Chapter 256

Analysis of Runs

Introduction

This procedure computes summary statistics and common non-parametric, single-sample runs tests for a series of *n* numeric, binary, or categorical data values. For numeric data, the exact and asymptotic *Wald-Wolfowitz Runs Tests for Randomness* are computed based on the number of runs above and below a reference value, along with exact and asymptotic *Runs Tests for Serial Randomness* based on the number of runs up and down. For binary data, the exact and asymptotic Wald-Wolfowitz Runs Tests for Randomness are computed based on the number of runs in each category. For categorical data, an asymptotic, *k*-category extension of the Wald-Wolfowitz Runs Test for Randomness is computed. Although most often used to test for randomness, these tests can also be used as goodness-of-fit tests.

The results in this procedure are based on the formulas given in chapters 11 and 12 of Bradley (1968) and chapter 10 of Sheskin (2011).

Experimental Design and Data Structure

A typical design for this scenario involves *n* numeric, binary, or categorical observations in a single column. All of the observations in the column must be of the same type (i.e. all numeric, all binary, or all categorical). Binary and categorical values may be text or numeric. Missing values are discarded and ignored.

Typical data might appear as

Numeric	Binary	Categorical
31	1	A
23	1	A
36	1	B
43	1	C
51	0	C
44	0	B
12	0	A
26	0	A
43	1	B
75	1	B
2	0	C
3	1	C

Technical Details

All details that follow assume a series of n non-missing data values of the same data type: numeric, binary, or categorical. The summaries and tests computed depend on the data type. For categorical data, we assume that there are k categories.

Computing Runs

The following sections describe how runs are calculated for binary, categorical, and numeric data.

Binary Data

In the case of binary data consisting of two distinct categories, a run is defined as a sequence where a single value is repeated one or more times. A new run occurs each time the data value changes. For example, in the following binary data series consisting of $n = 16$ values

$$0\ 0\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 1\ 1$$

there are two runs of 0's with length = 1 and two runs of 0's with length = 2. Similarly, there are two runs of 1's with length = 1, one run of 1's with length = 3, and one run of 1's with length = 5. In total, there are 4 runs of 0's and 4 runs of 1's, for a grand total of 8 distinct runs in all. This sequence of runs can be tested for randomness using the Wald-Wolfowitz Runs Test.

Categorical Data

For data consisting of k distinct categories, a run is defined as a sequence where a single value is repeated one or more times. A new run occurs each time the data value switches. For example, in the following categorical data series consisting of $k = 3$ categories and $n = 16$ values

$$A\ A\ B\ A\ C\ C\ C\ A\ B\ B\ B\ C\ A\ A\ C\ B$$

there are two runs of A's with length = 1 and two runs of A's with length = 2. There are two runs of B's with length = 1 and one run of B's with length = 3. Finally, there are two runs of C's with length = 1 and one run of C's with length = 3. In total, there are 4 runs of A's, 3 runs of B's, and 3 runs of C's for a grand total of 10 distinct runs in all. This sequence of runs can be tested for randomness using the k-Category extension of the Wald-Wolfowitz Runs Test.

Numeric Data

For numeric data, two different kinds of runs can be computed:

1. Runs above and below a reference value

2. Runs up and down

The runs above and below a reference value are used in the Wald-Wolfowitz test, while the runs up and down are used in the computing the runs test for serial randomness. We'll now describe how these runs are calculated.

Runs Above and Below a Reference Value

For numeric data, a reference value is used to determine the runs in the dataset relative to the reference. The reference value can be set as the mean, median, or mode of the data or as any custom user-defined value. The reference value is used to create a binary series from the numeric data by assigning a "1" to values above the reference and a "0" to values below the reference. For example, for the following numeric data series consisting of $n = 16$ values with a median of 28.5 used as the reference value

$$31\ 23\ 36\ 43\ 51\ 44\ 12\ 26\ 43\ 75\ 2\ 3\ 15\ 18\ 78\ 24$$

Analysis of Runs

we can rewrite the series as a sequence of binary values, indicating whether each raw numeric data value is above or below the reference as

$$1\ 0\ 1\ 1\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 1\ 0$$

The counting then proceeds just as in the case of binary data. In this example, there are three runs of single values below the reference (indicated as single 0's) and one run of four values below the reference (indicated as a string of four 0's). Similarly, there are two runs of single values (indicated as single 1's), one run of three (indicated as a string of three 1's), and one run of four values (indicated as a string of four 1's) above the reference. In all, there are four runs above and four runs below the reference for a grand total of 8 distinct runs. This sequence of runs can be tested for randomness using the Wald-Wolfowitz Runs Test.

Values Equal to the Reference

When a value is exactly equal to the reference, a decision must be made whether to count it as "above" or "below" or whether to skip the value and discard it from the analysis. In **NCSS**, a fourth option is given which allows you to count the value the same as the previous non-reference value was counted, allowing the current run to continue uninterrupted by values on the reference line. For example, if you have a simple numeric data series with $n = 7$ values and a median of 48 used as the reference value

$$47\ 44\ 49\ 48\ 50\ 51\ 45$$

the fourth value in the series, 48, is exactly equal to the reference. The resulting binary sequence is

$$0\ 0\ 1\ ?\ 1\ 1\ 0$$

Without counting the value in question but leaving it in place, it appears that there are 2 runs above and two runs below the reference. If we count it as "above"

$$0\ 0\ 1\ 1\ 1\ 1\ 0$$

then there is one run above and two runs below for a total of three runs. If we count it as "below"

$$0\ 0\ 1\ 0\ 1\ 1\ 0$$

then there are two runs above and three runs below for a total of five runs. If we count it the same as the previous, then it is counted as "above" since the third value was above the reference

$$0\ 0\ 1\ 1\ 1\ 1\ 0$$

and the runs totals would be one above and two below for of total of three runs. If we skip it and discard it

$$0\ 0\ 1\ 1\ 1\ 0$$

then there is one run above and two runs below for a total of three runs, but the number of values in the sequence is reduced to $n = 6$ for subsequent tests.

As we have illustrated, the treatment of these values on the reference line can have an impact on the run totals and, therefore, on the resulting hypothesis tests, particularly in the case of small samples. If the values before and after the value in question are on different sides of the reference line, it won't matter whether you count it as "above" or "below" since the run counts won't change. In this case the choice is called "non-critical." On the other hand, if the values before and after are on the same side, the run counts will depend on the way this value is treated, as demonstrated in the example. A common practice is to see if the results change dramatically when these values are counted as "above" as then as "below". It is also not uncommon to skip these values altogether.

Analysis of Runs

Runs Up and Down

To compute the runs up and down for n numeric data values, the sign of the difference between each value and the one previous is recorded and used to create a binary series of $n - 1$ signs. If the values are tied, then a zero is recorded. For example, for the following numeric data series consisting of $n = 10$ values

<p align="center">44 40 33 26 27 44 46 54 50 51</p>

the binary sequence of $n - 1 = 9$ difference signs is

<p align="center">– – – + + + + – +</p>

Runs up and down are counted as sequences of the same sign. In this example there are two runs down (one with length = 1 and one with length = 3) and two runs up (one with length = 1 and one with length = 4) for a total of 4 distinct runs up and down. This sequence of runs up and down can be tested for randomness using the Runs Test for Serial Randomness.

