Example 17.5.3

Detecting Randomness of a Set of Numbers

Are the following data random? Test at $\alpha = 0.05$.

16, 25, 52, 11, 38, 47, 12, 98, 4

SOLUTION

How do you test randomness with a numerical set? Create a new data set comparing each value to the median value. To do this, substitute each value in the original data set with an A if it is above the median value, a B if it is below the median value, and eliminate any values that equal the median.

H_0: The data are random.

H_a: The data are not random.

Median = 25

16	25	52	11	38	47	12	98	4
B	∅	A	B	A	A	B	A	B

$m = 4$ (the number of A's)

$n = 4$ (the number of B's)

$R = 7$ (the number of runs)

Using Appendix A, Table M the rejection region is $R \leq 1$ or $R \geq 9$ at the 0.05 level of significance. Since $R = 7$, we fail to reject H_0 and conclude that there is not sufficient evidence of non-randomness.

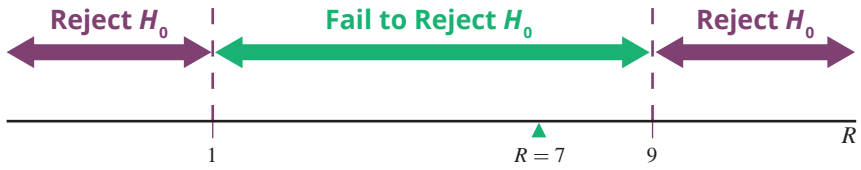

Figure 17.5.6

Technology

For instructions on performing the runs test for randomness using technology, please visit stat.hawkeslearning.com and navigate to **Discovering Business Statistics, Second Edition > Technology Instructions > Nonparametrics > Runs Test for Randomness**.

17.5 Exercises

Basic Concepts

1. Describe in your own words what is being tested with the runs test.
2. Consider the following sequence of 10 coin tosses.

   H, H, T, T, H, H, H, T, T, H

   Without performing any kind of test, do you believe this sequence is random? Explain why or why not.
3. What are the null and alternative hypotheses associated with the runs test?
4. What parameters need to be calculated in order to perform a runs test?
5. What is the rejection rule for a small sample runs test? How small is a *small* sample?
6. What is the rejection rule for a large sample runs test? How large is a *large* sample?
7. If a numerical set of data is under consideration, which parameter are the data points compared to in order to perform the runs test?

Exercises

8. Suppose that in your city the number of deaths due to traffic accidents involving drunk driving from 1999 to 2011 were 75, 91, 54, 85, 79, 63, 12, 55, 63, 49, 89, 98, and 71. Use the runs test to examine non-randomness at the 0.05 level.

9. A sociologist designs a study that involves a procedure of selecting families randomly from a phone book and then calling them to determine if they own or rent their residence. The results are recorded in the order of phone calls (O = Own, R = Rent).

O O O R R O R O R R O R R R R O R R R O O R R R O R

Does the sociologist have a random sequence of residential data at the 0.05 level?

10. A car tire manufacturer keeps track of the tires produced by one of the production lines. They observe the following sequence (D for defective items and N for non-defective items).

D D D N N D N D N D D D

Test the quality control manager's claim that there is no pattern in producing defective tires at the 0.05 level.

11. A marathon runner tries to run every day except when it is raining during the month of July. He observes the rainy (R) days and sunny (S) days to be able to predict the weather as follows.

S S S R R S S S R R R R S R S R R S S R S R S R R S R S R S S

Are the rainy days randomly scattered in the month of July at the 0.05 level?

17.6 The Kruskal–Wallis Test

In this section we present a procedure where k random samples are obtained, one from each of the k possibly different populations, and we are interested in testing whether all of the populations have identical distributions. Suitable hypotheses for this test would be as follows.

H_0: The populations from which the samples are drawn have identical distributions.

H_a: Not all populations have the same distribution.

Definition

Kruskal-Wallis Test

The **Kruskal-Wallis test** is a nonparametric procedure that can be used to determine if two or more distributions are different.

The **Kruskal-Wallis test** is a method that can be used instead of the ANOVA F-test. The Kruskal-Wallis test does not need the assumption of normality of the populations. It does require that independent, random samples be drawn.

The Kruskal-Wallis test is similar to the Wilcoxon rank-sum test in that the test statistic will be based on the sums of the ranks of the groups being compared.

The data consist of k random samples (not necessarily the same size) drawn from their respective populations. The data set may be arranged as follows.

Group 1	Group 2	…	Group k
$x_{1,1}$	$x_{2,1}$	…	$x_{k,1}$
$x_{1,2}$	$x_{2,2}$	…	$x_{k,2}$
…	…	…	…
x_{1,n_1}	x_{2,n_2}	…	x_{k,n_k}

Let N be the total number of observations, that is, $N = \sum_{i=1}^{k} n_i$.