SAGEstat:
program for the planning of SAGE analysis and the evaluation of SAGE data
version 4.0, Februari 2002

J.M. Ruijter,
Department of Anatomy and Embryology
Academic medical Center, Amsterdam
The Netherlands
e-mail: j.m.ruijter@amc.uva.nl

This file contains the texts that can also be found in the context sensitive help windows that can be opened from within every procedure in the program. Note that this text speaks of 'control' and 'experimental' library, whereas the program interface speaks of 'first' and 'second' library.

Update History
With version 4.0 SAGEstat the user interface of SAGEstat is completely re-designed. The result is an interface with
- one page per procedure.
- a clear distinction of inpu tand results per procedure.
- addition of a Z-test between 'all' tags in two libraries (input from and output to Microsoft Excel).
- independence of the configuration of Excel, decimal comma's and pionts are supported.
- graphic display of all matrices, with annotations on axes and graphs.

In version 4 of SAGEstat the normal distribution function for calculation of the probability of Z is based on: Abramovitz, M and Stegun, IA. Handbook of mathematical functions with formulas, graphs and mathematical tables. Dover, ISBN 0-486-61272-4. This function was copied from the website of the Australian Delphi User Group which gives a paper by Glenn Crouch about calculations with normal distributions (www.adug.org.au/MathsCorner/MathsCornerNDis.htm).

The implemented statistical procedures in SAGEstat version 4 are the same as in previous versions, except for the above mentioned Z probability function. The previously implemented function was only correct to 5 decimal places, the currently implemented function is correct to 7 decimal places. It is equivalent to the NORMSDIST function of Excel.

Overview
SAGEstat can be used for evaluation and planning of SAGE analysis. The Z-test proposed by Kal and co-workers (see References) for the EVALUATION of SAGE experiments focuses on the proportions of specific tags in each library. Since these proportions can be approximated to result from sampling with replacement, the probability of the resulting tag counts follows a binomial distribution. For the sample sizes involved in SAGE this binomial distribution can be approximated as a normal distribution and a test based on the normal approximation of the binomial distribution can be used. The test statistic Z is calculated as the observed difference between proportions of specific tags in both libraries divided by the standard error of this difference when the null hypothesis is true (see: equations). This Z-statistic is approximately normally distributed and can be compared to the critical Z-value for the two-sided significance level alpha.

The normal approximation of the binomial distribution allows the rearrangement of the equation of the Z-test in such a way that it can be used for the PLANNING of SAGE experiments (see: equations). This equation can be used in several ways:

1. Given N1 and N2 (the SAGE libraries are compiled), the critical values or the detectable differences can be calculated for a chosen significance level and power.

2. Given an observed difference, the total number of tags sequenced in both libraries and the chosen significance level, the power of the test can be determined.

3. Given an expected difference, a significance level, a power and the number of tags already sequenced in an existing SAGE library (N1), the number of tags that is needed in a new library (N2) can be calculated.

4. Given an expected difference, a chosen significance level and a required power, the number of tags that is needed in each library (N1 = N2) can be calculated.

References

A detailed description of the statistics implemented in this program can be found in:

Kal AJ, van Zonneveld AJ, Benes V, van den Berg M, Groot Koerkamp M, Albermann K, Strack N, Ruijter JM, Richter A, Dujon B, Ansorge W, Tabak HF (1999). Dynamics of gene expression revealed by comparison of SAGE transcript profiles from yeast grown on two different carbon sources. Mol Biol Cell 10, 1859-1872, 1999

Ruijter JM, van Kampen AHC, Baas F (2002) Statistical evaluation of Serial Analysis of Gene Expression (SAGE) libraries. Physiol Genomics 11: 37-44, 2002.

SAGEstat is written in Delphi 6.0 and will run under Windows 95 and later versions.
Excel is a trademark of Microsoft.

Disclaimer

SAGEstat is based on the Z-test which is a test between two proportions based on the normal approximation of the binomial distribution. A description of the Z-test can be found in almost every statistical textbook. Its use for comparing SAGE libraries is proposed and described in Kal et al., Mol Cell Biol 10: 1999 and in Ruijter et al., Physiol Genomics 11: 2002.

By opening and using this software you acknowledge that you have read these papers, understand them, and agree with their conclusions on the use of statistics in SAGE research..

You also acknowledge that you are aware of the fact that SAGEstat only calculates a P-value. The interpretation of this P-value, which may lead to false positives or false negatives results is not part of SAGEstat. The default alpha and beta levels in SAGEstat' s edit fields are not intended to serve as guide lines in those decisions.

