

DSTPLAN 4.2

Calculations for Sample Sizes and Related Problems

User's Guide

Barry W. Brown
Christopher Brauner
Albert Chan
David Gutierrez
Jay Herson
James Lovato
James Polsley
Kathy Russell
John Venier

The University of Texas M. D. Anderson Cancer Center
Department of Biostatistics and Applied Mathematics, Box 447
1515 Holcombe Boulevard
Houston, Texas 77030
©April 2000

DSTPLAN may be freely copied and (noncommercially) distributed

Contact: Barry W. Brown at above address or at bwb@mdanderson.org

This document typeset in L^AT_EX November 2005.

This work was supported by grant CA-16672 from the National Cancer Institute.

Contents

1	Introduction	8
1.1	Acknowledgments	8
1.2	Source code	9
1.3	Obtaining the software	9
1.4	Legalities (brief)	9
1.5	Overview of the package	11
1.6	Overview of the documentation	12
1.7	How DSTPLAN Works	12
2	Statistical hypothesis testing: ideas and terminology	14
2.1	Example	14
2.1.1	The null hypothesis, significance level, and critical region	16
2.1.2	Power	19
2.1.3	Alternative hypothesis	19
2.1.4	Planning a study	20
2.1.5	Single- versus two-sided tests	21
2.1.6	Note on the use of the term alternative	23
2.1.7	One-, two-, and k- sample tests	23
3	A brief introduction to some statistical distributions and a statistic: situations covered by DSTPLAN	24
3.1	Binomial	24
3.2	Normal Or Gaussian	25
3.3	Poisson	25

3.4	Exponential	26
3.5	A note on the difference between poisson and exponential testing	26
3.6	Correlation Coefficient	27
4	Usage summary for DSTPLAN	28
4.1	Introduction to the use of DSTPLAN	29
4.2	The table and repetition facility	35
4.3	Convention	39
4.4	Ranges of allowed values	40
5	One-sample binomial	42
5.1	The test	42
5.2	Parameters of the test	42
5.3	Method	43
5.4	Comments	43
5.5	Examples	44
6	Two-sample binomial	48
6.1	The test	48
6.2	Parameters of the test	48
6.3	Method	49
6.4	Examples	49
7	Binomial probabilities above and below the median	53
7.1	The test	53
7.2	Examples	54
8	Fisher's exact	59
8.1	Comment	59
8.2	The test	59
8.3	Parameters of the test	60
8.4	Method	60
8.5	Examples	60
9	The matched pairs test for equality of proportions	64

9.1	Description	64
9.2	The test	66
9.3	Parameters of the test	66
9.4	Method	66
9.5	Examples	67
10	Proportion response: conservative vs. expensive treatment	71
10.1	Explanation of the investigation	71
10.2	Parameters of the procedure	72
10.3	Method	73
10.4	Examples	73
11	Two-sample binomial with historical control group	76
11.1	The test	76
11.2	Parameters of the test	76
11.3	Method	77
11.4	Examples	77
12	k-sample binomial	81
12.1	The test	81
12.2	Situations for which calculations are provided	81
12.3	Parameters of the problem	82
12.4	Method	82
12.5	Examples	83
13	Case-control	87
13.1	The test	87
13.2	Parameters of the test	88
13.3	Method	89
13.4	Examples	89
14	Matched case-control	99
14.1	The test	99
14.2	Parameters of the test	99

14.3 Method	100
14.4 Examples	100
15 Sample size correction for loss of subjects over time	109
15.1 The calculation	109
15.2 Parameters	109
15.3 Method	110
15.4 Example	110
16 One-sample Poisson	112
16.1 The test	112
16.2 Parameters of the test	112
16.3 Units of time	113
16.4 Method	113
16.5 Comments	113
16.6 Examples	114
17 Two-sample Poisson distribution	117
17.1 The test	117
17.2 Parameters of the test	117
17.3 Units of time	118
17.4 Method	118
17.5 Examples	118
18 One-sample normal distribution	124
18.1 The test	124
18.2 Parameters of the test	124
18.3 Method	125
18.4 Examples	125
19 Two-sample normal distribution	130
19.1 The test	130
19.2 Parameters of the test	130
19.3 Method	131

19.4	Examples	131
20	Two-sample normal with unequal variances	137
20.1	The distribution and test	137
20.2	Parameters of the test	137
20.3	Method	138
20.4	Example	138
21	Two-sample log-normal	144
21.1	The distribution and test	144
21.2	Parameters of the test	144
21.3	Method	145
21.4	Example	145
22	One-sample exponential distribution	148
22.1	The test	148
22.2	Parameters of the test	148
22.3	Method	149
22.4	Examples	149
23	Two-sample exponential distribution	152
23.1	The test	152
23.2	Parameters of the test	152
23.3	Method	153
23.4	Examples	153
24	One-sample clinical trial	156
24.1	Description	156
24.2	The test	156
24.3	Parameters of the test	157
24.4	Choices for expressing the parameter of the exponential survival distribution	157
24.5	Example	158
24.6	The randomized clinical trial (exponential survival)	162

24.7	Description	162
24.8	The test	162
24.9	Parameters of the test	163
24.10	Choices for expressing the parameter of the exponential survival distribution	163
24.11	Method	164
24.12	Example	164
25	Exponential survival: comparison with historic control	173
25.1	Description	173
25.2	Information required to plan the trial	174
25.3	Choices for expressing the parameter of the exponential survival distribution	175
25.4	Method	176
25.5	Example	176
25.6	Another example	178
26	Two hazard rates depending on time	182
26.1	Description	182
26.2	Information required to plan the trial	183
26.3	Method	184
26.4	Example	184
27	One-sample correlation	189
27.1	The test	189
27.2	Parameters of the test	189
27.3	Method	190
27.4	Comment	190
27.5	Examples	190
28	Two-sample correlation	195
28.1	The test	195
28.2	Parameters of the test	195
28.3	Method	196

28.4 Examples 196

Chapter 1

Introduction

1.1 Acknowledgments

The creation of this code was supported in part by the Core Grant, CA16672, of the National Cancer Institute of NIH to The University of Texas M. D. Anderson Cancer Center. Other support was provided by the State of Texas, and the Larry and Pat McNeil Professorship in Cancer Research held by Dr. Barry W. Brown, and a cooperative study agreement with IBM.

Dr. Barry W. Brown is responsible for the overall design of DSTPLAN.

Mr. James Lovato implemented the current, greatly improved, user interface including the ability to obtain tables of values and the current answer reporting format. Mr. Lovato also wrote the routine for calculations involving a randomized clinical trial with two different hazards depending on time on study.

Mr. Christopher Brauner wrote the routines for binomial probabilities above and below a median, the two sample normal with unequal variance, the two sample log-normal, the two sample correlation, and the one-group clinical trial.

Dr. Jay Herson specified the matched pair design routines, which were coded by Mr. James Polsley. Mr. Polsley also wrote the one-sample correlation routine and the code for the cheap versus expensive treatment.

Mr. David Gutierrez wrote the k-sample binomial routines and the binomial comparison with historic controls routine. Mr. Gutierrez performed the detailed work that was needed to produce the present version of DSTPLAN.

Mr. Albert Chan wrote the routine for the comparison of survival curves using a historic control.

Mrs. Kathy Russell has maintained the program and incorporated new routines.

Mr. John Venier produced version 4.2 by fixing many bugs and cleaning up the source code extensively. He has converted all calculations to double precision and has expanded the allowed range of parameter values to reflect this. He has improved parameter range checking in the routines to produce more meaningful error messages. He has also revised this document and corrected some errors.

Mrs. Bet Moss and Ms. Elizabeth Schwarz provided invaluable assistance with both the documentation and an earlier version of the code.

1.2 Source code

DSTPLAN was written in FLECS (Fortran Language Extended Control Structures) which is a pre-processor similar in concept to RATFOR. The original version of FLECS was written by Terry Beyer of the University of Oregon and placed in the public domain. We added support for Fortran 77 and some new control structures. Those interested in acquiring FLECS should contact Barry Brown at the address below. The current version of DSTPLAN is being maintained in Fortran 77, although we hope to have a Fortran 90/95/2000 version at some point.

The Fortran produced by FLECS is as ugly as that produced by RATFOR. The code was polished using pol from Toolpack1.2, a public domain package that can be acquired from NAG or via anonymous ftp to lambda.lanl.gov (128.165.4.4). The current version is not quite as ugly, but it is still pretty bad. It was polished using the current version of NAG's Fortran source code polishing tools.

1.3 Obtaining the software

This software is available from

<http://biostatistics.mdanderson.org/SoftwareDownload>

The code is furnished in source form and as Win32 and Power PC Macintosh executables.

A document detailing methods used in DSTPLAN is included with the distribution.

1.4 Legalities (brief)

We want to make this program as widely available as possible. Hence, executable code derived from the source provided, the source itself, and the accompanying

documentation may be freely copied and distributed provided that this copyright section accompanies such copies. We encourage the distribution of this code.

Modifications and improvements to the code can similarly be freely copied and distributed. The authors would appreciate a copy of any such improvements for incorporation into later versions.

Source code may be extracted and used for any non-commercial purpose. Non-commercial means that a distribution fee may be charged, but the software is identified as freeware for which the authors receive no compensation from the users.

Commercial use of the source code requires written permission from the Department of Biomathematics at the address given below. Our policy is generally to grant such permission provided appropriate credit is given to us in the documentation.

