Chapter 485

Paired T-Tests

Introduction

The paired *t*-test may be used to test whether the mean difference of two populations is greater than, less than, or not equal to 0. Because the *t* distribution is used to calculate critical values for the test, this test is often called the paired *t*-test. The paired *t*-test assumes that the population standard deviation of paired differences is unknown and will be estimated by the data.

Other PASS Procedures for Testing One Mean or Median from Paired Data

Procedures in **PASS** are primarily built upon the testing methods, test statistic, and test assumptions that will be used when the analysis of the data is performed. You should check to identify that the test procedure described below in the Test Procedure section matches your intended procedure. If your assumptions or testing method are different, you may wish to use one of the other one-sample paired-data procedures available in **PASS**–the Paired Z-Tests and the nonparametric Wilcoxon Signed-Rank Test procedures. The methods, statistics, and assumptions for those procedures are described in the associated chapters.

If you wish to show that the mean of a population is larger (or smaller) than a reference value by a specified amount, you should use one of the clinical superiority procedures for comparing means. Non-inferiority, equivalence, and confidence interval procedures are also available.

The Statistical Hypotheses

In the usual paired *t*-test setting with δ defined as the mean paired difference, the null (H_0) and alternative (H_1) hypotheses for two-sided tests are defined as

$$H_0: \delta = 0 \quad \text{versus} \quad H_1: \delta \neq 0.$$

Rejecting H_0 implies that the mean paired difference is not equal to 0. The hypotheses for one-sided upper-tail tests are

$$H_0: \delta \leq 0 \quad \text{versus} \quad H_1: \delta > 0.$$

Rejecting H_0 implies that the mean is larger than the value μ_0. This test is called an *upper-tail test* because H_0 is rejected in samples in which the sample mean is larger than μ_0.

The *lower-tail test* is

$$H_0: \delta \geq 0 \quad \text{versus} \quad H_1: \delta < 0.$$

It will be convenient to adopt the following specialized notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
δ	δ	*Population mean paired difference*. This is the mean of paired differences. This parameter will be estimated by the study.
δ_1	$\delta 1$	*Actual paired difference at which power is calculated*. This is the value of the mean paired difference at which power is calculated.

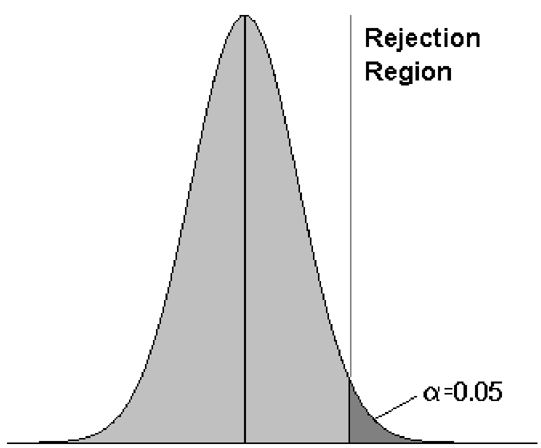
Test Procedure

1. **Find the critical value**. Assume that the true mean paired difference is 0. Choose a value T_α so that the probability of rejecting H_0 when H_0 is true is equal to a specified value called α. Using the *t* distribution, select T_α so that $\Pr(t > T_\alpha) = \alpha$. This value is found using a *t* probability table or a computer program (like **PASS**).
2. **Select a sample of *n* items from the population and compute the *t* statistic**. Call this value *T*. If $T > T_\alpha$ reject the null hypothesis that the mean paired difference equals 0 in favor of an alternative hypothesis that the mean is greater than 0.

Following is a specific example. Suppose we want to test the hypothesis that a variable, *X*, which is made up of paired differences, has a mean of 0 versus the alternative hypothesis that the mean is greater than 0. Suppose that previous studies have shown that the standard deviation of the paired differences, σ, is 40. A random sample of 100 pairs is used.

We first compute the critical value, T_α. The value of T_α that yields *α* = 0.05 is 6.6. If the paired mean difference computed from a sample is greater than 6.6, reject the hypothesis that the mean is 0. Otherwise, do not reject the hypothesis. We call the region greater than 6.6 the *Rejection Region* and values less than or equal to 6.6 the *Acceptance Region* of the significance test.

