Chapter 424

Two-Sample T-Tests Allowing Unequal Variance

Introduction

This procedure provides sample size and power calculations for one- or two-sided two-sample t-tests when no assumption of equal variances for the two population is made. This is commonly known as the Aspin-Welch test, Welch's t-test (Welch, 1937), or the Satterthwaite method. The as sumed difference between means can be specified by entering the means for the two groups and letting the software calculate the difference or by entering the difference directly.

The design corresponding to this test procedure is sometimes referred to as a *parallel-groups* design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

There are several statistical tests available for the comparison of the center of two populations. This procedure is specific to Aspin-Welch-Satterthwaite test. You can examine the sections below to identify whether the assumptions and test statistic you intend to use in your study match those of this procedure, or if one of the other **PASS** procedures may be more suited to your situation.

Other PASS Procedures for Comparing Two Means or Medians

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 two-sample procedures available in **PASS**. These procedures are Two-Sample T-Tests Assuming Equal Variance, Two-Sample Z-Tests Assuming Equal Variance, Two-Sample Z-Tests Allowing Unequal Variance, and the nonparametric Mann-Whitney-Wilcoxon (also known as the Mann-Whitney U or Wilcoxon rank-sum test) procedure. The methods, statistics, and assumptions for those procedures are described in the associated chapters.

If you wish to show that the mean of one population is larger (or smaller) than the mean of another population 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.

Test Assumptions

When running an Aspin-Welch-Satterthwaite t-test, the basic assumption is that the distributions of the two populations are normal. If that assumption is not likely to be met, another testing procedure, such as a non-parametric procedure could be used, and the corresponding procedure in PASS should be used for sample size or power calculations.

Test Procedure

If we assume that μ_1 and μ_2 represent the means of the two populations of interest, and that $\delta = \mu_1 - \mu_2$, the null hypothesis for comparing the two means is $H_0: \mu_1 = \mu_2$ (or $H_0: \delta = 0$). The alternative hypothesis can be any one of

Two-Sided: $H_1: \mu_1 \neq \mu_2$ (or $H_1: \delta \neq 0$)

Upper One-Sided: $H_1: \mu_1 > \mu_2$ (or $H_1: \delta > 0$)

Lower One-Sided: $H_1: \mu_1 < \mu_2$ (or $H_1: \delta < 0$)

depending upon the desire of the researcher or the protocol instructions. A suitable Type I error probability (α) is chosen for the test, the data is collected, and a t-statistic is generated using the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

This t-statistic follows a t distribution approximately, with estimated degrees of freedom

$$\frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2}\right)^2}$$

The null hypothesis is rejected in favor of the alternative if,

for $H_1: \mu_1 \neq \mu_2$ (or $H_1: \delta \neq 0$),

$$t < t_{\alpha/2} \quad \text{or} \quad t > t_{1-\alpha/2},$$

for $H_1: \mu_1 > \mu_2$ (or $H_1: \delta > 0$),

$$t > t_{1-\alpha},$$

or, for $H_1: \mu_1 < \mu_2$ (or $H_1: \delta < 0$),

$$t < t_{\alpha}.$$

Comparing the t-statistic to the cut-off t-value (as shown here) is equivalent to comparing the p-value to α.

Power Calculation

This section describes the procedure for computing the power from n_1 and n_2, α, the assumed μ_1 and μ_2, and the assumed standard deviations, σ_1 and σ_2. Two good references for these general methods are Julious (2010) and Chow, Shao, Wang, and Lokhnygina (2018), although these texts do not specifically cover the Aspin-Welch-Satterthwaite t-test methods.

The figure below gives a visual representation for the calculation of power for a one-sided test.
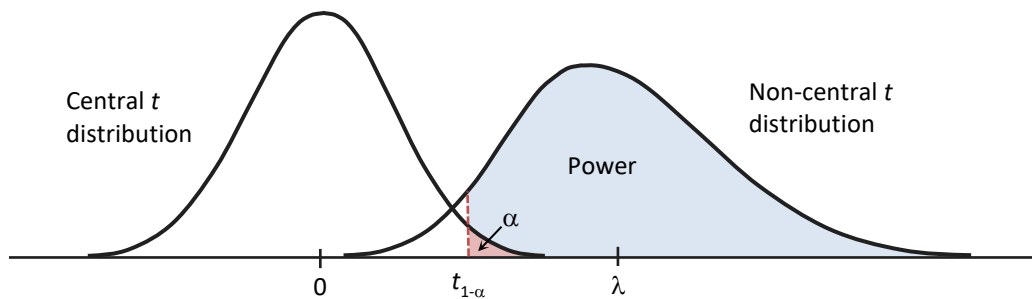
If we call the assumed difference between the means, $\delta = \mu_1 - \mu_2$, the steps for calculating the power are as follows:

1. Find $t_{1-\alpha}$ based on the central-t distribution with degrees of freedom,

$$df = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{\sigma_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{\sigma_2^2}{n_2}\right)^2}.$$

2. Calculate the non-centrality parameter:

$$\lambda = \frac{\delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

3. Calculate the power as the probability that the test statistic t is greater than $t_{1-\alpha}$ under the non-central-t distribution with non-centrality parameter λ:

$$Power = \Pr_{Non-central-t}(t > t_{1-\alpha}|df, \lambda).$$

The algorithms for calculating power for the opposite direction and the two-sided hypotheses are analogous to this method.

When solving for something other than power, **PASS** uses this same power calculation formulation, but performs a search to determine that parameter.

A Note on Specifying the Means or Difference in Means

When means are specified in this procedure, they are used to determine the assumed difference in means for power or sample size calculations. When the difference in means is specified in this procedure, it is the assumed difference in means for power or sample size calculations. It does not mean that the study will be powered to show that the mean difference is this amount, but rather that the design is powered to reject the null hypothesis of equal means if this were the true difference in means. If your purpose is to show that one mean is greater than another by a specific amount, you should use one of the clinical superiority procedures for comparing means.

A Note on Specifying the Standard Deviations

The sample size calculation for most statistical procedures is based on the choice of alpha, power, and an assumed difference in the primary parameters of interest – the difference in means in this procedure. An additional parameter that must be specified for means tests is the standard deviation. Here, we will briefly discuss some considerations for the choice of the standard deviation to enter.

If a number of previous studies of a similar nature are available, you can estimate the variance based on a weighted average of the variances, and then take the square root to give the projected standard deviation.

Perhaps more commonly, only a single pilot study is available, or it may be that no previous study is available. For both of these cases, the conservative approach is typically recommended. In **PASS**, there is a standard deviation estimator tool. This tool can be used to help select an appropriate value or range of values for the standard deviation.

If the standard deviation is not given directly from the previous study, it may be obtained from the standard error, percentiles, or the coefficient of variation. Once a standard deviation estimate is obtained, it may be useful to then use the confidence limits tab to obtain a confidence interval for the standard deviation estimate. With regard to power and sample size, the upper confidence limit will then be a conservative estimate of the standard deviation. Or a range of values from the lower confidence limit to the upper confidence limit may be used to determine the effect of the standard deviation on the power or sample size requirement.

If there is no previous study available, a couple of rough estimation options can be considered. You may use the data tab of the standard deviation estimator to enter some values that represent typical values you expect to encounter. This tool will allow you to see the corresponding population or sample standard deviation. A second rough estimation technique is to base the estimate of the standard deviation on your estimate of the range of the population or the range of a data sample. A conservative divisor for the population range is 4. For example, if you are confident your population values range from 45 to 105, you would enter 60 for the Population Range, and, say, 4, for 'C'. The resulting standard deviation estimate would be 15.

If you are unsure about the value you should enter for the standard deviation, we recommend that you additionally examine a range of standard deviation values to see the effect that your choice has on power or sample size.

Example 1 – Finding the Sample Size

Researchers wish to compare two types of local anesthesia to determine whether there is a difference in time to loss of pain. Subjects will be randomized to treatment, the treatment will be administered, and the time to loss of pain measured. The anticipated time to loss of pain for one of the types of anesthesia is 9 minutes. The researchers would like to generate a sample size for the study with 90% power to reject the null hypothesis of equal loss-of-pain time if the true difference is at least 2 minutes. How many participants are needed to achieve 90% power at significance levels of 0.01 and 0.05?

Past experiments of this type have had standard deviations in the range of 1 to 5 minutes for Group 1, and at about 2.5 for Group 2. It is unknown which treatment has lower time to loss of pain, so a two-sided test will be used.