Department of Biostatistics and Applied Mathematics, Box 447
The University of Texas M. D. Anderson Cancer Center
1515 Holcombe Boulevard
Houston, TX 77030

NO WARRANTY

WE PROVIDE ABSOLUTELY NO WARRANTY OF ANY KIND EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THIS PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

IN NO EVENT SHALL THE UNIVERSITY OF TEXAS OR ANY OF ITS COMPONENT INSTITUTIONS INCLUDING THE UNIVERSITY OF TEXAS M. D. ANDERSON CANCER CENTER BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY LOST PROFITS, LOST MONIES, OR OTHER SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA OR ITS ANALYSIS BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY THIRD PARTIES) THE PROGRAM.

(Above NO WARRANTY modified from the GNU NO WARRANTY statement.)

1.5 Overview of the package

DSTPLAN performs sample size and related calculations needed to plan studies. Parameters of a test include those necessary to describe the distributions under the null and alternative hypotheses, sample sizes, significance levels, and power. Given values for all but one parameter, DSTPLAN will calculate the value of that one parameter.

Following is a list of the study types for which the current version of DSTPLAN will perform calculations:

Studies with Outcomes Distributed According to the Binomial Distribution

- One-sample test
- Two-sample test
- Two-sample above and below the median
- Fisher's exact
- Matched pairs test for equality of proportions
- Conservative versus expensive treatment
- Two-sample test in which one group is a historic control
- K-sample
- Case-control
- Matched Case-Control
- Sample size correction for loss of subjects over time

Studies with Outcomes Distributed According to the Poisson Distribution:

- One-sample test
- Two-sample test

Studies with Outcomes Distributed According to the Normal Distribution:

- One-sample test
- Two-sample test, equal variances
- Two-sample test, unequal variances
- Two-sample log-normal

Studies with Outcomes Distributed According to the Exponential Distribution:

- One-sample test
- Two-sample test

Studies whose Outcome is a Set of Survival Times:

- One Sample Clinical Trials
- Randomized clinical trial for exponential survival times
- Comparison with a historic control - exponential survival
- Randomized clinical trial - two different hazard rates depending on time

Studies whose Outcome is the Correlation Coefficient:

- One-sample test
- Two-sample test

1.6 Overview of the documentation

Chapters 2 and 3 constitute an introduction to the statistical concepts involved in planning a study. It can safely be skipped by knowledgeable readers. Those totally ignorant of these concepts may well find this treatment inadequate and are referred to any text on statistics. These abound, but one we recommend is:

Colton, T.: *Statistics in Medicine*, Little Brown and Co.: Boston, 1974.

Chapter 4 is an introduction to the use of DSTPLAN and should be carefully read by users. The remaining chapters describe each study type and the routines implemented for dealing with it.

1.7 How DSTPLAN Works

The power of a particular test is usually easily calculated by the algorithm in the reference cited in the Methods section of the write-up devoted to a particular problem type. Computation of other parameters from the formula for power is done if this is straightforward.

If a parameter cannot be directly calculated from the formula for power, it is found numerically. Power generally either increases or decreases monotonically

with each parameter, which greatly simplifies the numeric problem. Power is evaluated at some initial approximation to the answer. Whether the approximation is above or below the correct answer is known due to the monotone change in power. Steps are taken in the appropriate direction until the answer is bounded. Each time that a step fails to bound the answer, the step size is doubled.

Once the answer is bounded, a sophisticated improvement to the interval halving method is used to refine the approximation to the answer. This method is Algorithm R from:

Bus, JCP and Dekker, T.J.: "Two Efficient Algorithms with Guaranteed Convergence for Finding a Zero of a Function." *ACM Transactions on Mathematical Software* 1:330, 1975.

Chapter 2

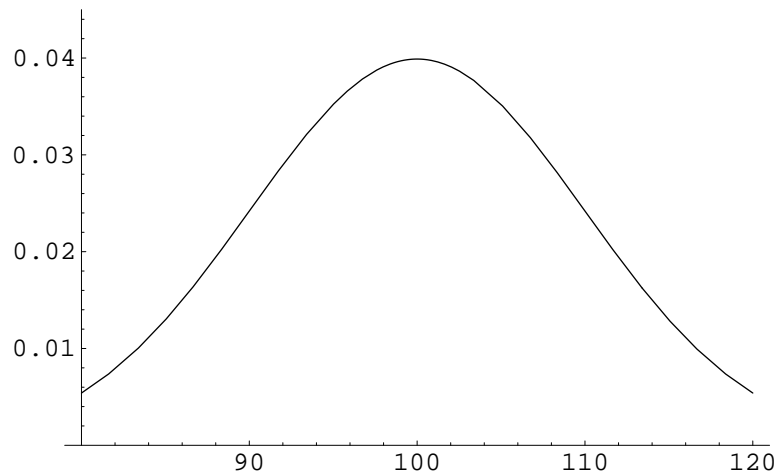
Statistical hypothesis testing: ideas and terminology

2.1 Example

A scientist is studying a drug that, he thinks, raises the intelligence of animals. There is a well standardized intelligence test for mice, and it is known that the tested scores of mice are distributed normally with a mean of 100 and a standard deviation of 10. The scientist proposes to administer the drug to a group of mice, to test their intelligence after administering the drug, and to compare the average test score to the known value of 100 to see if there is statistical evidence that the post-treatment mean exceeds 100.

This would be a very dangerous experiment, since it relies on many assumptions. These assumptions include the assumptions of the normal distribution of the scores and the values of the mean and of the standard deviation. Actual experiments are generally more sophisticated and allow the assumptions to be tested; however, the oversimplification is useful for demonstrating concepts.

The assumption of the normal or Gaussian distribution of scores on the intelligence test means that the graph of scores (x -axis) versus the frequency of the values (y -axis) is a curve of a particular mathematical function. The graph looks like Figure 1 and is bell shaped and symmetric about its mean.

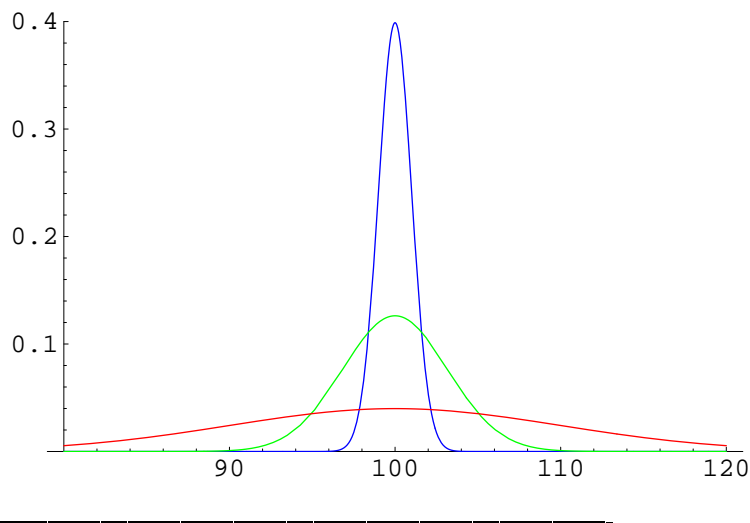


The assumption of normal distribution of scores is used to find the probability that an untreated mouse will yield an intelligence score that lies between two values, say 100 to 110. To determine this probability graphically from Figure 1, draw lines up from the x -axis to the curve at these values. Now measure the area bounded by the x -axis, the curve, and the two lines just drawn. This area is the desired probability – in this case, 0.34. Similarly, we can find the probability that the intelligence will be greater than 110 by finding the area of the curve bounded by the x -axis, the curve, and the rightmost line we just drew. Although the area so bounded extends forever to the right, the area is finite and, in this case, equals 0.1587. In practice tables, calculators, or computers are used to find such probabilities; they are not taken from areas measured on a graph.

The mean of the normal distribution occurs at the x -value of the highest point on the curve, in the case presented, 100.

The standard deviation is a measure of the probability of experimental values occurring far from the mean. If the standard deviation is small, then this probability is small; if it is large, then the probability of values falling far from the mean is large.

Figure 2 shows three normal distributions each with mean 100. The standard deviations of the three are 10 (red), 3.16 (green), and 1 (blue) respectively. Normal distributions with a low standard deviation are tall and skinny. Normal distributions with a high standard deviation are short and fat.



2.1.1 The null hypothesis, significance level, and critical region

The null hypothesis states the proposition that the study is trying to disprove; in our example, the null hypothesis can be stated: “the average intelligence score of drug treated mice is 100”.

An average test score of treated animals that is greater than 100 is evidence against the null hypothesis. How large must this average be to be (statistically) significant?

To answer this question, a small probability known as the significance level of the statistical test is specified. The significance level is the probability that if the null hypothesis is correct, the experiment will yield evidence that will cause a false conclusion that the null hypothesis is false. Restated, the significance level is the probability of falsely rejecting the null hypothesis due to chance alone.

It is desirable to make the significance level as small as possible. Unfortunately, smaller significance levels require more evidence to reject the null hypothesis when it is indeed false. Consequently, small significance levels require large and expensive trials. Larger significance levels allow smaller and cheaper trials but the strength of the evidence so obtained is less.

Popular choices for the significance level are 0.05 and 0.01.

Once the significance level has been chosen, the critical region of the test is determined. The critical region is the set of outcomes that will cause the null hypothesis to be rejected as untrue. If the null hypothesis is true, the

probability of an outcome in the critical region is precisely the significance level. This relation between the significance level and critical region allows the calculation of the critical region.

A small and cheap test for the example can now be designed. The drug will be given to only one mouse and its intelligence will be tested using a value of 0.05 as the significance level of the test. The critical region of the test will consist of all test scores greater than some number. The probability of a score greater than this number should be 0.05 if the true mean is 100. Using tables, the number is found to be 116.45. If the treated mouse scores above 116.45 on the test, there is evidence at a significance level of 0.05 that the drug has an effect, otherwise, there is no such evidence and the null hypothesis cannot be rejected.