Tied Values

When two values are tied, a decision must be made whether to count the zero as "up (+)" or "down (–)" or whether to skip it and discard it from the analysis. In **NCSS**, a fourth option is given which allows you to count the zero the same as the previous sign was counted, allowing the current run to continue uninterrupted by tied values. For example, if the data series above had a 50 at the second to last position with $n = 11$ values

<p align="center">44 40 33 26 27 44 46 54 50 50 51</p>

then the second- and third-to-last values are tied and the binary sequence of difference signs has a 0 in the second-to-last position

<p align="center">– – – + + + + – 0 +</p>

Without counting the zero in question but leaving it in place, it appears that there are 2 runs up and two runs down. If we count it as "up"

<p align="center">– – – + + + + – + +</p>

then there are still two runs up and two runs down for a total of four runs. If we count it as "down"

<p align="center">– – – + + + + – – +</p>

then there are still two runs up and two runs down for a total of four runs. This is called a "non-critical tie" since the runs are not influenced by the choice. If we count it the same as the previous, then it is counted as "down" since the previous direction was down

<p align="center">– – – + + + + – – +</p>

and there are still two runs up and two runs down for a total of four runs. If we skip it and discard it

<p align="center">– – – + + + + – +</p>

then there are still two runs up and two runs down for a total of four runs, but the number if signs is reduced to 9 instead of 10. This will influence test results.

Although not demonstrated by this example, the treatment of ties can have an impact on the run totals and, therefore, on the resulting hypothesis tests, particularly in the case of small samples. If the signs before and after the tied values are different (as was the case in this example) it won't matter whether you count the tie as "up" or "down" since the run counts won't change. This is called a "non-critical tie." On the other hand, if the signs before and after the tie are the same, the run counts will depend on the way the tie is treated. A common practice is to see if the results change dramatically when ties are counted as "up" as then as "down". It is also not uncommon to skip ties altogether and exclude them from runs tests.

Single-Sample Runs Test for Randomness (Wald-Wolfowitz Runs Test)

For n binary outcomes, Y_i ($= 0$ or 1), the non-parametric, Single-Sample Wald-Wolfowitz Runs Test can be used to test the null hypothesis that the binary series is random. The null hypothesis of randomness is rejected when the total number of runs of 0's and 1's is too many or too few. This test can be used for binary data and for numeric data that has been categorized as either above or below a reference value.

Given the number of 0's (n_1) and 1's (n_2) in a series, the minimum and maximum possible total number of runs are

$$R_{min} = 2$$
$$R_{max} = 2\text{Min}(n_1, n_2) + 1$$

The expected number of runs, $E(R)$, given n_1 and n_2 is

$$E(R) = \frac{2n_1 n_2}{n} + 1$$

Exact Test

To compute the exact test for a binary series, we must calculate the probabilities of obtaining given numbers of runs conditional on the number of 0's (n_1) and 1's (n_2) in the series. If the data are random, the probability that the total number if runs, R, will be equal to some <u>even number</u> $2u$ is

$$P(R = 2u) = \frac{2 \binom{n_1 - 1}{u - 1}\binom{n_2 - 1}{u - 1}}{\binom{n}{n_1}}.$$

The probability that the total number if runs, R, will be equal to some <u>odd number</u> $2u + 1$ is

$$P(R = 2u + 1) = \frac{\binom{n_1 - 1}{u - 1}\binom{n_2 - 1}{u} + \binom{n_1 - 1}{u}\binom{n_2 - 1}{u - 1}}{\binom{n}{n_1}}.$$

For the alternative hypothesis of "too many runs", the exact upper one-sided p-value for an observed total number of runs, r, is

$$P(R \geq r) = \sum_{v=r}^{R_{max}} P(R = v)$$

For the alternative hypothesis of "too few runs", the exact lower one-sided p-value for an observed total number of runs, r, is

$$P(R \leq r) = \sum_{v=R_{min}}^{r} P(R = v)$$

The two-sided exact p-value for an observed total number of runs, r, is

$$P(|R - E(R)| \geq |r - E(R)|) = \sum_{v=R_{min}}^{E(R)-|r-E(R)|} P(R = v) + \sum_{v=E(R)+|r-E(R)|}^{R_{max}} P(R = v)$$

The exact test is more accurate than the two asymptotic z tests and should always be used if available. The exact test is not computed for categorical data with more than two groups.

Asymptotic Z Test

For large n, the asymptotic standard normal z statistic is computed from the observed total number of runs, r, as

$$z = \frac{r - \mu_r}{\sigma_r}$$

where

$$\mu_r = E(R) = \frac{2n_1 n_2}{n} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n-1)}}$$

This test is less accurate than the exact test and usually less accurate than the continuity-corrected z test.

Asymptotic Z Test with Continuity Correction

For large n, the asymptotic continuity-corrected standard normal z statistic is computed from the observed total number of runs, r, as

$$z_{cc} = \begin{cases} \dfrac{r - \mu_r - 0.5}{\sigma_r} & \text{if } r \geq \mu_r \\ \dfrac{r - \mu_r + 0.5}{\sigma_r} & \text{if } r < \mu_r \end{cases}$$

where

$$\mu_r = E(R) = \frac{2n_1 n_2}{n} + 1$$

$$\sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n)}{n^2 (n-1)}}$$

This test is less accurate than the exact test but usually more accurate than the regular z test.

k-Category Extension of the Single-Sample Runs Test for Randomness

For n categorical outcomes, Y_i, the non-parametric k-Category Single-Sample Runs Test can be used to test the null hypothesis that the series is random. The null hypothesis of randomness is rejected when the total number of runs is too many or too few. This test is used for categorical data.

The minimum possible total number of runs given $n_j, j = 1, \ldots, k$, is

$$R_{min} = k$$

The maximum possible total number of runs is calculated using an algorithm that counts the total possible arrangements given n_j.

The expected number of runs, $E_k(R)$, given the n_j's is

$$E_k(R) = \frac{n(n+1) - \sum n_j^2}{n}$$

Asymptotic Z Test

For large n, the asymptotic standard normal z statistic is computed from the observed total number of runs, r, as

$$z = \frac{r - \mu_r}{\sigma_r}$$

where

$$\mu_r = E_k(R) = \frac{n(n+1) - \sum n_i^2}{n}$$

$$\sigma_r = \sqrt{\frac{\sum n_i^2 [\sum n_i^2 + n(n+1)] - 2n \sum n_i^3 - n^3}{n^2(n-1)}}$$

This test is usually less accurate than the continuity-corrected z test.

Asymptotic Z Test with Continuity Correction

For large n, the asymptotic continuity-corrected standard normal z statistic is computed from the observed total number of runs, r, as

$$z_{cc} = \begin{cases} \dfrac{r - \mu_r - 0.5}{\sigma_r} & \text{if } r \geq \mu_r \\[2ex] \dfrac{r - \mu_r + 0.5}{\sigma_r} & \text{if } r < \mu_r \end{cases}$$

where

$$\mu_r = E_k(R) = \frac{n(n+1) - \sum n_i^2}{n}$$

$$\sigma_r = \sqrt{\frac{\sum n_i^2 [\sum n_i^2 + n(n+1)] - 2n \sum n_i^3 - n^3}{n^2(n-1)}}$$

This test is usually more accurate than the regular z test.

