Chapter 551

Analysis of Covariance (ANCOVA) (Legacy)

Introduction

A common task in research is to compare the averages of two or more populations (groups). We might want to compare the income level of two regions, the nitrogen content of three lakes, or the effectiveness of four drugs. The one-way analysis of variance compares the means of two or more groups to determine if at least one mean is different from the others. The *F* test is used to determine statistical significance.

Analysis of Covariance (ANCOVA) is an extension of the one-way analysis of variance model that adds quantitative variables (covariates). When used, it is assumed that their inclusion will reduce the size of the error variance and thus increase the power of the design.

Covariates can only be used if the assumption of parallel slopes is viable.

Planned Comparisons

PASS performs power and sample size calculations for user-specified contrasts.

The usual *F* test tests the hypothesis that all means are equal versus the alternative that at least one mean is different from the rest. Often, a more specific alternative is desired. For example, you might want to test whether the treatment means are different from the control mean, the low dose is different from the high dose, a linear trend exists across dose levels, and so on. These questions are tested using planned comparisons.

We call the comparison *planned* because it was determined before the experiment was conducted. We planned to test the comparison.

A comparison is a weighted average of the means, in which the weights may be negative. When the weights sum to zero, the comparison is called a *contrast*. **PASS** provides results for contrasts. To specify a contrast, we need only specify the weights. Statisticians call these weights the *contrast coefficients*.

For example, suppose an experiment conducted to study a drug will have three dose levels: none (control), 20 mg., and 40 mg. The first question is whether the drug made a difference. If it did, the average response for the two groups receiving the drug should be different from the control. If we label the group means M0, M20, and M40, we are interested in comparing M0 with M20 and M40. This can be done in two ways. One way is to construct two tests, one comparing M0 and M20 and the other comparing M0 and M40. Another method is to perform one test comparing M0 with the average of M20 and M40. These tests are conducted using planned comparisons. The coefficients are as follows:

M0 vs. M20

To compare M0 versus M20, use the coefficients -1,1,0. When applied to the group means, these coefficients result in the comparison M0(-1)+M20(1)+M40(0) which reduces to M20-M0. That is, this contrast results in the difference between the two group means. We can test whether this difference is non-zero using the *t* test (or *F* test since the square of the *t* test is an *F* test).

M0 vs. M40

To compare M0 versus M40, use the coefficients -1,0,1. When applied to the group means, these coefficients result in the comparison M0(-1)+M20(0)+M40(1) which reduces to M40-M0. That is, this contrast results in the difference between the two group means.

M0 vs. Average of M20 and M40

To compare M0 versus the average of M20 and M40, use the coefficients -2,1,1. When applied to the group means, these coefficients result in the comparison M0(-2)+M20(1)+M40(1) which is equivalent to M40+M20-2(M0).

To see how these coefficients were obtained, consider the following manipulations. Beginning with the difference between the average of M20 and M40 and M0, we obtain the coefficients above—aside from a scale factor of one-half.

$$\begin{aligned}\frac{M20+M40}{2}-M0&=\frac{M20}{2}+\frac{M40}{2}-\frac{M0}{1}\\&=\frac{1}{2}M20+\frac{1}{2}M40-M0\\&=\frac{1}{2}(M20+M40-2M0)\end{aligned}$$

Assumptions

Using the *F* test requires certain assumptions. One reason for the popularity of the *F* test is its robustness in the face of assumption violation. However, if an assumption is not even approximately met, the significance levels and the power of the *F* test are invalidated. Unfortunately, in practice it often happens that several assumptions are not met. This makes matters even worse. Hence, steps should be taken to check the assumptions before important decisions are made.

The following assumptions are needed for a one-way analysis of variance:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution. Each group is normally distributed about the group mean.
3. The variances of the populations are equal.

4. The groups are independent. There is no relationship among the individuals in one group as compared to another.
5. Each group is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Additional assumptions are needed for an analysis of covariance:

1. The covariates have a linear relationship the response variable.
2. The slopes of these linear relationships between the covariate and the response variable are approximately equal across all groups.

Technical Details for ANCOVA

We found two, slightly different, formulations for computing power for analysis of covariance. Keppel (1991) gives results that modify the standard deviation by an amount proportional to its reduction because of the covariate. Borm et al. (2007) give results that use a normal approximation to the noncentral F distribution. We use the Keppel approach in **PASS**.

Suppose that each observation consists of a response measurement, Y, and one or more covariate measurements: X_1, X_2, ..., X_p. Further suppose that samples of n_1, n_2, ..., n_k observations will be obtained from each of k groups. The multiple regression equation relating Y to the X's within the i^{th} group is

$$Y_i = \beta_{0i} + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

The β's are the regression coefficients or slopes. Analysis of covariance assumes that, except for the intercept β_0, the slopes are equal across all groups. Thus, the difference between the means of any two groups is equal to the difference between their intercepts.

