Chapter 610

Multiple One-Sample or Paired T-Tests

Introduction

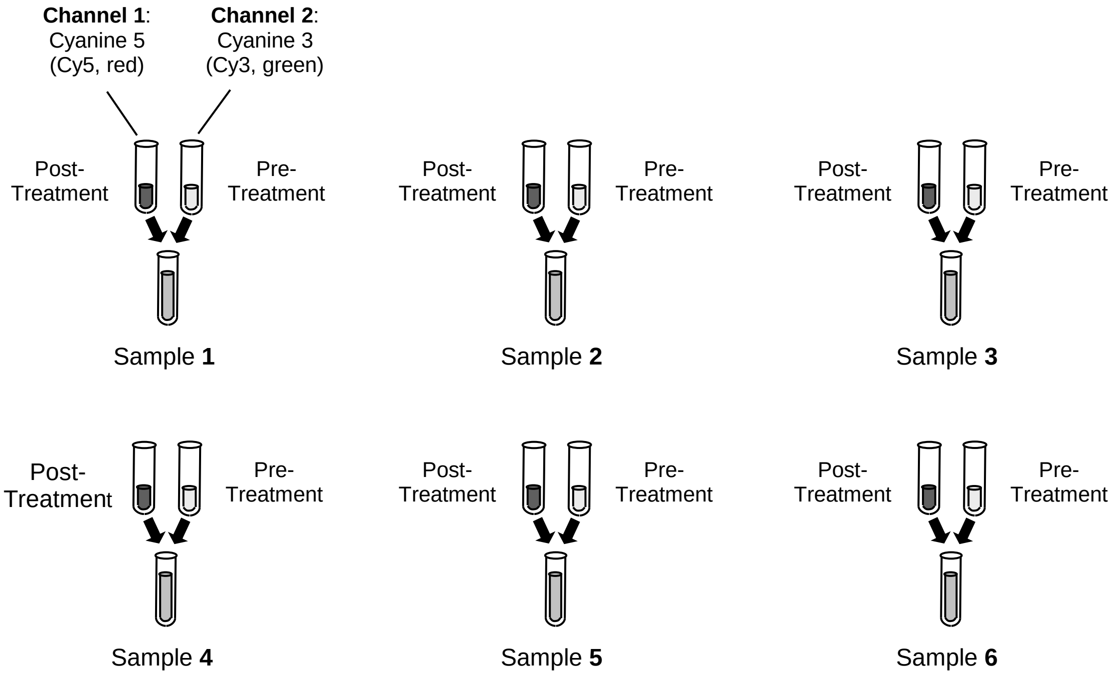
This chapter describes how to estimate power and sample size (number of arrays) for paired and one sample high-throughput studies using the Multiple One-Sample or Paired T-Tests. False discovery rate and experiment-wise error rate control methods are available in this procedure. Values that can be varied in this procedure are power, false discovery rate and experiment-wise error rate, sample size (number of arrays), the minimum mean difference detected, the standard deviation, and in the case of false discovery rate control, the number of genes with minimum mean difference.

Paired Design (Two-Channel Arrays)

The paired design is often used in two-channel experiments when the gene expression comparison to be made involves a natural pairing of experimental units.

As an example, suppose 6 cell samples will be available for comparison. A portion of each of the 6 cell samples (before treatment) is to be reserved as a control. The same treatment will then be given to each of the 6 remaining portions of the samples. It is of interest to determine the genes that are differentially expressed when the treatment is given. In this scenario there is a natural before/after treatment pairing for each sample. The reserved control portions of each sample will be labeled with Cyanine 3 (Cy3, green) dye, while the treatment portions are to be labeled with Cyanine 5 (Cy5, red) dye. From each sample, the labeled control and the labeled treatment portions will be mixed and exposed to an array. The control and treatment portions compete to bind at each spot. The expression of treatment and control samples for each gene will be measured with laser scanning. A pre-processing procedure is then used to obtain expression difference values for each gene. In this example, the result will be 6 relative expression values (e.g., $\text{Log}_2(\textit{Post} / \textit{Pre})$) for each gene represented on the arrays.

Paired Design, Six Arrays



Null and Alternative Hypotheses

The paired test null hypothesis for each gene is H_0: $\mu_{diff} = \mu_0$, where μ_{diff} is the true mean difference in expression for a particular gene, and μ_0 is the hypothesized difference in expression. The alternative hypothesis may be any one of the following: H_a: $\mu_{diff} < \mu_0$, H_a: $\mu_{diff} > \mu_0$, or H_a: $\mu_{diff} \neq \mu_0$. The choice of the alternative hypothesis depends upon the goals of the research. For example, if the goal of the experiment is to determine which genes are differentially expressed, without regard to direction, the alternative hypothesis would be H_a: $\mu_{diff} \neq \mu_0$. If, however, the goal is to identify only genes which have increased expression after the treatment is applied, the alternative hypothesis would be H_a: $\mu_{diff} > \mu_0$. A common value for μ_0 in a paired sample experiment is 0.

T-Test Formula

For testing the hypothesis H_0: $\mu_{diff} = \mu_0$, the formula for the paired T-statistic is:

$$T_{paired} = \frac{\bar{x}_{paired} - \mu_o}{\frac{s_{paired}}{\sqrt{n}}}$$

where $\bar{x}_{paired}$ is mean difference in expression of n replicates for a given gene, μ_0 is the hypothesized mean, and s_{paired} is standard deviation of the differences of the n replicates. If the assumptions (described below) of the test are met, the distribution of T_{paired} is the standard t distribution with $n - 1$ degrees of freedom. P-values are obtained from T_{paired} by finding the proportion of the t distribution that is more extreme than T_{paired} .

Assumptions

The assumptions of the paired t-test are:

1. The data are continuous (not discrete). Because of the large range of possible intensities, microarray expression values can be considered continuous.
2. The data, i.e., the expression differences for the pairs, follow a normal probability distribution. This assumption can be examined in microarray data only if the number of arrays in the experiment is reasonably large (>100).
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. If the samples used in the microarray experiment are not random, bias may easily be introduced into the results.

