nQuery Sample Size & Power Calculation Software | Validation Guidelines

---

Every nQuery sample size table, distribution function table, standard deviation table, and table-specific side table has been tested for computational accuracy across table-specific ranges of parameters, and has been tested for full functionality of editing options and menu items (side tables, plots, statements) for one or more table-specific parameter sets.

Methods for testing are described below. Following these general discussions are individual validation reports for the sample size and distribution function tables. nQuery sample size tables are indexed by goals (means, proportions, survival, agreement and regression), by number of groups (one, two, or more than two), by type of analysis (test, confidence interval, equivalence test), and by equal vs unequal n.

For a full list of validation reports, including screenshots. Please visit:
http://info.statsols.com/nquery-sample-size-validation-reports

Global testing of menu items

All menu items which are not table specific were tested to verify that they functioned appropriately. Each item within each menu was tested with examples for three sample size tables. The tester checked off each item for each table; when any problem was found, a description of the problem was recorded. Retesting took place after any necessary corrections were made.

Computational accuracy testing for a range of parameter values

Every nQuery table had one table-specific test problem tested for computational accuracy and for full functionality of editing, side table, and menu options, see next section. For every nQuery table, sets of table-specific parameter values were selected to cover a wide range of values for significance level, power, effect size, etc and computations were made for each set of parameter values and checked against other sources. The range of parameter values tested for each table is listed in that table's validation report. For the two- sample t-test for example, three sets of test problems were created : Set 1) the user entered effect size and power and nQuery solved for n, Set 2) the user entered effect size and n and nQuery solved for power, Set 3) the user entered power and n and nQuery solved for effect size. Printed tables with each of these sets of test values were included in the testing notebook. These sets of solutions were reviewed by Janet Elashoff, who checked for consistency, face validity, and for computational accuracy against other sources. The sources used for checking accuracy for each table are listed in that table's validation report. The results from other sources were recorded in the bottom section of the printed nQuery results tables. In spite of the extensive testing done, only a small subset of the values users might want to enter can be explicitly tested; for some tables reliable alternative sources were available to check only a few of the test problems.

Functionality testing for editing options, side tables, plots, statements

For every sample size table, plus the distribution function, standard deviation, and side tables, the full functionality of all options was verified for one test problem. Results for this problem were checked against one or more of hand calculation, a published table, other sample size software, or a published paper, depending on what alternate sources were available. For this example, tests were made to ensure that (1) the same results were obtained regardless of the order in which values are entered or edited, and (2) nQuery components specific to the table (plot, statement, side tables, unequal n version, and guide cards) functioned properly.

Note that answers may differ slightly when sample sizes are small depending on whether power is entered and n computed or n is entered and power computed since the integer sample sizes computed are those necessary to obtain power greater than or equal to that requested. In brief, the directions to the tester for sample size tables and nonspecific side-tables were:

- Check for consistency of results for each possible order of entry or editing of row values. For small tables, values were entered in all possible orders. For larger tables, a "minimum" example was defined and all possible orders investigated only for the minimum example. For example, for the two-sample t-test, the "minimum" example required filling in the significance level, one or two-sided, and two of effect size, power, sample size. The "maximum" example is the same numerical example with all rows filled in. For disallowed entry orders, check that the appropriate error messages are seen.
- Obtain and print statements, print the table, obtain and print plots.

- Edit all possible values with auto recalculation off, and with auto recalculation on.
- Edit row names where applicable.
- Review all Guide cards.
- For unequal n's tables, reproduce the equal n maximum example (using r = 1). Table specific side tables (usually obtained from Calculate effect size) were checked at the same time as the main table. The following items were checked for each specific side table:
- Test all buttons on the side table.
- Test that the data are saved appropriately.
- Enter values in all possible orders.
- Edit all possible values.
- Test for the effect of deleting values.
- Test with incorrect values to determine if an appropriate error message is displayed.
- Review Guide cards (in Help system). Testing was documented in two ways. Each item to be reviewed was checked off for each table. When any problems were found, a description of the problem was recorded with supporting printed table or screen print. In addition, the printed tables generated for specific testing steps were labeled as required and together with printed statements and plots were saved for review. All this documentation was reviewed by Janet Elashoff to ensure that the tables were internally consistent - i.e., the results for a given example did not differ depending on which row was computed, and that plots and statements were correct. Tables were re-tested after any necessary corrections were made.

