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.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Alternative Hypothesis .... **H1: ρ1 ≠ ρ0**
Alpha... **0.01 0.05 0.10**
N (Sample Size).. **20 60 100**
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 Reports

Numeric Results

Solve For: Power
Alternative Hypothesis: H1: ρ1 ≠ ρ0

		Correlation		
Power	Sample Size N	Baseline ρ0	Alternative ρ1	Alpha
0.09401	20	0	0.3	0.01
0.40755	60	0	0.3	0.01
0.68475	100	0	0.3	0.01
0.25394	20	0	0.3	0.05
0.65396	60	0	0.3	0.05
0.86524	100	0	0.3	0.05
0.37052	20	0	0.3	0.10
0.76282	60	0	0.3	0.10
0.92230	100	0	0.3	0.10

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The size of the sample drawn from the population.
ρ0 The value of the population correlation under the null hypothesis.
ρ1 The value of the population correlation under the alternative hypothesis.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A correlation (single group, Y versus X) design will be used to test whether the correlation is different from 0 (H0: ρ = 0 versus H1: ρ ≠ 0). The comparison will be made using a two-sided, one-sample Pearson correlation test, with a Type I error rate (α) of 0.01. To detect a correlation of 0.3 (corresponding to a difference of 0.3) with a sample size of 20, the power is 0.09401.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	20	25	5
20%	60	75	15
20%	100	125	25

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 25 subjects should be enrolled to obtain a final sample size of 20 subjects.

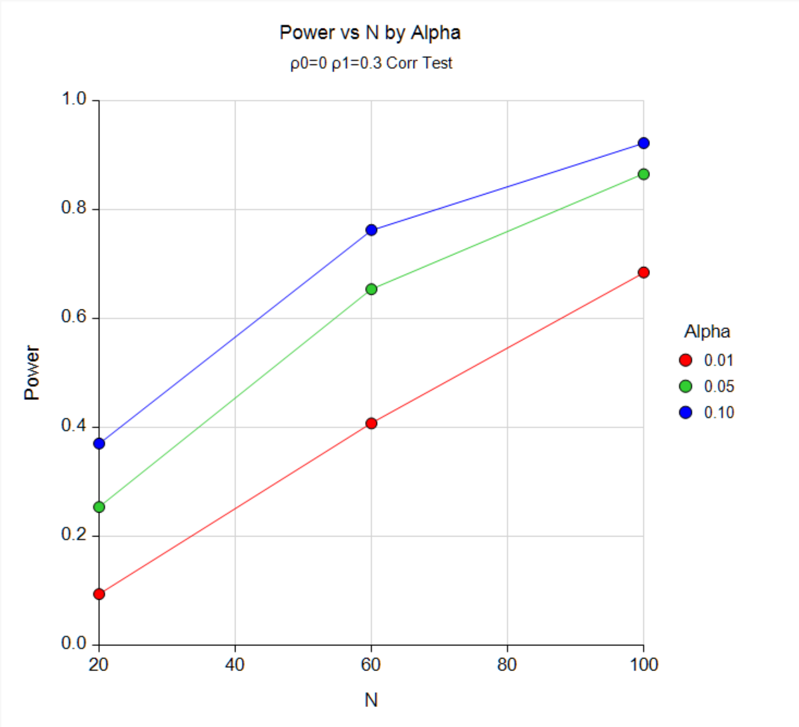
References

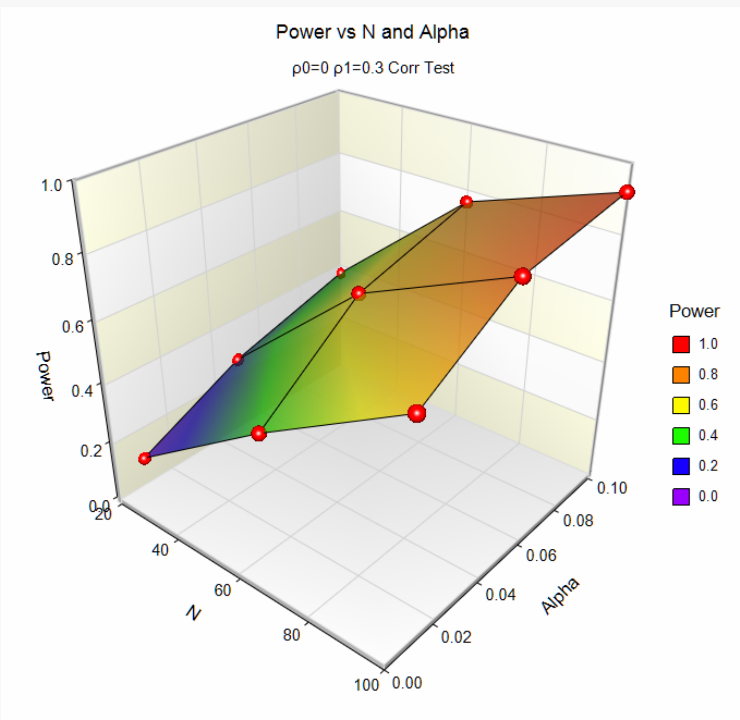
Graybill, Franklin. 1961. An Introduction to Linear Statistical Models. McGraw-Hill. New York, New York.
Guenther, William C. 1977. 'Desk Calculation of Probabilities for the Distribution of the Sample Correlation Coefficient', The American Statistician, Volume 31, Number 1, pages 45-48.
Zar, Jerrold H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.

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.

Plots Section

Plots





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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Sample Size**
Alternative Hypothesis ....**H1: ρ1 ≠ ρ0**
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

Solve For: Sample Size
Alternative Hypothesis: H1: ρ1 ≠ ρ0

Power	Sample Size N	Correlation Baseline ρ0	Correlation Alternative ρ1	Alpha
0.90081	112	0	0.3	0.05

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 (1984)

Zar (1984) page 312 presents an example in which the power of a correlation coefficient is calculated. If *N* = 12, *alpha* = 0.05, *ρ0* = 0, and *ρ1* = 0.866, Zar calculates a power of 98% for a two-sided test.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Alternative Hypothesis .... **H1: ρ1 ≠ ρ0**
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

Solve For: Power
Alternative Hypothesis: H1: ρ1 ≠ ρ0

	Sample	Correlation		
Power	Size N	Baseline ρ0	Alternative ρ1	Alpha
0.98398	12	0	0.866	0.05

The power of 0.98 matches Zar's results.

Example 4 – Validation using Graybill (1961)

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 *ρ1* = 0.5. Graybill calculates the power of a two-sided test when *ρ1* = 0.2 and 0.3 to be 0.363 and 0.193.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Power**
Alternative Hypothesis .... **H1: ρ1 ≠ ρ0**
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

Solve For: Power
Alternative Hypothesis: H1: ρ1 ≠ ρ0

		Correlation		
Power	Sample Size N	Baseline ρ0	Alternative ρ1	Alpha
0.36583	24	0.5	0.2	0.05
0.19950	24	0.5	0.3	0.05

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