One-Sample Design

The one-sample design is the simplest of all designs. A single mRNA or cDNA sample is obtained from each experimental unit of a single group. Each sample is exposed to a single microarray, resulting in a single expression value for each gene for each unit of the group. The goal is to determine for each gene whether there is evidence that the expression is different from some null value. This design may be useful for determining whether or not each gene is expressed at all, or for comparing expression of each gene to a hypothesized expression level.

Null and Alternative Hypotheses

The one-sample null hypothesis for each gene is H_0: $\mu = \mu_0$, where μ is the true mean expression for that particular gene, and μ_0 is the hypothesized mean, or the mean to be compared against. The alternative hypothesis may be any one of the following: H_a: $\mu < \mu_0$, H_a: $\mu > \mu_0$, or H_a: $\mu \neq \mu_0$. The choice of the alternative hypothesis depends upon the goals of the research. For example, if the goal of the experiment is to determine which genes are expressed above a certain level, the alternative hypothesis would be H_a: $\mu > \mu_0$.

T-Test Formula

For testing the hypothesis H_0: $\mu = \mu_0$, the formula for the one-sample T-statistic is:

$$T = \frac{\overline{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

where $\overline{x}$ is mean expression of n replicates for a given gene, μ_0 is the hypothesized mean, and s is standard deviation of the n replicates. If the assumptions (described below) of the test are met, the distribution of T is the standard t distribution with $n - 1$ degrees of freedom. P-values are obtained from T by finding the proportion of the t distribution that is more extreme than T.

Assumptions

The assumptions of the one-sample T-test are:

1. The data are continuous (not discrete). Because of the large range of possible intensities, microarray expression values can be considered continuous.
2. The data follow the normal probability distribution. This assumption can be examined in microarray data only if the number of arrays in the experiment is reasonably large (>300).
3. The sample is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample. If the samples used in the microarray experiment are not random, bias may easily be introduced into the results.

Technical Details

Multiple Testing Adjustment

When a one-sample/paired T-test is run for a replicated microarray experiment, the result is a list of P-values (Probability Level) that reflect the evidence of difference in expression. When hundreds or thousands of genes are investigated at the same time, many ‘small’ P-values will occur by chance, due to the natural variability of the process. It is therefore requisite to make an appropriate adjustment to the P-value (Probability Level), such that the likelihood of a false conclusion is controlled.

False Discovery Rate Table

The following table (adapted to the subject of microarray data) is found in Benjamini and Hochberg’s (1995) false discovery rate article. In the table, m is the total number of tests, m_0 is the number of tests for which there is no difference in expression, R is the number of tests for which a difference is declared, and U, V, T, and S are defined by the combination of the declaration of the test and whether or not a difference exists, in truth.

	Declared Not Different	Declared Different	Total
A true difference in expression does not exist	U	V	m_0
There exists a true difference in expression	T	S	$m - m_0$
Total	$m - R$	R	m

In the table, the m is the total number of hypotheses tested (or total number of genes) and is assumed to be known in advance. Of the m null hypotheses tested, m_0 is the number of tests for which there is no difference in expression, R is the number of tests for which a difference is declared, and U, V, T, and S are defined by the combination of the declaration of the test and whether or not a difference exists, in truth. The random variables U, V, T, and S are unobservable.

Need for Multiple Testing Adjustment

Following the calculation of a raw P-value (Probability Level) for each test, P-value adjustments need be made to account in some way for multiplicity of tests. It is desirable that these adjustments minimize the number of genes that are falsely declared different (V) while maximizing the number of genes that are correctly declared different (S). To address this issue the researcher must know the comparative value of finding a gene to the price of a false positive. If a false positive is very expensive, a method that focuses on minimizing V should be employed. If the value of finding a gene is much higher than the cost of additional false positives, a method that focuses on maximizing S should be used.

Error Rates – P-Value Adjustment Techniques

Below is a brief description of three common error rates that are used for control of false positive declarations. The commonly used P-value adjustment technique for controlling each error rate is also described.

Per-Comparison Error Rate (PCER) – No Multiple Testing Adjustment

The per-comparison error rate (PCER) is defined as

$$\text{PCER} = E(V)/m\,,$$

where $E(V)$ is the expected number of genes that are falsely declared different, and m is the total number of tests. Preserving the PCER is tantamount to ignoring multiple testing altogether. If a method is used which controls a PCER of 0.05 for 1,000 tests, approximately 50 out of 1,000 tests will falsely be declared significant. Using a method that controls the PCER will produce a list of genes that includes most of the genes for which there exists a true difference in expression (i.e., maximizes S), but it will also include a very large number of genes which are falsely declared to have a true difference in expression (i.e., does not appropriately minimize V). Controlling the PCER should be viewed as overly weak control of Type I error.

To obtain P-values (Probability Levels) that control the PCER, no adjustment is made to the P-value. To determine significance, the P-value is simply compared to the designated alpha.

Experiment-Wise Error Rate (EWER)

The experiment-wise error rate (EWER) is defined as

$$\text{EWER} = \Pr(V > 0),$$

where V is the number of genes that are falsely declared different. Controlling EWER is controlling the probability that a single null hypothesis is falsely rejected. If a method is used which controls a EWER of 0.05 for 1,000 tests, the probability that any of the 1,000 tests (collectively) is falsely rejected is 0.05. Using a method that controls the EWER will produce a list of genes that includes a small (depending also on sample size) number of the genes for which there exists a true difference in expression (i.e., limits S, unless the sample size is very large). However, the list of genes will include very few or no genes that are falsely declared to have a true difference in expression (i.e., stringently minimizes V). Controlling the EWER should be considered very strong control of Type I error.

Assuming the tests are independent, the well-known Bonferroni P-value adjustment produces adjusted P-values (Probability Levels) for which the EWER is controlled. The Bonferroni adjustment is applied to all m unadjusted P-values (p_j) as

$$\tilde{p}_j = \min(mp_j, 1)\,.$$

That is, each P-value (Probability Level) is multiplied by the number of tests, and if the result is greater than one, it is set to the maximum possible P-value of one.