Figure 1 - Finding Alpha

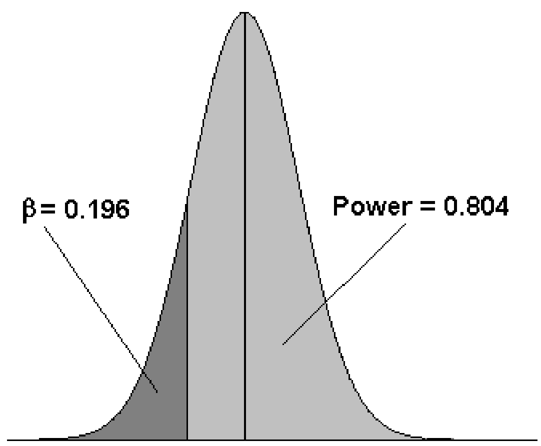


Now suppose that you want to compute the *power* of this testing procedure. In order to compute the power, we must specify an alternative value for the mean. We decide to compute the power if the true mean difference were 10. Figure 2 shows how to compute the power in this case.

The *power* is the probability of rejecting H_0 when the true mean is 10. Since we reject H_0 when the calculated mean difference is greater than 6.6, the probability of a Type-II error (called β) is given by the dark, shaded area of the second graph. This value is 0.196. The power is equal to 1 – β or 0.804.

Note that there are 5 parameters that may be varied in this situation: the mean paired difference, standard deviation of paired differences, alpha, power, and the sample size.

Figure 2 - Finding Power



Assumptions for Paired Tests

This section describes the assumptions that are made when you use one of these tests. The key assumption relates to normality or non-normality of the data. One of the reasons for the popularity of the *t*-test is its robustness in the face of assumption violation. However, if the assumptions are not met, the significance levels and the power of the *t*-test may be invalidated. Unfortunately, in practice it often happens that several assumptions are not met. Take the steps to check the assumptions before you make important decisions based on these tests.

Paired Z-Test Assumptions

The assumptions of the paired *z*-test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched pairs, follow a normal probability distribution.
3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.
4. The population standard deviation of paired differences is known.

Paired T-Test Assumptions

The assumptions of the paired *t*-test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched pairs, follow a normal probability distribution.
3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Wilcoxon Signed-Rank Test Assumptions

The assumptions of the Wilcoxon signed-rank test are as follows (note that the difference is between a data value and the hypothesized median or between the two data values of a pair):

1. The differences are continuous (not discrete).
2. The distribution of each difference is symmetric.
3. The differences are mutually independent.
4. The differences all have the same median.
5. The measurement scale is at least interval.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can often ignore the continuous variable assumption. Perhaps the greatest restriction is that your data come from a random sample of the population. If you do not have a random sample, your significance levels will probably be incorrect.

Paired T-Test Statistic

The paired *t*-test assumes that the paired differences, X_i, are a simple random sample from a population of normally distributed difference values that all have the same mean and variance. This assumption implies that the data are continuous, and their distribution is symmetric. The calculation of the *t*-test proceeds as follows

$$t_{n-1} = \frac{\bar{X}}{s/\sqrt{n}}$$

where

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}},$$

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

Population Size

This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

When a finite population size is specified, the standard deviation is reduced according to the formula:

$$\sigma'^2 = \left(1 - \frac{n}{N}\right)\sigma^2$$

where *n* is the sample size, *N* is the population size, σ is the original standard deviation, and σ' is the new standard deviation.

The quantity *n/N* is often called the sampling fraction. The quantity $\left(1 - \frac{n}{N}\right)$ is called the *finite population correction factor*.

The Standard Deviation of Paired Differences (σ)

If you have results from a previous (or pilot) study, use the estimate of the standard deviation of paired differences, σ, from the study. Another reasonable (but somewhat rough) estimate of σ may be obtained using the range of paired differences as

$$\sigma = \frac{Range}{4}$$

If you have estimates of the expected standard deviations of the paired variables (σ_1 and σ_2) and the Pearson correlation between the paired variables (ρ), the standard deviation of paired differences (σ) may be calculated using the equation

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

such that

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}\,.$$

If $\sigma_1 = \sigma_2 = \sigma_x$, then this formula reduces to

$$\sigma^2 = 2\sigma_x^2(1 - \rho)$$

such that

$$\sigma = \sqrt{2\sigma_x^2(1 - \rho)}\,.$$

If you have an estimate of the within-subject population standard deviation (σ_w), then σ may be calculated using the equation

$$\sigma^2 = 2\sigma_w^2$$

such that

$$\sigma = \sqrt{2\sigma_w^2}\,.$$

σ_w is often estimated by the square root of the within mean square error (WMSE) from a repeated measures ANOVA.

Power Calculation for the Paired T-Test

When the standard deviation is unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which $\delta_1 > 0$.