Therefore, you assume all responsibility and liability for the selection of this software program to achieve your intended results, and for the conclusions you draw from these results.

The authors can not be hold responsible for any consequences of the use of this program.

Save matrix to text file

Whenever the program calculates a matrix of results you can save this matrix to a text file by choosing the menu option File -Save matrix. The matrix is then saved as a space-delimited text file and can be imported into a spreadsheet program. Most spreadsheet programs offer the option File - Open -Text file- Space delimited, in a series of dialog windows.
Note that in the spreadsheet the string ' *******' appears in some cells of the matrix dueor certain conditions. This string has to be removed when plotting a graph because it will be plotted as value 0.

Show graph of matrix

Whenever a matrix is calculated the "show graph" button will be enabled. Pressing this button gives you a plot of the data in the first column of the matrix.

From the listbox at the top-left of the Graph-window you can choose the other columns; the graph will be updated automatically.

You can also change the appearance of the graph by editing the upper and lower limits of the Y-axis and X-axis or you can choose to display one or both axis on a logarithmic scale. Press the update button to effectuate these changes.

The "copy to clipboard" button copies a bitmap image of the graph on the Windows Clipboard. From there you can paste the graph into a drawing program or into a presentation. Note that these copied graphs are drafts. A better-looking graph can be generated by saving the matrix (menu option: File - Save matrix) and importing the file into a spreadsheet program (File - Open - Text file - Space delimited). The big matrix then gives the best graphic result.

Procedure choice

TESTING
When you have sequenced your tags and know the number of specific tags you are interested in, SAGEstat includes two procedures for testing the statistical significance of the results:

1. Test the significance of the observed difference for a specific tag. This procedure will optionally use the tag numbers to give an estimate of the number of copies per cell.

2. Test the significance of the observed differences for each tag in two libraries

3. Calculate a matrix of critical differences between observed numbers of specific tags or copies per cell in both conditions based on the numbers of tags used.

PLANNING
When you are planning a SAGE analysis and want to know how many tags you have to sequence to detect an expected difference in the number of transcripts between the control condition and an experimental condition you can choose between the following procedures:

4. Calculate N for both libraries when you plan to sequence both groups yourself.

5. Calculate N for the experimental library when you plan to compare your new experimental condition with a control already sequenced by yourself or somebody else.

6. Calculate the difference you can detect as significant given the number of tags you are planning or willing to sequence.

When you have no indication of the difference to expect, you can calculate a matrix of tag numbers needed to detect a range of differences. These matrices will give you an idea of what you are up against. There are three matrices available:

7. A matrix of N for both groups for when you plan to sequence both groups yourself. Note that you will have to sequences 2 times N tags.

8. A matrix for N experimental for when you plan to compare your new experimental condition with a control condition already sequenced.

9. A matrix of the power that can be reached for a given abundance and library size in the control condition and increasing abundance with varying library size in the experimental condition

Test significance of observed difference

PURPOSE:
Use this procedure when you have done a SAGE experiment and want to test the significance of the difference in tag counts you have found for a specific mRNA in each library
Given the number of tags you have found in the control and experimental condition and the number of tags that you sequenced in each condition, this calculation gives you the statistical significance of the difference using a Z-test on proportions.

Use ' calculate matrix of critical differences' when you want to test more then one specific mRNA.

Use ' calculate detectable difference' when you are planning a SAGE experiment but have already decided how many tags you are able to sequence.

INPUT
You have to give the following input:
· the number of tags you have sequenced in the control condition
· the number of tags you have sequenced in the experimental condition
· the number of specific tags you have found in the control condition
· the number of specific tags you have found in the experimental condition

Optional input:
You may give an estimate of the total number of mRNA copies in the cell. This will give you as additional result an estimate of the difference between control and experimental conditions translated into numbers of transcripts per cell.

RESULT
The result of this calculation is the p-value of the observed difference in number of mRNA tags. This p-value is given in the ' p=' field on the top left of the window. When you have given an estimate of the total number of mRNA copies per cell, the fields control and experimental under ' number of mRNA copies per cell' give the calculated number of transcripts per cell that are present, based on the number of tags found.

The power of the test (in case of a non-significant results) and the confidence limits of the difference are also given.

Please note that performing a lot of pairwise tests will lead to an accumulation of Type I error. To be on the save side, divide the significance level you want for your whole experiment (often 0.05) by the number of tests you do and use the result as the alpha level for each test (this is known as a Bonferroni correction).

