

Chapter 428

Two-Sample Z-Tests Allowing Unequal Variance

Introduction

This procedure provides sample size and power calculations for one- or two-sided two-sample z-tests when the variances are assumed to be known and unequal. The assumed 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 the two-sample z-test assuming known and unequal variance. 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 Allowing Unequal Variance, Two-Sample T-Tests Assuming Equal Variance, Two-Sample Z-Tests Assuming Equal 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 a two-sample unequal-variance z-test, the basic assumptions are that the distributions of the two populations are normal, and that the variances of the two distributions are known and are different. If those assumptions are not likely to be met, another testing 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 z-statistic is generated using the formula:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

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

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

$$z < z_{\alpha/2} \quad \text{or} \quad z > z_{1-\alpha/2},$$

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

$$z > z_{1-\alpha},$$

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

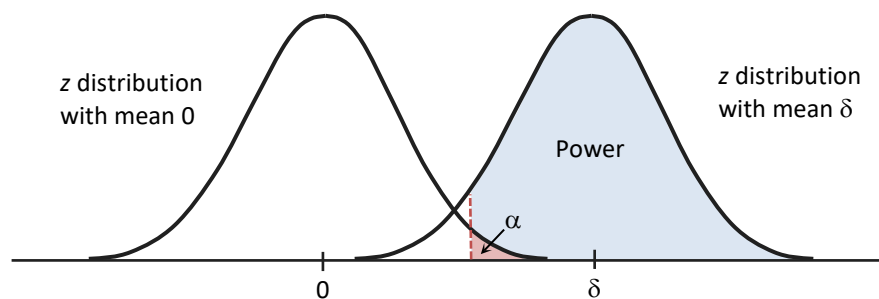
$$z < z_{\alpha}.$$

Comparing the z-statistic to the cut-off z-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 known standard deviations, σ_1 and σ_2 . Two good references for these general methods are Julious (2010) and Chow, Shao, and Wang (2008), although these texts do not specifically cover the unequal variance z-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 $z_{1-\alpha}$ based on the z distribution.
2. Calculate the power as the probability that the test statistic z is greater than $z_{1-\alpha}$ under the z distribution with mean δ :

$$\text{Power} = \Pr(z > z_{1-\alpha} | \delta).$$

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.

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 between 2 and 10 minutes. How many participants are needed to achieve 90% power at significance levels of 0.01 and 0.05?

It is known that the standard deviation is 4.1 minutes for Group 1, and 5.3 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 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	Sample Size
Alternative Hypothesis	Two-Sided
Power.....	0.90
Alpha.....	0.01 0.05
Group Allocation	Equal (N1 = N2)
Input Type.....	Means
μ_1	11 to 19 by 1
μ_2	9
σ_1	4.1
σ_2	5.3

Two-Sample Z-Tests Allowing Unequal Variance

Output

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

Numeric Reports

Numeric Results

Solve For: [Sample Size](#)
 Test Type: Two-Sample Unequal-Variance Z-Test
 Difference: $\delta = \mu_1 - \mu_2$
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power		Sample Size			Mean		Difference δ	Standard Deviation		Alpha
Target	Actual	N1	N2	N	μ_1	μ_2		σ_1	σ_2	
0.9	0.90197	168	168	336	11	9	2	4.1	5.3	0.01
0.9	0.90013	118	118	236	11	9	2	4.1	5.3	0.05
0.9	0.90345	75	75	150	12	9	3	4.1	5.3	0.01
0.9	0.90310	53	53	106	12	9	3	4.1	5.3	0.05
0.9	0.90197	42	42	84	13	9	4	4.1	5.3	0.01
0.9	0.90484	30	30	60	13	9	4	4.1	5.3	0.05
0.9	0.90345	27	27	54	14	9	5	4.1	5.3	0.01
0.9	0.90192	19	19	38	14	9	5	4.1	5.3	0.05
0.9	0.90778	19	19	38	15	9	6	4.1	5.3	0.01
0.9	0.91780	14	14	28	15	9	6	4.1	5.3	0.05
0.9	0.90872	14	14	28	16	9	7	4.1	5.3	0.01
0.9	0.91045	10	10	20	16	9	7	4.1	5.3	0.05
0.9	0.91680	11	11	22	17	9	8	4.1	5.3	0.01
0.9	0.92174	8	8	16	17	9	8	4.1	5.3	0.05
0.9	0.92697	9	9	18	18	9	9	4.1	5.3	0.01
0.9	0.90825	6	6	12	18	9	9	4.1	5.3	0.05
0.9	0.91506	7	7	14	19	9	10	4.1	5.3	0.01
0.9	0.91576	5	5	10	19	9	10	4.1	5.3	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. $N = N_1 + N_2$.
μ_1 and μ_2	The assumed population means.
δ	The difference between population means at which power and sample size calculations are made. $\delta = \mu_1 - \mu_2$.
σ_1 and σ_2	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 ($H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 \neq 0$). The comparison will be made using a two-sided, two-sample, unequal-variance Z-test, with a Type I error rate (α) of 0.01. The known standard deviation for Group 1 is assumed to be 4.1 and the known standard deviation for Group 2 is assumed to be 5.3. To detect a difference in means of $\mu_1 - \mu_2 = 11 - 9 = 2$ with 90% power, the number of needed subjects will be 168 in Group 1 and 168 in Group 2.

