

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

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

Paired T-Tests

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

<u>Parameter</u>	<u>PASS Input/Output</u>	<u>Interpretation</u>
δ	δ	<i>Population mean paired difference.</i> This is the mean of paired differences. This parameter will be estimated by the study.
δ_1	δ_1	<i>Actual paired difference at which power is calculated.</i> This is the value of the mean paired difference at which power is calculated.

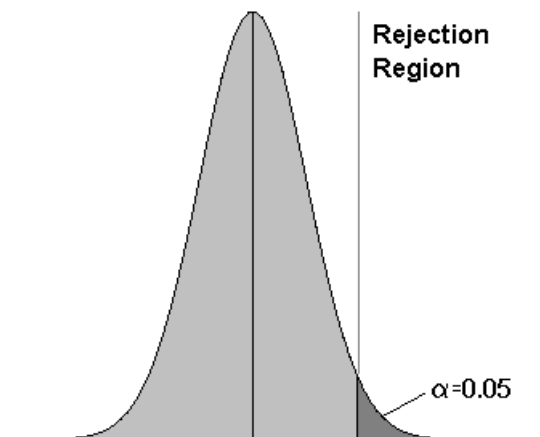
Test Procedure

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 $\alpha = 0.05$ 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



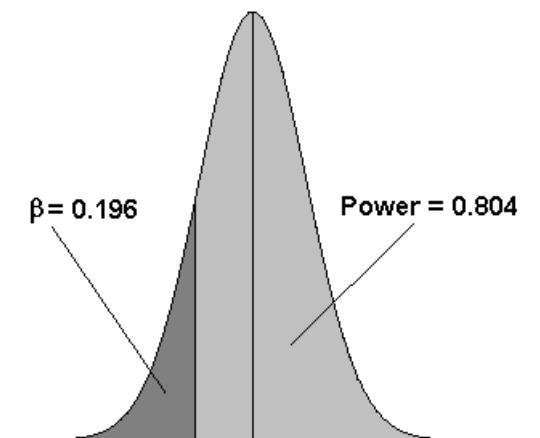
Paired T-Tests

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 $1 - \beta$ 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.

Paired T-Tests

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.

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 .

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 select either *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 of an experiment that has already been run.

Test

Alternative Hypothesis

Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite of the alternative, specifying the alternative is all that is needed. Usually, the two-tailed (\neq) option is selected.

The options containing only $<$ or $>$ are one-tailed tests. When you choose one of these options, you must be sure that the input parameters match this selection.

Possible selections are:

- **Two-Sided (H1: $\delta \neq 0$)**
This is the most common selection. It yields the *two-tailed t-test*. Use this option when you do not want to specify beforehand the direction of the test. Many scientific journals require two-tailed tests.
- **One-Sided (H1: $\delta < 0$)**
This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which δ is less than 0.
- **One-Sided (H1: $\delta > 0$)**
This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which δ is greater than 0.

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:

$$\sigma_1^2 = \left(1 - \frac{n}{N}\right) \sigma^2$$

where n is the sample size, N is the population size, σ is the original standard deviation, and σ_1 is the new standard deviation.

Paired T-Tests

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*.

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.

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 a true null hypothesis is rejected.

Values 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)

This option specifies one or more values of the sample size, the number of pairs in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax *50 100 150 200 250* or *50 to 250 by 50*.

Effect Size

δ_1 (Mean of Paired Differences)

Enter a value (or range of values) for the mean of paired differences at which power and sample size are calculated. This value indicates the minimum detectible paired difference.

σ (Std Dev of Paired Differences)

This option specifies one or more values of the standard deviation. This must be a positive value. Be sure to use the standard deviation of the paired differences and not the standard deviation of the mean paired difference (the standard error).

When this value is not known, you must supply an estimate of it. **PASS** includes a special tool for estimating the standard deviation. This tool may be loaded by pressing the *SD* button. Refer to the Standard Deviation Estimator chapter for further details.