Use ' calculate matrix of critical differences' when you want to test a series of specific differences.

Test differences between two libraries

PURPOSE:
Use this procedure when you have done a SAGE experiment and want to test the significance of the differences in tag counts between two libraries for each specific tag.
Given the number of tags you have found in the control and experimental condition and the number of tags that you sequenced in each condition, this calculation gives you the statistical significance of the difference using a Z-test on proportions for each tag.

INPUT
SAGEstat will read the input from Excel
The Excel sheet has to contain the following information:
· the number of tags you have sequenced in the control condition
· the number of tags you have sequenced in the experimental condition
for each tag:
· the number of specific tags you have found in the control condition
· the number of specific tags you have found in the experimental condition

In the input window you have to give the column and row for the cells containing the above input

See the 'Example' tab for an illustration of the Excel sheet and input window

RESULT
The result of this procedure is a column of the p-values for each of the observed differences in tag counts.

Please note that performing a lot of pairwise tests will lead to an accumulation of Type I error. To be on the save side, divide the significance level you want for your whole experiment (often 0.05) by the number of tests you do and use the result as the alpha level for each test (this is known as a Bonferroni correction).

Calculate matrix of critical values

PURPOSE:
Use this procedure when you have done a SAGE experiment and want to have a table of critical values to easily judge the differences you have found for a series of specific mRNAs. Critical values are defined as the number of tags that must have been found in the experimental library for the difference with the control library to be statistically significant.
Given the number of tags you have sequenced in the control and experimental condition, this calculation gives you a matrix of critical differences in the form of upper and lower limits for the number of specific tags. Tag numbers outside those limits are statistically significant. The matrix can be plotted and used as a nomogram.
The matrix can also be saved to file and imported in a spreadsheet to plot a graph that can serve as a nomogram. Please note that the matrix is only valid for the specified number of tags sequenced in each condition.

Use 'test significance of observed difference' when you want to test the difference for one specific mRNA.

INPUT
You have to give the following input:
· the number of tags you have sequenced in the control condition
· the number of tags you have sequenced in the experimental condition

Use the radio button to choose:
1) Display critical values as:
O when you want critical values as number of tags: choose as number of tags
O when you want critical values translated to number of copies per cell: choose as copies per cell. In this case you also have to give an estimate of the total number of mRNA copies per cell
2) Matrix Size:
O A small matrix uses only 10 different levels of expression in the control condition and 6 levels of enrichment in the experimental condition.
O A big matrix gives you 65 levels of control expression (increasing from 1 to almost 1200 in increasing steps) and more different levels of enrichment. This matrix can be used to plot a smooth graph when you save the matrix and import the file into a spreadsheet program.

RESULT
This calculation results in a small or big matrix of critical differences i.e. limits that should be exceeded in the experimental condition for an observed difference to be statistically significant. This matrix is displayed in the bottom part of the window. Both upper and lower critical limits are calculated for three significance levels: 0.05, 0.01 and 0.001, respectively (two grey rows on top of the matrix). These limits are given in three columns on both sides of the value supposedly found in the control condition (seven columns of white cells). The control value is also given in the first grey column. The string '*******' is printed when no significant result can be reached given the control value and the number of tags sequenced in each condition. The string '*******' is also printed when the number of specific tags exceeds the library size.

Please note that this matrix can be used to plot a GRAPH when you save the matrix (menu option: File - Save matrix) and import the file into a spreadsheet program (File - Open - Text file - Space delimited). Use the first ' control' column as X-axis and the other as one of the Y-axis series in a scatter plot to obtain an easy to use nomogram, specific for the number of tags sequenced in both groups. The big matrix gives the best graphic result. Interpretation of this graph can be difficult when the library sizes are not equal.

Be aware of accumulation of Type I error when you are performing multiple pairwise tests.

Use ' test significance of observed difference' to test one specific difference found.

Calculate N for both libraries

PURPOSE:
Use this procedure when you are planning a SAGE experiment and want to calculate the number of tags needed in both the control and the experimental condition.
Given an estimate of the total number of mRNA copies in a cell and the numbers of specific transcripts per cell you expect to find in the control and experimental conditions, this calculation gives you the number of tags that need to be sequenced in both conditions.
The resulting number of tags is the number needed in BOTH conditions to detect the difference in the numbers of specific copies per cell with a two-sided probability of a Type I error (incorrect rejection of the Null hypothesis that there is no difference) of less then alpha, and a probability of a Type II error (failure to detect a true difference = failure to reject the Null hypothesis) of less then beta.

