observations taken on the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ individual.

### Formula for Computing the Spearman Rank Correlation Coefficient

If there are no ties in ranks:

where

difference in ranks

number of data pairs

### Spearman’s Rank Correlation Coefficient

**Step 1** State the hypothesis.

**Step 2** Find the critical value.

**Step 3** Find the test value.

1. Rank each data set.
2. Subtract the rankings for each individual.
3. Square the differences.
4. Find the sum of the squares.
5. Substitute in the formula for .

**Step 4** Make the decision.

**Step 5** Summarize the results.

Use Table L for critical value for all up to 30 and the normal distribution for any

If there are ties in the ranks, use the formula previously shown in Chapter 10:

The hypotheses are:

*H0*: (There is not a linear relationship …)

*H1*: (There is a linear relationship …)

where ρ is the population correlation coefficient.

### Example 13-8. Subway and Commuter Rail Passengers

Six cities are randomly selected, and the number of daily passenger trips (in thousands) for subways and commuter rail service is obtained.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **City** | 1 | 2 | 3 | 4 | 5 | 6 |
| **Subway** | 845 | 494 | 425 | 313 | 108 | 41 |
| **Rail** | 39 | 291 | 142 | 103 | 33 | 38 |

At , is there a relationship between the variables?

*Solution*:

**Step 1** State the hypothesis.

*H0*:

*H1*:

**Step 2** Find the critical value.

and , so the critical value we find on Table L is \_\_\_\_\_\_\_\_\_\_\_\_.

**Step 3** Find the test value.

1. Rank each data set separately.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **City** | 1 | 2 | 3 | 4 | 5 | 6 |
| **Subway** | 845 | 494 | 425 | 313 | 108 | 41 |
| ***Rank*** |  |  |  |  |  |  |
| **Rail** | 39 | 291 | 142 | 103 | 33 | 38 |
| ***Rank*** |  |  |  |  |  |  |

1. Subtract the rankings for each individual.

| **City** | **Subway**  **Rank** | **Rail**  **Rank** | **Difference** |
| --- | --- | --- | --- |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |
| 5 |  |  |  |
| 6 |  |  |  |

1. Square the differences.
2. Find the sum of the squares.
3. Substitute in the formula for .

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4** Make the decision.

**Step 5** Summarize the results.

## Objective 7. Test Hypotheses, Using the Runs Test.

One way to determine if the data obtained from a sample are actually \_\_\_\_\_\_\_\_\_\_\_\_\_\_

is to use the runs test.

### Definition: Run

A run is a succession of identical letters preceded or followed by a different letter or no letter at all, such as the beginning or end of a succession.

### Find the number of runs for nominal data:

The data are:

FMMFFMFMMFFMMMFFFFFM

Group 1 are F and Group 2 are M.

The number in Group 1 is 11. The number in Group 2 is 9.

represents the number of runs.

1. F

2. MM

3. FF

4. M

5. F

6. MM

7. FF

8. MMM

9. FFFFF

10. M

There are 10 runs, that is, .

### Runs Test for Randomness

The runs test for randomness is a nonparametric test that is used to determine if a \_\_\_\_\_\_\_\_\_\_\_\_\_\_ of data values occurs at random.

### Assumptions for the Runs Test for Randomness

1. The data from the sample are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in the order in which they were selected.
2. Each letter, number, or event can be \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ into one or two mutually exclusive categories.

### Formulas for the Test Statistic Value for the Runs Test

When and , use the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, as the test statistic value.

When and , or when and , use

where

### The Runs Test

**Step 1** State the hypotheses and identify the claim.

**Step 2** Find the critical values.

1. When and , use Table M.
2. When or , or when and , use Table E.

**Step 3** Find the test value.

1. When and , use \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
2. When or , or when and , use \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 13-9. Exam Scores

An instructor wishes to see whether grades of students who finish an exam occur at random. Show here are the grades of 30 students in the order that they finished the exam (from left to right across each row, then proceed to the next row).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 87 | 93 | 82 | 77 | 64 | 98 | 100 | 93 |
| 88 | 65 | 72 | 73 | 56 | 63 | 85 | 92 |
| 95 | 91 | 88 | 63 | 72 | 79 | 55 | 53 |
| 65 | 68 | 54 | 71 | 73 | 72 |  |  |

Test for randomness, at .

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0*:

*H1*:

**Step 2** Find the critical values.

First determine the \_\_\_\_\_\_\_\_\_\_\_ of each sample.

Find the median of the data set. The median = ­­­­­­\_\_\_\_\_\_\_

Determine the number of values above the median and the number of values below the median. Ignore any values that are equal to the median:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 87(­­\_\_) | 93(\_\_) | 82(\_\_) | 77(\_\_) | 64(\_\_) | 98(\_\_) | 100(\_\_) | 93(\_\_) |
| 88(\_\_) | 65(\_\_) | 72(\_\_) | 73(\_\_) | 56(\_\_) | 63(\_\_) | 85(\_\_) | 92(\_\_) |
| 95(\_\_) | 91(\_\_) | 88(\_\_) | 63(\_\_) | 72(\_\_) | 79(\_\_) | 55(\_\_) | 53(\_\_) |
| 65(\_\_) | 68(\_\_) | 54(\_\_) | 71(\_\_) | 73(\_\_) | 72(\_\_) |  |  |

There are \_\_\_ data values above the median (A) and \_\_\_ data values below the median (B). \_\_\_ values are equal to the median and are not included.

Thus, \_\_\_\_ and \_\_\_\_.

Since and , we will use Table M.

The critical values of are \_\_\_ and \_\_\_. The null hypothesis will not be rejected when *G* falls between \_\_\_\_ and \_\_\_\_.

**Step 3** Find the test value.

When and , use the number of runs denoted by *.*

The runs are:

1. AAAA

2. B

3.

4.

5.

6.

7.

8.

There are \_\_\_\_\_ runs.

The test value is \_\_\_\_.

**Step 4** Make the decision.

**Step 5** Summarize the results.

### Example 13-10. Speeding Tickets

A police chief records the gender of the drivers who receive speeding tickets.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| M | M | M | F | F | M | F | M | F | M |
| M | F | M | M | M | F | M | M | F | F |
| F | M | M | F | M | M | F | M | M | M |
| M | M | F | M | F | F | F | M | M | M |
| F | F | M | F | F | F | M | M | M | M |

Test the claim at that the gender of the ticketed drivers is random.

*Solution:*

**Step 1** State the hypotheses and identify the claim.

*H0*: Gender of drivers who receive tickets does not occurs randomly.

*H1*:

**Step 2** Find the critical values.

The number of males who receive tickets ( = \_\_\_\_\_

The number of females who receive tickets (= \_\_\_\_\_

When and , use Table E.

At , the critical values are .

**Step 3** Find the test value.

The number of males who receive tickets ( = \_\_\_\_\_

The number of females who receive tickets (= \_\_\_\_\_

The number of runs is

1. MMM
2. FF
3. M
4. F
5. M
6. F
7. MM
8. F
9. MMM

The number of runs is \_\_\_\_\_\_\_

When and , or when and , use

where

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Step 4** Make the decision.

Since the test value of (is / is not) between the critical values of , the null hypothesis (is / is not) rejected.

**Step 5** Summarize the results.