1. Find t_α such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(t_\alpha)$ is the area under a central-*t* curve to the left of *x* and *df* = *n* – 1.
2. Calculate: $X_1 = t_\alpha \frac{\sigma}{\sqrt{n}}$.
3. Calculate the noncentrality parameter: $\lambda = \frac{\delta_1}{\frac{\sigma}{\sqrt{n}}}$.
4. Calculate: $t_1 = \frac{X_1 - \delta_1}{\frac{\sigma}{\sqrt{n}}} + \lambda$.
5. Power = $1 - T'_{df,\lambda}(t_1)$, where $T'_{df,\lambda}(x)$ is the area to the left of *x* under a noncentral-*t* curve with degrees of freedom *df* and noncentrality parameter λ.

Example 1 – Computing Power

Usually, a researcher designs a study to compare two or more groups of subjects, so the one sample case described in this chapter occurs infrequently. However, there is a popular research design that does lead to the single mean test: *paired observations*.

For example, suppose researchers want to study the impact of an exercise program on the individual's weight. To do so they randomly select *N* individuals, weigh them, put them through the exercise program, and weigh them again. The variable of interest is not their actual weight, but how much their weight changed.

In this design, the data are analyzed using a one-sample *t*-test on the differences between the paired observations. The null hypothesis is that the average difference is zero. The alternative hypothesis is that the average difference is some nonzero value.

To study the impact of an exercise program on weight loss, the researchers decide to conduct a study that will be analyzed using the paired *t*-test. A sample of individuals will be weighed before and after a specified exercise program that will last three months. The difference in their weights will be analyzed.

Past experiments of this type have had standard deviations in the range of 10 to 15 pounds. The researcher wants to detect a difference of 5 pounds or more with an alpha of 0.05. What is the power for sample sizes between 30 and 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 .... **Two-Sided (H1: δ ≠ 0)**
Population Size... **Infinite**
Alpha... **0.05**
N (Sample Size).. **30 to 100 by 10**
δ1 (Mean of Paired Differences) .... **-5**
Standard Deviation Input Type .... **Enter the SD of Paired Differences**
σ (SD of Paired Differences).......................... **10 12.5 15**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: δ = 0 vs. H1: δ ≠ 0

		Paired Differences				
Power	Sample Size N	Mean δ1	Standard Deviation σ	Effect Size |δ1|/σ	Alpha	Beta
0.75396	30	-5	10.0	0.50000	0.05	0.24604
0.86940	40	-5	10.0	0.50000	0.05	0.13060
0.93390	50	-5	10.0	0.50000	0.05	0.06610
0.96779	60	-5	10.0	0.50000	0.05	0.03221
0.98478	70	-5	10.0	0.50000	0.05	0.01522
0.99300	80	-5	10.0	0.50000	0.05	0.00700
0.99685	90	-5	10.0	0.50000	0.05	0.00315
0.99861	100	-5	10.0	0.50000	0.05	0.00139
0.56281	30	-5	12.5	0.40000	0.05	0.43719
0.69399	40	-5	12.5	0.40000	0.05	0.30601
0.79179	50	-5	12.5	0.40000	0.05	0.20821
0.86162	60	-5	12.5	0.40000	0.05	0.13838
0.90984	70	-5	12.5	0.40000	0.05	0.09016
0.94225	80	-5	12.5	0.40000	0.05	0.05775
0.96355	90	-5	12.5	0.40000	0.05	0.03645
0.97730	100	-5	12.5	0.40000	0.05	0.02270
0.42291	30	-5	15.0	0.33333	0.05	0.57709
0.53833	40	-5	15.0	0.33333	0.05	0.46167
0.63709	50	-5	15.0	0.33333	0.05	0.36291
0.71898	60	-5	15.0	0.33333	0.05	0.28102
0.78521	70	-5	15.0	0.33333	0.05	0.21479
0.83770	80	-5	15.0	0.33333	0.05	0.16230
0.87860	90	-5	15.0	0.33333	0.05	0.12140
0.91002	100	-5	15.0	0.33333	0.05	0.08998

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N The sample size, the number of subjects (or pairs) in the study.
δ The population mean of paired differences.
δ1 The value of the mean of paired differences at which power and sample size are calculated.
σ The standard deviation of paired differences for the population.
|δ1|/σ The Effect Size, i.e., the relative magnitude of the effect.
Alpha The probability of rejecting a true null hypothesis.
Beta The probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Summary Statements

A paired design will be used to test whether the paired mean difference (δ) is different from 0 (H0: δ = 0 versus H1: δ ≠ 0). The comparison will be made using a two-sided, paired-difference t-test, with a Type I error rate (α) of 0.05. The underlying standard deviation of the paired difference distribution is assumed to be 10. To detect a paired mean difference of -5 with a sample size of 30 pairs, the power is 0.75396.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	30	38	8
20%	40	50	10
20%	50	63	13
20%	60	75	15
20%	70	88	18
20%	80	100	20
20%	90	113	23
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.

