Chapter 800

Pearson's Correlation Tests

Introduction

The correlation coefficient, ρ (rho), is a popular statistic for describing the strength of the relationship between two variables. The correlation coefficient is the slope of the regression line between two variables when both variables have been standardized by subtracting their means and dividing by their standard deviations. The correlation ranges between plus and minus one.

When ρ is used as a descriptive statistic, no special distributional assumptions need to be made about the variables (Y and X) from which it is calculated. When hypothesis tests are made, you assume that the observations are independent and that the variables are distributed according to the bivariate-normal density function. However, as with the t-test, tests based on the correlation coefficient are robust to moderate departures from this normality assumption.

The population correlation ρ is estimated by the sample correlation coefficient *r*. Note we use the symbol *R* on the screens and printouts to represent the population correlation.

Difference between Linear Regression and Correlation

The correlation coefficient is used when both X and Y are from the normal distribution (in fact, the assumption actually is that X and Y follow a bivariate normal distribution). The point is, X is assumed to be a random variable whose distribution is normal. In the linear regression context, no statement is made about the distribution of X. In fact, X is not even a random variable. Instead, it is a set of fixed values such as 10, 20, 30 or -1, 0, 1. Because of this difference in definition, we have included both Linear Regression and Correlation algorithms. This module deals with the Correlation (random X) case.

Test Procedure

The testing procedure is as follows. H_0 is the null hypothesis that the true correlation is a specific value, ρ_0 (usually, $\rho_0 = 0$). H_A represents the alternative hypothesis that the actual correlation of the population is ρ_1, which is not equal to ρ_0. Choose a value R_α, based on the distribution of the sample correlation coefficient, so that the probability of rejecting H_0 when H_0 is true is equal to a specified value, α. Select a sample of *n* items from the population and compute the sample correlation coefficient, r_S. If $r_S > R_\alpha$ reject the null hypothesis that $\rho = \rho_0$ in favor of an alternative hypothesis that $\rho = \rho_1$, where $\rho_1 > \rho_0$. The power is the probability of rejecting H_0 when the true correlation is ρ_1.

All calculations are based on the algorithm described by Guenther (1977) for calculating the cumulative correlation coefficient distribution.

Calculating the Power

Let $R(r|N,\rho)$ represent the area under a correlation density curve to the left of *r*. *N* is the sample size and ρ is the population correlation. The power of the significance test of $\rho_1 > \rho_0$ is calculated as follows:

1. Find r_α such that $1 - R(r_\alpha|N,\rho_0) = \alpha$.

2. Compute the power $= 1 - R(r_\alpha|N,\rho_1)$.

Notice that the calculations follow the same pattern as for the t-test. First find the rejection region by finding the critical value (r_α) under the null hypothesis. Next, calculate the probability that a sample of size *N* drawn from the population defined by setting the correlation to ρ_1 is in this rejection region. This is the power.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would either select *Power* or *Sample Size*.

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Power* when you want to calculate the power.

Test

Alternative Hypothesis

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is $H_0 : \rho_0 = \rho_1$.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Ha: ρ0 ≠ ρ1**

  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the correlation values are different, but you do not want to specify beforehand which correlation is larger.

- **Ha: ρ0 < ρ1**

  This option yields a *one-tailed* test.

- **Ha: $\rho_0 > \rho_1$**

  This option also yields a *one-tailed* test.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal correlations when in fact they are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when you reject the null hypothesis of equal correlations when in fact they are equal.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Sample Size)

The number of observations in the sample. Each observation is made up of two values: one for *X* and one for *Y*.

Effect Size

ρ_0 (Baseline Correlation)

Specify the value of ρ_0. Note that the range of the correlation is between plus and minus one. This value is usually set to zero.

ρ_1 (Alternative Correlation)

Specify the value of ρ_1, the population correlation under the alternative hypothesis. Note that the range of the correlation is between plus and minus one. The difference between ρ_0 and ρ_1 is being tested by this significance test.

You can enter a range of values separated by blanks or commas.

