Power and Sample Size Examples

Example 1: Retrospective calculation of power for a two-sample t-test. A study was performed in order to evaluate the effectiveness of two devices for improving the efficiency of gas home-heating systems. Energy consumption in houses was measured after one of the two devices was installed. The two devices were an electric vent damper (Damper 1) and a thermally activated vent damper (Damper 2). The energy consumption data for each damper type were measured. Compare the effectiveness of these two devices by determining whether or not there is any evidence of a difference in energy use when using an electric vent damper versus a thermally activated vent damper.

Row	Damper1	Damper2
1	7.87	12.28
2	9.43	7.23
3	7.16	2.97
4	8.67	8.81
5	12.31	9.27
6	9.84	11.29
7	10.04	8.29
8	12.62	9.96
9	7.62	10.30
10	11.12	16.06
11	13.43	14.24
12	9.07	11.43
13	6.94	10.28
14	10.28	13.60
15	9.37	5.94
16	7.93	10.36
17	13.96	6.85
18	6.80	6.72
19	4.00	10.21
20	8.58	8.61
21	8.00	11.62
22	5.98	11.21
23	15.24	10.95
24	8.54	7.62
25	11.09	10.40
26	11.70	12.92
27	12.71	15.12
28	6.78	13.47
29	9.82	8.47
30	12.91	11.70
31	10.35	7.73
32	9.60	8.37
33	9.58	7.29
34	9.83	10.49
35	9.52	8.69
36	10.64	8.26
37	6.62	7.69
38	5.20	12.19
39		5.56
40		9.76
41		7.15
42		12.69
43		13.38
44		13.11
45		10.50
46		14.35
47		13.42
48		6.35
49		9.83
50		12.16

Abbreviated write-up:

H_o: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 \neq 0$

p-value = .2598

Retain Ho. The sample results do not show significant evidence of a difference in energy use for the two dampers.

Since the data analysis did not reveal any statistically significant difference, we may wish to calculate the power of detecting a difference.

- If the power is low, then there is little chance of finding a significant difference (and nonsignificant results are likely even if real differences exist). We may want to get more observations to increase power and continue to evaluate the problem.

- If the power is high, then significant differences should be detectable, and so we can conclude that there is no meaningful difference.

The power depends on the magnitude of the difference I wish to detect. Maybe I am interested in detecting a difference in energy usage as small as 2 units, for example. Actually, we can investigate the power for several different effect sizes at the same time. In the Minitab output below, we have requested power calculations for differences in energy usage of 1.0, 1.5, 2.0, 2.5, 3.0, . . ., 4.5, 5.0 units.