There are two different uses of the term significance level. Those who have analyzed data will recognize the significance level as the p-value, and will note that it is obtained when the data is analyzed, not before the study has begun. For example, when the treated mouse scores 116.45 on the intelligence test, then one refers to a table and finds that the p-value is 0.05. This use in analysis is very closely related to the usage described for planning, but the steps of the process are taken in reverse order. In analysis, the outcome of the experiment determines the critical region for the p-value. The critical region consists of the observed outcome and all outcomes more extreme than that observed. The probability of an outcome in the critical region assuming that the null hypothesis is true determines the p-value. In planning a study, in contrast, the desired significance level is specified and the critical region is calculated from it.

There is an asymmetry in the testing framework with respect to the null hypothesis. The null hypothesis can be rejected, or it can fail to be rejected, but it cannot be proved. Failing to reject the null hypothesis is not at all the same as proving that it is correct. The null hypothesis is usually a statement of equality, and exact equality cannot be demonstrated. Two objects cannot be shown to be the same size by measuring them, because their sizes may differ by less than the measurement error. Similarly, a statistical study cannot demonstrate that two averages are the same, because a larger study might show a small difference that the given study did not reveal. Fortunately, this situation does not obviate the worth of either measurement or of statistical investigation. Both can show that two things probably don't differ much.

Returning to the example, the reader should be appalled at the thought of a one mouse test for drug effectiveness. One objection to the test might be that with only one animal, a significant outcome might be a fluke, and due to chance; however, this is not so. The mathematics takes care of that problem. With one animal, we have only one chance in twenty of rejecting the null hypothesis if it is true. With a thousand animals and a test at the 0.05 significance level, we

would still have the same chance of rejecting the null hypothesis if it were true.

One problem with the single mouse experiment is that it provides no way to check the assumptions used to construct the test. These assumptions are powerful indeed. They include the assertions that the drug treated animals have normally distributed intelligence scores and that the standard deviation of these scores is 10. With one mouse these assumptions cannot be tested at all. With 1000 animals, frequency histograms could be examined to see if they look normal. The sample standard deviation could be calculated and examined for proximity to the assumed value of 10. These objections to the single mouse experiment are not part of the sample size calculation process presented here, although such concerns should be most carefully considered in the design of any trial.

An issue in the single mouse trial that is relevant to sample size consideration is that the drug effect would have to be huge to have any probability of demonstrating that it is real. Hence, an outcome that fails to demonstrate a drug effect would not necessarily indicate that the effect is probably too small to be of interest. For example, if the treated mouse scores 110, this might well be considered to be enough evidence of an effect to perform a larger study (although not enough evidence to publish the result, since it is not significant at the 0.05 level). The necessity for a larger study implies that the expense of the single mouse study would be wasted. Here, the waste is not large, but if the initial study were large and expensive, or if it involved putting humans at risk, the consequences of having to abandon the data obtained to perform a larger trial are substantial. The primary purpose of determining a needed sample size is to assure that such waste does not occur; that is, to assure that the probability of missing an important difference is small.

Let's examine the effect of significance level on the critical region. For the single mouse experiment, if the significance level is lowered to 0.01 then the mouse must score above 123.3 for evidence of effectiveness; at a significance level of 0.001 the minimal score becomes 131. Clearly, as the significance level goes down, the minimal score to be achieved to detect effectiveness goes up.

Increasing the sample size has the opposite effect: it moves the critical value toward the null hypothesis value. Figure 2, discussed earlier, also displays the distribution of mean scores, assuming a true mean of 100, for three different sized groups of mice: one mouse (short fat curve), ten mice (middle curve), and 100 mice (tall skinny curve). As the number of mice increases, the distribution of the mean score concentrates near the value of 100, and the critical region lies closer to the mean. For a significance level of 0.05, the critical region for one mouse starts at 116.45; for ten mice it starts at 105.2; and for 100 mice, 101.64.¹

¹These numbers are true assuming that the true standard deviations of the sample means are as given. In practice one usually does not know the true standard deviation and instead

Because the critical region is closer to the mean as more mice are tested, smaller drug effects can be detected.

2.1.2 Power

The power of a test is the probability of rejecting the null hypothesis if it is, in fact, incorrect. Power depends on the magnitude of the deviation from the null hypothesis. With a fixed number of mice, the scientist stands a much better chance of demonstrating the existence of an effect if the drug raises the intelligence by twenty points than if it only raises it by one-half point. Power also depends on sample size. For a particular effect level, the probability of obtaining statistically significant evidence increases with the number of mice tested. Power also increases with increasing significance level; this is in accordance with intuition. If the level of evidence needed for a conclusion is lessened (increasing significance level), then the probability of obtaining that evidence (power) should increase.

From the definition of power, it is evident that high values are desired. High values require large and expensive studies compared to low values if all other parameters are fixed; in this respect high values of power are similar to low values of significance.

The value of 0.8 is quite popular for power, and this figure has been cited as the minimally acceptable value for clinical trials.

2.1.3 Alternative hypothesis

Since the power of a test depends on the true effect of the drug, power can only be specified at particular values of this effect. In planning a study, the smallest effect that should have a reasonable chance of detection is used. This value, used for calculating power, is called the alternative hypothesis value. The smallest effect that should be detectable is determined by the subject matter of the test, not by statistical considerations.

uses the standard deviation calculated from the sample itself. In this case, a t distribution must be used to calculate the critical value or values. To evaluate the experimental results, the t statistic would be calculated from the observed sample and this number would be compared to the critical value or values, thereby performing a "t-test". The t-test critical values are generally more extreme than those calculated using a normal distribution, *i.e.* the upper t-test critical values are generally higher than the corresponding values from a normal distribution, and the lower t-test critical values are generally lower. In the examples given, the t distribution critical values would be 105.8 for ten mice and 101.66 for 100 mice. The t distribution critical value for one mouse can not be calculated; a sample of at least two mice would be required in order to calculate the sample standard deviation. DSTPLAN assumes that standard deviations are always estimated from samples, and uses the t distribution accordingly.

In the example, the experimenter may choose a minimal detectable value of the true drug effect as that raises intelligence by two points, *i.e.*, from 100 to 102. The alternative hypothesis is that the true mean intelligence of treated mice is 102.

The definition of power can now be stated somewhat more precisely. The power of a test at the alternative hypothesis value is the probability that the outcome will lie in the critical region if the alternative hypothesis is true.

2.1.4 Planning a study

All of the steps involved in planning a study have now been discussed; they are briefly reviewed here.

The first step consists of deciding on a reasonable statistical model that will represent the possible outcomes of the trial. In the example, the model is that the intelligence scores of the mice will be distributed in a normal or Gaussian fashion.

The null hypothesis is specified; this is usually a statement that the trial is designed to disprove. In the example, the null hypothesis is that the true mean intelligence score on the test is 100.

The minimal effect that should be detectable is specified. The size and cost of the test depends greatly on the size of this effect. In the example, this alternate hypothesis is that the true mean intelligence score on the test is 102.

In some cases, such as the example, there is a nuisance parameter, which doesn't change between the null and alternate hypothesis but which does affect the outcome of the study. This parameter must be specified. The standard deviation of the normal distribution is the nuisance parameter of the example. In the example, the standard deviation is assumed to be 10.

Finally, the significance level and power of the trial must be specified in order to determine the sample size. The traditional value of 0.05 will be used for the significance level, but a high value of 0.99 will be taken for the power.

The minimal sample size which meets these requirements is 395 mice.²

The sample size is frequently beyond the resources of the investigator, who must then decide whether a feasible test is worth pursuing. The investigator

²As in [2.1.1], the number 395 was obtained using normal distribution calculations which assume that the standard deviation of mouse intelligence scores is known to be 10. In practice, the standard deviation is not known and must be estimated from the sample of mouse intelligence scores. If the standard deviation is to be estimated from the sample, t distributions must be used to calculate critical values and hence minimal sample sizes. If this were done, a sample size of 396 would be obtained. DSTPLAN uses t distributions where appropriate and hence will report that a sample size of 396 would be needed for this example.

may only be able to afford to test 100 mice, not 395.[2.1.4] If the sample size is fixed, other conditions obviously must change. Using DSTPLAN, we find that if only one hundred mice are tested the power of the test, holding the other parameters fixed, is 0.63. This is not acceptable, since experiments are rarely knowingly undertaken when the probability of detecting an important difference is only a little over one-half. If the power is reduced to 0.9, however, the smallest intelligence score that would be detectable with this power is 102.95. It is likely that this combination of conditions is satisfactory. If the drug raises intelligence by three points there is a ninety percent chance of proving that fact (at the 0.05 significance level) in an experiment with sample size 100.

DSTPLAN facilitates the examination of alternatives in a testing situation in which resources limit the size of the trial. Any parameter involved in the test can be calculated given the values of all of the other parameters.

2.1.5 Single- versus two-sided tests

In the example, the investigator has a drug which is believed to increase intelligence. The test described was constructed to demonstrate this effect. The investigator is not interested in the drug if it does not change intelligence at all or even if the drug lowers intelligence. In either of these cases, the drug will not be applied to human subjects.

If the drug were not being investigated solely for possible beneficial effects, but rather because it had been found to be a significant contaminant of a city water supply, the situation would be considerably changed. In this circumstance mean scores appreciably lower than 100 could not be ignored; they might indicate harm to the consumers of the water supply. Instead of asking only whether there were any evidence of beneficial effect of the drug, results would be examined for both beneficial and malign effects.

These cases illustrate the difference between a one-sided and a two-sided test. When the drug is being tested only for increasing intelligence then any evidence that it lessens intelligence will be ignored. Consequently, a one-sided test is desired. The entire critical region is concentrated in the area above mean scores of 100 so as to minimize the chance of missing an effect in this direction should there be one.