Validation reports for each table

The validation reports for each table contain a list of parameter values tested and the sources they were tested against. In addition a screen shot shows one or more examples. Validation reports for the table- specific side table and the unequal n's version of the table appear following the main validation report for the table.

For a full list of validation reports, including screenshots. Please visit: http://info.statsols.com/nquery-sample-size-validation-reports

Distribution Functions	• z (Normal/Gaussian) • t distribution • t distribution (non-central) • Chi-square distribution
Means Tests ↓	
One group	• MOT0 One group t-test • MOT1 Paired t test for difference in means
Two group	• MTT0 Two group t-test (equal n's) • MTT0U Two group t-test (unequal n's)
Confidence Interval Tests ↓	
Two group	• MTC0 Confidence interval for difference in means (large equal n's) • MTC1 Confidence interval for difference in means (coverage probability)
Equivalence Tests ↓	
Two Group	• MTE0 Two group t-test for equivalence in means (equal n's) • MTE0U Two group t-test for equivalence in means (unequal n's)
Proportions Tests ↓	
Two group	• PTT0 Two group Chi-square test of equal proportions (equal n's) • PTT0U Two group Chi-square test of equal proportions (unequal n's) • PTT1 Two group continuity corrected Chi-square

	test of equal proportions (equal n's) ● PTT1U Two group continuity corrected Chi-square test of proportions (unequal n's)
Equivalence Tests ↓	
Two group	● PTE0 Two group test of equivalence in proportions (equal n's) ● PTE0U Two group test of equivalence in proportions (unequal n's)
Survival Tests ↓	
Two group	● STT0 Log rank test of survival in two groups (equal n's)
Regression Tests ↓	
One-group	● ROT0 Logistic regression for one normally distributed covariate

Table DOT0. z (Gaussian/Normal) distribution

See nQuery manual Appendix for computing method. Figure DOT0 shows the two most commonly used z-value computations. Range of values tested: probability .0005, .001, .01, .5, .99, .999, .9995 to z, and corresponding values of z back to probability. Values checked against tables in Dixon WJ, Massey FJ (1983) Introduction to Statistical Analysis. 4th edition. McGraw-Hill. New York and/or StaTable. Electronic Tables for Scientists and Engineers. Version 1. 1994 Cytel Software Corporation.

Table DOT1. t distribution

See nQuery manual Appendix for computing method. Figure DOT1 shows t-values for 1 and 3000 degrees of freedom for probability .95. Additional values tested: Degrees of freedom 1, 2, 10, 30, 120, 10000 and probability .001, .005, .01, .5, .99, .995, .999 to t and those values of t

back to probability. Values checked against tables in Dixon WJ, Massey FJ (1983) Introduction to Statistical Analysis. 4th edition. McGraw-Hill. New York and/or StaTable. Electronic Tables for Scientists and Engineers. Version 1. 1994 Cytel Software Corporation.

Table DOT2. t distribution (non-central)

See nQuery manual Appendix for computing method. Figure DOT2 shows t-values for 6 degrees of freedom to reflect example in which two-sided alpha is .05 and power requested is .90. Additional values tested: for non-centrality of 0, degrees of freedom 1, 2, 10, 10000 and probability .001, .999 to t and those values of t back to probability, plus the paired values of probability and t back to non-centrality. For degrees of freedom 2, 100, 10000 compute tvalue for probability of .999 and non-centrality of 0, then change probability to .001 and compute non-centrality. Using non-centrality value and probability of .001 recompute t and then using t and non-centrality recompute probability. Values checked against tables in Dixon WJ, Massey FJ (1983) Introduction to Statistical Analysis. 4th edition. McGraw-Hill. New York and/or StaTable. Electronic Tables for Scientists and Engineers. Version 1. 1994 Cytel Software Corporation.