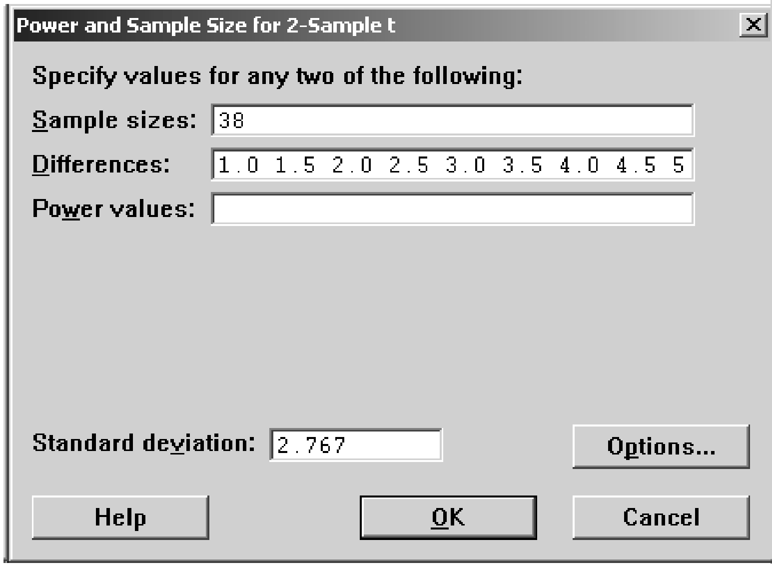
Power and Sample Size

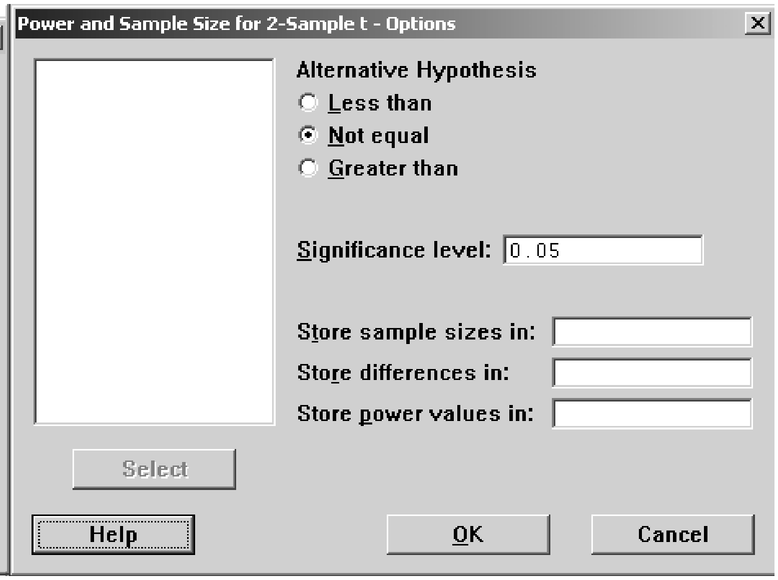
```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 2.767


            Sample
Difference    Size    Power
       1.0      38  0.34293
       1.5      38  0.64516
       2.0      38  0.87483
       2.5      38  0.97299
       3.0      38  0.99657
       3.5      38  0.99975
       4.0      38  0.99999
       4.5      38  1.00000
       5.0      38  1.00000

The sample size is for each group.
```





<u>Specifying differences</u>
For a two-sample t-test, express the "Difference" as the difference between the population means that you would like to be able to detect.

If you choose "Less than" as your alternative hypothesis, then you must enter a negative value in "Differences." If you choose "Greater than," you must enter a positive value. This is because with a one-tailed test, you have no power to detect an effect in the opposite direction.

Example: one-sided test.
Suppose instead that we wish to test the hypothesis that the electric vent damper (Damper 1) consumes less energy than the thermally activated vent damper (Damper 2).

H_o: $\mu_1 - \mu_2 = 0$ Alternative Hypothesis = "Less than"
H_a: $\mu_1 - \mu_2 < 0$

If you want to check the power of detecting a difference of 2 units, then set "Difference" = −2, or, if desired, we can investigate the power for several sizes of differences.

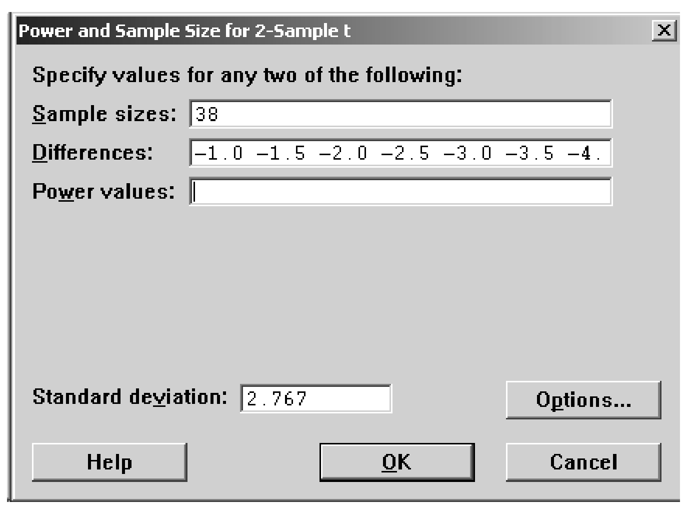
Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus <)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 2.767

            Sample
Difference    Size    Power
      -1.0      38  0.46654
      -1.5      38  0.75692
      -2.0      38  0.93013
      -2.5      38  0.98800
      -3.0      38  0.99881
      -3.5      38  0.99993
      -4.0      38  1.00000
      -4.5      38  1.00000
      -5.0      38  1.00000

The sample size is for each group.
```



Comments:

(1) Notice that a one-sided test has more power than a two-sided test for the same number of observations. Compare the entries below for a two-sided test and a one-sided test.

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 2.767

             Sample
Difference     Size    Power
       1.0       38  0.34293
       1.5       38  0.64516
       2.0       38  0.87483
       2.5       38  0.97299
       3.0       38  0.99657
       3.5       38  0.99975
       4.0       38  0.99999
       4.5       38  1.00000
       5.0       38  1.00000
```