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 1a** 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**
Power	**0.90**
Alpha	**0.01 0.05**
Group Allocation	**Equal (N1 = N2)**
Input Type	**Means**
$\mu 1$	**11**
$\mu 2$	**9**
$\sigma 1$	**1 to 5 by 1**
$\sigma 2$	**2.5**

Output

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

Numeric Results

Solve For: Sample Size
Test Type: Two-Sample Welch's Unequal-Variance T-Test
Difference: δ = μ1 - μ2
Hypotheses: H0: δ = 0   vs.   H1: δ ≠ 0

Power		Sample Size			Mean			Standard Deviation		
							Difference			
Target	Actual	N1	N2	N	μ1	μ2	δ	σ1	σ2	Alpha
0.9	0.90538	30	30	60	11	9	2	1	2.5	0.01
0.9	0.90607	21	21	42	11	9	2	1	2.5	0.05
0.9	0.90085	40	40	80	11	9	2	2	2.5	0.01
0.9	0.90032	28	28	56	11	9	2	2	2.5	0.05
0.9	0.90315	59	59	118	11	9	2	3	2.5	0.01
0.9	0.90637	42	42	84	11	9	2	3	2.5	0.05
0.9	0.90097	85	85	170	11	9	2	4	2.5	0.01
0.9	0.90187	60	60	120	11	9	2	4	2.5	0.05
0.9	0.90140	119	119	238	11	9	2	5	2.5	0.01
0.9	0.90202	84	84	168	11	9	2	5	2.5	0.05

Target Power   The desired power value (or values) entered in the procedure. Power is the probability of rejecting a false null
                hypothesis.
Actual Power   The power obtained in this scenario. Because N1 and N2 are discrete, this value is often (slightly) larger than
                the target power.
N1 and N2   The number of items sampled from each population.
N   The total sample size. N = N1 + N2.
μ1 and μ2   The assumed population means.
δ   The difference between population means at which power and sample size calculations are made. δ = μ1 -
                μ2.
σ1 and σ2   The assumed population standard deviations for groups 1 and 2, respectively.
Alpha   The probability of rejecting a true null hypothesis.

Summary Statements

A parallel two-group design will be used to test whether the Group 1 mean is different from the Group 2 mean (H0:
μ1 - μ2 = 0 versus H1: μ1 - μ2 ≠ 0). The comparison will be made using a two-sided, two-sample, unequal-variance
t-test (Welch, 1938), with a Type I error rate (α) of 0.01. The standard deviation for Group 1 is assumed to be 1 and
the standard deviation for Group 2 is assumed to be 2.5. To detect a difference in means of μ1 - μ2 = 11 - 9 = 2
with 90% power, the number of needed subjects will be 30 in Group 1 and 30 in Group 2.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	30	30	60	38	38	76	8	8	16
20%	21	21	42	27	27	54	6	6	12
20%	40	40	80	50	50	100	10	10	20
20%	28	28	56	35	35	70	7	7	14
20%	59	59	118	74	74	148	15	15	30
20%	42	42	84	53	53	106	11	11	22
20%	85	85	170	107	107	214	22	22	44
20%	60	60	120	75	75	150	15	15	30
20%	119	119	238	149	149	298	30	30	60
20%	84	84	168	105	105	210	21	21	42

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.
N1, N2, and N	The evaluable sample sizes at which power is computed. If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N'	The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. After solving for N1 and N2, N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' 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.)
D1, D2, and D	The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

Dropout Summary Statements

Anticipating a 20% dropout rate, 38 subjects should be enrolled in Group 1, and 38 in Group 2, to obtain final group sample sizes of 30 and 30, respectively.

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.

Julious, S. A. 2010. Sample Sizes for Clinical Trials. Chapman & Hall/CRC. Boca Raton, FL.

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.

These reports show the values of each of the parameters, one scenario per row.