False Discovery Rate (FDR)

The false discovery rate (FDR) (Benjamini and Hochberg, 1995) is defined as

$$\text{FDR} = E(\frac{V}{R}1_{\{R>0\}}) = E(\frac{V}{R} \mid R>0)\Pr(R>0),$$

where R is the number of genes that are declared significantly different, and V is the number of genes that are falsely declared different. Controlling FDR is controlling the expected *proportion* of falsely declared differences (false discoveries) to declared differences (true and false discoveries, together). If a method is used which controls a FDR of 0.05 for 1,000 tests, and 40 genes are declared different, it is expected that 40*0.05 = 2 of the 40 declarations are false declarations (false discoveries). Using a method that controls the FDR will produce a list of genes that includes an intermediate (depending also on sample size) number of genes for which there exists a true difference in expression (i.e., moderate to large S). However, the list of genes will include a small number of genes that are falsely declared to have a true difference in expression (i.e., moderately minimizes V). Controlling the FDR should be considered intermediate control of Type I error.

Multiple Testing Adjustment Comparison

The following table gives a summary of the multiple testing adjustment procedures and error rate control. The power to detect differences also depends heavily on sample size.

Common Adjustment Technique	Error Rate Controlled	Control of Type I Error	Power to Detect Differences
None	PCER	Minimal	High
Bonferroni	EWER	Strict	Low
Benjamini and Hochberg	FDR	Moderate	Moderate/High

Type I Error: Rejection of a null hypothesis that is true.

Calculating Power

When the standard deviation is unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which $Diff > 0$. Additional details of calculating power in the one-sample/paired scenario are found in the PASS chapter for One Mean.

1. Find t_α such that $1 - T_{df}\left(t_\alpha\right) = \alpha$, where $T_{df}\left(t_\alpha\right)$ is the area under a central-*t* curve to the left of *x* and $df = n - 1$.

2. Calculate: $x_a = t_\alpha \frac{\sigma}{\sqrt{n}}$.

3. Calculate the non-centrality parameter: $\lambda_a = \frac{Diff}{\frac{\sigma}{\sqrt{n}}}$.

4. Calculate: $t_a = \frac{x_a - Diff}{\frac{\sigma}{\sqrt{n}}} + \lambda_a$.

5. Calculate: Power = $1 - T'_{df,\lambda}(t_a)$, where $T'_{df,\lambda}(x)$ is the area under a noncentral-*t* curve with degrees of freedom *df* and noncentrality parameter λ to the left of *x*.

Adjusting Alpha

Experiment-wise Error Rate

When the Bonferroni method will be used to control the experiment-wise error rate, α_{EWER}, of all tests, the adjusted α, α_{ADJ}, for each test is given by

$$\alpha_{ADJ} = \frac{\alpha_{EWER}}{Number\ of\ Tests}$$

α_{ADJ} is the value that is used in the power and sample size calculations.

False Discovery Rate

When a false discovery rate controlling method will be used to control the false discovery rate for the experiment, *fdr*, the adjusted α, α_{ADJ}, for each test is given by Jung (2005) and Chow, Shao, and Wang (2008):

$$\alpha_{ADJ} = \frac{(K)(1-\beta)(fdr)}{(N_T - K)(1 - fdr)}$$

where *K* is the number of genes with differential expression, β is the probability of a Type II error (not declaring a gene significant when it is), and N_T is the total number of tests.

α_{ADJ} is the value that is used in the power and sample size calculations. Because α_{ADJ} depends on β, α_{ADJ} must be solved iteratively when the calculation of power is desired.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design and Options tabs. 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 involved in the power and sample size calculations.

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 false discovery rate (or alpha) error level.

Select *Power* when you want to calculate the power of an experiment.

Test

Test Type

Select the type of test that will be used when the analysis of the gene expression data is carried out.

- **T**

  The T-Test assumes the paired differences come from a normal distribution with UNKNOWN standard deviation (i.e., a standard deviation that will be estimated from the data).

- **Z**

  The Z-Test assumes the paired differences come from a normal distribution with KNOWN standard deviation.

Recommendation: Because it very rare to know the true standard deviation of paired differences in advance, T is the recommended test statistic.

Alternative Hypothesis

Specify whether the hypothesis test for each gene is one-sided (directional) or two-sided (non-directional).

Recommendation: In most paired experiments, differential expression in either direction (up-regulation or down-regulation) is of interest. Such experiments should have the Two-Sided alternative hypothesis.

For experiments for determining whether there is expression above some threshold, a One-Sided alternative hypothesis is recommended. Often regulations dictate that the FDR or EWER level be divided by 2 for One-Sided alternative tests.

Error Rates

Power for each Test

Power is the probability of rejecting each null hypothesis when it is false. Power is equal to 1-Beta.

The POWER for each gene represents that probability of detecting differential expression when it exists.

RANGE: The valid range is from 0 to 1.

RECOMMENDED: Popular values for power are 0.8 and 0.9.

NOTES: You can enter a range of values such as *.70 .80 .90* or *.70 to .95 by .05*.

False Discovery (Alpha) Method

A type I error is declaring a gene to be differentially expressed when it is not. The two most common methods for controlling type I error in microarray expression studies are false discovery rate (FDR) control and Experiment-wise Error Rate (EWER) control.

- **FDR**

  Controlling the false discovery rate (FDR) controls the PROPORTION of genes that are falsely declared differentially expressed. For example, suppose that an FDR of 0.05 is used for 10000 tests (on 10000 genes). If differential expression is declared for 100 of the 10000 genes, 5 of the 100 genes are expected to be false discoveries.

- **EWER**

  Controlling the experiment-wise error rate (EWER) controls the PROBABILITY of ANY false declarations of differential expression, across all tests. For example, suppose that an EWER of 0.05 is used for 10000 tests (on 10000 genes). If differential expression is declared for 100 of the 10000 genes, the probability that even one of the 100 declarations is false is 0.05.