Single-Sample Runs Test for Serial Randomness

For n numeric outcomes, Y_i, the non-parametric, Single-Sample Runs Test for Serial Randomness can be used to test the null hypothesis that the numeric series is random. The null hypothesis of randomness is rejected when the total number of runs up (+) and down (–) is too many or too few. This test is can be applied only to numeric data.

The minimum and maximum possible total number of runs are

$$R_{\pm min} = 1$$
$$R_{\pm max} = n - 1$$

The expected number of runs, $E(R_\pm)$, is

$$E(R_\pm) = \frac{2n - 1}{3}$$

Exact Test

The required probabilities for the exact runs test for serial randomness are given with precision of four decimal places in Table X on page 363 of Bradley (1968) for $n \leq 25$. **NCSS** simply uses these table values in the probability calculations for the exact test.

For the alternative hypothesis of "too many runs", the exact upper one-sided p-value for an observed total number of runs up and down, $r_\pm$, is

$$P(R_\pm \geq r_\pm) = \sum_{v=r_\pm}^{R_{\pm max}} P(R_\pm = v)$$

For the alternative hypothesis of "too few runs", the exact lower one-sided p-value for an observed total number of runs up and down, $r_\pm$, is

$$P(R_\pm \leq r_\pm) = \sum_{v=R_{\pm min}}^{r_\pm} P(R_\pm = v)$$

The two-sided exact p-value for an observed total number of runs up and down, $r_\pm$, is

$$P\big(\big|R_\pm - E(R_\pm)\big| \geq \big|r_\pm - E(R_\pm)\big|\big) = \sum_{v=R_{\pm min}}^{E(R_\pm)-|r_\pm-E(R_\pm)|} P(R_\pm = v) + \sum_{v=E(R_\pm)+|r_\pm-E(R_\pm)|}^{R_{\pm max}} P(R_\pm = v)$$

The exact test is only calculated for $n \leq 25$. For larger n an asymptotic z test should be used. The exact test is more accurate than the two asymptotic z tests and should always be used if available.

Asymptotic Z Test

For large n, the asymptotic standard normal z statistic is computed from the observed total number of runs, $r_\pm$, as

$$z = \frac{r_\pm - \mu_{r_\pm}}{\sigma_{r_\pm}}$$

where

$$\mu_{r_\pm} = E(R_\pm) = \frac{2n - 1}{3}$$

$$\sigma_{r_\pm} = \sqrt{\frac{16n - 29}{90}}$$

This test is less accurate than the exact test and usually less accurate than the continuity-corrected z test.

Asymptotic Z Test with Continuity Correction

For large n, the asymptotic continuity-corrected standard normal z statistic is computed from the observed total number of runs, $r_\pm$, as

$$z_{cc} = \begin{cases} \dfrac{r_\pm - \mu_{r_\pm} - 0.5}{\sigma_{r_\pm}} & \text{if } r_\pm \geq \mu_{r_\pm} \\ \dfrac{r_\pm - \mu_{r_\pm} + 0.5}{\sigma_{r_\pm}} & \text{if } r_\pm < \mu_{r_\pm} \end{cases}$$

where

$$\mu_{r_{\pm}} = E(R_{\pm}) = \frac{2n-1}{3}$$

$$\sigma_{r_{\pm}} = \sqrt{\frac{16n-29}{90}}$$

This test is less accurate than the exact test but usually more accurate than the regular z test.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies the variables used in the analysis and some options about the calculation of runs.

Input Data Type

This procedure can handle **Numeric**, **Binary**, and **Categorical** data. Select the type of data you want to analyze. This choice will affect which summaries and tests are output since some are available only for numeric data. All of the variables you choose should match the selected data type. If you want to analyze both numeric and categorical (or binary) data, you should split up the analysis into separate runs.

Variables

Variable(s)

Specify one or more columns of data for the analysis. If more than one column is specified, a separate analysis will be performed for each variable.

Reference Value (Displayed when Input Data Type = Numeric)

Select the reference value to use for computing runs from numeric data. Runs above and below this value are counted and used in corresponding tests.

The reference value is used to transform a numeric series into a binary series, indicating whether each numeric value is above or below the reference.

The options are

- **Mean**

  Use the data mean as the reference value.

- **Median**

  Use the data median as the reference value.

- **Mode**

  Use the data mode as the reference value.

- **User-Defined Value**

  Use a custom value for the reference. When this option is selected, you must enter a value in the box to the right.

User-Defined Reference Value (Displayed when Reference Value = User-Defined Value)

Enter a custom value to be used as the reference. This value must be numeric.

Counting Options for Runs Above and Below a Reference Value
(Displayed when Input Data Type = Numeric)

When a Data Value is Equal to the Reference

Specify how data values that are exactly equal to the reference value are treated when computing runs above and below the reference.

The options are

- **Count as Above**

  Values equal to the reference are counted as being "Above" the reference.

- **Count as Below**

  Values equal to the reference are counted as being "Below" the reference.

- **Count the Same as the Previous**

  Values equal to the reference are counted in the same direction (above or below) as the previous value in the series.

  For example, if observation #4 is above the reference and observation #5 is equal to the reference, then observation #5 would be counted as being above since observation #4, which immediately precedes value #5, is above. On the other hand, if observation #4 is below the reference, then observation #5 will also be counted as below.

  If the first value is equal to the reference, then it is skipped and not used when computing runs. All subsequent observations equal to the reference are also skipped until a non-reference value is encountered in the series. This would reduce the overall sample size used in the analyses of runs above and below the reference.

- **Skip the Value**

  Values equal to the reference are skipped and not used when computing runs above and below the reference. Selecting this option may reduce the overall sample size used in the analyses.

Counting Options for Runs Up and Down
(Displayed when Input Data Type = Numeric)

When a Data Value is Tied with the Previous

Specify how ties are treated when computing runs up and down.

The options are

- **Count as Up** (+)

  When tied values are encountered the direction between them is always counted as "Up".

- **Count as Down** (-)

  When tied values are encountered the direction between them is always counted as "Down".

- **Count the Same as the Previous**

  The direction (up or down) between tied values is counted the same as the previous direction in the series.

  For example, if observation #3 < observation #4 and observation #4 = observation #5, then the direction between observation #4 and #5 would be counted as "up" since the direction of observation #3 to #4, which immediately precedes #4 to #5, is "up". On the other hand, if observation #3 > observation #4, then the direction between observation #4 and #5 would be counted as "down".

  If the second value is tied with the first, then it is skipped and not used when computing runs. All subsequent tied observations are also skipped until a non-tie occurs. This would reduce the overall sample size used in the analyses of runs up and down.

- **Skip the Value**

  When tied values are encountered, the second value is skipped and not used when computing runs up and down. Selecting this option may reduce the overall sample size used in the analyses.