Let σ^2 denote the common variance of all groups ignoring the covariates and σ_ε^2 the within-group variance after considering the covariates. These values are related according to the formula

$$\sigma_\varepsilon^2 = (1 - \rho^2)\sigma^2$$

where ρ^2 is the coefficient of multiple determination (estimated by R^2).

Given the above terminology, the ratio of the mean square between groups to the mean square within groups follows a central F distribution with two parameters matching the degrees of freedom of the numerator mean square and the denominator mean square. When the null hypothesis of mean equality is rejected, the above ratio has a noncentral F distribution which also depends on the noncentrality parameter, λ. This parameter is calculated as

$$\lambda = \bar{n}k\frac{\sigma_m^2}{\sigma_\varepsilon^2}$$

where

$$\sigma_m^2 = \sum_{i=1}^{k} \frac{n_i(\mu_i - \bar{\mu})^2}{N},$$

$$\bar{\mu} = \sum_{i=1}^{k} \frac{n_i u_i}{N},$$

$$N = \sum_{i=1}^{k} n_i,$$

$$\bar{n} = \frac{N}{k}.$$

Some authors use the symbol ϕ for the noncentrality parameter. The relationship between the two noncentrality parameters is

$$\phi = \sqrt{\frac{\lambda}{k}}$$

The process of planning an experiment should include the following steps:

1. Determine an estimate of the within group standard deviation, σ. This may be done from prior studies, from experimentation with the Standard Deviation Estimation module, from pilot studies, or from crude estimates based on the range of the data. See the chapter on estimating the standard deviation for more details.
2. Determine a set of means that represent the group differences that you want to detect.
3. Determine the R-squared value between the response and the covariates.
4. Determine the appropriate group sample sizes that will ensure desired levels of α and β.

Power Calculations for ANCOVA

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{k-1,N-k-p,\alpha}$ where α is the probability of a type-I error and *k, p,* and *N* are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.
2. From a hypothesized set of $\mu_i{}'s$, calculate the noncentrality parameter λ based on the values of *N*, *k*, σ_m, ρ^2, and σ.
3. Compute the power as the probability of being greater than $F_{k-1,N-k-p,\alpha}$ on a noncentral-*F* distribution with noncentrality parameter λ.

Technical Details for a Planned Comparison

The terminology of planned comparisons is identical to that of the one-way AOV, so the notation used above will be repeated here.

Suppose you want to test whether the contrast *C*

$$C = \sum_{i=1}^{k} c_i \, \mu_i$$

is significantly different from zero. Here the $c_i's$ are the contrast coefficients.

Define

$$\sigma_{mc} = \frac{\left|\sum_{i=1}^{k} c_i \, \mu_i\right|}{\sqrt{N \sum_{i=1}^{k} \frac{c_i^2}{n_i}}}$$

Define the noncentrality parameter λ_c, as

$$\lambda_c = \bar{n} k \frac{\sigma_{mc}^2}{\sigma_\varepsilon^2}$$

Power Calculations for Planned Comparisons

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{1,N-k-p,\alpha}$ where α is the probability of a type-I error and *k* and *N* are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.
2. From a hypothesized set of μ_i's, calculate the noncentrality parameter λ_c based on the values of *N*, *k*, σ_{mc}, ρ^2, and σ.
3. Compute the power as the probability of being greater than $F_{1,N-k-p,\alpha}$ on a noncentral-*F* distribution with noncentrality parameter λ_c.

Example 1 – Finding the Statistical Power

An experiment is being designed to compare the means of four groups using an *F* test with a significance level of 0.05. A covariate is available that is estimated to have an R-squared of 0.4 with the response. Previous studies have shown that the standard deviation within a group is 18. Note that this value ignores the covariate.

Treatment means of 40, 10, 10, and 10 represent clinically important treatment differences. To better understand the relationship between power, sample size, and R-squared, the researcher wants to compute the power for R-squared's of 0.2, 0.3, 0.4, and 0.5, and for several group sample sizes between 2 and 10. The sample sizes will be equal across all groups.