Recommendation: For exploratory studies where a list of candidate genes for further study is the goal, FDR is the recommended Type I error control method, because of its higher power.

For confirmatory studies where final determination of differential expression is the goal, EWER is the recommended Type I error control method, because of its strict control of false discoveries.

FDR or EWER Value

Specify the value for the False Discovery (Alpha) Method selected above.

RANGE: These levels are bounded by 0 and 1. Commonly, the chosen level is between 0.001 and 0.250

RECOMMENDED: FDR or EWER is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

NOTE: You can enter a list of values such as *.05 .10 .15* or *.05 to .15 by .01*.

Sample Size

N (Number of Arrays)

Enter a value for the sample size (N). This is the number of arrays in the experiment. For two-channel paired experiments, this is the number of arrays, not the number of samples. You may enter a range such as *10 to 100 by 10* or a list of values separated by commas or blanks.

Effect Size

D (Minimum Mean Difference Detected)

Specify the true mean difference in expression (D) such that genes with true mean difference above D will be detected at the given power and corresponding sample size.

In paired expression studies, it is very common that the difference in expression is measured on the log scale (e.g., log2(A) – log2(B)). Values of D should reflect the differences that will be used in testing. For example, if log2 differences are used, D = 1 implies a two-fold difference in expression, while D = 2 implies a four-fold difference in expression.

When D is large, the resulting sample size will only detect the genes with extreme differential expression.

When D is small, a larger sample size is required to have power sufficient to detect these small differences in expression.

You can enter a range of values such as *1 2 3* or *0.2 to 2 by 0.1*.

S (Standard Deviation of Paired Differences in Expression)

Specify the standard deviation of paired differences. This standard deviation is assumed for all tests.

S should be on the same scale as D.

To obtain the standard deviation of paired differences from the standard deviation of expression, use SDpaired = (sqrt(2))*(SDexpression).

Because the true variation in paired differences will vary from gene to gene, it is recommended that a range of values be entered here.

You can enter a range of values such as *1 2 3 4 5* or *0.2 to 2 by 0.1*.

Number of Tests

Number of Tests

Specify the number of hypothesis tests that will be made.

This number will usually be the number of genes summarized on each array minus the number of housekeeping genes.

Only one number may be entered in this box.

Number of Tests – FDR Only

K (Number of Tests with Mean Difference > D)

Specify the number of tests for which a true mean difference in expression greater than D is expected.

K for EWER

The choice of K has no direct effect on the calculation of power or sample size when the False Discovery (Alpha) Method is set to EWER. K is not used when False Discovery (Alpha) method is set to EWER.

K for FDR

The choice of K has direct effect on the calculation of power or sample size when the False Discovery (Alpha) Method is set to FDR.

You can enter a range of values such as *10 20 30 40 50* or *20 to 100 by 10*.

Options Tab

The Options tab contains convergence options that are rarely changed.

Convergence Options

FDR Power Convergence

When FDR is selected for False Discovery (Alpha) Method, and Find (Solve For) is set to Power, the corresponding search algorithm will converge when the search criteria is below this value.

This value will rarely be changed from the default value.

RECOMMENDED: 0.0000000001

Example 1 – Finding Power

This example examines the power to detect differential expression for an experiment that involved 22 two-channel arrays. Two samples were obtained from each of 22 individuals. One of the two samples was randomly assigned the treatment and the other remained as the control. Following treatment, the two samples were exposed to a single microarray. Each microarray produced intensity information for 10,000 genes. The 22 arrays were pre-processed by subtracting the control intensity (Log2) from the treatment intensity for each gene on each array. Thus, a positive value implies upward expression in the treatment, while a negative value implies down-regulation in the treatment. In this example, the paired T-test was used to determine which genes were differentially expressed (upward or downward) following exposure to the treatment.

The researchers found very few differentially expressed genes, and wish to examine the power of the experiment to detect two-fold differential expression (Log2-scale difference of 1). Typical standard deviations of the Log2 paired differences ranged from 0.2 to 2.0.

The researchers guess the number of genes with at least 2-fold differential expression to be around 50, but will examine the effect of this estimate on power by trying 10 and 100 genes as well. A false discovery rate of 0.05 was used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple One-Sample or Paired T-Tests** procedure window by expanding **Means**, then **One Mean**, then clicking on **Multiple Tests**, and then clicking on **Multiple One-Sample or 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**.

Option	Value
Design Tab	
Solve For	**Power**
Test Type	**T**
Alternative Hypothesis	**Two-Sided**
False Discovery (Alpha) Method	**FDR (False Discovery Rate)**
FDR or EWER Value	**0.05**
N (Number of Arrays)	**22**
D (Difference)	**1.0**
S (Standard Deviation)	**0.2 to 2 by 0.2**
Number of Tests	**10000**
K (Number > D)	**10 50 100**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output. The calculations should take a few moments.