Two-Sample Z-Tests Allowing Unequal Variance

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%	168	168	336	210	210	420	42	42	84
20%	118	118	236	148	148	296	30	30	60
20%	75	75	150	94	94	188	19	19	38
20%	53	53	106	67	67	134	14	14	28
20%	42	42	84	53	53	106	11	11	22
20%	30	30	60	38	38	76	8	8	16
20%	27	27	54	34	34	68	7	7	14
20%	19	19	38	24	24	48	5	5	10
20%	19	19	38	24	24	48	5	5	10
20%	14	14	28	18	18	36	4	4	8
20%	14	14	28	18	18	36	4	4	8
20%	10	10	20	13	13	26	3	3	6
20%	11	11	22	14	14	28	3	3	6
20%	8	8	16	10	10	20	2	2	4
20%	9	9	18	12	12	24	3	3	6
20%	6	6	12	8	8	16	2	2	4
20%	7	7	14	9	9	18	2	2	4
20%	5	5	10	7	7	14	2	2	4

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, 210 subjects should be enrolled in Group 1, and 210 in Group 2, to obtain final group sample sizes of 168 and 168, respectively.

References

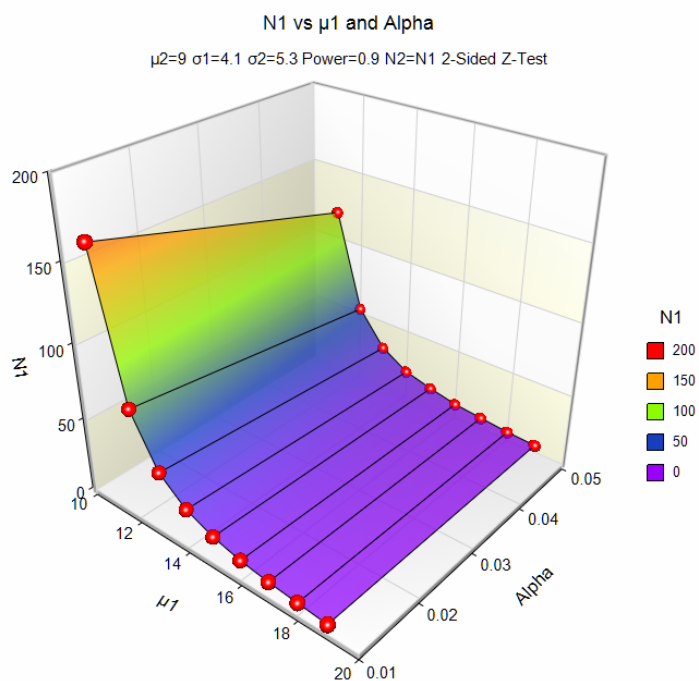
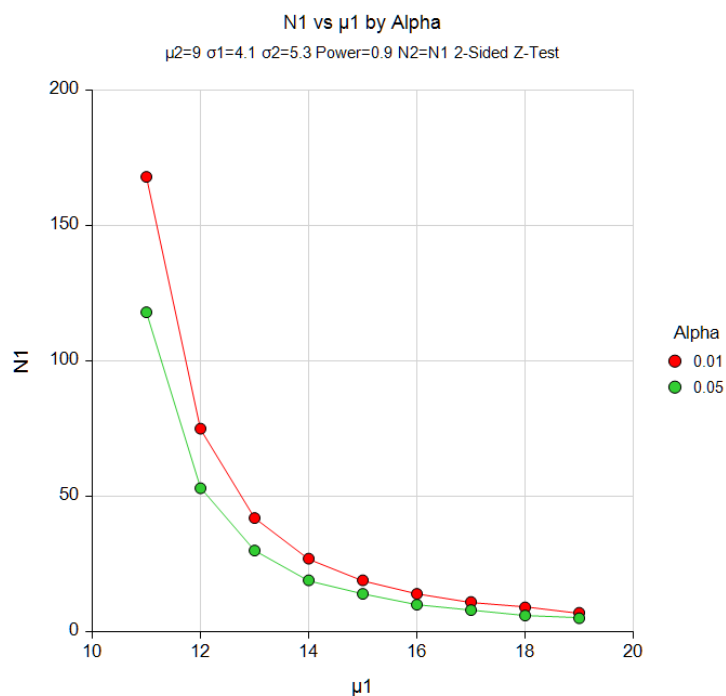
- Julious, S. A. 2010. Sample Sizes for Clinical Trials. Chapman & Hall/CRC. Boca Raton, FL.
 Chow, S. C., Shao, J., and Wang, H. 2008. Sample Size Calculations in Clinical Research (Second Edition). Chapman & Hall/CRC. Boca Raton, FL.
 Machin, D., Campbell, M., Fayers, P., and Pinol, A. 1997. Sample 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.