Example 1 – Finding the Power

Suppose a study will be run to test whether the correlation between forced vital capacity (X) and forced expiratory value (Y) in a particular population is 0.30. Find the power when alpha is 0.01, 0.05, and 0.10 and the $N = 20$, 60, 100.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Pearson's Correlation Tests** procedure window by expanding **Correlation**, then **Correlation**, then clicking on **Test (Inequality)**, and then clicking on **Pearson's Correlation Tests**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	**Power**
Alternative Hypothesis	**H1: ρ0 ≠ ρ1**
Alpha	**0.01 0.05 0.10**
N (Sample Size)	**20 60 100**
ρ0 (Baseline Correlation)	0.0
ρ1 (Alternative Correlation)	**0.3**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for H1: ρ0 ≠ ρ1

Power	N	Alpha	Beta	ρ0	ρ1
0.09401	20	0.01000	0.90599	0.00000	0.30000
0.40755	60	0.01000	0.59245	0.00000	0.30000
0.68475	100	0.01000	0.31525	0.00000	0.30000
0.25394	20	0.05000	0.74606	0.00000	0.30000
0.65396	60	0.05000	0.34604	0.00000	0.30000
0.86524	100	0.05000	0.13476	0.00000	0.30000
0.37052	20	0.10000	0.62948	0.00000	0.30000
0.76282	60	0.10000	0.23718	0.00000	0.30000
0.92230	100	0.10000	0.07770	0.00000	0.30000

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
ρ0 is the value of the population correlation under the null hypothesis.
ρ1 is the value of the population correlation under the alternative hypothesis.

Summary Statements
A sample size of 20 achieves 9% power to detect a difference of -0.30000 between the null hypothesis correlation of 0.00000 and the alternative hypothesis correlation of 0.30000 using a two-sided hypothesis test with a significance level of 0.01000.

This report shows the values of each of the parameters, one scenario per row. The values from this table are plotted in the chart below.

Pearson's Correlation Tests

Plots Section
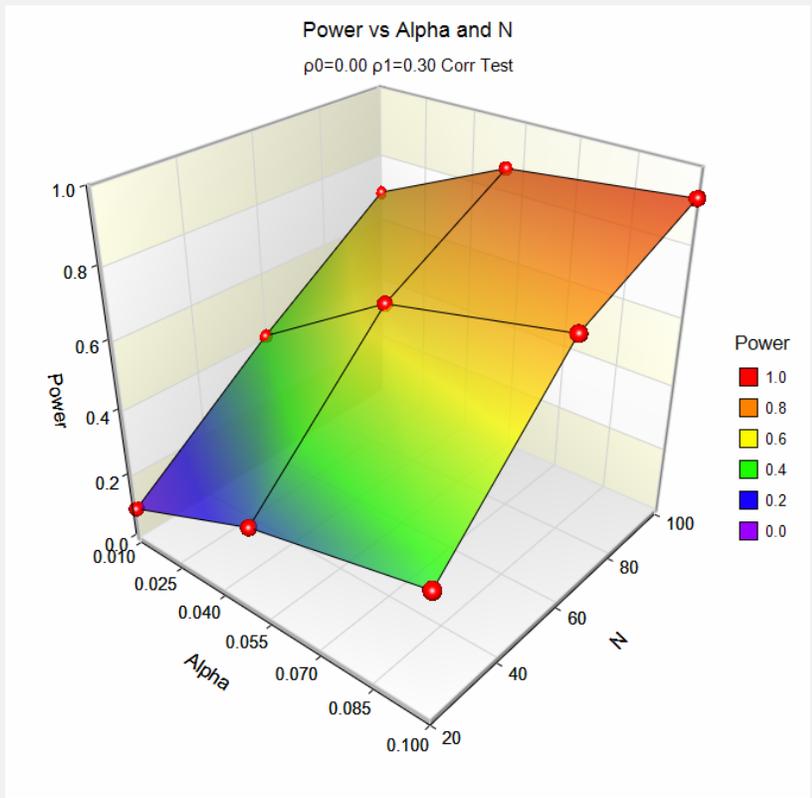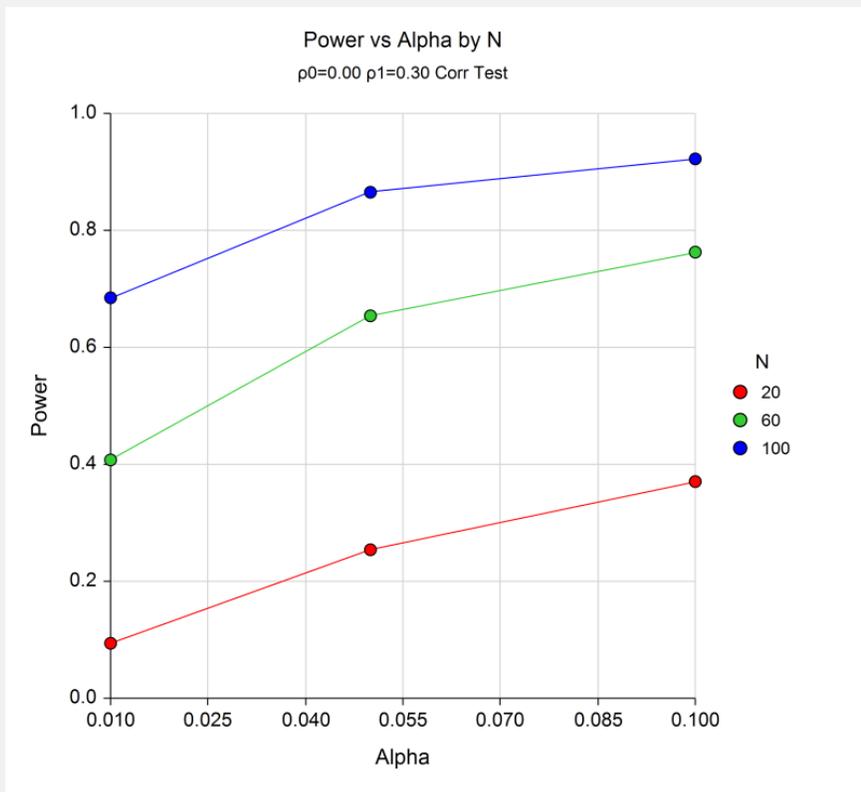
These plots show the relationship between alpha, power, and sample size in this example.