Table DOT3. Chi-square distribution

See nQuery manual Appendix for computing method. Figure DOT3 shows computations for probability .95 for 1 and 30 degrees of freedom. Additional values tested: Degrees of freedom 1, 2, 7, 30 and probability .005, .01, .1, .95, .99, .995, .999, .9995 and the resulting values of Chi-square back to the probabilities. Checked against tables in Dixon WJ, Massey FJ (1983) Introduction to Statistical Analysis. 4th edition. McGraw-Hill. New York

Table MOT0. One group t test

Computed using t and non-central t distributions; the non-centrality parameter is (delta)(sqrt(n)) where delta is defined in the table row title. See nQuery manual Appendix for details of computing method for t. Figure MOT0 shows computations for one example. Additional values tested: significance level .001, .20 one and two-sided by power of 50% and 99% and effect size of .05 and 6 to n, values of n equal to 6 and 2000 and effect sizes of .05 and 6 back to power. Checked against tables in Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford, and DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc and Statistical System Power Analysis by Hintze JL (1991) BMDP Statistical Software.

Table MOT1. Paired t test for difference in means

Computed using t and non-central t distributions; the non-centrality parameter is (delta)(sqrt(n)) where delta is defined in the table row title. See nQuery manual Appendix for details of computing method for t. Figure MOT1 shows computations for one example. Additional values tested: significance level .001, .20 one and two-sided by power of 50% and 99% and effect size of .05 and 6 to n, values of n equal to 6 and 2000 and effect sizes of .05 and 6 back to power. Checked against tables in Machin D, Campbell MJ (1987) Statistical Tables for the Design of

Clinical Trials, Blackwell Scientific Publications Oxford, and DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc and Statistical System Power Analysis by Hintze JL (1991) BMDP Statistical Software.

Table MTT0. Two group t-test (equal n's)

Computed using t and non-central t distributions; the non-centrality parameter is (delta)(sqrt(n)) where delta is defined in the table row title. See nQuery manual Appendix for details of computing method for t. Figure MOT1 shows computations for one example. Additional values tested: significance level .001, .20 one and two-sided by power of 50% and 99% and effect size of .05 and 6 to n, values of n equal to 6 and 2000 and effect sizes of .05 and 6 back to power. Checked against tables in Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford, and DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc and Statistical System Power Analysis by Hintze JL (1991) BMDP Statistical Software.

Table MTT0U. Two group t-test (unequal n's)

Computed using t and non-central t distributions. The non-centrality parameter is (delta)(sqrt(n1))(sqrt(r))/sqrt(1+r) where delta is defined in the table row title and r = n2/n1 , the ratio of sample sizes. See nQuery manual Appendix for details of computing method for t. Figure MTT0U shows computations for two examples. Additional values tested: parameter values same as for MTT0 with sample size ratios of 1, 2, 4 and 1/4. Checked against MTT0 for equal n's and against DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc and SOLO Statistical System Power Analysis by Hintze JL (1991) BMDP Statistical Software.

Table MTC0. Confidence interval for difference of two means (large n, equal n's)

Computed using the large sample normal approximation with the z distribution. See nQuery manual Appendix or the table guide cards for the computing formula. Figure MTC0 shows computations for one example. Additional values tested: confidence levels from 0.6 to 0.99 with interval width .2(sigma), sigma to n and n of 16, 2000 to interval width. Checked against hand calculations and published examples.