Numeric Results

```
Numeric Results for Multiple One-Sample T-Tests
Null Hypothesis: MeanDiff = 0     Alternative Hypothesis: MeanDiff ≠ 0
Number of Tests: 10000

                           Minimum      Std.                 Number       False
             Number           Mean      Dev.    Effect        Tests   Discovery
          of Arrays     Difference  of Diff.      Size    To Detect        Rate
Power           (N)            (D)       (S)      (ES)          (K)       (FDR)       Beta
1.00000          22            1.0       0.2     5.000           10      0.0500    0.00000
1.00000          22            1.0       0.2     5.000           50      0.0500    0.00000
1.00000          22            1.0       0.2     5.000          100      0.0500    0.00000
1.00000          22            1.0       0.4     2.500           10      0.0500    0.00000
1.00000          22            1.0       0.4     2.500           50      0.0500    0.00000
1.00000          22            1.0       0.4     2.500          100      0.0500    0.00000
0.98617          22            1.0       0.6     1.667           10      0.0500    0.01383
0.99793          22            1.0       0.6     1.667           50      0.0500    0.00207
0.99924          22            1.0       0.6     1.667          100      0.0500    0.00076
0.71696          22            1.0       0.8     1.250           10      0.0500    0.28304
0.89092          22            1.0       0.8     1.250           50      0.0500    0.10908
0.93538          22            1.0       0.8     1.250          100      0.0500    0.06462
.                 .              .         .         .            .           .          .
.                 .              .         .         .            .           .          .
.                 .              .         .         .            .           .          .

References
Jung, S.-H. 2005. Sample size for FDR-control in microarray data analysis. Bioinformatics: Vol. 21 no. 14, pp.
3097-3104. Oxford University Press.
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 individual probability of detecting each test with true mean difference > D.
N is the number of arrays required to achieve the corresponding power.
D is the smallest difference in expression for which this power and sample size are valid.
S is the standard deviation estimate for the paired differences used in each test.
ES, or D/S, is the relative magnitude of the true mean expression difference for the test where the true mean
difference > D.
K is the number of tests with true mean difference > D.
FDR is the expected proportion of false declarations of differential expression to total declarations of
differential expression.
Beta is the individual probability of failing to detect each test with true mean difference > D.

Summary Statements
A sample size of 22 achieves 100.00% power for each test to detect a true difference in
expression of at least 1.0 with an estimated standard deviation of 0.2 with a false discovery
rate of 0.0500 using a two-sided one-sample T-Test. Of the 10 tests with anticipated true mean
difference in expression > 1.0, 9 are expected to be detected.
```

This report shows the values of each of the parameters, one scenario per row. The values of power and beta were calculated from the other parameters.

The definitions of each column are given in the Report Definitions section.

Additional Numeric Result Detail

Additional Numeric Result Detail for Multiple One-Sample T-Tests
Number Tests: 10000

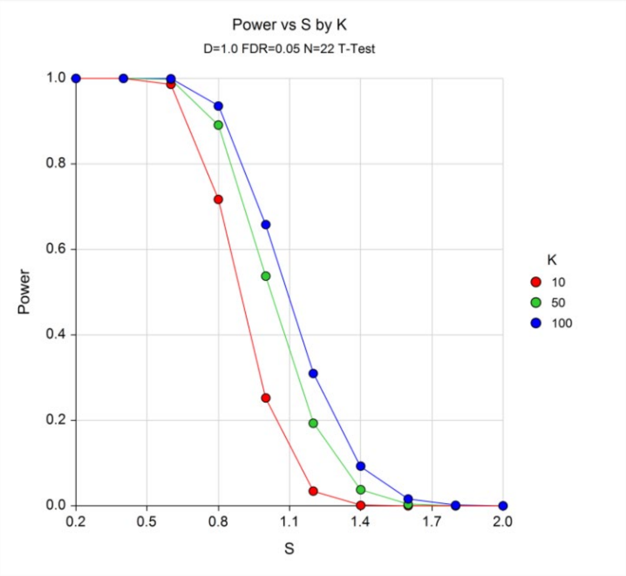
Power	Number of Arrays (N)	Minimum Mean Difference (D)	Std. Dev. of Diff. (S)	Number Tests To Detect (K)	False Discovery Rate (FDR)	Single Test Alpha	Prob To Detect All K
1.00000	22	1.0	0.2	10	0.0500	0.0000527	1.00000
1.00000	22	1.0	0.2	50	0.0500	0.0002645	1.00000
1.00000	22	1.0	0.2	100	0.0500	0.0005316	1.00000
1.00000	22	1.0	0.4	10	0.0500	0.0000527	1.00000
1.00000	22	1.0	0.4	50	0.0500	0.0002645	1.00000
1.00000	22	1.0	0.4	100	0.0500	0.0005316	1.00000
0.98617	22	1.0	0.6	10	0.0500	0.0000520	0.86996
0.99793	22	1.0	0.6	50	0.0500	0.0002639	0.90158
0.99924	22	1.0	0.6	100	0.0500	0.0005312	0.92649
0.71696	22	1.0	0.8	10	0.0500	0.0000378	0.03589
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

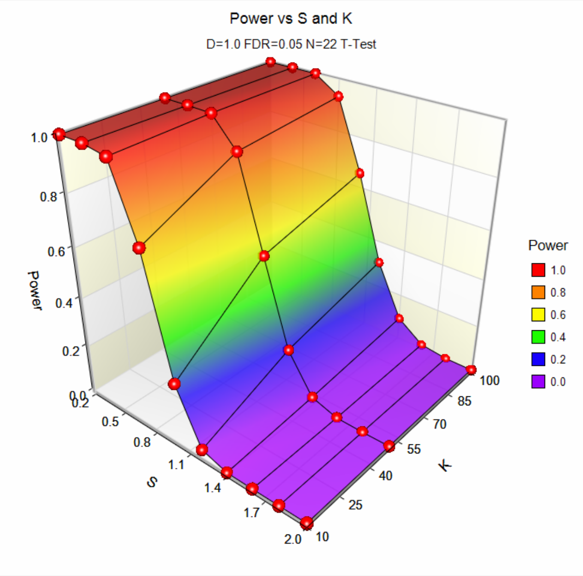
Report Definitions
Power is the individual probability of detecting each test where the true mean difference > D.
N is the number of arrays required to achieve the corresponding power.
D is the smallest difference in expression for which this power and sample size are valid.
S is the standard deviation estimate for the paired differences used in each test.
K is the number of genes with true mean difference > D.
FDR is the expected proportion of false declarations of differential expression to total declarations of differential expression.
Single Test Alpha is the probability of falsely declaring differential expression for an individual test.
Prob to Detect All K is the probability of declaring differential expression for all K tests that have true mean difference > D.

This report shows additionally the single gene alpha and the probability of detecting all K differentially expressed genes.

The definitions of each column are given in the Report Definitions section.

Plots Section





These plots show the relationship between power and the standard deviation of the differences for various three values of K.

Example 2 – Finding the Sample Size

This example determines the number of two-channel arrays needed to achieve 80% power to detect differential expression for each gene. Two samples will be obtained from each of the sampled individuals. One of the two samples will be randomly assigned the treatment and the other will remain as the control. Following treatment, the two samples will be exposed to a single microarray. Each microarray will produce intensity information for 12,682 genes. The arrays will be pre-processed by subtracting the control intensity (Log2) from the treatment intensity for each gene on each array. Thus, a positive value implies upward expression in the treatment, while a negative value implies down-regulation in the treatment. The paired T-test will be used to determine which genes are differentially expressed (upward or downward) following exposure to the treatment.