Use ' calculate matrix of N for both libraries' to get an idea about the number of tags needed for a range of differences.

INPUT
You have to give the following input:
· alpha (default: 0.05)
· beta (default: 0.1, equivalent to a Power of 0.9)
· an estimate of the total number of mRNA molecules per cell
· the number of specific mRNA transcripts in the control condition
· the number of specific mRNA transcripts in the experimental condition
Please note:
· that the number of specific transcripts in the experimental condition can be either higher or lower than in the control condition.

RESULT
The result of this calculation is the number of tags that need to be sequenced in the control AND in the experimental condition. These numbers are displayed in the fields ' control' and ' experimental' under ' number of tags to be sequenced in SAGE' .
Please note that you need to sequence BOTH numbers of tags to detect the requested difference at significance level alpha and power 1-beta.

Use ' calculate matrix of N for both libraries' to get an idea about the number of tags needed for a range of differences.

Calculate N experimental

PURPOSE
Use this procedure when you want to calculate the number of tags needed in the experimental condition when you want to compare the results to a control condition that has already been sequenced. Make sure that the difference between your new experimental condition and the ' old' control condition is really only the factor you are interested in.
This approach has the advantage that you only have to sequence one condition. However, when the number of tags already sequenced is low, you may be better off when you sequence both libraries yourself.
Given an estimate of the total number of mRNA copies in a cell, the numbers of specific transcripts per cell in the control and the experimental conditions and the number of tags already sequenced in the

control condition this calculation gives you the number of tags that need to be sequenced in the experimental condition.
The resulting number of tags is the number needed to detect the difference in the numbers of specific copies per cell with a two-sided probability of a Type I error (incorrect rejection of the Null hypothesis that there is no difference) of less then alpha, and a probability of a Type II error (failure to detect a true difference = failure to reject the Null hypothesis) of less then beta.

Use ' calculate matrix of N experimental' to get an idea about the number of tags needed for a range of differences.

Use ' calculate matrix of N for both groups' to find out whether you are really better of by just sequencing one library.

INPUT
You have to give the following input:
· alpha (default: 0.05)
· beta (default: 0.1, equivalent to a Power of 0.9)
· an estimate of the total number of mRNA molecules per cell
· the number of specific mRNA transcripts in the control condition
· the number of specific mRNA transcripts in the experimental condition
· the number of tags already sequenced in the control condition
Please note:
· that the number of specific transcripts in the experimental condition can be either higher or lower than in the control condition.
· that exchanging the number of expected transcripts between control and experimental conditions will lead to another number of tags needed in the experimental condition.

RESULT
The result of this calculation is the number of tags that need to be sequenced in the experimental condition. This number is displayed in the field ' experimental' under ' number of tags sequenced in SAGE' . ' *******' is printed when no number of tags in the experimental condition would be high enough to reach statistical significance given the number of tags already sequenced in the control condition.

Please note:
· that you need to sequence this number of tags in your experimental library to detect the requested difference at significance level alpha and power 1-beta.
· that exchanging the number of transcripts between control and experimental will lead to another number of tags needed in the experimental condition.

Use ' calculate matrix of N experimental' to get an idea about the number of tags needed for a range of differences. This approach has the advantage that you only have to sequence one condition. However, when the number of tags already sequenced is low, you may be better off when you sequence both libraries yourself.

Use ' calculate matrix of N for both groups' to find out whether you are really better of by just sequencing one library.

Calculate detectable difference

PURPOSE:
Use this procedure when you are planning a SAGE experiment and want to calculate the sensitivity of this experiment. You can calculate the difference you are able to detect as significant with the number of tags you are planning or willing to sequence in the control and the experimental condition. You can calculate for a number of transcripts in the experimental condition that is either higher or lower than the control condition.
Given the number of tags you are planning to sequence in both conditions and the number of specific mRNA transcripts you expect to find in the control condition, this calculation gives you the number of specific transcripts that have to be present in the experimental condition to be statistically significant.

This number of transcripts is the number needed to be present to detect the difference in the numbers of specific copies per cell with a two-sided probability of a Type I error (incorrect rejection of the Null hypothesis that there is no difference) of less then alpha, and a probability of a Type II error (failure to detect a true difference = failure to reject the Null hypothesis) of less then beta.

Note that calculation of the detectable difference leads to a wider interval than the calculation of the critical difference. This is because in the last calculation you have done your experiment and know how many tags are observed while in the first calculation you are still uncertain about the outcome of the experiment.