Plots Section
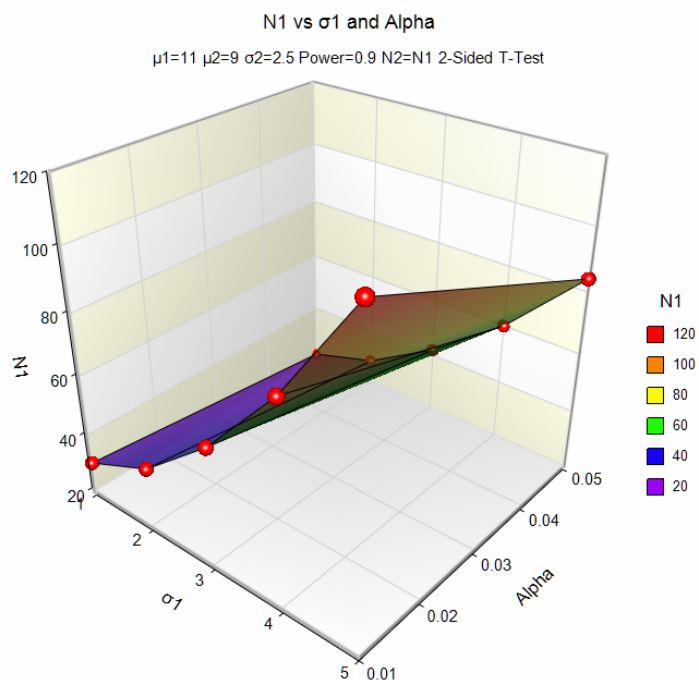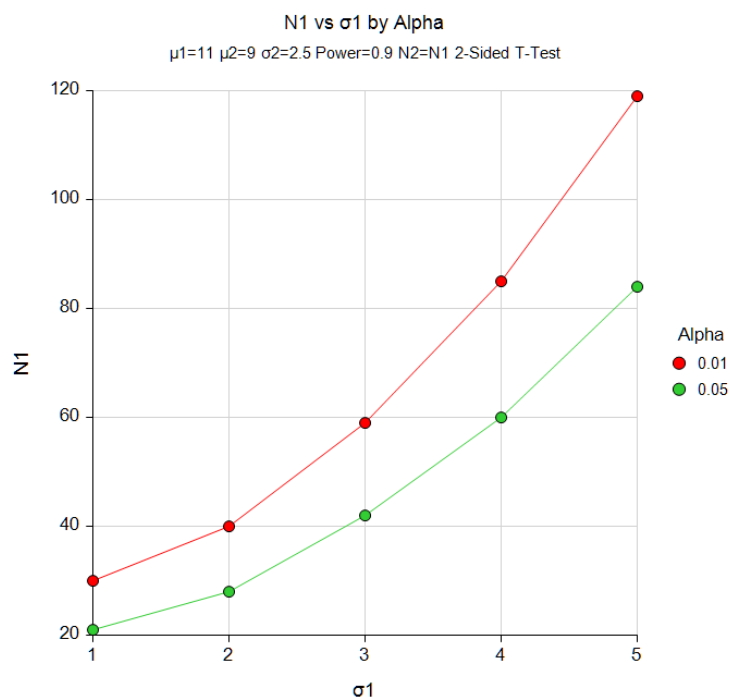
Plots
_____

These plots show the relationship between the standard deviation for Group 1 and sample size for the two alpha levels.

To specify the difference directly to get the same results, change **Input Type** to **Difference** and enter **2** for **δ**, leaving everything else the same. You may then make the appropriate entries as listed below, or open **Example 1b** by going to the **File** menu and choosing **Open Example Template**.

Design Tab
_____

Input Type...**Difference**
δ...**2**

Output

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

Numeric Results
_____

Solve For: Sample Size
Test Type: Two-Sample Welch's Unequal-Variance T-Test
Difference: $\delta = \mu1 - \mu2$
Hypotheses: H0: $\delta = 0$ vs. H1: $\delta \neq 0$
_____

Power		Sample Size			Mean Difference	Standard Deviation		
Target	Actual	N1	N2	N	δ	σ1	σ2	Alpha
0.9	0.90538	30	30	60	2	1	2.5	0.01
0.9	0.90607	21	21	42	2	1	2.5	0.05
0.9	0.90085	40	40	80	2	2	2.5	0.01
0.9	0.90032	28	28	56	2	2	2.5	0.05
0.9	0.90315	59	59	118	2	3	2.5	0.01
0.9	0.90637	42	42	84	2	3	2.5	0.05
0.9	0.90097	85	85	170	2	4	2.5	0.01
0.9	0.90187	60	60	120	2	4	2.5	0.05
0.9	0.90140	119	119	238	2	5	2.5	0.01
0.9	0.90202	84	84	168	2	5	2.5	0.05

The results are the same as those obtained when the means are entered.

Example 2 – Finding the Power

Suppose a new corn fertilizer is to be compared to a current fertilizer. The current fertilizer produces an average of about 74 lbs. per plot. The researchers need only show that there is difference in yield with the new fertilizer. They would like to consider the effect of the number of plots used on the power of the test if the improvement in yield is at least 10 lbs.

Researchers plan to use a one-sided two-sample Aspin-Welch-Satterthwaite t-test with alpha equal to 0.05. Previous studies indicate the standard deviation for plot yield to be 24 lbs. for Group 1, and between 20 and 30 lbs for Group 2. The plot group sizes of interest are 10 to 100 plots per group.