But if the drug is being investigated for either a beneficial or harmful effect, a two-sided test is needed. The two-sided test splits the critical region with half going above the null hypothesis value and half below, so that a change in either direction can be detected. In practice, most tests are two-sided because there is interest in a change in either direction. Only in tests of a new process for improvement over an old process are one-sided tests used.

As an example of the change caused by a two-sided test, the 100 mouse experiment is again considered. In a one-sided test using a significance level of 0.05, the critical region consists of all scores higher than a certain number; if the true mean score is 100, then the probability of the mean of the tested animals falling above the number must be 0.05. From tables, the number is found to be 101.645. In contrast, in a two-sided test, the critical region consists of both a set of numbers lower than one point and also the set of numbers higher than another point. Both portions of the critical region have a probability of 0.025 under the null hypothesis. The two points are thus 98.04 and 101.96. A drug effect will be judged real if the mean score of the 100 treated animals either lies below 98.04 or above 101.96.³

If the drug increases intelligence by an average of 3 points, then the probability that the mean of the tested animals lies in the critical region for a one-sided test is 0.91. The probability that it lies above 101.96 (the bound on the upper portion of the critical region for a two-sided test) is 0.85. The two-sided test thus has a power of 0.06 less than the one-sided test.⁴

Power in this example has been slightly redefined from its usual meaning in two-sided tests, and this redefinition is used throughout DSTPLAN. The usual definition is the probability of rejecting the null hypothesis when the alternative is true; this would include both the probability of a mean tested score above 101.96 and a mean tested score below 98.04. However, a mean score below 98.04 would cause the investigator to conclude that the drug lowered intelligence when in fact it raises it. The redefinition excludes this erroneous outcome from the calculation of power. In practice, this change is usually negligible. In the example, the change in power from that of the traditional definition is only 0.0000004.⁵

³As in [2.1.1] and [2.1.4], the values of the critical regions in this paragraph are true assuming the standard deviation of the distribution of mouse intelligence scores is known to be 10. In this case, the experimenter may look up the critical value in normal distribution tables. In practice, since the standard deviation of the sample must be estimated from the sample, t distribution tables must be used to calculate the critical values as t statistics must be calculated to evaluate the studies. In this example, t distribution tables would produce a critical value of 101.660 for the one-sided case and critical values of 98.02 and 101.98 for the two-sided case. DSTPLAN, although it does not always output critical values directly, nevertheless does use them in its calculations. When it does it uses critical values from t distributions where appropriate.

⁴As in [2.1.1], [2.1.4], and [2.1.5], this is true for a normally distributed sample mean, while DSTPLAN follows the typical practice of performing these calculations using t distribution critical values. DSTPLAN also uses the non-central t distribution for power calculations. As a consequence, DSTPLAN calculates the two-sided power to be 0.84, 0.07 less than the one-sided power, which it also calculates to be 0.91.

⁵Following the discussion in [2.1.1], [2.1.4], [2.1.5] and [2.1.5], the reader may be interested to note that the change in power is 0.0000004 whether one uses normal or t distribution calculations.

2.1.6 Note on the use of the term alternative

The term alternative hypothesis has been described as a specific parameter value used in the process of study planning. In describing statistical tests, the term is used to describe generally whether the test is one- or two-sided, and, in some cases whether the side being tested is above or below the null hypothesis value.

The test of our example could be described as having a one-sided alternative or being a test of the null hypothesis that the mean score of treated mice is 100 versus an alternative that the mean is greater than 100. A test of whether the drug affected intelligence by either raising or lowering it could be described as being a test of the null hypothesis that the mean score of treated mice is 100 versus a two-sided alternative.

This use of the term alternative correctly indicates that the test is attempting to demonstrate any deviation from the null hypothesis in the direction or directions noted. The use of the term alternative hypothesis as a specific value is appropriate in planning a study, because power is a function of the specific alternative value chosen.

2.1.7 One-, two-, and k- sample tests

In our example so far, the investigator is testing a single group against a known value. Consequently, there is only one experimental sample and the test compares the mean value of this sample to a particular value. Had the experiment used a treated group and a control group, a two-sample test would be appropriate. The two-sample test tests the groups for equality of mean values.

It is possible to have more than two groups in an experiment (the letter k is used to denote the number of groups in this document). In such a case, the test is again for the null hypothesis of equality of all of the groups.

Chapter 3

A brief introduction to some statistical distributions and a statistic: situations covered by DSTPLAN

A very brief description of the statistical distributions for which sample size determinations can be performed by DSTPLAN is given in this chapter.

3.1 Binomial

The binomial distribution can be described as follows. A number, n , trials of some nature are performed; for each trial the outcome is whether or not some predefined event occurs. The probability of this event is the same value, p , for each of the trials. The binomial distribution describes the probability that there are precisely s events observed in the n trials.

There are cases which this abstract definition corresponds precisely to the concrete realization of an experiment. For example, the trials might consist of administering a drug to a series of subjects and observing whether or not the subjects respond in some predetermined fashion.

There are other cases in which the trial is the selection and observation of a subject and the event consists of the subject's having some predefined property. In this case, it is more usual to speak of the proportion of subjects

with the property than of the probability that the subject has the property. One example might be the proportion of defects in a lot of manufactured items. Another would be the proportion of students passing a standardized test.

The probability or proportion, p , is the parameter of the binomial distribution that is most frequently tested, and the testing situations for which DSTPLAN provides calculations all involve p .

3.2 Normal Or Gaussian

The normal distribution is a bell shaped curve that is symmetric about its mean. Graphs of several normal distributions are presented in Chapter 2. The normal distribution is frequently used as a model for continuous data. One theoretical reason for its popularity is that any characteristic of a population that is influenced by a large number of random factors, none of which predominate in its determination, is approximately normally distributed.

Biological characteristics are usually influenced by a large number of genetic and environmental influences. Consequently, the distribution of characteristics over the population is approximately normal (Gaussian). Height, weight, and scores on tests are examples.

The normal distribution has two parameters: its mean and its standard deviation. The mean is the point about which the distribution is symmetric and is the natural location parameter of the distribution. (A location parameter determines whether values from a distribution tend to be large or small – *i.e.* it determines the location of the distribution on the real line.) The standard deviation measures the spread of values from the mean. Distributions with larger standard deviations are more likely to produce observations far from the mean than are distributions with smaller standard deviations.

The parameter of the normal distribution most frequently tested is the mean, and the testing situations for which DSTPLAN provides calculations all test the values of means.

3.3 Poisson

The Poisson process is a mathematical representation of pure randomness; according to this process an event is as likely to occur at any one time as any other time. The probability of a particular number of events in a fixed period of time is obtained from the Poisson distribution.

The Poisson process is an excellent approximation to the binomial distribution in which the number of trials is very large and the probability of success

per trial is small.

The Poisson process has been found to be a good model for a number of empirical observations including the number of misprints per page of a book, the number of raisins in a loaf of bread, the number of radioactive events per unit of time, and the number of mutations in a colony of bacteria.

The Poisson process has only one parameter, the rate (mean number of events per unit time). This is the parameter that is tested.

3.4 Exponential

The exponential distribution is associated with a continuous measurement such as time. The exponential distribution arises from its connection with the Poisson process – the time between successive events in a Poisson process is distributed exponentially. An example of data approximately distributed exponentially is the number of miles traveled by an automobile between occurrences of flat tires.

The exponential distribution can be obtained from a different viewpoint that is frequently encountered in clinical research and industrial engineering. The hazard rate of a process at a particular time is the probability of an event at that time. If this hazard rate is constant, then the distribution of times between events is distributed exponentially.

In this formulation, an example of the event is death and time can be considered to be the age of a human being. The hazard of death in humans is not constant; newborns and the elderly are more at risk than the remainder of the population. So the distribution of ages at death is not approximately exponential. However, if we consider populations who have been diagnosed with a dangerous disease such as cancer, and examine the time from diagnosis to death, then frequently the distribution of this time is exponential. The approximately unchanging high risk of death from the disease greatly overshadows the slowly changing risk due to aging, so the composite risk is nearly constant.

The exponential distribution has only one parameter, the mean, so this is the parameter tested.

3.5 A note on the difference between poisson and exponential testing

In tests for one or two means of an exponential distribution, a fixed number of test units are observed until an event is observed for each unit. The total time

of the test is thus a random variable.

In testing Poisson rates, the process or processes are observed for a predetermined amount of time and the number of events is the random variable that is observed.

Because of the close connection between the Poisson and exponential distributions, it matters little which testing procedure is used. The test statistic will be the mean time to an event in either case. Both exponential and Poisson calculations are provided in DSTPLAN for the convenience of the user who may conceptualize a testing situation in one or the other fashion.

3.6 Correlation Coefficient

The correlation coefficient is not a statistical distribution, but rather a measure of the degree of linear association of one factor with another. If a graph of the two factors is a straight line that increases from left to right then the correlation coefficient is 1.0; if it is a line that decreases from left to right then the correlation coefficient is -1.0. If the values of one factor do not change systematically with the values of the other, then the correlation coefficient is 0.0. In most cases, the values are scattered about a line with positive or negative slope and the correlation coefficient lies between 0 and 1 or 0 and -1, respectively.

Chapter 4

Usage summary for DSTPLAN

This section is for those who don't read documentation.

1. At the beginning of the program you are given the choice of receiving either a full report of the calculation or only the answer. Choose the full report unless there is compelling reason not to.
2. Two levels of menus lead to a particular calculation.
3. When queried for values, at least one entry must be the character '?' for the item or items to be calculated. In two-sample calculations, both sample sizes may be entered '?' in which case equal sample sizes for the two groups to provide the desired power will be calculated. There is one exception to this rule. For the normal, unequal variance, calculation when the two question marks are entered for sample size, you are given the option of whether to use equal sample sizes for the groups or whether to use an efficient allocation. The efficient allocation reduces the overall sample size for a given power.
4. An entry of the character 't' in place of a parameter value allows a table of values to be specified for that parameter. The calculation is performed on each value in the table. In two-sample calculations, both sample sizes may have 't' entries in which case a table of equal sample sizes will be entered.