Use ' test significance of observed difference' when you have already done your sequencing.

INPUT
You have to give the following input:
· alpha (default: 0.05)
· beta (default: 0.1, equivalent to a Power of 0.9)
· an estimate of the total number of mRNA transcripts in the cell
· the number of specific mRNA transcripts in the control condition
· the number of tags you want to sequence in the control condition
· the number of tags you want to sequence in the experimental condition

Use the radiobuttons to choose the Direction of Effect:
O when you expect the experimental condition to give higher expression: choose experimental up.
O when you expect the experimental condition to give lower expression: choose experimental down.

Please note:
· that when the numbers of tags sequenced in both conditions are not equal, exchanging the number of specific tags found and the choice for the direction of the effect will give a different result.

RESULT
The result of this calculation is the number of mRNA transcripts that should be present in the experimental condition for the difference to be statistically significant. This number is displayed in the field ' experimental' under ' number of mRNA copies per cell' . The string ' ******** ' is printed when given the number of tags sequenced in both conditions and the number of transcripts in the control condition the minimum number of transcripts in the experimental condition (0) is not small enough to be statistically significant.
Please note that this is the minimum difference that should be present between both conditions to detect the difference as significant at significance alpha and power 1-beta with the number of tags you are planning or willing to sequence.

Use ' test significance of observed difference' when you are not planning an experiment but have already done your sequencing and know the number of observed tags.

Calculate matrix of N for both libraries

PURPOSE
Use this procedure when you are planning a SAGE analysis and want to calculate the number of tags needed in both the control and the experimental condition that you want to compare but have no estimate of the differences in expression in both conditions.
Given an estimate of the total number of mRNA copies in a cell this calculation gives you a matrix of numbers of tags that need to be sequenced in both conditions. This matrix of tag numbers will be calculated for a range of specific transcripts per cell in the control condition and a range of fold increased expression levels in the experimental condition.
The resulting number of tags is the number needed in BOTH conditions, to detect the difference in the numbers of specific copies per cell with a two-sided probability of a Type I error (incorrect rejection of the Null hypothesis that there is no difference) of less then alpha, and a probability of a Type II error (failure to detect a true difference = failure to reject the Null hypothesis) of less then beta.

INPUT
You have to give the following input:

· alpha (default: 0.001; Bonferroni correction)
· beta (default: 0.1, equivalent to a Power of 0.9)
· an estimate of the total number of mRNA molecules per cell

Use the radiobuttons to choose:
Matrix Size
O A small matrix uses only 10 different levels of expression in the control condition and 6 levels of enrichment in the experimental condition.
O A big matrix gives you 65 levels of control expression (increasing from 1 to almost 1200 in increasing steps) and more different levels of enrichment. This matrix can be used to plot a smooth graph when you save the matrix and import the file into a spreadsheet program.

RESULT
This calculation results in a small or big matrix of tag numbers needed in BOTH libraries. This matrix is displayed in the bottom part of the window. Each white cell in the matrix gives you the number of tags needed to be sequenced in BOTH the control and the experimental condition to detect a significant difference in expression given by the combination of control expression (given in the first grey column) and an enrichment in the experimental group (given as a multiplication factor in the top grey row). When the abundance in the control condition multiplied by the fold enrichment leads to more then the total number of transcripts that can be present in a cell ' *******' is printed.

Please note:
· that you need to sequence BOTH numbers of tags, in control AND experimental conditions, to detect the requested difference at significance alpha and power 1-beta.
· that in this matrix you can exchange the words ' control' and ' experimental' when you are interested in the number of tags needed to detect a difference in which the experimental condition is lower than the control.

This matrix can be used to plot a GRAPH when you save the matrix (menu option: File - Save matrix) and import the file into a spreadsheet program (File - Open - Text file - Space delimited). The big matrix gives the best graphic result.

Use ' calculate N for both groups' to get the number of tags needed for a specific difference between control an experimental groups that is not included in the matrix.