The researchers wish to detect differential expression that is two-fold or greater (Log2-scale difference of 1). Typical standard deviations of the Log2 paired differences for this experiment are expected to range from 0.2 to 2.0.

The researchers guess the number of genes with at least 2-fold differential expression to be around 50, but will examine the effect of this estimate on sample size by trying 10 and 100 genes as well. A false discovery rate of 0.05 will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple One-Sample or Paired T-Tests** procedure window by expanding **Means**, then **One Mean**, then clicking on **Multiple Tests**, and then clicking on **Multiple One-Sample or 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**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Test Type	**T**
Alternative Hypothesis	**Two-Sided**
Power for each Test	**0.80**
False Discovery (Alpha) Method	**FDR (False Discovery Rate)**
FDR or EWER Value	**0.05**
D (Difference)	**1**
S (Standard Deviation)	**0.2 to 2 by .2**
Number of Tests	**12682**
K (Number > D)	**10 50 100**

Output

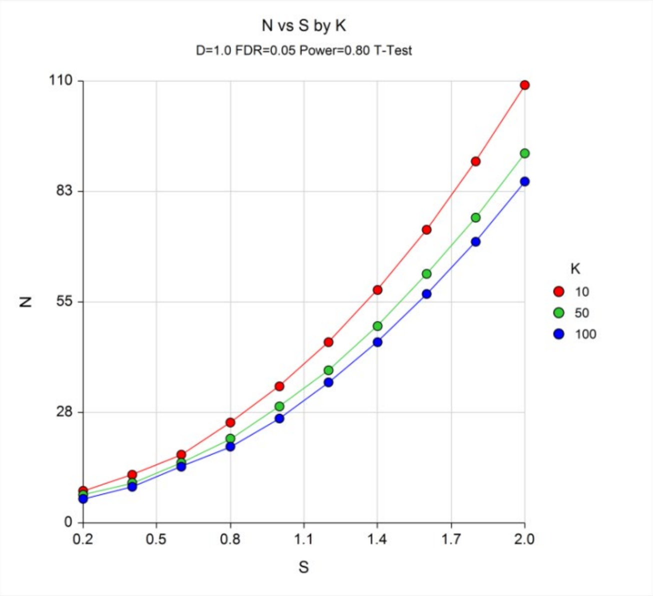
Click the Calculate button to perform the calculations and generate the following output. The calculations may take a few moments.

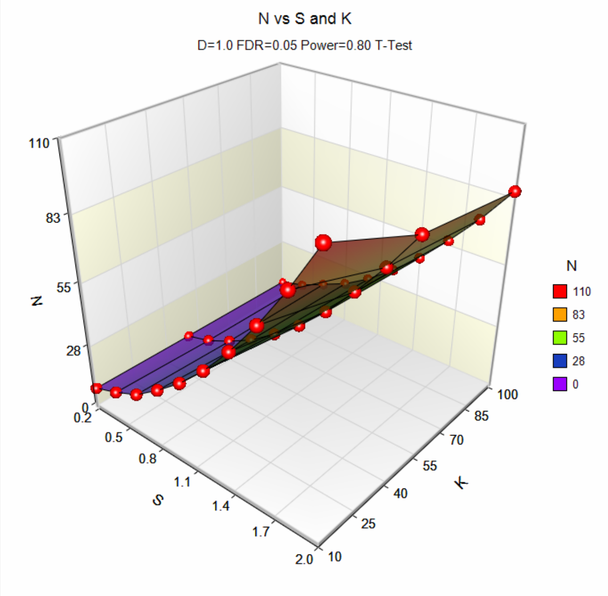
Numeric Results

```
Numeric Results for Multiple One-Sample T-Tests
Null Hypothesis: MeanDiff = 0     Alternative Hypothesis: MeanDiff ≠ 0
Number of Tests: 12682

                          Minimum      Std.                Number      False
            Number           Mean      Dev.     Effect      Tests  Discovery
          of Arrays    Difference   of Diff.      Size  To Detect       Rate
Power           (N)           (D)       (S)       (ES)        (K)      (FDR)       Beta
0.96741           8           1.0       0.2      5.000         10     0.0500    0.03259
0.97509           7           1.0       0.2      5.000         50     0.0500    0.02491
0.91190           6           1.0       0.2      5.000        100     0.0500    0.08810
0.88530          12           1.0       0.4      2.500         10     0.0500    0.11470
0.86231          10           1.0       0.4      2.500         50     0.0500    0.13769
0.83278           9           1.0       0.4      2.500        100     0.0500    0.16722
0.81531          17           1.0       0.6      1.667         10     0.0500    0.18469
0.85472          15           1.0       0.6      1.667         50     0.0500    0.14528
0.86398          14           1.0       0.6      1.667        100     0.0500    0.13602
0.83661          25           1.0       0.8      1.250         10     0.0500    0.16339
0.82805          21           1.0       0.8      1.250         50     0.0500    0.17195
.                 .             .         .          .          .          .          .
.                 .             .         .          .          .          .          .
.                 .             .         .          .          .          .          .
```

This report shows the values of each of the parameters, one scenario per row. The sample size (number of arrays) estimates were calculated from the other parameters. The power is the actual power produced by the given sample size.

Plots Section





These plots show the relationship between sample size and the standard deviation of the differences for three values of K.

Example 3 – Finding the Minimum Detectable Difference

This example finds the minimum difference in expression that can be detected with 90% power from a microarray experiment with 14 two-channel arrays. The 14 arrays permit tests on 5,438 genes. The arrays will be pre-processed by subtracting the control intensity (Log2) from the treatment intensity for each gene on each array. Thus, a positive value implies upward expression in the treatment, while a negative value implies down-regulation in the treatment. The paired T-test will be used to determine which genes are differentially expressed (upward or downward) following exposure to the treatment. Standard deviations of the Log2 paired differences for this experiment range from 0.2 to 1.8.