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**
Alpha.. **0.05**
k (Number of Groups) **4**
Group Allocation Ratios .... **Equal**
n (Sample Size per Group) .... **2 to 14 by 2**
Hypothesized Means **40 10 10 10**
S (Standard Deviaton of Subjects)................. **18**
Number of Covariates **1**
R2 (R-Squared with Covariates) .... **0.2 to 0.5 by 0.1**
Contrast Coefficients..................................... **None**

Output

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

Numeric Results

Numeric Results

Solve For: Power
Number of Groups (k): 4
Number of Covariates: 1

	Sample Size		Standard Deviation				
Power	**Group n**	**Total N**	**Means Sm**	**Subjects S**	**Effect Size**	**R-Squared with Covariates R2**	**Alpha**
0.17245	2	8	12.99	18	0.7217	0.2	0.05
0.19041	2	8	12.99	18	0.7217	0.3	0.05
0.21428	2	8	12.99	18	0.7217	0.4	0.05
0.24742	2	8	12.99	18	0.7217	0.5	0.05
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
n The average group sample size.
N The total sample size of all groups combined.
Sm The standard deviation of the hypothesized group means under the alternative hypothesis.
S The within group standard deviation.
Effect Size The ratio of σm and σε. Effect Size = Sm / S.
R2 R-Squared. This value gives the strength of the relationship between the response and the covariates.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

In an analysis of covariance study, sample sizes of 2, 2, 2, and 2 will be obtained for each of the 4 groups in order to compare the group means. The comparison will be made using an F-test with a Type I error rate (α) of 0.05. The covariate has an R-squared of 0.2. The total sample size of 8 subjects achieves 17% power to detect differences among the means. The size of the differences in the means is represented by the standard deviation of the means, which is 12.99. The common standard deviation within each group is assumed to be 18.

.
.
.

References

Desu, M. M. and Raghavarao, D. 1990. Sample Size Methodology. Academic Press. New York.
Fleiss, Joseph L. 1986. The Design and Analysis of Clinical Experiments. John Wiley & Sons. New York.
Kirk, Roger E. 1982. Experimental Design: Procedures for the Behavioral Sciences. Brooks/Cole. Pacific Grove, California.
Borm, Fransen, and Lemmens. 2007. 'A simple sample size formula for analysis of covariance in randomized clinical trials.' J of Clinical Epidemiology, 60, 1234-1238.
Keppel, Geoffrey. 1991. Design and Analysis - A Researcher's Handbook. Third Edition. Prentice Hall. Englewood Cliffs, New Jersey. See pages 323 - 324.

This report shows the numeric results of this power study. Following are the definitions of the columns of the report.

Detailed Results Report

Details when Alpha = 0.05, Power = 0.17245, SM = 12.99, S = 18, Cov's = 1, R2 = 0.2

Group	Ni	Percent Ni of Total Ni	Mean	Deviation From Mean	Ni Times Deviation
1	2	25	40.0	22.5	45
2	2	25	10.0	7.5	15
3	2	25	10.0	7.5	15
4	2	25	10.0	7.5	15
ALL	8	100	17.5		

(More Reports Follow)

These reports show details of each row of the previous report.

Group

The number of the group shown on this line. The last line, labeled *ALL*, gives the average or the total as appropriate.

Ni

This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

Percent Ni of Total Ni

This is the percentage of the total sample that is allocated to each group.

Mean

The is the value of the Hypothesized Mean. The final row gives the average for all groups.

Deviation From Mean

This is the absolute value of the mean minus the overall mean. Since *Sm* is the sum of the squared deviations, these values show the relative contribution to *Sm*.

Ni Times Deviation

This is the group sample size times the absolute deviation. It shows the combined influence of the size of the deviation and the sample size on Sm.

Dropout-Inflated Sample Size Report

Dropout-Inflated Sample Size

Average Group Sample Size n	Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
2	1 - 4	20%	2	3	1
	Total		8	12	4
4	1 - 4	20%	4	5	1
	Total		16	20	4
6	1 - 4	20%	6	8	2
	Total		24	32	8
8	1 - 4	20%	8	10	2
	Total		32	40	8
10	1 - 4	20%	10	13	3
	Total		40	52	12
12	1 - 4	20%	12	15	3
	Total		48	60	12
14	1 - 4	20%	14	18	4
	Total		56	72	16

n The average group sample size.
Group Lists the group numbers.
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.
Ni The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
Ni' The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula Ni' = Ni / (1 - DR), with Ni' 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.)
Di The expected number of dropouts in each group. Di = Ni' - Ni.

