Since our sample size is 12 ($n < 30$), the value of the statistic is compared with the critical value obtained using Appendix A, Table L.

n	$\alpha = 0.10$	$\alpha = 0.05$	…	$\alpha = 0.01$
…				
11	0.523	0.623		0.818
12	0.497	0.591		0.780
13	0.475	0.566		0.745
14	0.457	0.545		0.716
15	0.441	0.525		0.689
…				

Note that Table L in Appendix A is constructed for a two-tailed test. So, we want to reject H_0 if $r_s \le -0.497$ or $r_s \ge 0.497$. Since $0.5944 > 0.497$, we reject the null hypothesis at the 10% level of significance. Hence, there seems to be an association between SAT scores and GPAs.

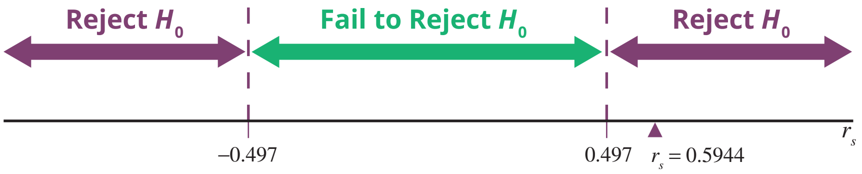

Figure 17.4.1

17.4 Exercises

Basic Concepts

1. What is the correlation coefficient? How is this different from the Spearman rank correlation coefficient?
2. What is the formula for calculating Spearman's rho?
3. Can you calculate Spearman's rho if there are ties in the rank data?
4. Identify the difference in notation between Spearman's rho for population and sample data.
5. Explain the similarities in the behavior of the parametric correlation coefficient and Spearman's rho.
6. Identify one main advantage of the Spearman's rank correlation coefficient versus the parametric correlation coefficient.
7. Explain the procedure for ranking data when calculating Spearman's rho.
8. What are the null and alternative hypotheses for the rank correlation test?
9. Consider the value $r_s = 0.12$. Interpret this value in terms of the x and y variables used to calculate Spearman's rho.

Exercises

10. Chris is a new cashier assigned to a cash register in a supermarket. Each day a sample of purchases at that register is examined and a percent of pricing errors is recorded along with the total number of customers who used that register. Do the following data indicate an association between Chris' performance and how busy his register was? Use $\alpha = 0.05$.

% Pricing Errors and Total Customers

Number of Customers	Errors (%)	Number of Customers	Errors (%)
57	4.2	67	2.5
44	5.5	71	2.9
32	5.7	69	2.6
60	3.9	56	1.0
55	3.2	51	2.0
59	4.1	70	1.7
63	3.3		

11. Twelve new runners were randomly assigned to different training programs, where they were required to run a certain number of miles every week for a year prior to a major race. After the training, the participants ran the race and their finishing times were recorded.

Miles of Training and Race Times

Miles Logged	Race Time (Minutes)	Miles Logged	Race Time (Minutes)
35	198	30	189
25	165	29	240
45	155	42	224
60	148	24	201
70	135	19	246
21	243	55	166

a. With 95% confidence, is there evidence that the number of miles logged in a week during training affects the runner's race time?

b. Can the linear correlation coefficient, r, be calculated in order to fit a least squares regression line to the data in the table in an effort to predict the finish time of runners based on the number of miles logged during training? Why or why not?

12. The following data consist of college rankings of five universities by two different magazines. Is there a correlation between the rankings of the magazines? Use $\alpha = 0.10$.

College Rankings by Magazines

College	A	B	C	D	E
Magazine 1	1	4	2	3	5
Magazine 2	4	3	1	5	2

13. An anthropologist records the heights (in inches) of ten fathers and their sons. Do the following data support (at the 5% level) that taller fathers tend to have taller sons?

Heights of Fathers and Sons (Inches)			
Son's Height	Father's Height	Son's Height	Father's Height
72	70	65	71
68	73	70	78
74	72	69	67
66	68	67	65
71	69	80	66

14. After a mother-daughter golf tournament, mothers and daughters were ranked among themselves. Do the following data show (at the 5% level) a correlation between the daughters' and mothers' golf skills?

Golf Rankings			
Daughter's Ranking	Mother's Ranking	Daughter's Ranking	Mother's Ranking
1	5	5	3
9	4	3	6
10	8	7	7
2	2	6	10
4	1	8	9

17.5 The Runs Test for Randomness

Randomness is an important concept in probability and statistics. In this section we are going to discuss a method for determining whether a sequence of observations exhibits randomness. To illustrate this concept, we will use the familiar coin-tossing experiment. One characteristic of the coin-tossing experiment is that in the long run there should be approximately equal numbers of heads and tails. In an ordered sequence, however, randomness implies more than compliance with this frequency criterion. For example, if the outcomes of 20 tosses of a coin were recorded as

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

we would suspect that the process was flawed. We would be equally surprised if the ordered outcomes were

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

but be reasonably happy with the sequence

H H T H T T T H T H H T H T H H H T T H.

A characteristic that reflects our reservations about the first two sequences is the number of **runs**, where a run is a subsequence of one or more heads (or tails).

Definition

Run

A **run** is a series of increasing values, a series of decreasing values, or a sequence of at least one symbol.

In the first sequence, there are three runs: a run of 5 heads, then 10 tails, then 5 heads.

H H H H H T T T T T T T T T T H H H H H

Run Run Run

In the second sequence, there are 20 runs, each consisting of a single head or tail.

H T H T H T H T H T H T H T H T H T H T

R R R R R R R R R R R R R R R R R R R R