Paired T-Tests

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

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **T-Test (Inequality)**, and then clicking on **Paired T-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	Two-Sided (H1: $\delta \neq 0$)
Population Size	Infinite
Alpha	0.05
N (Sample Size)	30 to 100 by 10
δ_1 (Mean of Paired Differences)	-5
σ (Std Dev of Paired Differences)	10 12.5 15

Paired T-Tests

Annotated Output

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

Numeric Results and Plots

Numeric Results

Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.75396	30	-5.0	10.0	0.500	0.050	0.24604
0.86940	40	-5.0	10.0	0.500	0.050	0.13060
0.93390	50	-5.0	10.0	0.500	0.050	0.06610
0.96779	60	-5.0	10.0	0.500	0.050	0.03221
0.98478	70	-5.0	10.0	0.500	0.050	0.01522
0.99300	80	-5.0	10.0	0.500	0.050	0.00700
0.99685	90	-5.0	10.0	0.500	0.050	0.00315
0.99861	100	-5.0	10.0	0.500	0.050	0.00139
0.56281	30	-5.0	12.5	0.400	0.050	0.43719
0.69399	40	-5.0	12.5	0.400	0.050	0.30601
0.79179	50	-5.0	12.5	0.400	0.050	0.20821
0.86162	60	-5.0	12.5	0.400	0.050	0.13838
0.90984	70	-5.0	12.5	0.400	0.050	0.09016
0.94225	80	-5.0	12.5	0.400	0.050	0.05775
0.96355	90	-5.0	12.5	0.400	0.050	0.03645
0.97730	100	-5.0	12.5	0.400	0.050	0.02270
0.42291	30	-5.0	15.0	0.333	0.050	0.57709
0.53833	40	-5.0	15.0	0.333	0.050	0.46167
0.63709	50	-5.0	15.0	0.333	0.050	0.36291
0.71898	60	-5.0	15.0	0.333	0.050	0.28102
0.78521	70	-5.0	15.0	0.333	0.050	0.21479
0.83770	80	-5.0	15.0	0.333	0.050	0.16230
0.87860	90	-5.0	15.0	0.333	0.050	0.12140
0.91002	100	-5.0	15.0	0.333	0.050	0.08998

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.

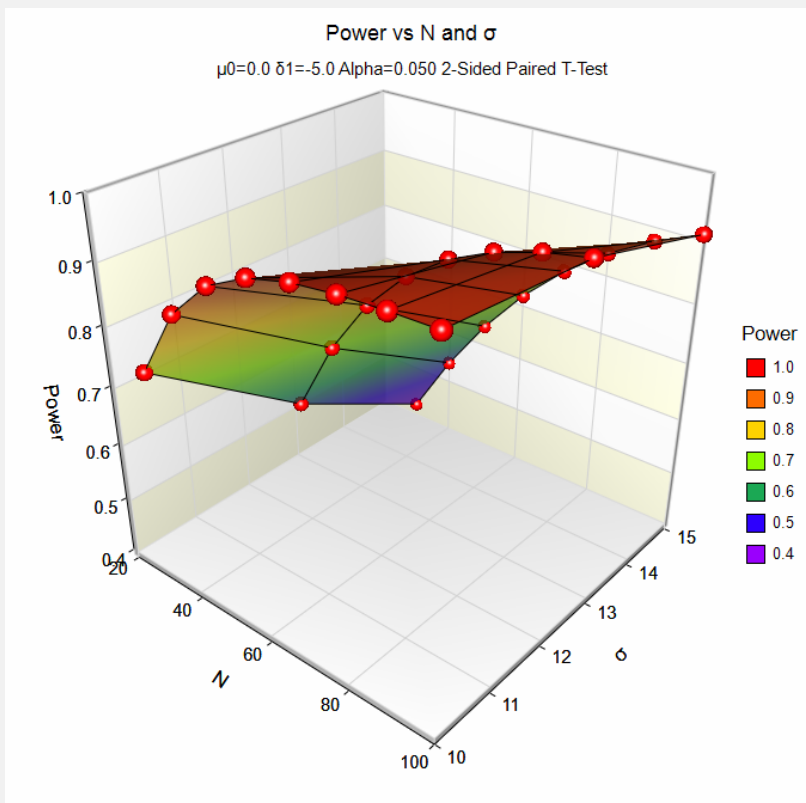
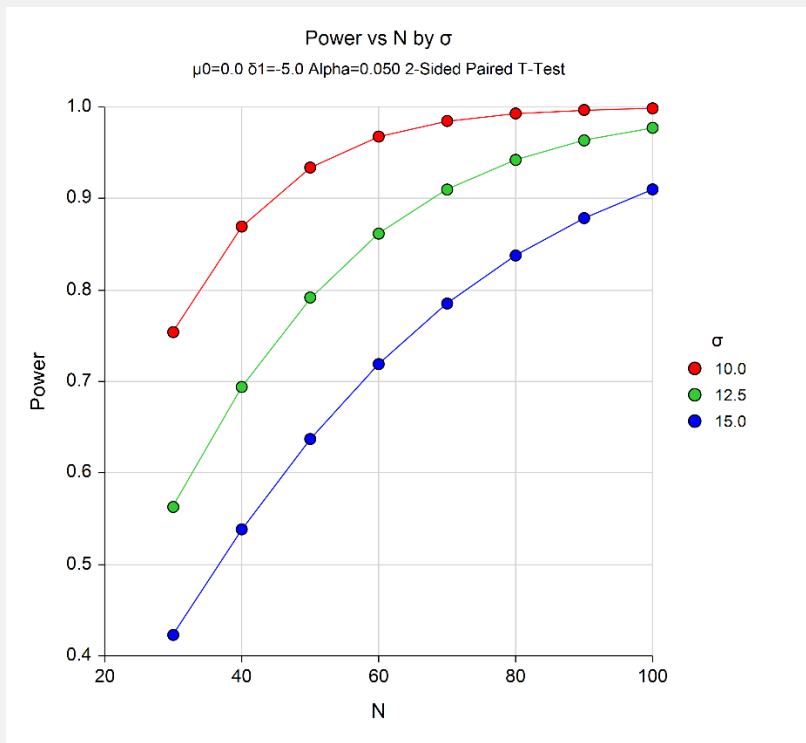
Report Definitions

- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- N is the sample size, the number of subjects (or pairs) in the study.
- δ is the mean of paired differences.
- δ_1 is the value of the mean of paired differences at which power and sample size are calculated.
- σ is the standard deviation of paired differences for the population.
- Effect Size = $|\delta_1|/\sigma$ is the relative magnitude of the effect.
- 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.

Summary Statements

A sample size of 30 achieves 75% power to detect a mean of paired differences of -5.0 with an estimated standard deviation of paired differences of 10.0 and with a significance level (alpha) of 0.050 using a two-sided paired t-test.

Paired T-Tests



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

Paired T-Tests

Example 2 – Finding the Sample Size

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

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **T-Test (Inequality)**, and then clicking on **Paired T-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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Population Size.....	Infinite
Power.....	0.80
Alpha.....	0.05
δ_1 (Mean of Paired Differences).....	-5
σ (Std Dev of Paired Differences).....	10 12.5 15

Output

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

Numeric Results and Plots

Numeric Results

Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80778	34	-5.0	10.0	0.500	0.050	0.19222
0.80779	52	-5.0	12.5	0.400	0.050	0.19221
0.80230	73	-5.0	15.0	0.333	0.050	0.19770

The required sample sizes for each scenario are displayed.

Paired T-Tests

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 $\mu_0 = 1.5$, $\mu_1 = 2.0$, $\sigma = 1.0$, $\alpha = 0.05$, and $\text{power} = 0.80$. They obtain a sample size of 34.

If we set $\delta_1 = 2.0 - 1.5 = 0.5$ then we should get the same result because the one-sample t -test and the paired t -test use the same fundamental calculations.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **T-Test (Inequality)**, and then clicking on **Paired T-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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\mu \neq \mu_0$)
Population Size	Infinite
Power	0.80
Alpha	0.05
μ_0 (Null or Baseline Mean)	1.5
μ_1 (Actual Mean)	2
σ (Standard Deviation)	1

Output

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

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80778	34	0.5	1.0	0.500	0.050	0.19222

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

Paired T-Tests

Example 4 – Validation using Zar (1984)

Zar (1984) pages 111-112 presents an example in which $\delta_1 = 1.0$, $\sigma = 1.25$, $\alpha = 0.05$, and $N = 12$. Zar obtains an approximate power of 0.72.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **T-Test (Inequality)**, and then clicking on **Paired T-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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Population Size.....	Infinite
Alpha.....	0.05
N (Sample Size).....	12
δ_1 (Mean of Paired Differences).....	1
σ (Std Dev of Paired Differences).....	1.25

Output

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

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.71366	12	1.0	1.3	0.800	0.050	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.

Paired T-Tests

Example 5 – Validation using Machin (1997)

Machin, Campbell, Fayers, and Pinol (1997) page 37 presents an example in which $\delta_1 = 0.2$, $\sigma = 1.0$, $\alpha = 0.05$, and $\beta = 0.20$. They obtain a sample size of 199.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests** procedure window by expanding **Means**, then **Paired Means**, then clicking on **T-Test (Inequality)**, and then clicking on **Paired T-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**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided (H1: $\delta \neq 0$)
Population Size	Infinite
Power	0.80
Alpha	0.05
δ_1 (Mean of Paired Differences)	0.2
σ (Std Dev of Paired Differences)	1

Output

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

Numeric Results						
Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$						
Power	N	δ_1	σ	Effect Size	Alpha	Beta
0.80169	199	0.2	1.0	0.200	0.050	0.19831

The sample size of 199 matches Machin's result.