Calculate matrix of N for experimental library

PURPOSE
Use this procedure when you are planning a SAGE analysis and want to calculate the number of tags needed in the experimental condition and aim to compare the results to a control condition that has already been sequenced. You have, however, no estimate of the differences in expression in both conditions.
Given an estimate of the total number of mRNA copies in a cell and the number of tags already sequenced in the control condition this calculation gives you a matrix of numbers of tags that need to be sequenced in the experimental condition. This matrix of tag numbers will be calculated for a range of specific transcripts per cell in the control condition and a range of increased or decreased expression levels in the experimental condition. Make sure that the difference between your new experimental condition and the ' old' control condition is really only the factor you are interested in. This approach has the advantage that you only have to sequence one condition. However, when the number of tags already sequenced is low, you may be better off when you do both groups yourself. The resulting number of tags is the number needed to detect the difference in the numbers of specific copies per cell with a two-sided probability of a Type I error (incorrect rejection of the Null hypothesis that there is no difference) of less then alpha, and a probability of a Type II error (failure to detect a true difference = failure to reject the Null hypothesis) of less then beta.

INPUT
You have to give the following input:
· alpha (default: 0.001; Bonferroni correction)
· beta (default: 0.1, equivalent to a Power of 0.9)
· an estimate of the total number of mRNA molecules per cell

· the number of tags already sequenced in the control condition

Use the radiobuttons to choose:
1. Direction of Effect:
O when you expect the experimental condition to give higher expression : choose experimental up.
O when you expect the experimental condition to give lower expression : choose experimental down.
2) Matrix Size:
O A small matrix uses only 10 different levels of expression in the control condition and 6 levels of enrichment in the experimental condition.
O A big matrix gives you 65 levels of control expression (increasing from 1 to almost 1200 in increasing steps) and more different levels of enrichment. This matrix can be used to plot a smooth graph when you save the matrix and import the file into a spreadsheet program.

RESULT
This calculation results in a small or big matrix of tag numbers needed in the experimental group. This matrix is displayed in the bottom part of the window. Each white cell in the matrix gives you the number of tags needed to be sequenced in the experimental condition to detect a significant difference in expression given by the combination of control expression (given in the first grey column) and an ' enrichment' in the experimental library (given as a multiplication factor in the top grey row). In case you choose for the ' experimental down' radio button, ' control' and ' experimental' in the matrix are interchanged.

The string ' *******' is printed when
1) no number of tags in the experimental condition would be high enough to reach statistical significance given the number of tags sequenced in the control condition. This occurs in the upper left corner of the matrix.
2) the abundance in the control condition multiplied by the grade of enrichment lead to more then the total number of transcripts that can be present in a cell. This occurs in the bottom-right corner of the matrix.

Please note:
· that you need to sequence these numbers of tags in your new experimental library to detect the requested difference at significance level alpha and power 1-beta.
· that in this matrix you cannot just exchange the words ' control' and ' experimental' when you are interested in the number of tags needed to detect the opposite difference.
· that you have to use the radio button under ' Direction of Effect' when you are interested in the number of tags needed to detect a difference in the opposite direction.

This matrix can be used to plot a GRAPH when you save the matrix (menu option: File - Save matrix) and import the file into a spreadsheet program (Open - Space delimited text file). The big matrix gives the best graphic result.

Use ' calculate N for experimental library' to get the number of tags needed for a specific difference between control an experimental libraries that is not included in the matrix.

Calculate matrix of reachable power

PURPOSE
Use this procedure when you are planning a SAGE experiment an want to calculate the power of the resulting test. The power of a test is defined as the chance of observing a given difference as significant when it is a true difference.
The power of a test can only be calculated for a given difference between the control and experimental library, and for given library sizes. The SAGEstat program calculates the power of the Z-test a given tag abundance in the control condition with a given size of the first library and a range of abundances and library sizes in the experimental condition.

INPUT
You have to give the following input:
· the number of tags in the control library
· the number of specific tags in the control library

· the number of tags in the experimental library
The first and second input serve to calculate the abundance in the control condition.

The program calculates the power for the given size of the second library and for libraries 2, 3, 5 and 10 times that size,

Matrix Size:
O A small matrix uses only 10 different levels of expression in the experimental condition
O A big matrix gives you 65 levels of experimental expression (increasing from 1 to almost 1200 in increasing steps.
The big matrix can be used to plot a smooth graph when you save the matrix and import the file into a spreadsheet program.

RESULT
This calculation gives you a table of 6 columns:
The first column gives the number of tags in the experimental library. The top rows of the second till sixth column give the size of the experimental libraries. The body of the table contains for each library size and number of specific experimental tags the power of the Z-test when the difference is tested between these experimental tags and the given tag count in the control library.
The matrix can be used to plot a GRAPH when you save the matrix (menu option: File - Save matrix) and import the file into a spreadsheet program (File - Open - Text file - Space delimited). The big matrix gives the best graphic result.