Two-Sample Z-Tests Allowing Unequal Variance

Plots Section

Plots



These plots show the relationship between the sample size and mean for Group 2 for the two alpha levels.

Two-Sample Z-Tests Allowing Unequal Variance

To specify the difference directly to get the same results, change **Input Type** to **Difference** and enter **2 to 10 by 1** 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 to 10 by 1**

Output

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

Numeric Results

Solve For: [Sample Size](#)
 Test Type: Two-Sample Unequal-Variance Z-Test
 Difference: $\delta = \mu_1 - \mu_2$
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power		Sample Size			Mean Difference δ	Standard Deviation		Alpha
Target	Actual	N1	N2	N		σ_1	σ_2	
0.9	0.90197	168	168	336	2	4.1	5.3	0.01
0.9	0.90013	118	118	236	2	4.1	5.3	0.05
0.9	0.90345	75	75	150	3	4.1	5.3	0.01
0.9	0.90310	53	53	106	3	4.1	5.3	0.05
0.9	0.90197	42	42	84	4	4.1	5.3	0.01
0.9	0.90484	30	30	60	4	4.1	5.3	0.05
0.9	0.90345	27	27	54	5	4.1	5.3	0.01
0.9	0.90192	19	19	38	5	4.1	5.3	0.05
0.9	0.90778	19	19	38	6	4.1	5.3	0.01
0.9	0.91780	14	14	28	6	4.1	5.3	0.05
0.9	0.90872	14	14	28	7	4.1	5.3	0.01
0.9	0.91045	10	10	20	7	4.1	5.3	0.05
0.9	0.91680	11	11	22	8	4.1	5.3	0.01
0.9	0.92174	8	8	16	8	4.1	5.3	0.05
0.9	0.92697	9	9	18	9	4.1	5.3	0.01
0.9	0.90825	6	6	12	9	4.1	5.3	0.05
0.9	0.91506	7	7	14	10	4.1	5.3	0.01
0.9	0.91576	5	5	10	10	4.1	5.3	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 z-test with alpha equal to 0.05. The standard deviation is known to be 24 lbs. for Group 1, and 26 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
μ_1	84
μ_2	74
σ_1	24
σ_2	26

Two-Sample Z-Tests Allowing Unequal Variance

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 Unequal-Variance Z-Test
 Difference: $\delta = \mu_1 - \mu_2$
 Hypotheses: $H_0: \delta \leq 0$ vs. $H_1: \delta > 0$