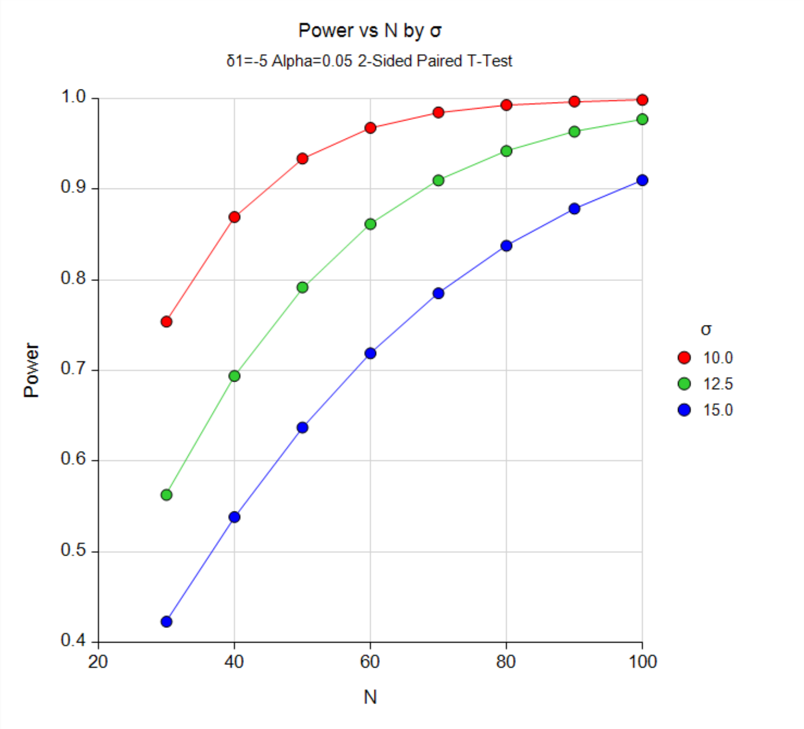
Dropout Summary Statements

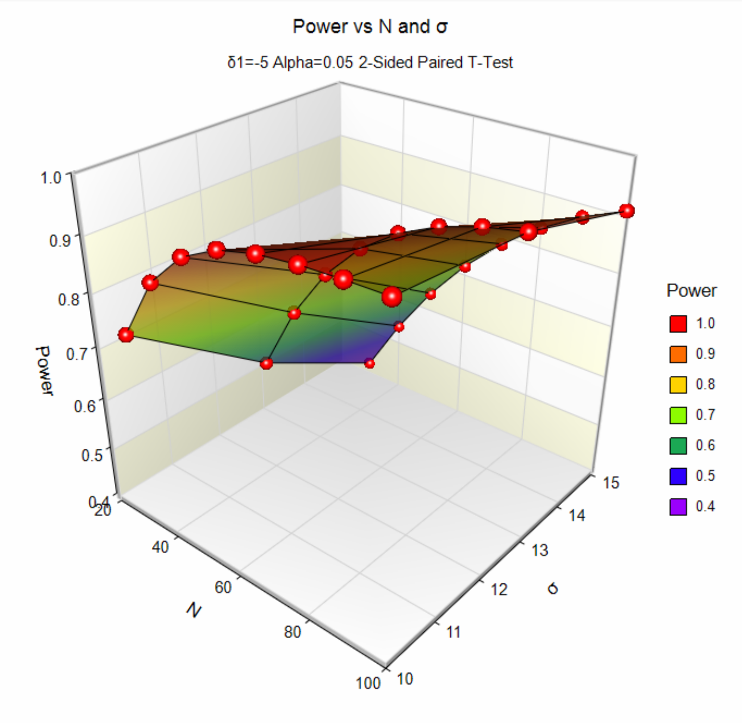
Anticipating a 20% dropout rate, 38 subjects should be enrolled to obtain a final sample size of 30 subjects.

References

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, Third Edition. Taylor & Francis/CRC. Boca Raton, Florida.
Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample Size Tables for Clinical Studies, 2nd Edition. Blackwell Science. Malden, MA.
Zar, Jerrold H. 1984. Biostatistical Analysis (Second Edition). Prentice-Hall. Englewood Cliffs, New Jersey.