The most used choice in the construction of tables is the specification of equally spaced values via low value, high value, and number of intervals. To obtain a set of nice values, the number of intervals should divide the high value minus the low value.

5. The entries that are not '?' or 't' must be numbers and they must make sense. Thus, significance and power values must be between 0 and 1, the mean of a Poisson must be nonnegative, and so on.
6. The form <number>:<new value> can be used to change exactly one parameter value between different calculations. The <number> is the ordinal number of the item whose value is to be changed, <new value> is the new value of this parameter. All other values remain the same as in the previous calculation.

4.1 Introduction to the use of DSTPLAN

The use of the system will be presented through annotation of a computer session. The session that the reader obtains on his machine may differ slightly from that presented as minor details of DSTPLAN vary slightly from machine to machine.

===== Begin Echo of Computer Session =====

DSTPLAN
Version 4.2: April, 2000
Calculations for Sample Sizes and Related Problems

Barry W. Brown
Christopher Brauner
Albert Chan
David Gutierrez
Jay Herson
James Lovato
James Polsley
John Venier

This work was supported by grant CA-16672 from the National Cancer Institute

Copyright 2000 for:

The University of Texas M. D. Anderson Cancer Center
Department of Biomathematics, Box 237
1515 Holcombe Boulevard
Houston, TX 77030 (USA)

Contact: BWB at above address or bwb@mdanderson.org

This program may be freely copied and (noncommercially) distributed

```

Press Return / Enter to continue:
$?
                                     # to continue past banner
Enter
    (1) To print only answers to the screen
    (2) To print full problem report including
        answers to the screen
    (3) As (2) but also write to a file
Select one of the above:
$?
3                                     # print full report to screen and a file

```

=====
===== End Echo of Computer Session =====

The program initially presents a banner identifying itself and requests the user to press the enter or return key to continue.

The program then offers several options for how the results should be displayed and an option to save the results to a file. Option one displays only the calculated result. Option two displays a list of the input parameters as well as the calculated result. Generally, option 2 is preferred to option 1.

Option three displays a list of the input parameters and the calculated result and also writes this list to a file (whose name is requested if this option is chosen). The file can be printed after the DSTPLAN session is finished for a permanent record of the session.

=====
===== Begin Echo of Computer Session =====

```

Tests for trials with:
(0) Exit from DSTPLAN
(1) BINOMIALlly distributed outcomes
(2) POISSON distributed outcomes
(3) NORMALlly distributed outcomes
(4) EXPONENTIALlly distributed outcomes
(5) SURVIVAL time studies
(6) Outcome is the CORRELATION coefficient

```

```

Tests for trials with BINOMIALlly distributed outcomes:
(0) Return to main menu
(1) One-sample (exact method)
(2) Two-sample (arcsin approximation)
(3) Two-sample binomial above and below the median
(4) Fisher's exact - equal numbers each group

```

(method of Casagrande, Pike and Smith)

WARNING - Very conservative for two-sample binomial

- (5) Matched pairs test for equality of proportions
(method of Miettinen)
- (6) Proportion responders, conservative vs expensive treatment
(normal approximation to binomial)
- (7) Two-sample binomial with historical control group
(arcsin approximation)
- (8) K-sample (non-central chi-square)
- (9) Case-control (arcsin approximation)
- (10) Matched case-control
- (11) Sample size correction for loss of subjects over time

Tests for trials with POISSON distributed outcomes:

- (0) Return to main menu
- (1) One-sample (exact method)
- (2) Two-sample (Cox's approximation)

Tests for trials with NORMALLY distributed outcomes:

- (0) Return to main menu
- (1) One-sample (exact method)
- (2) Two-sample with equal variances (exact method)
- (3) Two-sample with unequal variances (Welch approximation)
- (4) Two-sample log-normal (exact method)

Tests for trials with EXPONENTIALLY distributed outcomes:

- (0) Return to main menu
- (1) One-sample (exact method)
- (2) Two-sample (exact method)

Tests for trials of SURVIVAL time:

- (0) Return to main menu
- (1) One-sample clinical trial
- (2) Randomized clinical trial - exponential survival
(method of George and Desu)
- (3) Comparison with a historical control - exponential survival
(method of Dixon and Simon)
- (4) Randomized clinical trial - two different hazard rates
depending on time (method of Schoenfeld)

Tests for trials in which outcome is CORRELATION coefficient:

- (0) Return to main menu
- (1) One-sample correlation (normal approximation)
- (2) Two-sample correlation

=====
 ===== End Echo of Computer Session =====

The above edited session shows a listing of all planning situations for which DSTPLAN can perform calculations. The main menu, which is the first menu in the list, presents a list of distributions for which calculations are available. All choices within each available distribution are then listed along with the option to return to the main menu. The user interaction needed to generate this listing as well as redundant main menu displays have been eliminated from the above session for a cleaner presentation.

Now a selection of the one-sample normal distribution will be made for the following example:

=====
 ===== Begin Echo of Computer Session =====

Tests for trials with:

- (0) Exit from DSTPLAN
- (1) BINOMIALlly distributed outcomes
- (2) POISSON distributed outcomes
- (3) NORMALlly distributed outcomes
- (4) EXPONENTIALlly distributed outcomes
- (5) SURVIVAL time studies
- (6) Outcome is the CORRELATION coefficient

Select one of the above:

\$?

3 # NORMALlly distributed outcomes

Tests for trials with NORMALlly distributed outcomes:

- (0) Return to main menu
- (1) One-sample (exact method)
- (2) Two-sample with equal variances (exact method)
- (3) Two-sample with unequal variances (Welch approximation)
- (4) Two-sample log-normal (exact method)

Select one of the above:

\$?

1 # One-sample normal (exact method)

=====
 ===== End Echo of Computer Session =====

In the previous section, the assumption was made that the standard deviation of the treated group of mice is known to be 10. In practice, however, the standard deviation is almost always estimated from the data when performing the test. However, the standard deviation is important in sample size calculations so DSTPLAN requires an initial estimate to perform its calculations. The estimation of the standard deviation from the data is taken into consideration in the DSTPLAN routines for the normal distribution. Let's see what difference this makes.

A previous calculation gave the sample size necessary to detect a treatment effect of raising the mean intelligence score to 102 versus a null hypothesis value of 100. The standard deviation of tested intelligence scores is 10, and a significance level of 0.05 and a power of 0.99 were chosen. A sample size of 395 is required to meet these criteria. Note that the standard deviation is not estimated in this example. The corresponding sample size using the estimate of the standard deviation will be calculated using DSTPLAN.

```
===== Begin Echo of Computer Session =====
```

```
***** One-sample normal (exact method) *****
```

```
Problem Initialization:-----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
1 # 1-sided test
```

```
===== End Echo of Computer Session =====
```

The program tells us to enter 1 or 2 for a one- or two-sided test. A one-sided test is chosen as per the example.

A problem has now been specified. This designation consists of the routine (one-sample normal), and whether the test is one- or two-sided (one-sided).

```
===== Begin Echo of Computer Session =====
```

```
Parameter Specification: -----
```

```
Enter: 1: Null hypothesis mean  
       2: Alternative hypothesis mean  
       3: Standard deviation  
       4: Sample size
```

5: Significance level

6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

100 102 10 ? 0.05 0.99 # One-sample normal ex. #1

=====
End Echo of Computer Session
=====

The null hypothesis mean is 100, the alternative is 102, the common standard deviation is 10. The '?' indicates that the sample size is to be calculated. The desired significance level is 0.05 and the desired probability of detecting the difference between means of 100 and 102 is 0.99. (The 't' option will be discussed later.)

=====
Begin Echo of Computer Session
=====

***** One-sample normal (exact method) *****

These calculations are for a one sided test.

Null hypothesis mean 100.000

Alternative hypothesis mean 102.000

Standard deviation 10.000

Significance level 0.050

Power 0.990

The Sample size is calculated to be 395.619

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # another calculation

=====
===== End Echo of Computer Session =====

The program requests the parameters' values. When these are provided, the program determines that the requisite sample size is 395.6. Since only integral sample sizes are feasible, this number should be rounded up to 396. The change in required sample size according to whether the standard deviation is assumed known or estimated from the data is only 1; in problems in which smaller samples are used, the difference will be greater.

After calculating an answer, the program offers several options for the next action. More calculations of the same type problem may be performed. This provides an opportunity, for example, to see the sample size required for different treatment effects. The problem specification can be changed: the problem specification in this case is whether the test is one- or two-sided. The third option is to return to the previous menu listing planning situations with the Normal distribution type. The final option is to exit DSTPLAN.

4.2 The table and repetition facility

Several calculations that involve a change in only one parameter are frequently needed, so DSTPLAN has a facility to lessen the amount of typing that would otherwise be required. When the list of parameters is to be entered after this question, instead of entering all of them if only one is to be changed, enter its ordinal number from the list, then a colon (:), and finally the new value. The values of the modified set of parameters will be printed followed by the answer to the new problem.

Here is an example in which the power of the test is desired for a variety of sample sizes. The values from the previous example are used. Both the table facility using the 't' option and the repetition facility using the colon (:) will be illustrated. This session picks up where the last session left off, *i.e.*, with having just chosen to perform another calculation of the same sort:

=====
===== Begin Echo of Computer Session =====

Parameter Specification: -----

Enter: 1: Null hypothesis mean
2: Alternative hypothesis mean

- 3: Standard deviation
- 4: Sample size
- 5: Significance level
- 6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

100 102 10 t 0.05 ? # One-sample normal ex. #2 (t)

You are specifying a list for values of Sample size

The list currently contains 0 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
 (2) ADD values equally spaced between bounds
 (3) ADD values as (2) with logarithmic spacing
 (4) PRINT list
 (5) DELETE a single element of list
 (6) DELETE consecutive elements of list
 (7) SORT list in ascending order and eliminate duplicate values
 (8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

7 # The number of values

Enter 7 values

\$?