When tied values are present, a common practice is to run the analysis twice counting them as "up" in the first and "down" in the second to see if the results change.

Reports Tab

This tab controls which statistical reports and tables are displayed in the output.

Descriptive Reports

Data Summary

This report presents a summary of the data and input options used in the analysis.

Runs Summary

This report presents a summary of the runs in the data, including the number and longest run length in each category.

Detailed Runs Table(s)

This report presents complete information about the runs in the data. The table outlines the number of runs of each length in each category. The items that can be included in the table are

- **Number of Runs**
- **Percentages within Run Length (Row)**
- **Percentages within Category (Column)**
- **Percentages of Total Runs**

Runs Tests for Randomness using the Number of Runs Above and Below a Reference Value

Exact Test

Check this option to output the non-parametric exact test of the number of runs. This test is often used to test the null hypothesis of randomness and is sometimes referred to as the Wald-Wolfowitz single-sample runs test. It can also be used as a goodness-of-fit test.

If the data are binary or categorical then this is a test of the number of runs in each category. If the data are numeric, then this is a test of the number of runs above and below a reference value.

This test is more accurate than the two asymptotic z tests and should always be used if available.

This test is only available for binary and numeric data (which is effectively reduced to a binary series). It is not available for data with more than two categories.

This can be used for both small and large sample sizes.

Asymptotic Z Test

Check this option to output the non-parametric asymptotic *z* test of the number of runs. This test is often used to test the null hypothesis of randomness and is sometimes referred to as the Wald-Wolfowitz single-sample runs test. It can also be used as a goodness-of-fit test.

If the data are binary or categorical then this is a test of the number of runs in each category. If the data are numeric, then this is a test of the number of runs above and below a reference value.

This test is less accurate than the exact test and usually less accurate than the continuity-corrected Z test. This test should be used for large sample sizes only.

Asymptotic Z Test with Continuity Correction

Check this option to output the non-parametric asymptotic *z* test of the number of runs with a continuity correction. This test is often used to test the null hypothesis of randomness and is sometimes referred to as the Wald-Wolfowitz single-sample runs test. It can also be used as a goodness-of-fit test.

If the data are binary or categorical then this is a test of the number of runs in each category. If the data are numeric, then this is a test of the number of runs above and below a reference value. This test is less accurate than the exact test but usually more accurate than the regular *z* test. This test should be used for large sample sizes only.

Runs Tests for Serial Randomness using the Number of Runs Up and Down (Numeric Data)

Exact Test (Only Available for n ≤ 25)

Check this option to output the non-parametric exact test of the number of runs up and down. This test is often used to test the null hypothesis of randomness and is sometimes called the runs test for serial randomness.

This test is more accurate than the two asymptotic *z* tests and should always be used if available.

This test is only available for numeric data with total sample size $n \le 25$. When $n > 25$, the asymptotic tests results (particularly the continuity-corrected test) are very close to the exact test.

Asymptotic Z Test

Check this option to output the non-parametric asymptotic *z* test of the number of runs up and down. This test is often used to test the null hypothesis of randomness and is sometimes called the runs test for serial randomness.

This test is less accurate than the exact test and usually less accurate than the continuity-corrected *z* test.

This test should be used for large sample sizes only.

Asymptotic Z Test with Continuity Correction

Check this option to output the non-parametric asymptotic *z* test of the number of runs up and down. This test is often used to test the null hypothesis of randomness and is sometimes called the runs test for serial randomness.

This test is less accurate than the exact test but usually more accurate than the regular *z* test.

This test should be used for large sample sizes only.

Alpha for Tests

Alpha

Enter the value of alpha to be used for all hypothesis tests in this procedure. The probability level (p-value) is compared to alpha to determine whether to reject the null hypothesis.

Exact Test

Maximum N for Exact Tests

Specify the maximum allowable value of the total sample size, n, for exact hypothesis tests. When n is greater than this amount, the "exact" results are not calculated.

Because of the (sometimes prohibitively) long running time needed for exact calculations with larger sample sizes, this option allows you to set a cap for n for such tests.

Fortunately, the results of many of the asymptotic (non-exact) tests are very close to the exact test results for larger sample sizes.

Report Options Tab

The following options control the format of the reports.

Report Options

Variable Names

Specify whether to use variable names, variable labels, or both to label output reports. In this discussion, the variables are the columns of the data table.

- **Names**

  Variable names are the column headings that appear on the data table. They may be modified by clicking the Column Info button on the Data window or by clicking the right mouse button while the mouse is pointing to the column heading.

- **Labels**

  This refers to the optional labels that may be specified for each column. Clicking the Column Info button on the Data window allows you to enter them.

- **Both**

  Both the variable names and labels are displayed.

Comments

1. Most reports are formatted to receive about 12 characters for variable names.

2. Variable Names cannot contain blanks or math symbols (like + - * / . ,), but variable labels can.

Value Labels

Value Labels are used to make reports more legible by assigning meaningful labels to numbers and codes.

The options are

- **Data Values**

  All data are displayed in their original format, regardless of whether a value label has been set or not.

- **Value Labels**

  All values of variables that have a value label variable designated are converted to their corresponding value label when they are output. This does not modify their value during computation.

- **Both**

  Both data value and value label are displayed.

Example

A variable named GENDER (used as a grouping variable) contains 1's and 2's. By specifying a value label for GENDER, the report can display "Male" instead of 1 and "Female" instead of 2. This option specifies whether (and how) to use the value labels.

Detailed Runs Table Formatting

Column Justification

Specify whether data columns in the contingency tables will be left or right justified.

Column Widths

Specify how the widths of columns in the contingency tables will be determined.

The options are

- **Autosize to Minimum Widths**

  Each data column is individually resized to the smallest width required to display the data in the column. This usually results in columns with different widths. This option produces the most compact table possible, displaying the most data per page.

- **Autosize to Equal Minimum Width**

  The smallest width of each data column is calculated and then all columns are resized to the width of the widest column. This results in the most compact table possible where all data columns have the same with. This is the default setting.

- **Custom (User-Specified)**

  Specify the widths (in inches) of the columns directly instead of having the software calculate them for you.

Custom Widths (Single Value or List)

Enter one or more values for the widths (in inches) of columns in the contingency tables. This option is only displayed if Column Widths is set to "Custom (User-Specified)".

- **Single Value**

    If you enter a single value, that value will be used as the width for all data columns in the table.

- **List of Values**

    Enter a list of values separated by spaces corresponding to the widths of each column. The first value is used for the width of the first data column, the second for the width of the second data column, and so forth. Extra values will be ignored. If you enter fewer values than the number of columns, the last value in your list will be used for the remaining columns.

    Type the word "Autosize" for any column to cause the program to calculate it's width for you. For example, enter "1 Autosize 0.7" to make column 1 be 1 inch wide, column 2 be sized by the program, and column 3 be 0.7 inches wide.

Wrap Column Headings onto Two Lines

Check this option to make column headings wrap onto two lines. Use this option to condense your table when your data are spaced too far apart because of long column headings.