Table MTC1. Confidence interval for difference of two means (coverage probability)

Computed using t and non-central t and Chi-square distributions. See nQuery manual Appendix for details of computing method for t and Chi-square. Figure MTC1 shows computations for one example with varying tolerance probabilities. Additional values tested: confidence levels from 0.6 to 0.99 with interval width .2(sigma), sigma to n and n of 16, 2000 to interval width. Using starting values from MTC0, checked against tables in Kupper, L.L., and K.B. Hafner, 1989. How

appropriate are popular sample size formulas? The American Statistician, 43: 101-105. Results for tolerance probability of 50% checked for consistency against MTC0.

Table MTE0. Two group t-test of equivalence in means (equal n's)

This sample size table is for a one-sided two-group t-test of the null hypothesis that the means are not equivalent (i.e. that the test drug mean is "worse than" the standard drug mean by an amount /\0 or more). The alternative hypothesis is that the means are equivalent, defined as differing by only /\1 (usually 0). Computed using t and non-central t distributions; the non-centrality parameter is (delta)(sqrt(n))/sqrt(2) where delta is defined in the table row title. See nQuery manual Appendix for details of computing method for t. Results for this table then will agree exactly with those for Table MTT0, the two-sample t-test, when a one-sided test is selected in MTT0 and the numeric values for effect size are the same. Figure MTE0 shows computations for two examples. Additional values tested: significance level .001, .20 one- sided by power of 50% and 99% and effect size of .05 and 6 to n, values of n equal to 3 and 2000 and effect sizes of .05 and 6 back to power. Checked against MTT0 and tables for two-group t-test in Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford, and two-group t-test in DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc and Statistical System Power Analysis by Hintze JL (1991) BMDP Statistical Software.

Table MTE0U. Two group t-test of equivalence in means (unequal n's)

This sample size table is for a one-sided two-group t-test of the null hypothesis that the means are not equivalent (i.e. that the test drug mean is "worse than" the standard drug mean by an amount /\0 or more). The alternative hypothesis is that the means are equivalent, defined as differing by only /\1 (usually 0). Computed using t and non-central t distributions; the non-centrality parameter is (delta)(sqrt(n1))(sqrt(r))/sqrt(1+r) where delta is defined in the table row title. See nQuery manual Appendix for details of computing method for t. Results for this table then will agree exactly with those for Table MTT0U, the two-sample t-test for unequal n's, when a one-sided test is selected in MTT0U and the numeric values for effect size are the same. Figure MTE0U shows computations for two examples. Additional values tested: parameter values same as for MTT0 with sample size ratios of 1, 2, 4 and 1/4. Checked against MTT0U and two-group t- test in DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc and Statistical System Power Analysis by Hintze JL (1991) BMDP Statistical Software..

Table PTT0. Two group Chi-square test of equal proportions (odds ratio = 1) (equal n's)

Computed using the large sample normal approximation, equation (10) in the nQuery manual appendix. Formula quoted widely, see for example, Fleiss JL, Tytun A, Ury SHK (1980) A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics. 36:343-346. Figure PTT0 shows computations for two examples. Solutions for n for significance

level .01,.1, one and two-sided, proportion in group 1.05, proportion in group 2 of .1,.5, .8, power of 50%, 80%, 99% checked for exact agreement against Table 3.1 in Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford. Additional values tested: significance level .001, .20 one and two-sided for proportion in group 1 of .05, .5, .95 by proportions in group 2 of .1,.6,.9 for power of 50% and 99% to sample size and sample size of 20, 500 to power checked for general consistency with DOS software Stat-Power version 2.0 by Bavry JL (1991) Scientific Software, Inc which uses a different calculation routine.