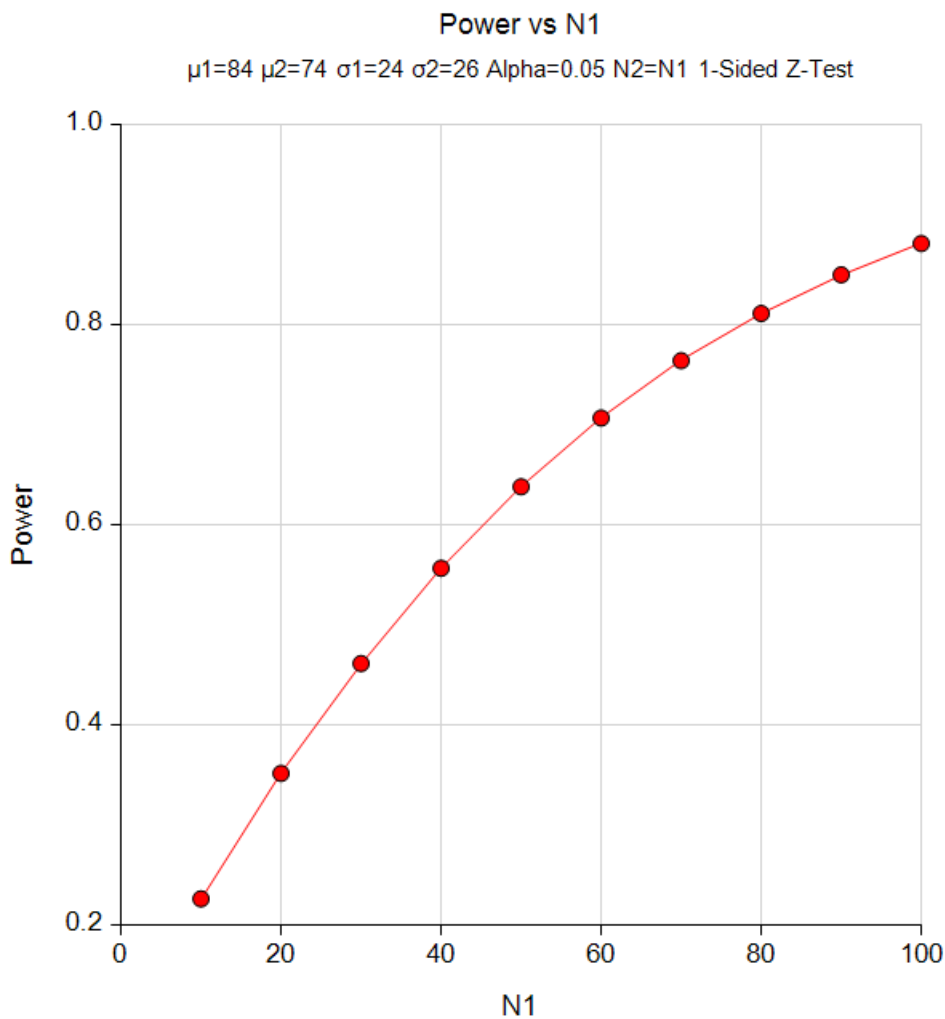
Power	Sample Size			Mean		Difference δ	Standard Deviation		Alpha
	N1	N2	N	μ_1	μ_2		σ_1	σ_2	
0.22628	10	10	20	84	74	10	24	26	0.05
0.35162	20	20	40	84	74	10	24	26	0.05
0.46140	30	30	60	84	74	10	24	26	0.05
0.55669	40	40	80	84	74	10	24	26	0.05
0.63816	50	50	100	84	74	10	24	26	0.05
0.70688	60	60	120	84	74	10	24	26	0.05
0.76414	70	70	140	84	74	10	24	26	0.05
0.81137	80	80	160	84	74	10	24	26	0.05
0.84997	90	90	180	84	74	10	24	26	0.05
0.88126	100	100	200	84	74	10	24	26	0.05

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

Two-Sample Z-Tests Allowing Unequal Variance

Plots Section

Plots



This plot shows the relationship between the power and sample size.

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 known 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
σ_1	6
σ_2	8

Reports Tab

Means, Difference Decimals	3
----------------------------------	----------

Two-Sample Z-Tests Allowing Unequal Variance

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 Unequal-Variance Z-Test
 Difference: $\delta = \mu_1 - \mu_2$
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power	Sample Size			Mean Difference δ	Standard Deviation		Alpha
	N1	N2	N		σ_1	σ_2	
0.9	40	40	80	5.125	6	8	0.05

If the true population mean weight loss for one group is 5.125 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 Chow, Shao, and Wang (2008)

Chow, Show, and Wang (2008) pages 63 and 64 present an example for a two-sided test in which the mean difference is 0.05, the standard deviation for each group is known to be 0.10, the power is 80%, and the significance level is 0.05. They calculate the per group sample size as 63.

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..... **0.80**
 Alpha..... **0.05**
 Group Allocation **Equal (N1 = N2)**
 Input Type..... **Difference**
 δ **0.05**
 σ_1 **0.1**
 σ_2 **0.1**

Output

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

Numeric Results

Solve For: **Sample Size**
 Test Type: Two-Sample Unequal-Variance Z-Test
 Difference: $\delta = \mu_1 - \mu_2$
 Hypotheses: $H_0: \delta = 0$ vs. $H_1: \delta \neq 0$

Power		Sample Size			Mean Difference	Standard Deviation		
Target	Actual	N1	N2	N	δ	σ_1	σ_2	Alpha
0.8	0.8013	63	63	126	0.05	0.1	0.1	0.05

The sample size of 63 per group matches Chow, Show, and Wang's result exactly.