Omit Percent Sign after Percentages

The program normally adds a percent sign, %, after each percentage. Checking this option will cause this percent sign to be omitted.

Decimal Places

Item Decimal Places

These decimal options allow the user to specify the number of decimal places for items in the output. Your choice here will not affect calculations; it will only affect the format of the output.

- **Auto**

    If one of the "Auto" options is selected, the ending zero digits are not shown. For example, if "Auto (0 to 7)" is chosen,

    0.0500 is displayed as 0.05
    1.314583689 is displayed as 1.314584

    The output formatting system is not designed to accommodate "Auto (0 to 13)", and if chosen, this will likely lead to lines that run on to a second line. This option is included, however, for the rare case when a very large number of decimals is needed.

Example 1 – Runs Analysis of a Numeric Variable

In this example we'll show you how to run an analysis on a single numeric variable. The median will be used as the reference value for computing runs for the Wald-Wolfowitz Runs Test. The data for this example are contained in the "RunsAnalysis" example dataset and are from Siegel and Castellan (1988) page 61. The data consist of aggression scores from 24 children.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the procedure window.

1   Open the RunsAnalysis example dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **RunsAnalysis.NCSS**.
- Click **Open**.

2   Open the Analysis of Runs procedure window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Runs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- Select the **Variables** tab.
- Leave **Input Data Type** as **Numeric**.
- Double-click in the **Variable(s)** text box. This will bring up the variable selection window.
- Select **Aggression** from the list of variables and then click **OK**. **"Aggression"** will appear in the **Variable(s)** box.
- Leave all other options at their default values.

4   Specify the reports.
- Click on the **Reports** tab.
- Check **all three descriptive reports**.
- Check **Percentages within Run Length (Row)** to include it on the **Detailed Runs Table**.
- Check **all hypothesis tests**.

5   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Data Summary Report

Data Summary

Input Data Type	Numeric
Rows Processed	24
Rows with Missing Values	0
Total Values In the Sequence	24
Reference Type	Median
Median	25
Values Equal to the Median	0
Values Tied with the Previous	0
n (Total Values Used to Compute Runs Above and Below the Median)	24
n1 (Values Above)	12 (50.0%)
n2 (Values Below)	12 (50.0%)
n+- (Total Values Used to Compute Runs Up and Down)	24

The Data Summary report gives a summary description of the data used in the analysis. Half of the 24 total values were above the median (=25) and half were below. There were no values on the median and no sequential ties.

Runs Chart

Runs Chart

Runs Chart of Aggression
Reference = Median (25)
Runs: Above = 5, Below = 5, Up = 5, Down = 5
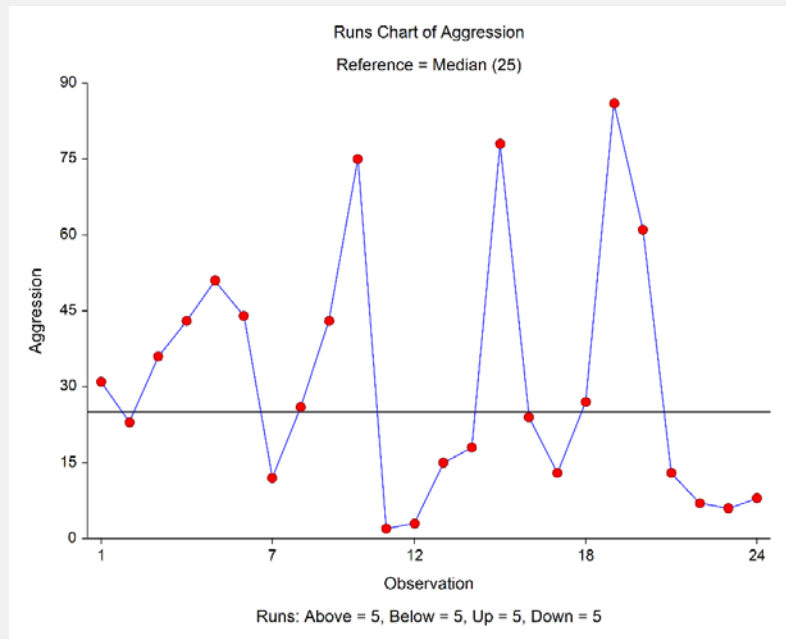
This chart shows the data series with the reference line. It can be used to verify the runs above and below the median as well as the runs up and down.

Runs Summary Report

Runs Summary

Runs Above and Below the Median		Runs Up and Down	
Min Possible Total Runs	2	Min Possible Total Runs	1
Max Possible Total Runs	24	Max Possible Total Runs	23
Expected Total Runs	13	Expected Total Runs	15.66667
r (Observed Total Runs)	10	r+- (Observed Total Runs)	10
r1 (Runs Above)	5 (50.0%)	r+ (Runs Up)	5 (50.0%)
r2 (Runs Below)	5 (50.0%)	r- (Runs Down)	5 (50.0%)
Longest Run Length	4	Longest Run Length	4
Longest Above	4	Longest Up	4
Longest Below	4	Longest Down	4

This report summarizes the runs in the data. There were 5 runs above and 5 runs below the median. There were 5 runs up and 5 runs down. The longest run above or below was 4 and the longest run up or down was also 4.

Exact Test of Runs Above and Below the Median

Exact Test of Runs Above and Below the Median
H0: Data Sequence is Random
Expected Total Runs under H0 = 13

Alternative Hypothesis	n1	n2	n	Total Runs (r)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	12	12	24	10	0.93010	No
H1: Too Few Runs	12	12	24	10	0.15044	No
H1: Two-Sided	12	12	24	10	0.30089	No

The test has a one-sided p-value of 0.15044 for too few runs and a two-sided p-value of 0.30089. This indicates that there is not enough evidence to reject the null hypothesis of randomness.

Asymptotic Z Test of Runs Above and Below the Median

Asymptotic Z Test of Runs Above and Below the Median
H0: Data Sequence is Random
Expected Total Runs under H0 = 13

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	12	12	24	10	2.396	-1.2523	0.89476	No
H1: Too Few Runs	12	12	24	10	2.396	-1.2523	0.10524	No
H1: Two-Sided	12	12	24	10	2.396	-1.2523	0.21047	No

The test has a one-sided p-value of 0.10524 for too few runs and a two-sided p-value of 0.21047. These results are in line with what was observed in the exact test.

Asymptotic Z Test of Runs Above and Below the Median with Continuity Correction

Asymptotic Z Test of Runs Above and Below the Median with Continuity Correction
H0: Data Sequence is Random
Expected Total Runs under H0 = 13

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	12	12	24	10	2.396	-1.0436	0.85166	No
H1: Too Few Runs	12	12	24	10	2.396	-1.0436	0.14834	No
H1: Two-Sided	12	12	24	10	2.396	-1.0436	0.29669	No

The test has a one-sided p-value of 0.14834 for too few runs and a two-sided p-value of 0.29669. These results are in line with what was observed in the exact test. Notice that the p-values for the continuity-corrected z test are much closer to the exact test than those in the regular z test.