In this example we will examine a range for K (the number of genes with mean difference greater than the minimum detectable difference), since this should vary with the mean difference chosen. A false discovery rate of 0.05 will be used.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple One-Sample or Paired T-Tests** procedure window by expanding **Means**, then **One Mean**, then clicking on **Multiple Tests**, and then clicking on **Multiple One-Sample or 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**.

Option	**Value**
Design Tab	
Solve For	**Mean Difference**
Test Type	**T**
Alternative Hypothesis	**Two-Sided**
Power for each Test	**0.90**
False Discovery (Alpha) Method	**FDR (False Discovery Rate)**
FDR or EWER Value	**0.05**
N (Number of Arrays)	**14**
S (Standard Deviation)	**0.2 to 1.8 by .4**
Number of Tests	**5438**
K (Number > D)	**10 to 50 by 10**

Output

Click the Calculate button to perform the calculations and generate the following output. The calculations may take a few moments.

Numeric Results

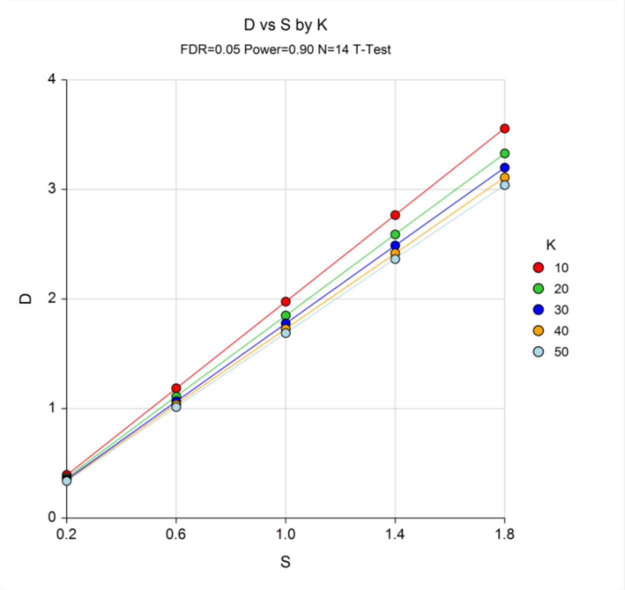
```
Numeric Results for Mulitple One-Sample T-Tests
Null Hypothesis: MeanDiff = 0     Alternative Hypothesis: MeanDiff ≠ 0
Number of Tests: 5438

                          Minimum        Std.                Number        False
               Number        Mean        Dev.     Effect      Tests    Discovery
             of Arrays  Difference    of Diff.      Size   To Detect        Rate
Power              (N)         (D)         (S)      (ES)         (K)       (FDR)       Beta
0.90000             14      0.3951      0.2000     1.976          10      0.0500    0.10000
0.90000             14      0.3699      0.2000     1.849          20      0.0500    0.10000
0.90000             14      0.3555      0.2000     1.777          30      0.0500    0.10000
0.90000             14      0.3454      0.2000     1.727          40      0.0500    0.10000
0.90000             14      0.3377      0.2000     1.689          50      0.0500    0.10000
```

(Numeric Results – Continued)

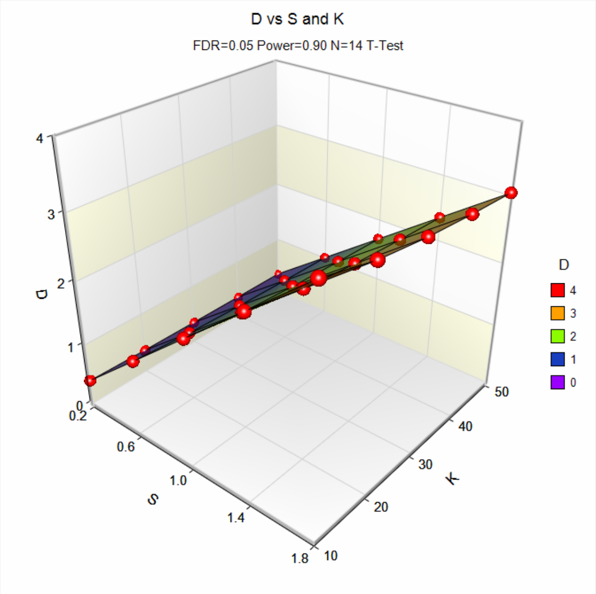
0.90000	14	1.1854	0.6000	1.976	10	0.0500	0.10000
0.90000	14	1.1096	0.6000	1.849	20	0.0500	0.10000
0.90000	14	1.0664	0.6000	1.777	30	0.0500	0.10000
0.90000	14	1.0363	0.6000	1.727	40	0.0500	0.10000
0.90000	14	1.0132	0.6000	1.689	50	0.0500	0.10000
0.90000	14	1.9756	1.0000	1.976	10	0.0500	0.10000
0.90000	14	1.8493	1.0000	1.849	20	0.0500	0.10000
0.90000	14	1.7774	1.0000	1.777	30	0.0500	0.10000
0.90000	14	1.7272	1.0000	1.727	40	0.0500	0.10000
0.90000	14	1.6887	1.0000	1.689	50	0.0500	0.10000
0.90000	14	2.7658	1.4000	1.976	10	0.0500	0.10000
0.90000	14	2.5890	1.4000	1.849	20	0.0500	0.10000
0.90000	14	2.4884	1.4000	1.777	30	0.0500	0.10000
0.90000	14	2.4181	1.4000	1.727	40	0.0500	0.10000
0.90000	14	2.3642	1.4000	1.689	50	0.0500	0.10000
0.90000	14	3.5561	1.8000	1.976	10	0.0500	0.10000
0.90000	14	3.3287	1.8000	1.849	20	0.0500	0.10000
0.90000	14	3.1993	1.8000	1.777	30	0.0500	0.10000
0.90000	14	3.1090	1.8000	1.727	40	0.0500	0.10000
0.90000	14	3.0397	1.8000	1.689	50	0.0500	0.10000

This report shows the values of each of the parameters, one scenario per row. The Minimum Mean Difference (D) estimates were calculated from the other parameters.