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus <)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 2.767

             Sample
Difference     Size    Power
      -1.0       38  0.46654
      -1.5       38  0.75692
      -2.0       38  0.93013
      -2.5       38  0.98800
      -3.0       38  0.99881
      -3.5       38  0.99993
      -4.0       38  1.00000
      -4.5       38  1.00000
      -5.0       38  1.00000
```

(2) Notice that when you demand a higher degree of evidence in order to reject Ho (i.e., demand a lower p-value), the power for such a test is lower. Compare the entries below for significance levels set at $\alpha = .05$ and $\alpha = .01$.

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 2.767

             Sample
Difference     Size    Power
       1.0       38  0.34293
       1.5       38  0.64516
       2.0       38  0.87483
       2.5       38  0.97299
       3.0       38  0.99657
       3.5       38  0.99975
       4.0       38  0.99999
       4.5       38  1.00000
       5.0       38  1.00000
```

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.01  Assumed standard deviation = 2.767

             Sample
Difference     Size    Power
       1.0       38  0.15021
       1.5       38  0.39523
       2.0       38  0.69287
       2.5       38  0.89859
       3.0       38  0.97947
       3.5       38  0.99754
       4.0       38  0.99983
       4.5       38  0.99999
       5.0       38  1.00000
```

Example 2: Prospective calculation of sample size to ensure high power for detecting effects that you have determined to be important.

Ordinary corn doesn't have as much of the amino acid lysine as animals need in their feed. Plant scientists have developed varieties of corn that have increased amounts of lysine. Suppose you are planning an experiment to test the effectiveness of the high-lysine corn feed. You plan to measure the weight gains of one-day-old chicks after 21 days for a group of chicks fed the high lysine corn and for a control group fed the normal corn.

You will carry out the analysis with a two-sample t test

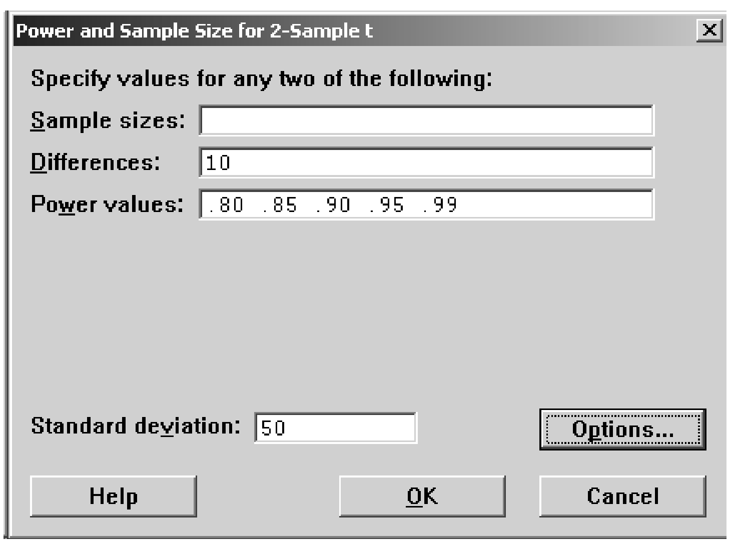
H_o: $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 - \mu_2 > 0$

where μ_1 = mean weight gain for the high lysine group and μ_2 = mean weight gain for the control group.

(a) You would like to be able to detect a difference in average weight gain as small as 10 grams with at least 80% power. You plan to use a significance level of .05 (i.e., you will reject H_o if p-value < .05). Investigate the sample sizes needed for such a study. Past experience shows that the standard deviation in weight gains for chicks fed normal corn is about 40 grams, and about 50 grams for chicks fed high lysine corn.

Comment:

The calculation of power requires knowledge of the standard deviations of the populations from which we sample, which we don't have. So we estimate the population standard deviations from knowledge of the process or a pilot study. Use the larger of the two in the calculation of power (conservative).



Power and Sample Size

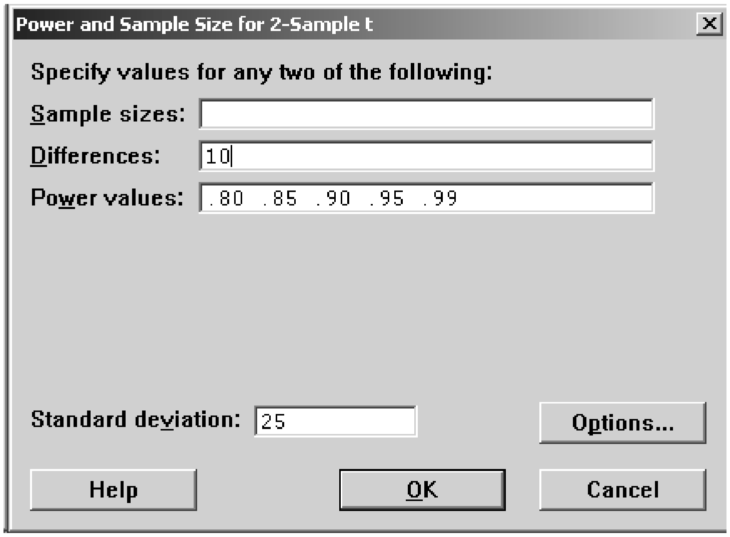
```
2-Sample t Test