Exact Test for Serial Randomness

Exact Test for Serial Randomness
H0: Data Sequence is Random
Expected Total Runs under H0 = 15.66667

Alternative Hypothesis	n+-	Total Runs (r+-)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	24	10	0.99920	No
H1: Too Few Runs	24	10	0.00440	Yes
H1: Two-Sided	24	10	0.00510	Yes

The test has a one-sided p-value of 0.00440 for too few runs and a two-sided p-value of 0.00510. This indicates that the sequence is not random with respect to the number of runs up and down. There are fewer runs in the data series that expected if the series is truly random.

Asymptotic Z Test for Serial Randomness

Asymptotic Z Test for Serial Randomness
H0: Data Sequence is Random
Expected Total Runs under H0 = 15.66667

Alternative Hypothesis	n+-	Total Runs (r+-)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	24	10	1.986	-2.8532	0.99784	No
H1: Too Few Runs	24	10	1.986	-2.8532	0.00216	Yes
H1: Two-Sided	24	10	1.986	-2.8532	0.00433	Yes

The test has a one-sided p-value of 0.00216 for too few runs and a two-sided p-value of 0.00433. These results are in line with what was observed in the exact test.

Asymptotic Z Test for Serial Randomness with Continuity Correction

Asymptotic Z Test for Serial Randomness with Continuity Correction
H0: Data Sequence is Random
Expected Total Runs under H0 = 15.66667

Alternative Hypothesis	n+-	Total Runs (r+-)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	24	10	1.986	-2.6015	0.99536	No
H1: Too Few Runs	24	10	1.986	-2.6015	0.00464	Yes
H1: Two-Sided	24	10	1.986	-2.6015	0.00928	Yes

The test has a one-sided p-value of 0.00464 for too few runs and a two-sided p-value of 0.00928. These results are in line with what was observed in the exact test.

Detailed Table of Runs Above and Below the Median

Detailed Table of Runs Above and Below the Median
Minimum Possible Total Runs = 2
Maximum Possible Total Runs = 24
Expected Total Number of Runs = 13

Run Length		Category Runs Above the Median	Runs Below the Median	Total
1	Number of Runs	2	2	4
	% within Run Length	50.0%	50.0%	100.0%
2	Number of Runs	0	1	1
	% within Run Length	0.0%	100.0%	100.0%
3	Number of Runs	2	0	2
	% within Run Length	100.0%	0.0%	100.0%
4	Number of Runs	1	2	3
	% within Run Length	33.3%	66.7%	100.0%
Total	Number of Runs	5	5	10
	% within Run Length	50.0%	50.0%	100.0%

This report displays the number of runs of each length in each category (above and below the median) and their corresponding percentages.

Detailed Table of Runs Up and Down

Detailed Table of Runs Up and Down
Minimum Possible Total Runs = 1
Maximum Possible Total Runs = 23
Expected Total Number of Runs = 15.66667

Run Length		Category Runs Up	Runs Down	Total
1	Number of Runs	1	2	3
	% within Run Length	33.3%	66.7%	100.0%
2	Number of Runs	1	2	3
	% within Run Length	33.3%	66.7%	100.0%
3	Number of Runs	2	0	2
	% within Run Length	100.0%	0.0%	100.0%
4	Number of Runs	1	1	2
	% within Run Length	50.0%	50.0%	100.0%
Total	Number of Runs	5	5	10
	% within Run Length	50.0%	50.0%	100.0%

This report displays the number of runs of each length in each category (up and down) and their corresponding percentages.

Example 2 – Runs Analysis of a Binary Variable (Validation using Sheskin (2011))

In this example we'll show you how to run an analysis on a single binary variable. The data for this example are from page 410 of Sheskin (2011) and represent the results of 20 coin flips. The goal is to determine if the flipping was indeed random. Sheskin (2011) computes the expected number of runs as 11, the observed number of runs as 11, and an asymptotic standard deviation of 2.18. They conclude that the exact test fails to reject the null hypothesis although no p-value is given. The z value for the regular asymptotic runs test is 0, and the z value for the continuity-corrected runs test is -0.23.

You may follow along here by making the appropriate entries or load the completed template **Example 2** by clicking on Open Example Template from the File menu of the procedure window.

1   Open the RunsAnalysis example dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **RunsAnalysis.NCSS**.
- Click **Open**.

2   Open the Analysis of Runs procedure window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Runs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- Select the **Variables** tab.
- Change **Input Data Type** to **Binary**.
- Double-click in the **Variable(s)** text box. This will bring up the variable selection window.
- Select **CoinFlip** from the list of variables and then click **OK**. **"CoinFlip"** will appear in the **Variable(s)** box.

4   Specify the reports.
- Click on the **Reports** tab.
- Uncheck **Detailed Runs Table(s)**.
- Check **all available hypothesis tests**.

5   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.
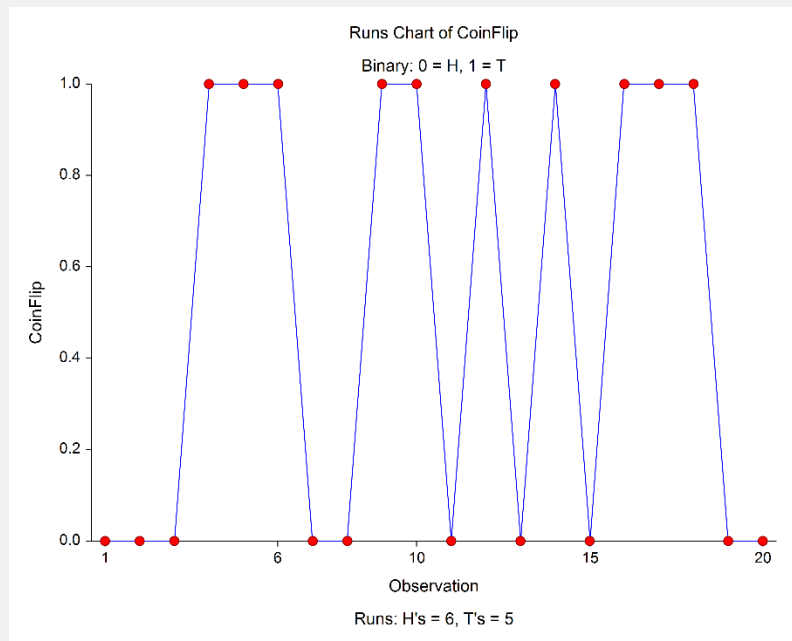
Output

Data Summary

Input Data Type	Binary
Rows Processed	20
Rows with Missing Values	0
k (Unique Values)	2 (H,T)
Total Values In the Sequence	20
n (Total Values Used to Compute Runs)	20
n1 (Number of H's)	10 (50.0%)
n2 (Number of T's)	10 (50.0%)

Runs Chart

Runs Chart of CoinFlip
Binary: 0 = H, 1 = T
Runs: H's = 6, T's = 5

Runs Summary

Minimum Possible Total Runs	2
Maximum Possible Total Runs	20
Expected Total Runs	**11**
r (Observed Total Runs)	**11**
r1 (Runs of H's)	6 (54.5%)
r2 (Runs of T's)	5 (45.5%)
Longest Run Length	3
Longest Run of H's	3
Longest Run of T's	3