Plots Section


D vs S by K

FDR=0.05 Power=0.90 N=14 T-Test


D vs S and K

FDR=0.05 Power=0.90 N=14 T-Test

These plots show the relationship between D (the minimum detectable difference on the Log2 scale) and the standard deviation of the differences for five values of K.

Example 4 – Validation (EWER) using Stekel

Stekel (2003), pp. 226-228, gives an example in which N = 20, D = 1, and S = 0.68 for a two-sided paired T-Test. The number of genes tested is 6500. The control of false discoveries is “no more than one false positive.” This corresponds to an EWER value of 0.975. The power obtained for this example is 0.94.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple One-Sample or Paired T-Tests** procedure window by expanding **Means**, then **One Mean**, then clicking on **Multiple Tests**, and then clicking on **Multiple One-Sample or 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**.

Option	Value
Design Tab	
Solve For	**Power**
Test Type	**T**
Alternative Hypothesis	**Two-Sided**
False Discovery (Alpha) Method	**EWER (Experiment-wise Error Rate)**
FDR or EWER Value	**0.975**
N (Number of Arrays)	**20**
D (Difference)	**1**
S (Standard Deviation)	**0.68**
Number of Tests	**6500**

Output

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

Numeric Results

```
Numeric Results for Multiple One-Sample T-Tests
Null Hypothesis: MeanDiff = 0     Alternative Hypothesis: MeanDiff ≠ 0
Number of Tests: 6500

                           Minimum        Std.             Experiment
               Number         Mean        Dev.     Effect        -Wise        Single
            of Arrays   Difference    of Diff.       Size   Error Rate          Test
Power             (N)          (D)         (S)       (ES)      (EWER)          Alpha        Beta
0.93591            20         1.00        0.68      1.471       0.9750     0.0001500     0.06409
```

The power of 0.93591 matches Stekel’s result.

Example 5 – Validation (EWER) using Lee

Lee (2004), pp. 218-220, gives an example in which Power = 0.90, D = 1.0 1.5 2.0 2.5 and S = 1.0 for a two-sided paired Z-Test. The number of genes tested is 1000. The control of false discoveries is 0.5. This corresponds to an EWER value of 0.5. This setup corresponds to the upper left corner of Table 14.3 on page 219. The sample sizes obtained for this setup are 23, 11, 6, and 4, respectively.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple One-Sample or Paired T-Tests** procedure window by expanding **Means**, then **One Mean**, then clicking on **Multiple Tests**, and then clicking on **Multiple One-Sample or Paired T-Tests**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Test Type	**Z**
Alternative Hypothesis	**Two-Sided**
Power for each Test	**0.90**
False Discovery (Alpha) Method	**EWER (Experiment-wise Error Rate)**
FDR or EWER Value	**0.5**
D (Difference)	**1.0 1.5 2.0 2.5**
S (Standard Deviation)	**1.0**
Number of Tests	**1000**

Output

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

Numeric Results

```
Numeric Results for Multiple One-Sample Z-Tests
Null Hypothesis: MeanDiff = 0     Alternative Hypothesis: MeanDiff ≠ 0
Number of Tests: 1000

                          Minimum        Std.             Experiment
              Number         Mean        Dev.     Effect        -Wise      Single
           of Arrays   Difference    of Diff.       Size   Error Rate        Test
Power            (N)          (D)         (S)       (ES)       (EWER)       Alpha        Beta
0.90576           23         1.00        1.00      1.000       0.5000   0.0005000     0.09424
0.93244           11         1.50        1.00      1.500       0.5000   0.0005000     0.06756
0.92194            6         2.00        1.00      2.000       0.5000   0.0005000     0.07806
0.93565            4         2.50        1.00      2.500       0.5000   0.0005000     0.06435
```

Sample sizes of 23, 11, 6, and 4 match the results shown in Lee (2004).

Example 6 – Validation (FDR) using Jung

Jung (2005), page 3100, gives an example for the sample size needed to control FDR in a two-sample Z-Test. This example is repeated in Chow, Shao, and Wang (2008). We adapt the effect size in this validation to correspond to a one-sample test. Namely, the effect size is reduced by one half. In the example, Power = 0.60 (from 24/40), D = 1.0, and S = 1.0 for a one-sided two-sample Z-Test. We use S = 2.0 to correspond to the equivalent in the one-sample test. The number of genes tested is 4000. The FDR level is 1%. This setup corresponds to Example 1 on page 3100. The required sample size obtained for this setup is 68.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Multiple One-Sample or Paired T-Tests** procedure window by expanding **Means**, then **One Mean**, then clicking on **Multiple Tests**, and then clicking on **Multiple One-Sample or Paired T-Tests**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Test Type	**Z**
Alternative Hypothesis	**One-Sided**
Power for each Test	**0.60**
False Discovery (Alpha) Method	**FDR (False Discovery Rate)**
FDR or EWER Value	**0.01**
D (Difference)	**1.0**
S (Standard Deviation)	**2.0**
Number of Genes Tested	**4000**
K (Number > D)	**40**

Output

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

Numeric Results

```
Numeric Results for Multiple One-Sample Z-Tests
Null Hypothesis: MeanDiff = 0    Alternative Hypothesis: MeanDiff > 0
Number of Tests: 4000

                          Minimum      Std.                Number       False
              Number         Mean      Dev.     Effect      Tests   Discovery
            of Arrays  Difference  of Diff.       Size  To Detect        Rate
Power             (N)         (D)       (S)       (ES)        (K)       (FDR)        Beta
0.61099            68        0.50      1.00      0.500         40      0.0100     0.38901
```

A sample size of 68 matches the result shown in Jung (2005). For Example 3 in Jung (2005), the alternative hypothesis is two-sided and results in a sample size of 73. This result may be validated in **PASS** by changing Alternative to Two-Sided in this example.