Example 2 – Finding the Sample Size

Continuing with the last example, find the sample size necessary to achieve a power of 90% with a 0.05 significance level.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Pearson's Correlation Tests** procedure window by expanding **Correlation**, then **Correlation**, then clicking on **Test (Inequality)**, and then clicking on **Pearson's Correlation Tests**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Sample Size**
Alternative Hypothesis .... **H1: ρ0 ≠ ρ1**
Power **0.90**
Alpha **0.05**
ρ0 (Baseline Correlation) **0.0**
ρ1 (Alternative Correlation) **0.3**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for H1: ρ0 ≠ ρ1

Power	N	Alpha	Beta	ρ0	ρ1
0.90081	112	0.05000	0.09919	0.00000	0.30000

The required sample size is 112. You would now experiment with the parameters to find out how much varying each will influence the sample size.

Example 3 – Validation using Zar

Zar (1984) page 312 presents an example in which the power of a correlation coefficient is calculated. If $N = 12$, $alpha = 0.05$, $\rho0 = 0$, and $\rho1 = 0.866$, Zar calculates a power of 98% for a two-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Pearson's Correlation Tests** procedure window by expanding **Correlation**, then **Correlation**, then clicking on **Test (Inequality)**, and then clicking on **Pearson's Correlation Tests**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Alternative Hypothesis	**H1: ρ0 ≠ ρ1**
Alpha	**0.05**
N (Sample Size)	**12**
ρ0 (Baseline Correlation)	**0.0**
ρ1 (Alternative Correlation)	**0.866**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for H1: ρ0 ≠ ρ1

Power	N	Alpha	Beta	ρ0	ρ1
0.98398	12	0.05000	0.01602	0.00000	0.86600

The power of 0.98 matches Zar's results.

Example 4 – Validation using Graybill

Graybill (1961) pages 211-212 presents an example in which the power of a correlation coefficient is calculated when the baseline correlation is different from zero. Let $N = 24$, $alpha = 0.05$, and $\rho 1 = 0.5$. Graybill calculates the power of a two-sided test when $\rho 1 = 0.2$ and 0.3 to be 0.363 and 0.193.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Pearson's Correlation Tests** procedure window by expanding **Correlation**, then **Correlation**, then clicking on **Test (Inequality)**, and then clicking on **Pearson's Correlation Tests**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	**Value**
Design Tab	
Solve For	**Power**
Alternative Hypothesis	**H1: ρ0 ≠ ρ1**
Alpha	**0.05**
N (Sample Size)	**24**
ρ0 (Baseline Correlation)	**0.5**
ρ1 (Alternative Correlation)	**0.2 0.3**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for H1: ρ0 ≠ ρ1

Power	N	Alpha	Beta	ρ0	ρ1
0.36583	24	0.05000	0.63417	0.50000	0.20000
0.19950	24	0.05000	0.80050	0.50000	0.30000

The power values match Graybill's results to two decimal places.