Analysis of Runs

Exact Test of the Number of Runs
H0: Data Sequence is Random
Expected Total Runs under H0 = **11**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	10	10	20	11	0.58593	No
H1: Too Few Runs	10	10	20	11	0.58593	No
H1: Two-Sided	10	10	20	11	1.00000	No

Asymptotic Z Test of the Number of Runs
H0: Data Sequence is Random
Expected Total Runs under H0 = **11**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	10	10	20	11	2.176	0.0000	0.50000	No
H1: Too Few Runs	10	10	20	11	2.176	0.0000	0.50000	No
H1: Two-Sided	10	10	20	11	2.176	0.0000	1.00000	No

Asymptotic Z Test of the Number of Runs with Continuity Correction
H0: Data Sequence is Random
Expected Total Runs under H0 = **11**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	10	10	20	11	2.176	-0.2297	0.59085	No
H1: Too Few Runs	10	10	20	11	2.176	-0.2297	0.40915	No
H1: Two-Sided	10	10	20	11	2.176	-0.2297	1.00000	No

The observed number of runs is equal to the expected number of runs for random data. The p-values are all very large. The results match those given in Sheskin (2011).

Example 3 – Runs Tests for a Binary Variable (Validation using Gibbons and Chakraborti (2011))

The data for this example are from page 84 of Gibbons and Chakraborti (2011) and represent the comparison of temperatures on 10 consecutive days in Florida with the average of each day in previous years. An "A" is recorded if the temperature is above average and a "B" is recorded if it is below average. They compute the expected number of runs to be 5.8, the observed number of runs as 6, and an asymptotic standard deviation of 1.424. They compute an exact upper one-sided p-value of 0.5952 and an exact two-sided p-value of 1.000. They also compute an upper one-sided asymptotic p-value of 0.5835 and a two-sided asymptotic p-value of 1.000 using the continuity-corrected z test.

You may follow along here by making the appropriate entries or load the completed template **Example 3** by clicking on Open Example Template from the File menu of the procedure window.

1   Open the RunsAnalysis example dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **RunsAnalysis.NCSS**.
- Click **Open**.

2   Open the Analysis of Runs procedure window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Runs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- Select the **Variables** tab.
- Change **Input Data Type** to **Binary**.
- Double-click in the **Variable(s)** text box. This will bring up the variable selection window.
- Select **Temperature** from the list of variables and then click **OK**. **"Temperature"** will appear in the **Variable(s)** box.

4   Specify the reports.
- Click on the **Reports** tab.
- Uncheck **all descriptive reports.**
- Leave the **default hypothesis tests** checked.

5   Specify the plots.
- Click on the **Plots** tab.
- Uncheck **Runs Chart.**

6   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Output

Exact Test of the Number of Runs
H0: Data Sequence is Random
Expected Total Runs under H0 = **5.8**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	6	4	10	6	0.59524	No
H1: Too Few Runs	6	4	10	6	0.69048	No
H1: Two-Sided	6	4	10	6	1.00000	No

Asymptotic Z Test of the Number of Runs with Continuity Correction
H0: Data Sequence is Random
Expected Total Runs under H0 = **5.8**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	6	4	10	6	1.424	-0.2107	0.58345	No
H1: Too Few Runs	6	4	10	6	1.424	-0.2107	0.41655	No
H1: Two-Sided	6	4	10	6	1.424	-0.2107	1.00000	No

There is not enough evidence to reject the null hypothesis that the data are random. The results match those given in Gibbons and Chakraborti (2011).

Example 4 – Runs Analysis of a Categorical Variable (Validation using Sheskin (2011))

The data for this example are from page 414 of Sheskin (2011) and represent the results of 20 rolls of a three-sided die with sides A, B, and C. They compute the expected number of runs to be 14.1, the observed number of runs as 12, and an asymptotic standard deviation of 1.98. They compute z values of -1.06 and -0.81 for the regular and continuity corrected tests, respectively.

You may follow along here by making the appropriate entries or load the completed template **Example 4** by clicking on Open Example Template from the File menu of the procedure window.

1   Open the RunsAnalysis example dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **RunsAnalysis.NCSS**.
- Click **Open**.

2   Open the Analysis of Runs procedure window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Runs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- Select the **Variables** tab.
- Change **Input Data Type** to **Categorical**.
- Double-click in the **Variable(s)** text box. This will bring up the variable selection window.
- Select **Dice** from the list of variables and then click **OK**. **"Dice"** will appear in the **Variable(s)** box.

4   Specify the reports.
- Click on the **Reports** tab.
- Uncheck **Detailed Runs Table(s)**.
- Check **all available hypothesis tests**.

5   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.
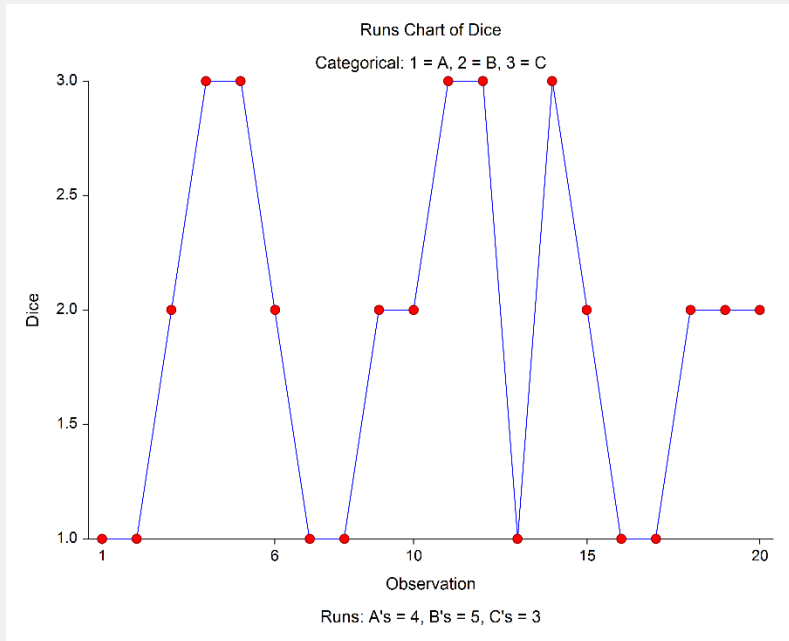
Output

Data Summary

Input Data Type	Categorical
Rows Processed	20
Rows with Missing Values	0
k (Unique Values)	3 (A,B,C)
Total Values In the Sequence	20
n (Total Values Used to Compute Runs)	20
n1 (Number of A's)	7 (35.0%)
n2 (Number of B's)	8 (40.0%)
n3 (Number of C's)	5 (25.0%)

Runs Chart

Runs Chart of Dice
Categorical: 1 = A, 2 = B, 3 = C
Runs: A's = 4, B's = 5, C's = 3

Runs Summary

Minimum Possible Total Runs	3
Maximum Possible Total Runs	20
Expected Total Runs	**14.1**
r (Observed Total Runs)	**12**
r1 (Runs of A's)	4 (33.3%)
r2 (Runs of B's)	5 (41.7%)
r3 (Runs of C's)	3 (25.0%)
Longest Run Length	3
Longest Run of A's	2
Longest Run of B's	3
Longest Run of C's	2