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 ....**Power**

Alternative Hypothesis ....**One-Sided**

Alpha..**0.05**

Group Allocation ....**Equal (N1 = N2)**

Sample Size Per Group ....**10 to 100 by 10**

Input Type...**Means**

μ1...**84**

μ2...**74**

σ1 ....**24**

σ2 ....**20 25 30**

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Test Type: Two-Sample Welch's Unequal-Variance T-Test
Difference: $\delta = \mu_1 - \mu_2$
Hypotheses: H0: $\delta \leq 0$ vs. H1: $\delta > 0$

	Sample Size			Mean		Difference	Standard Deviation		
Power	N1	N2	N	μ1	μ2	δ	σ1	σ2	Alpha
0.25087	10	10	20	84	74	10	24	20	0.05
0.40528	20	20	40	84	74	10	24	20	0.05
0.53474	30	30	60	84	74	10	24	20	0.05
0.64110	40	40	80	84	74	10	24	20	0.05
0.72653	50	50	100	84	74	10	24	20	0.05
0.79385	60	60	120	84	74	10	24	20	0.05
0.84605	70	70	140	84	74	10	24	20	0.05
0.88598	80	80	160	84	74	10	24	20	0.05
0.91618	90	90	180	84	74	10	24	20	0.05
0.93880	100	100	200	84	74	10	24	20	0.05
0.22168	10	10	20	84	74	10	24	25	0.05
0.35293	20	20	40	84	74	10	24	25	0.05
0.46698	30	30	60	84	74	10	24	25	0.05
0.56523	40	40	80	84	74	10	24	25	0.05
0.64855	50	50	100	84	74	10	24	25	0.05
0.71817	60	60	120	84	74	10	24	25	0.05
0.77562	70	70	140	84	74	10	24	25	0.05
0.82251	80	80	160	84	74	10	24	25	0.05
0.86041	90	90	180	84	74	10	24	25	0.05
0.89080	100	100	200	84	74	10	24	25	0.05
0.19657	10	10	20	84	74	10	24	30	0.05
0.30765	20	20	40	84	74	10	24	30	0.05
0.40647	30	30	60	84	74	10	24	30	0.05
0.49456	40	40	80	84	74	10	24	30	0.05
0.57234	50	50	100	84	74	10	24	30	0.05
0.64028	60	60	120	84	74	10	24	30	0.05
0.69905	70	70	140	84	74	10	24	30	0.05
0.74946	80	80	160	84	74	10	24	30	0.05
0.79235	90	90	180	84	74	10	24	30	0.05
0.82861	100	100	200	84	74	10	24	30	0.05

These reports show the values of each of the parameters, one scenario per row.

Plots Section
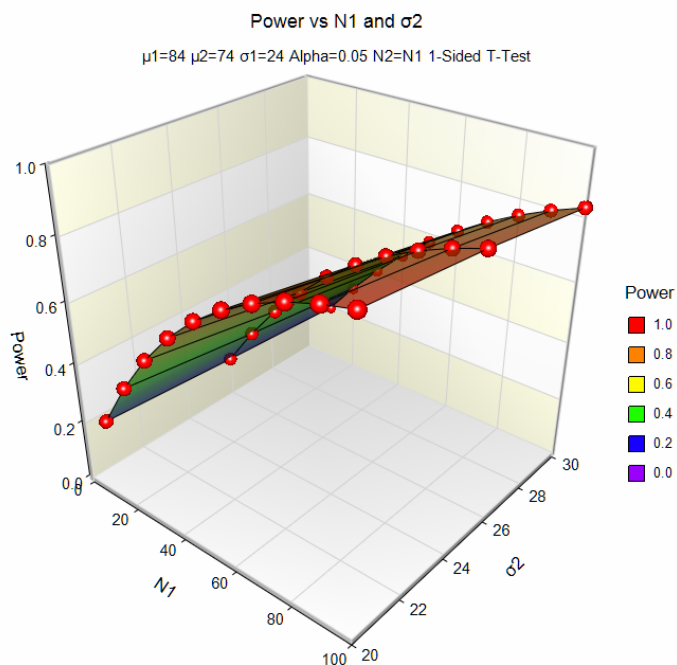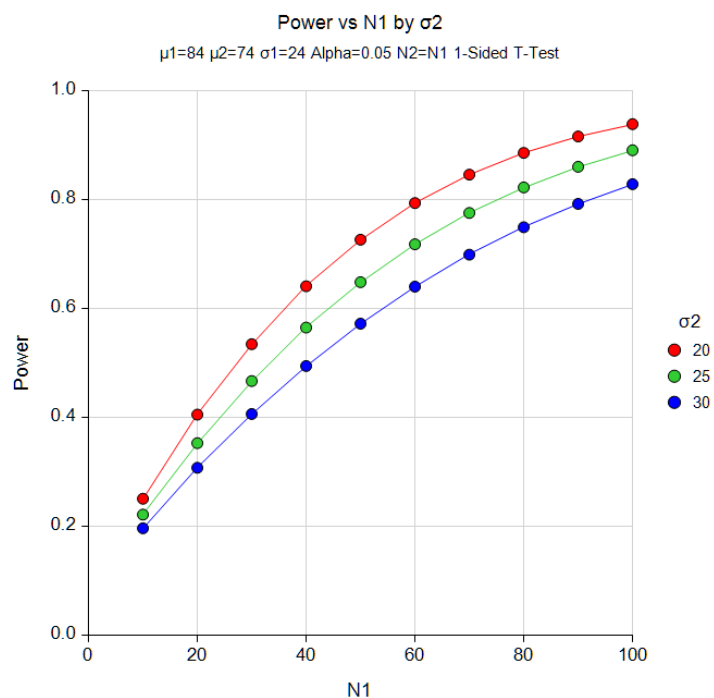
Plots
_____