50 100 150 200 250 300 350 # The 7 values

The list currently contains 7 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
 (2) ADD values equally spaced between bounds
 (3) ADD values as (2) with logarithmic spacing
 (4) PRINT list
 (5) DELETE a single element of list
 (6) DELETE consecutive elements of list
 (7) SORT list in ascending order and eliminate duplicate values
 (8) QUIT

\$?

8 # QUIT

***** One-sample normal (exact method) *****

These calculations are for a one sided test.

Null hypothesis mean 100.000

Alternative hypothesis mean 102.000

Standard deviation 10.000

Significance level 0.050

Sample size	Power
50.000	0.401
100.000	0.634
150.000	0.786
200.000	0.880
250.000	0.934
300.000	0.965
350.000	0.982

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # another calculation

Parameter Specification: -----

- Enter:
- 1: Null hypothesis mean
 - 2: Alternative hypothesis mean
 - 3: Standard deviation
 - 4: Sample size

5: Significance level

6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

2:103 # One-sample normal ex. #3 (:)

Current parameter values are:

100.000 103.000 10.0000 T 0.500000E-01 ?

You are specifying a list for values of Sample size

The list currently contains 0 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

7 # The number of values

Enter 7 values

\$?

50 100 150 200 250 300 350 # The 7 values

The list currently contains 7 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values

```

      (8) QUIT
$?
8          #          QUIT

      ***** One-sample normal (exact method) *****

These calculations are for a one sided test.

Null hypothesis mean ..... 100.000

Alternative hypothesis mean ..... 103.000

Standard deviation ..... 10.000

Significance level ..... 0.050

                Sample size          Power

                50.000                0.673
                100.000               0.909
                150.000               0.978
                200.000               0.995
                250.000               0.999
                300.000               1.000
                350.000               1.000

```

===== End Echo of Computer Session =====

This command facility is actually a little more fancy than has been described. The need for the extra features is rare, however, and readers who are not familiar with computers are advised to skip the remainder of this section.

There are two types of responses to the prompt for new sets of parameter values for calculation. One is to type in the entire set of values. The other is to type a command of the form <integer>:<value>. This command causes the parameter whose ordinal position is <integer> to assume the value, <value>.

4.3 Convention

In the two-sample routines in which the location parameter of one of the groups is to be calculated, the computer must have a way of knowing whether it is

to be greater or less than the location parameter of the other group. (The location parameter for the normal and exponential distribution is the mean, for the binomial it is the probability of an event at each trial, for the Poisson it is the rate or number of events per unit time.) If the parameter to be calculated belongs to group one it will be less than that of group two; if the calculated parameter belongs to group two it will be larger than that of group one.

4.4 Ranges of allowed values

Two types of errors may be encountered by the user. First, an input parameter may be flagged as being out of range. DSTPLAN puts limits on the range of numbers it will accept. These limitations do not affect the usefulness of DSTPLAN very much and they do allow it to run on a variety of computers.

Numbers such as exponential means, normal standard deviations, etc. which must be positive are restricted to the range one ten-billionth to ten billion (0.0000000001 to 10,000,000,000.0). Note that a lower limit of zero on a number's value is allowed only in those rare cases when it is theoretically justified, when it allows the problems to remain well-defined, and when it will not cause overflow errors. Numbers such as normal means, etc. which may be either negative or positive are restricted to the range negative ten billion to ten billion (-10,000,000,000.0 to 10,000,000,000.0).

Sample sizes are typically restricted to a range of 2.0 to ten billion (2.0 to 10,000,000,000.0). Occasionally the lower limit may be lowered to 0.0 or 1.0 if allowed by the problem at hand, or raised to 3.0 or 4.0 if required by statistical theory.

Probabilities, significance levels, and powers are restricted to the range 0.00000001 to 0.99999999 (one hundred-millionth to one minus one hundred-millionth).

User specification of parameters and, in most cases, answers calculated are constrained to lie in these ranges.

In spite of the wide ranges of numbers provided, the user may be confronted with a message that indicates that there is no answer within the allowed range of values. Generally, this indicates that no such answer exists, even if the range were infinite. For example, in a two-sample problem in which the sample size of one group is being calculated, the size of the other group may be specified to be so small that the desired power cannot be achieved even with an infinite sized sample.

Even when the input values for particular parameters do not make the problem impossible, they may change the allowed input range for other parameters.

If this happens, it will always be in the direction of raising the lower bound or lowering the upper bound, or both. Thus for one set of input parameters the full theoretically allowed input range may be valid, but for a different set of input parameters only a restricted range may be valid. In any case, if an input parameter is out of bounds, the user will be confronted with an error message to that effect, including the allowed input range. For some problems this is also true for certain one dimensional functions of more than one input parameter, for example the difference between two means.

Chapter 5

One-sample binomial

5.1 The test

The null hypothesis is that the probability of an event in each trial of the group being tested is a specified value, p_s .

The test associated with these calculations is the exact test for the one-sample binomial distribution.

SIDES: Only a one-sided alternative is available.

5.2 Parameters of the test

Null hypothesis probability: This is the probability, p_s , of an event in each trial according to the null hypothesis. This probability is used to obtain the critical region.

Alternative hypothesis probability: The probability, p_a , of an event in each trial according to the alternative hypothesis. This probability is used to calculate power.

Sample Size: The number of trials or observations in the single sample.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

5.3 Method

The critical region is calculated as the high or low number of events (high if p_a is greater than p_s , low if p_a is less than p_s) such that the probability of the number of events falling in the critical region is less than or equal to the significance level and as close to its value as possible, given that the null hypothesis is correct. The power is calculated as the probability of the number of events falling in the critical region if the alternative hypothesis is correct.

5.4 Comments

The outcome of a trial is an integral number of events. The critical region of a test is consequently bounded by an integer, so there are only a fixed number of significance levels available. A user-specified significance level is thus usually not exactly attainable. DSTPLAN deals with this problem by choosing the critical region as large as possible under the constraint that the specified significance value may not be exceeded. This choice frequently causes the significance level to be less than that specified. For similar reasons, the power may be increased over that requested. To give the user some insight, the critical values bounding the desired significance level are printed with the associated significance and power values whenever either value is changed from that requested.

The inability to exactly match requested significance levels is the reason that two-sided tests are not considered in the one-sample binomial case. Generally, in two-sided tests, when the null hypothesis is true, half of the probability of an outcome in the critical region is in the top portion of this region and half in the bottom. Because significance levels are not generally attained, this condition cannot be exactly achieved in the one-sample binomial case. The user is responsible for choosing the suitable approximation by considering the two sides of the test separately.

The fact that only integer valued outcomes are possible causes another difficulty: power does not necessarily increase with the sample size. For large increases in sample size, the power does become greater, but for small changes it may actually decrease. However, the method used for calculating sample sizes assumes that the power does increase with sample size. A consequence is that the method may not find the smallest sample size with the specified significance level and power. To ameliorate this difficulty, if the computed sample size is less than 100, the next twenty smaller sample sizes are examined to see if they provide the desired significance level and power.

5.5 Examples

In a Phase II trial for a new cancer treatment, a number of patients are given the treatment and the response rate is observed. If the response rate is so small as to be inconsistent with a true 20% rate then the treatment is rejected as ineffective.

The smallest such trial which has a significance level of 0.05 consists of 14 patients. What true response rate must the treatment have so that there will be an 80% chance of rejecting it?

If it is desired to have an 80% chance of rejecting treatments whose response rate is less than 5%, how large must the study be?

If it is desired to have an 80% chance of rejecting treatments whose response rate is less than 10%, how large must the study be?

```
===== Begin Echo of Computer Session =====
```

```
***** One-sample binomial (exact method) *****
```

```
This routine does calculations only for one-sided tests
```

```
Parameter Specification: -----
```

```
Enter:  1: Null hypothesis probability
        2: Alternative hypothesis probability
        3: Sample size
        4: Significance level
        5: Power
```

```
Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.
```

```
$?
```

```
0.2 ? 14 0.05 0.8          #          One-sample binomial ex. #1
```

```
Is the null hypothesis probability less than
the alternative hypothesis probability? (y/n)
```

```
$?
```

```
n          #          Null hyp. NOT < Alt. hyp.
```

```
***** One-sample binomial (exact method) *****
```

Null hypothesis probability 0.200

Sample size 14.000

Significance level 0.050

Power 0.800

The Alternative hypothesis probability is calculated to be 0.016

The new significance level is 0.044

The new power is 0.800

Press Return / Enter to continue:

\$?

done seeing answer

Bracketing values:

Critical value	Significance	Power
0.0	0.0440	0.8000
1.0	0.1979	0.9799

Choice of Next Action: -----

Enter:

(1) To perform another calculation of the same sort

(2) To return to higher level menu

(3) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Null hypothesis probability

2: Alternative hypothesis probability

3: Sample size

4: Significance level

5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter

the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.2 0.05 ? 0.05 0.8 # One-sample binomial ex. #2

***** One-sample binomial (exact method) *****

Null hypothesis probability 0.200

Alternative hypothesis probability 0.050

Significance level 0.050

Power 0.800

The Sample size is calculated to be 30.000

The new significance level is 0.044

The new power is 0.812

Press Return / Enter to continue:

\$?

done seeing answer

Bracketing values:

Critical value	Significance	Power
2.0	0.0442	0.8122
3.0	0.1227	0.9392

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To return to higher level menu
- (3) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Null hypothesis probability

- 2: Alternative hypothesis probability
- 3: Sample size
- 4: Significance level
- 5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

2:0.1 # One-sample binomial ex. #3

Current parameter values are:

0.200000 0.100000 ? 0.500000E-01 0.800000

***** One-sample binomial (exact method) *****

Null hypothesis probability 0.200

Alternative hypothesis probability 0.100

Significance level 0.050

Power 0.800

The Sample size is calculated to be 82.000

The new significance level is 0.046

The new power is 0.806

Press Return / Enter to continue:

\$?

done seeing answer

Bracketing values:

Critical value	Significance	Power
10.0	0.0458	0.8057
11.0	0.0836	0.8847

=====
 ===== End Echo of Computer Session =====
 =====

Chapter 6

Two-sample binomial

6.1 The test

The null hypothesis is that the probability of an event in each trial is the same in both of the groups.