Analysis of Runs

Asymptotic Z Test of the Number of Runs
H0: Data Sequence is Random
Expected Total Runs under H0 = **14.1**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	7	8	20	12	1.980	-1.0604	0.85553	No
H1: Too Few Runs	7	8	20	12	1.980	-1.0604	0.14447	No
H1: Two-Sided	7	8	20	12	1.980	-1.0604	0.28894	No

Additional Category Sizes: N3=5

Asymptotic Z Test of the Number of Runs with Continuity Correction
H0: Data Sequence is Random
Expected Total Runs under H0 = **14.1**

Alternative Hypothesis	n1	n2	n	Total Runs (r)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	7	8	20	12	1.980	-0.8080	0.79044	No
H1: Too Few Runs	7	8	20	12	1.980	-0.8080	0.20956	No
H1: Two-Sided	7	8	20	12	1.980	-0.8080	0.41911	No

Additional Category Sizes: N3=5

There is not enough evidence to reject the null hypothesis that the data are random. The results match those given in Sheskin (2011).

Example 5 – Runs Test for Serial Randomness (Validation using Sheskin (2011))

The quality control data for this example are from page 416 of Sheskin (2011) and represent the amount of milk (in liters) dispensed by a machine into 21 consecutive containers. They compute the expected number of runs up and down to be 13.67, the observed number of runs as 12, and an asymptotic standard deviation of 1.85. They compute an asymptotic z values of -0.90 for the regular z test. They further state that the exact test fails to reject the null hypothesis, but do not provide the p-value.

You may follow along here by making the appropriate entries or load the completed template **Example 5** by clicking on Open Example Template from the File menu of the procedure window.

1   Open the RunsAnalysis example dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **RunsAnalysis.NCSS**.
- Click **Open**.

2   Open the Analysis of Runs procedure window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Runs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- Select the **Variables** tab.
- Leave **Input Data Type** to **Numeric**.
- Double-click in the **Variable(s)** text box. This will bring up the variable selection window.
- Select **Milk** from the list of variables and then click **OK**. **"Milk"** will appear in the **Variable(s)** box.

4   Specify the reports.
- Click on the **Reports** tab.
- Uncheck **all descriptive reports.**
- Uncheck **all hypothesis tests that use a reference.**
- Under **Runs Tests for Serial Randomness**, check **Exact Test** and **Asymptotic Z Test**.

5   Specify the plots.
- Click on the **Plots** tab.
- Uncheck **Runs Chart.**

6   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Output

Exact Test for Serial Randomness
H0: Data Sequence is Random
Expected Total Runs under H0 = **13.66667**

Alternative Hypothesis	n+-	Total Runs (r+-)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	21	12	0.87980	No
H1: Too Few Runs	21	12	0.26220	No
H1: Two-Sided	21	12	0.42240	No

Asymptotic Z Test for Serial Randomness
H0: Data Sequence is Random
Expected Total Runs under H0 = **13.66667**

Alternative Hypothesis	n+-	Total Runs (r+-)	SD	Z	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	21	12	1.847	-0.9024	0.81658	No
H1: Too Few Runs	21	12	1.847	-0.9024	0.18342	No
H1: Two-Sided	21	12	1.847	-0.9024	0.36684	No

There is not enough evidence to reject the null hypothesis that the data are random. The results match those given in Sheskin (2011).

Example 6 – Runs Test for Serial Randomness on Data with Ties (Validation using Bradley (1968))

The quality control data for this example are from page 279 of Bradley (1968) and represent measurements from 25 articles of clothing. They find 2 tied values. They count 7 runs and compute an exact lower, one-sided p-value of 0.0000 if the direction between the tied values is counted as "up". They also count 9 runs and compute an exact lower, one-sided p-value of 0.0003 if the direction between the tied values is counted as "down".

You may follow along here by making the appropriate entries or load the completed template **Example 6** by clicking on Open Example Template from the File menu of the procedure window.

1   Open the RunsAnalysis example dataset.
- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **RunsAnalysis.NCSS**.
- Click **Open**.

2   Open the Analysis of Runs procedure window.
- Using the Analysis menu or the Procedure Navigator, find and select the **Analysis of Runs** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3   Specify the variables.
- Select the **Variables** tab.
- Leave **Input Data Type** to **Numeric**.
- Double-click in the **Variable(s)** text box. This will bring up the variable selection window.
- Select **Article** from the list of variables and then click **OK**. **"Article"** will appear in the **Variable(s)** box.

4   Specify the reports.
- Click on the **Reports** tab.
- Uncheck **all descriptive reports.**
- Uncheck **all hypothesis tests that use a reference.**
- Check only the **Exact Runs Tests for Serial Randomness**.

5   Specify the plots.
- Click on the **Plots** tab.
- Uncheck **Runs Chart.**

6   Run the procedure.
- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Output

Data Summary

Input Data Type	Numeric
Rows Processed	25
Rows with Missing Values	0
Total Values In the Sequence	25
Reference Type	Median
Median	655
Values Equal to the Median	2 (Counted As "Above" when Computing Runs)
Values Tied with the Previous	**2** (Counted As "Up (+)" when Computing Runs)
n (Total Values Used to Compute Runs	25
Above and Below the Median)	
n1 (Values Above)	13 (52.0%)
n2 (Values Below)	12 (48.0%)
n+- (Total Values Used to Compute Runs	25
Up and Down)	

Exact Test for Serial Randomness
H0: Data Sequence is Random
Expected Total Runs under H0 = 16.33333

Alternative Hypothesis	n+-	Total Runs (r+-)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	25	7	1.00000	No
H1: Too Few Runs	25	**7**	**0.00000**	Yes
H1: Two-Sided	25	7	0.00000	Yes

The data summary indicates that there were 2 ties that were counted as "Up" when computing runs. A total of 7 runs up and down were counted. The lower one-sided p-value of 0.00000 indicates that there is evidence that the series is not random. If we change the way **NCSS** handles ties to count the direction as "Down", then the following results are obtained.

Data Summary

.	.
.	.
.	.
Values Tied with the Previous	**2** (Counted As "Down (-)" when Computing Runs)
.	.
.	.
.	.

Exact Test for Serial Randomness
H0: Data Sequence is Random
Expected Total Runs under H0 = 16.33333

Alternative Hypothesis	n+-	Total Runs (r+-)	Prob Level	Reject H0 at α = 0.05?
H1: Too Many Runs	25	9	1.00000	No
H1: Too Few Runs	25	**9**	**0.00030**	Yes
H1: Two-Sided	25	9	0.00030	Yes

The data summary indicates that there were 2 ties that were counted as "Down" when computing runs. A total of 9 runs up and down were counted. At least one of the ties is a critical tie, since the number of runs counted depends on how the tie is handled. The lower one-sided p-value of 0.00030 indicates that there is evidence that the series is not random. The final results from the hypothesis tests were not affected by the critical tie. In both cases, the results match those given in Bradley (1968).