Plots





These reports and plots show the relationship between sample size and power for various values of *N* and *σ*.

Example 2 – Finding the Sample Size

Continuing with Example 1, how many pairs are required for each scenario to achieve 80% power?

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 .... **Two-Sided (H1: δ ≠ 0)**
Population Size .... **Infinite**
Power... **0.8**
Alpha.. **0.05**
δ1 (Mean of Paired Differences) .... **-5**
Standard Deviation Input Type .... **Enter the SD of Paired Differences**
σ (SD of Paired Differences).......................... **10 12.5 15**

Output

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

Numeric Results

Solve For: Sample Size
Hypotheses: H0: δ = 0 vs. H1: δ ≠ 0

		Paired Differences				
Power	Sample Size N	Mean δ1	Standard Deviation σ	Effect Size |δ1|/σ	Alpha	Beta
0.80778	34	-5	10.0	0.50000	0.05	0.19222
0.80779	52	-5	12.5	0.40000	0.05	0.19221
0.80230	73	-5	15.0	0.33333	0.05	0.19770

The required sample sizes for each scenario are displayed.

Example 3 – Validation using Chow, Shao, Wang, and Lokhnygina (2018)

Chow, Shao, Wang, and Lokhnygina (2018) presents an example on pages 45 and 46 of a two-sided one-sample *t*-test sample size calculation in which μ0 = 1.5, μ1 = 2.0, σ = 1.0, alpha = 0.05, and power = 0.80. They obtain a sample size of 34.

If we set δ1 = 2.0 – 1.5 = 0.5 then we should get the same result because the one-sample *t*-test and the paired *t*-test use the same fundamental calculations.

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 .... **Sample Size**
Alternative Hypothesis .... **Two-Sided (H1: δ ≠ 0)**
Population Size .... **Infinite**
Power.. **0.8**
Alpha.. **0.05**
δ1 (Mean of Paired Differences) **0.5**
Standard Deviation Input Type .... **Enter the SD of Paired Differences**
σ (SD of Paired Differences).......................... **1**

Output

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

Numeric Results

Solve For: Sample Size
Hypotheses: H0: δ = 0 vs. H1: δ ≠ 0

		Paired Differences				
Power	Sample Size N	Mean δ1	Standard Deviation σ	Effect Size |δ1|/σ	Alpha	Beta
0.80778	34	0.5	1	0.5	0.05	0.19222

The sample size of 34 matches Chow, Shao, Wang, and Lokhnygina (2018) exactly.

Example 4 – Validation using Zar (1984)

Zar (1984) pages 111-112 presents an example in which δ1 = 1.0, σ = 1.25, alpha = 0.05, and N = 12. Zar obtains an approximate power of 0.72.

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 ....**Two-Sided (H1: δ ≠ 0)**
Population Size...**Infinite**
Alpha..**0.05**
N (Sample Size)..**12**
δ1 (Mean of Paired Differences) **1**
Standard Deviation Input Type ....**Enter the SD of Paired Differences**
σ (SD of Paired Differences)..........................**1.25**

Output

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

Numeric Results

Solve For: Power
Hypotheses: H0: δ = 0 vs. H1: δ ≠ 0

		Paired Differences				
Power	Sample Size N	Mean δ1	Standard Deviation σ	Effect Size |δ1|/σ	Alpha	Beta
0.71366	12	1	1.25	0.8	0.05	0.28634

The difference between the power computed by **PASS** of 0.71366 and the 0.72 computed by Zar is due to Zar’s use of an approximation to the noncentral *t* distribution.

Example 5 – Validation using Machin (1997)

Machin, Campbell, Fayers, and Pinol (1997) page 37 presents an example in which δ1 = 0.2, σ = 1.0, alpha = 0.05, and beta = 0.20. They obtain a sample size of 199.

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 5** 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 .... **Two-Sided (H1: δ ≠ 0)**
Population Size .... **Infinite**
Power **0.8**
Alpha **0.05**
δ1 (Mean of Paired Differences) **0.2**
Standard Deviation Input Type .... **Enter the SD of Paired Differences**
σ (SD of Paired Differences) **1**

Output

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

Numeric Results

Solve For: Sample Size
Hypotheses: H0: δ = 0 vs. H1: δ ≠ 0

		Paired Differences				
Power	Sample Size N	Mean δ1	Standard Deviation σ	Effect Size |δ1|/σ	Alpha	Beta
0.80169	199	0.2	1	0.2	0.05	0.19831

The sample size of 199 matches Machin's result.