Table PTT0U. Two group Chi-square test of equal proportions (odds ratio = 1) (unequal n's)

Computed using the large sample normal approximation with adjustment for unequal n's. Formula quoted widely, see for example, Fleiss JL, Tytun A, Ury SHK (1980) A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics. 36:343-346. Figure PTT0U shows computations for one example for three values of r. Note that since the average of the two proportions differs for r and 1/r and the variance terms are PIi (1-PIi)/ni , the required total sample size will differ depending on whether the more extreme proportion is to have the smaller or the larger sample size. Solutions for r=1 checked against PTT0. Sample sizes giving a power of 80% for significance level .05 two- sided, proportion in group 1, .05, proportion in group 2 of .1,.4, .8, for r=1,2,3,4 used to compute power in DOS software Statistical Power Analysis: A Computer Program by Borenstein, M, Cohen J (1988) Lawrence Erlbaum Associates.

Table PTT1. Two group continuity corrected Chi-square test of equal proportions (Fisher's exact) (equal n's)

Computed using the large sample normal approximation, equations (10) and (11) in the nQuery manual appendix. See Fleiss JL, Tytun A, Ury SHK (1980) A simple approximation for calculating sample sizes for comparing independent proportions. Biometrics. 36:343-346. Figure PTT1 shows computations for two examples, second example is from Fleiss JL (1981) Statistical Methods for Rates and Proportions, 2nd edition, John Wiley & Sons, Inc. New York. Solutions for range of parameter values tested in PTT0 checked against answers from PTT0 modified using correction factors in Table 3.2 from Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford.

Table PTT1U. Two group continuity corrected Chi-square test of equal proportions (Fisher's exact) (unequal n's)

Computed using the large sample normal approximation, equations (10) and (11) in the nQuery manual appendix modified for unequal n's; see Fleiss JL, Tytun A, Ury SHK (1980) A simple approximation for calculating sample sizes for comparing independent proportions, Biometrics 36:343-346 for details. Figure PTT1U shows computations for one example from Fleiss JL

(1981) Statistical Methods for Rates and Proportions, 2nd edition, John Wiley & Sons, Inc. New York. Solutions for range of parameter values used for PTT0U checked against answers from PTT1 and PTT0U.

Table PTE0. Two group test of equivalence in proportions (equal n's)

Computed using the large sample normal approximation, equation (14) in the nQuery manual appendix, see Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford for formula, references and table. Figure PTE0 shows computations for three examples. Solutions for n for significance level .01,.1, one and two-sided, proportion in group 1 .05, proportion in group 2 of .1,.5, .8, power of 50%, 80%, 99% checked for exact agreement against Table 4.1 in Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford.

Table PTE0U. Two group test of equivalence in proportions (unequal n's)

Computed using the large sample normal approximation, equation (14) in the nQuery manual appendix, modified by multiplying the term PIs(1-PIs) by r, see Makuch R, Simon R (1978) Sample size requirements for evaluating a conservative therapy, Cancer Treatment Reports 62:1037-1040. Figure PTE0U shows computations for five examples. Solutions for r=1 checked against PTE0 for same parameter values. Solutions for unequal sample sizes checked against examples in Makuch and Simon (1978) and Farrington CP, Manning G (1990) Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non non-zero risk difference or non-unity relative risk. Statistics in Medicine 9:1447-1454.

Table STT0. Log rank test of survival in two groups: all followed for fixed time, constant hazard ratio

Computed using equation (17) in nQuery manual Appendix. Figure STT0 shows computations for one example. Solutions for significance level .01 and .2 two-sided, group one proportion .05, .55, .95, group two proportion .1, .5, .9 and power 50 and 99 to sample size and sample sizes back to power checked for exact agreement against Table 9.2 in Machin D, Campbell MJ (1987) Statistical Tables for the Design of Clinical Trials, Blackwell Scientific Publications Oxford.

Table ROT0. Logistic regression test that =0 for one normally distributed covariate x

Computed using normal approximation formula given in F.Y. Hsieh (1989) Sample size tables for logistic regression. Statistics in Medicine 8: 795-802. Figure ROT0 shows computations for one example. Range of parameter values evaluated by Hsieh checked for agreement against tables in his paper.