Testing mean 1 = mean 2 (versus >)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 50


            Sample  Target
Difference    Size   Power  Actual Power
        10     310    0.80      0.800218
        10     361    0.85      0.850743
        10     429    0.90      0.900078
        10     542    0.95      0.950067
        10     790    0.99      0.990054

The sample size is for each group.
```

(b) The researcher is disappointed that the required sample sizes are so large due to the cost of doing such a large study. She considers modifying the experiment so as to reduce the variability in the chicks' weight gains. She finds that she can cut the standard deviation in half by switching to a different supplier of chicks and a different research facility. However, the cost of each observation would be doubled. Would these measures be cost effective? That is, would the modified experiment be less costly?



Power and Sample Size

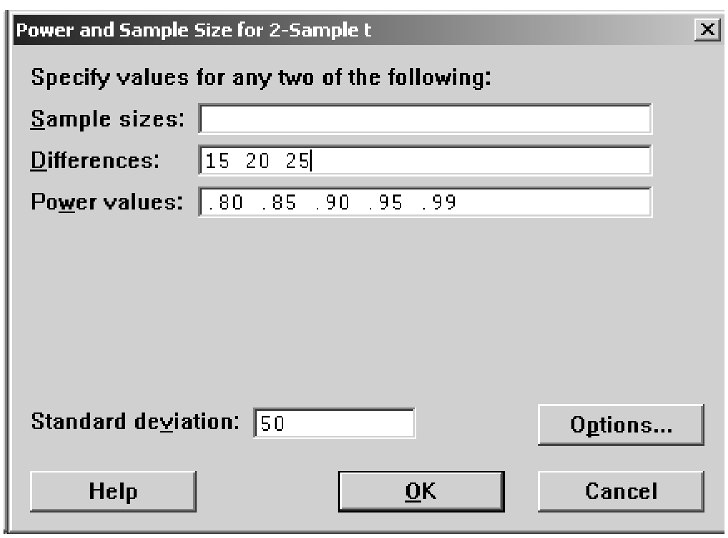
```
2-Sample t Test

Testing mean 1 = mean 2 (versus >)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 25


             Sample  Target
Difference     Size   Power  Actual Power
        10       78    0.80      0.800147
        10       91    0.85      0.851556
        10      108    0.90      0.900643
        10      136    0.95      0.950051
        10      198    0.99      0.990051

The sample size is for each group.
```

(c) After more consideration, the researcher decides that it's probably not important to be able to detect such a small difference in the average weight gains. She decides to investigate the power associated with a difference in weight gains of 15, 20, and 25 g, using the original supplier and the original facility.



Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus >)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 50


             Sample  Target
Difference     Size   Power  Actual Power
        15      139    0.80      0.802340
        15      161    0.85      0.851086
        15      191    0.90      0.900016
        15      242    0.95      0.950581
        15      352    0.99      0.990130
        20       78    0.80      0.800147
        20       91    0.85      0.851556
        20      108    0.90      0.900643
        20      136    0.95      0.950051
        20      198    0.99      0.990051
        25       51    0.80      0.805899
        25       59    0.85      0.854276
        25       70    0.90      0.902966
        25       88    0.95      0.951425
        25      127    0.99      0.990064

The sample size is for each group.
```