These plots show the relationship between the power and sample size for the three values of σ2.

Again, to specify the difference directly to get the same results, change **Input Type** to **Difference** and enter **10** for **δ**, leaving everything else the same. You'll get the same results as those above.

Design Tab

Input Type...**Difference**

δ..**10**

Example 3 – Finding the Difference

In some cases, it may be useful to determine how different the means of two populations would need to be to achieve the desired power with a specific constrained sample size.

Suppose, for example, that 80 subjects are available for a study to compare weight loss regimens. Researchers would like to determine which of the two regimens is better. The anticipated weight loss after two months is anticipated to be about 20 lbs. The researchers would like to know how different the mean weight loss must be to have a study with 90% power when alpha is 0.05 and the standard deviation is assumed to be 6 lbs. for Group 1 and 8 lbs. for Group 2.

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 ....**Effect Size (Means or Difference)**
Alternative Hypothesis ....**Two-Sided**
Power...**0.90**
Alpha..**0.05**
Group Allocation ....**Equal (N1 = N2)**
Sample Size Per Group **40**
Input Type..**Difference**
δ..**Search > 0**
σ1 ....**6**
σ2 ....**8**

Reports Tab

Means, Difference Decimals **3**

Output

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

Numeric Results

Solve For: Effect Size (Means or Difference)
Test Type: Two-Sample Welch's Unequal-Variance T-Test
Difference: δ = μ1 - μ2
Hypotheses: H0: δ = 0   vs.   H1: δ ≠ 0

	Sample Size			Mean Difference	Standard Deviation		
Power	N1	N2	N	δ	σ1	σ2	Alpha
0.9	40	40	80	5.195	6	8	0.05

If the true population mean weight loss for one group is 5.195 lbs. more or less than the other, the researchers will have 90% power to show a difference between the groups.

Example 4 – Validation of Sample Size using Simulation

We used the Tests for Two Means (Simulation) procedure to find the sample size needed for a two-sided test with alpha equal to 0.05, where the difference in means is 3.5, the standard deviation for Group 1 is 1 and the standard deviation for Group 2 is 3. The sample size result was 11 per group.

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 ....**Sample Size**
Alternative Hypothesis ....**Two-Sided**
Power...**0.90**
Alpha..**0.05**
Group Allocation ....**Equal (N1 = N2)**
Input Type...**Difference**
δ..**3.5**
σ1 ....**1**
σ2 ....**3**

Output

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

Numeric Results

Solve For: Sample Size
Test Type: Two-Sample Welch's Unequal-Variance T-Test
Difference: $δ = μ1 - μ2$
Hypotheses: $H0: δ = 0$ vs. $H1: δ ≠ 0$

Power		Sample Size			Mean Difference	Standard Deviation		
Target	Actual	N1	N2	N	δ	σ1	σ2	Alpha
0.9	0.92066	11	11	22	3.5	1	3	0.05

The sample size of 11 per group matches the simulation result exactly.