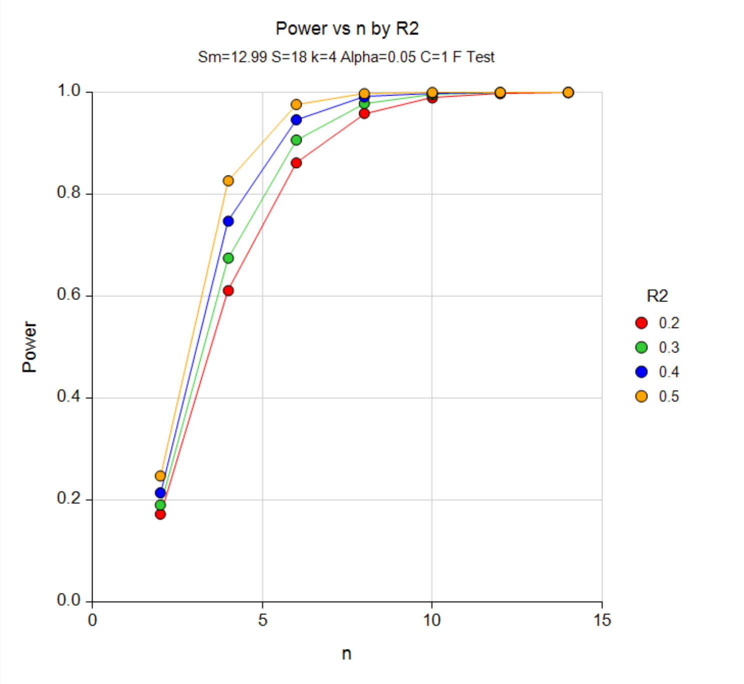
Dropout Summary Statements

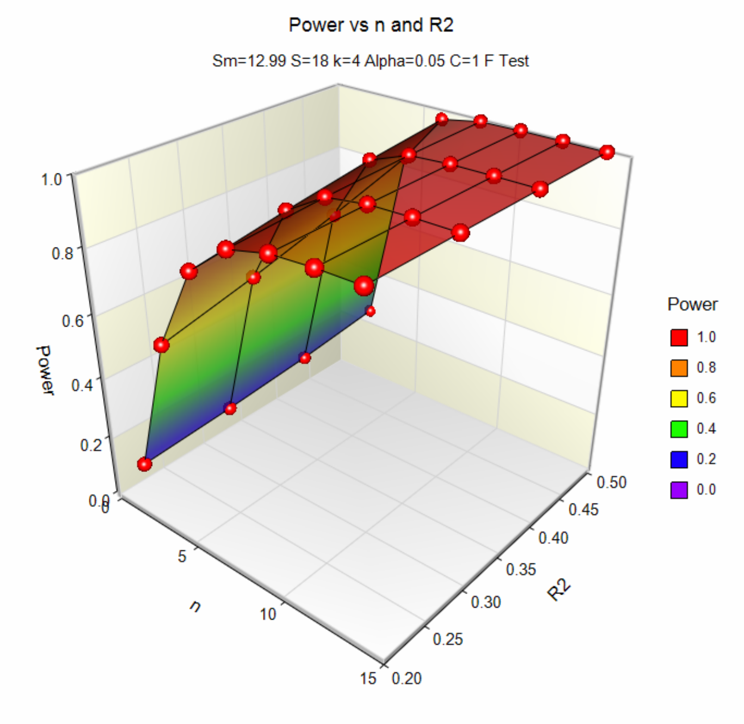
Anticipating a 20% dropout rate, group sizes of 3, 3, 3, and 3 subjects should be enrolled to obtain final group sample sizes of 2, 2, 2, and 2 subjects.

This report shows the sample sizes adjusted for dropout. In this example, dropout is assumed to be 20%. You can change the dropout rate on the Reports tab.

Plots Section

Plots





These plots give a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size and increasing the R-squared.

Note that the value of R-squared has a large impact on power for small sample sizes, but low for larger ones.

Example 2 – Validation using Borm, et al. (2007)

Borm, Fransen, and Lemmens (2007) page 1237 presents an example of determining a sample size in an experiment with 2 groups, mean difference of 0.6, standard deviation of 1.2, alpha of 0.05, one covariate with an R-squared of 0.25, and power of 0.80. They find a total sample size of 95.

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**
Power.. **0.80**
Alpha.. **0.05**
k (Number of Groups) **2**
Group Allocation Ratios .... **Equal**
Hypothesized Means **0 0.6**
S (Standard Deviaton of Subjects)................ **1.2**
Number of Covariates **1**
R2 (R-Squared with Covariates) **0.25**
Contrast Coefficients.................................... **None**

Output

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

Numeric Results

Solve For: Sample Size
Number of Groups (k): 2
Number of Covariates: 1

	Sample Size		Standard Deviation				
Power	Group n	Total N	Means Sm	Subjects S	Effect Size	R-Squared with Covariates R2	Alpha
0.80752	49	98	0.3	1.2	0.25	0.25	0.05

PASS also found *N* = 98. Note that Borm (2007) used calculations based on a normal approximation and obtained *N* = 95, but **PASS** uses exact calculations based on the non-central *F* distribution.