The test for which these calculations are performed is the χ^2 test comparing two binomial probabilities.

SIDES: Both one- and two-sided alternatives are available.

6.2 Parameters of the test

Group 1 probability: The probability under the alternative hypothesis of an event in each trial of group 1. If this probability is to be calculated, it will be less than the probability for group 2.

Group 2 probability: The probability under the alternative hypothesis of an event in each trial of group 2. If this probability is to be calculated, it will be greater than the probability for group 1.

Number in group 1: The number of individual trials constituting group 1.

Number in group 2: The number of individual trials constituting group 2.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

Number for equal sample size: The equal number in each group necessary

to achieve the specified power for the given values of the other parameters.

6.3 Method

The normal approximation to the arcsin transformation of the binomial distribution is used.

Reference: Mace, Arthur E.: *Sample-Size Determination*, Robert E. Krieger Publishing Company, Huntington, NY. (1974) Pages 101-104.

6.4 Examples

The calculations shown determine the sample size necessary to detect a difference in probability of 0.1 for varying values of the lower probability. As is well known, the sample size needed is maximal when the two probabilities straddle 0.5 equally.

```
===== Begin Echo of Computer Session =====
```

```
***** Two-sample Binomial (Arcsin approximation) *****
```

```
Problem Initialization: -----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
1          #          1-sided tests
```

```
Parameter Specification: -----
```

```
Enter:  1: Probability of an event in group 1
```

```
        2: Probability of an event in group 2
```

```
        3: Number of trials in group 1
```

```
        4: Number of trials in group 2
```

```
        5: Significance level
```

```
        6: Power
```

```
Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.
```

\$?
0.1 0.2 ? ? 0.05 0.8 # Two-sample binomial ex. #1

***** Two-sample Binomial (Arcsin approximation) *****

These calculations are for a one sided test.

Probability of an event in group 1 0.100

Probability of an event in group 2 0.200

Significance level 0.050

Power 0.800

The Number of equal trials in each group is calculated to be 153.529

Press Return / Enter to continue:

\$?
Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?
1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Probability of an event in group 1
 - 2: Probability of an event in group 2
 - 3: Number of trials in group 1
 - 4: Number of trials in group 2
 - 5: Significance level
 - 6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table

of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?
0.25 0.35 ? ? 0.05 0.8 # Two-sample binomial ex. #2

***** Two-sample Binomial (Arcsin approximation) *****

These calculations are for a one sided test.

Probability of an event in group 1 0.250

Probability of an event in group 2 0.350

Significance level 0.050

Power 0.800

The Number of equal trials in each group is calculated to be 258.037
Press Return / Enter to continue:

\$?
Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?
1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Probability of an event in group 1
 - 2: Probability of an event in group 2
 - 3: Number of trials in group 1
 - 4: Number of trials in group 2

5: Significance level

6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

0.45 0.55 ? ? 0.05 0.8 # Two-sample binomial ex. #3

***** Two-sample Binomial (Arcsin approximation) *****

These calculations are for a one sided test.

Probability of an event in group 1 0.450

Probability of an event in group 2 0.550

Significance level 0.050

Power 0.800

The Number of equal trials in each group is calculated to be 308.095

===== End Echo of Computer Session =====

Chapter 7

Binomial probabilities above and below the median

7.1 The test

This calculation relates to a continuous variable that is thought to affect a value of a binomial probability. As a concrete example, consider a binomial probability of complete tumor disappearance following radiation therapy. A new measure of tumor radiosensitivity is discovered, and it is desired to test whether this measure is predictive of the probability of tumor disappearance. One would assume that radiation sensitive tumors would be more likely to be cured than would radioresistant tumors. We do not have much experience with the new measure, but we do know that overall, for example, 60% of all tumors are now cured with radiation - a figure we will call P_O (for overall probability).

By definition, half the subjects will have a sensitivity measure below the median value found in the experiment and half will have a measure above the median value. If P_L is the probability of cure for the group with measures below the median and P_H the probability of those with measures above the median, it follows that $(P_L + P_H)/2 = P_O$.

Suppose that the maximum sized feasible trial contains 100 subjects. It is of interest to know the minimal P_H (above 60%) that provides an 80% probability of showing that P_H is larger than P_L . To find this number we could try $(P_L=0.55, P_H=0.65)$, $(P_L=0.50, P_H=0.70)$, ..., $(P_L=0.20, P_H=1.00)$ and calculate the power for each pair of values using the two-sample binomial routine. Then the pair of values providing a power of 0.8 would be chosen. The calcula-

tion provided in DSTPLAN automates this process.

Note that the calculation of the previous paragraph is not easily accomplished using the two-sample binomial and the table of values facility. The reason is that both P_H and P_L change in the above calculation, whereas the table of values facility only allows a table of one value.

7.2 Examples

1. What is the minimal detectable difference in probabilities according to the setup of the above write-up?
2. If the total sample size is 50, what is the minimal detectable probability?
3. If the overall probability were 0.9, what would be the minimal detectable probability?

```
===== Begin Echo of Computer Session =====
```

```
***** Above and Below the Median *****
```

```
Problem Initialization: -----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
1 # 1-sided tests
```

```
Enter overall probability:
```

```
$?
```

```
0.6 # Overall probability
```

```
Parameter Specification: -----
```

```
Enter: 1: Above Median Pr - Overall Pr
```

```
2: Total Sample Size
```

```
3: Significance level
```

```
4: Power
```

```
Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.
```

```
$?
```

? 100 0.05 0.8 # Abv. Med. Pr. - Ov. Pr. ex. #1

***** Above and Below the Median *****

These calculations are for a one sided test.

Overall Probability 0.600

Total Sample Size 100.000

Significance level 0.050

Power 0.800

The Above Median Pr - Overall Pr is calculated to be 0.120

Press Return / Enter to continue:

\$?

Done seeing answer

=====
End Echo of Computer Session
=====

The greater of the two minimally detectable probabilities is $0.6 + 0.120 = 0.720$, and by symmetry, the lower probability is $0.6 - 0.120 = 0.480$. The minimal detectable difference is $2 \times 0.120 = 0.240$.

=====
Begin Echo of Computer Session
=====

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Above Median Pr - Overall Pr

2: Total Sample Size
3: Significance level
4: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?
2:50 # Change the Sample Size
Current parameter values are:
? 50.0000 0.500000E-01 0.800000

***** Above and Below the Median *****

These calculations are for a one sided test.

Overall Probability 0.600

Total Sample Size 50.000

Significance level 0.050

Power 0.800

The Above Median Pr - Overall Pr is calculated to be 0.168
Press Return / Enter to continue:

\$?
Done seeing answer

=====
End Echo of Computer Session
=====

If the sample size is only 50, the minimal detectable difference in probabilities increases to $2 \times 0.168 = 0.336$.

=====
Begin Echo of Computer Session
=====

Choice of Next Action: -----

Enter:
(1) To perform another calculation of the same sort

- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

2 # Change setup

Problem Initialization: -----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Enter overall probability:

\$?

0.9 # Overall probability

Parameter Specification: -----

Enter: 1: Above Median Pr - Overall Pr

2: Total Sample Size

3: Significance level

4: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

? 100 0.05 0.8 # Abv. Med. Pr. - Ov. Pr. ex. #2

***** Above and Below the Median *****

These calculations are for a one sided test.

Overall Probability 0.900

Total Sample Size 100.000

Significance level 0.050

Chapter 8

Fisher's exact

8.1 Comment

Fisher's exact test is less powerful than the χ^2 test for detecting a difference in binomial proportions. The χ^2 test approximates a test specifically designed to test for the difference in proportions. Fisher's exact test, in contrast, conditions on the marginal distributions of observations. Fisher's test inquires as to the probability of the outcome observed or one more extreme within the set of experiments in which the number of observations in each group is the same as the current experiment and also in which the total number of events is the same as in the current experiment. Experience shows that the power of Fisher's test is less than that of the χ^2 .

The calculations provided here should be used only if Fisher's test will be used to analyze the data or if the continuity correction to the χ^2 test will be used. The continuity correction usually changes the significance level to correspond closely to that of Fisher's test. The use of the two-sample binomial calculations is recommended in most cases.

8.2 The test

The null hypothesis is that the probability of an event in each trial is the same in both of the groups.

The tests for which these calculations are performed are the Fisher's exact test and the χ^2 test with the continuity correction.

SIDES: Both one- and two-sided alternatives are available.

8.3 Parameters of the test

Group 1 probability: The probability under the alternative hypothesis of an event in each trial of group 1. If this probability is to be calculated, it will be less than the probability for group 2.

Group 2 probability: The probability under the alternative hypothesis of an event in each trial of group 2. If this probability is to be calculated, it will be greater than the probability for group 1.

Number in each group: The number of individual trials performed in each group.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

8.4 Method

Reference: Cassagrande, JT, Pike, MC, and Smith, PG: "An Improved Approximate Formula for Calculating Sample Sizes for Comparing Two Binomial Distributions," *Biometrics* 34: pp. 483-486, 1978.

8.5 Examples

The calculations shown determine the sample size necessary to detect a difference in probability of 0.1 for varying values of the lower probability. As is well known, the sample size needed is maximal when the two probabilities straddle 0.5 equally.

These examples are the same as those used for the two-sample binomial. Note that the sample sizes required are appreciably greater in the current calculations.

```
===== Begin Echo of Computer Session =====
```

```
***** Fisher's exact - equal numbers each group *****  
          (method of Cassagrande, Pike and Smith)  
WARNING - very conservative for two-sample binomial
```

```
Problem Initialization:-----
```

Enter 1 or 2 for 1- or 2-sided test:
\$?
1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Probability of an event in group 1
2: Probability of an event in group 2
3: Number in each group
4: Significance level
5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?
0.1 0.2 ? 0.05 0.8 # Eq. no. in each group ex. #1

***** Fisher's exact (equal numbers each group) *****

These calculations are for a one sided test.

Probability of an event in group 1 0.100

Probability of an event in group 2 0.200

Significance level 0.050

Power 0.800

The Number in each group is calculated to be 176.037

Press Return / Enter to continue:

\$?
Done seeing answer

Choice of Next Action: -----

Enter:
(1) To perform another calculation of the same sort
(2) To change the setup for this type of problem

(3) To return to higher level menu

(4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Probability of an event in group 1

2: Probability of an event in group 2

3: Number in each group

4: Significance level

5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.25 0.35 ? 0.05 0.8 # Eq. no. in each group ex. #2

***** Fisher's exact (equal numbers each group) *****

These calculations are for a one sided test.

Probability of an event in group 1 0.250

Probability of an event in group 2 0.350

Significance level 0.050

Power 0.800

The Number in each group is calculated to be 278.260

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

(1) To perform another calculation of the same sort

- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Probability of an event in group 1
- 2: Probability of an event in group 2
- 3: Number in each group
- 4: Significance level
- 5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.45 0.55 ? 0.05 0.8 # Eq. no. in each group ex. #3

***** Fisher's exact (equal numbers each group) *****

These calculations are for a one sided test.

Probability of an event in group 1 0.450

Probability of an event in group 2 0.550

Significance level 0.050

Power 0.800

The Number in each group is calculated to be 327.775

=====
 ===== End Echo of Computer Session =====
 =====

Chapter 9

The matched pairs test for equality of proportions

9.1 Description

NOTE: This test and its description may be removed from future versions of DSTPLAN and probably made into a separate application; it is one of the most complex tests we support. Also, I think the following description may well be inaccurate in many places, at least judging as well as I can from the source code and Miettinen's original paper. I am leaving the description essentially unchanged from the version 4.1 documentation; however, the computer session is taken from the 4.2 version of DSTPLAN.
– John Venier

This test is more complex than most of the others considered by DSTPLAN. Consequently, some notation must be established before the test can be explained.

		Y	
		1	0
X	1	Z11	Z10
	0	Z01	Z00

X and Y are the values of two different characteristics of each matched pair. Each characteristic is dichotomized into two values that will be denoted as 0 or 1. The issue is whether the proportion of Y's which has characteristic 1 is the same for X's with characteristic 0 as it is for X's with characteristic 1.

To illustrate this setup, consider a case-control study of saccharin consumption in bladder cancer. Pairs of cases who have bladder cancer and age and sex matched controls who don't have bladder cancer are examined. The variable X is then 1 for the cases who have bladder cancer, and 0 for the controls who don't have cancer. The variable Y is 1 if the individual has a history of significant consumption of saccharin, and 0 otherwise. The question examined is whether the proportion with a history of significant saccharin consumption is the same for cases and controls.

Z_{11} , Z_{10} , Z_{01} , and Z_{00} are frequencies of paired observations by their values on the X and Y characteristics, *e.g.*, Z_{11} is the number of pairs with both X and Y equal to 1, Z_{10} is the number of pairs with X equal to 1 and Y equal to 0, etc.

θ_1 : Proportion of individuals with $X = 1$ who have $Y = 1$.

θ_2 : Proportion of individuals with $X = 0$ who have $Y = 1$. PSI: Proportion of pairs for which $X = 1$ and $Y = 0$ or else $X = 0$ and $Y = 1$.

δ : $\theta_1 - \theta_2$ (for a one-sided alternative).

In terms of the example mentioned, these quantities are as follows.

$X = 1$ case: patient with cancer of the bladder.

$X = 0$ control: person of same sex, race and age group as case, who lives in same neighborhood as case and who has no history of cancer or bladder disorders.

$Y = 1$ history of significant saccharin consumption.

$Y = 0$ no history of significant saccharin consumption.

Z_{11} = number of pairs where both case and control were saccharin consumers.

Z_{10} = number of pairs where case was a saccharin consumer and control was not.

Z_{01} = number of pairs where case was not a saccharin consumer but the matched control was.

Z_{00} = number of pairs where neither the case nor the matched control was a saccharin consumer.

θ_1 = proportion of saccharin consumers among cases.

θ_2 = proportion of saccharin consumers among controls.

ψ = proportion of case-control pairs where the case and control had differing saccharin histories.

δ = difference in proportion of saccharin consumers in cases and controls.

Investigators wish to test the hypothesis $\theta_1 = \theta_2$, (or $\delta = 0$) *i.e.*, that the proportion of saccharin consumers is the same among cases and controls.

9.2 The test

The null hypothesis is $\theta_1 = \theta_2$.

Test Statistic: $Q = (Z_{10} - Z_{01})/\sqrt{Z_{10} + Z_{01}}$ (McNemar's Test)

SIDES: Both one- and two-sided alternatives are available.

9.3 Parameters of the test

δ to be detected: The minimum value of δ that is to be detectable with the test.

Sample size: The number of pairs that must be studied to achieve the specified power.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

The user of this routine must be able to provide estimates of frequencies Z_{11} , Z_{10} , Z_{01} , and Z_{00} from some preliminary sample. This is necessary to estimate ψ which is used in all calculations.

If a preliminary sample is not available the program will compute sample size estimates only with knowledge of the other 3 parameters. The user will be asked if he/she can provide estimates of Theta1 and Theta2. If these estimates are provided the program will compute a conservative estimate of sample size (N). The program will inform the user that instead of sampling N pairs immediately he/she may take a preliminary sample size of N/4 pairs and re-enter the program with the preliminary data.

If the user cannot estimate P_1 , P_2 the program will assume $\theta_1 = 0.09$, $\theta_2 = 0.1$, compute an ultra-conservative estimate of sample size (N) and inform the user that rather than sample N pairs immediately he/she may take a preliminary sample of N/6 pairs and re-enter the program later with the preliminary data.

9.4 Method

Reference: Miettinen, OS: "The Matched Pairs Design in the Case of All or None Responses," *Biometrics* 24: pp. 339-352, 1968.

9.5 Examples

Suppose that a preliminary study has been done and that there are 30 pairs in which both cases and controls consumed saccharin; 40 pairs in which the cases consumed saccharin and the controls didn't; 10 pairs in which the controls consumed saccharin and the cases didn't; and 20 pairs in which neither consumed saccharin. Find the sample size necessary to produce an 0.80 power using a one-sided test at significance level 0.05 for a difference in proportions of 0.1. Do the same calculation for the case in which the preliminary sample is not available but an estimate of proportion of saccharin consumers among cases is 0.7 and among controls, 0.4. Finally do the same calculation assuming that not even this information is available.

Using the same preliminary data and setup, find the minimal detectable difference with a sample size of 150.

```
===== Begin Echo of Computer Session =====
```

```
***** Matched pairs test for equality of proportions *****
              (method of Miettinen)
```

```
Problem Initialization:-----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
1                #          1-sided tests
```

```
Parameter Specification: -----
```

```
Enter:  1: Delta to be detected
        2: Sample size
        3: Significance level
        4: Power
```

```
Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action. The Znn will
be held constant all calculations.
```

```
$?
```

```
0.1 ? 0.05 0.8                #          Equality of proportions ex. #1
```

```
Have you collected data from a preliminary sample? (y/n)
```

```
$?
```

```
n                #          Have NOT collected data
```

Can you estimate theta1, theta2 ? (y/n)
 \$?
 y # CAN est. theta1, theta2
 Enter: (1) theta1, (2) theta2
 \$?
 0.7 0.4 # theta1, theta2

***** Matched pairs test for equality of proportions *****

These calculations are for a one sided test.

Delta to be detected 0.100
 Significance level 0.050
 Power 0.800
 Theta1 0.700
 Theta2 0.400

A conservative estimate of required sample size is 330.411
 It is recommended that you select a preliminary sample of size 82.603
 and re-enter the program with the preliminary data.

Choice of Next Action: -----

Enter:
 (1) To perform another calculation of the same sort
 (2) To change the setup for this type of problem
 (3) To return to higher level menu
 (4) To quit DSTPLAN

Select one of the above:

\$?
 1 # Another calculation

Parameter Specification: -----

Enter: 1: Delta to be detected
 2: Sample size
 3: Significance level

4: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action. The Znn will be held constant all calculations.

```
$?
0.1 ? 0.05 0.8          #          Equality of proportions ex. #2
Have you collected data from a preliminary sample? (y/n)
$?
n                        #          Have NOT collected data
Can you estimate theta1, theta2 ? (y/n)
$?
n                        #          Can NOT estimate thetas
```

***** Matched pairs test for equality of proportions *****

These calculations are for a one sided test.

Delta to be detected 0.100

Significance level 0.050

Power 0.800

A conservative estimate of required sample size is 504.527

It is recommended that you select a preliminary sample of size 84.088 and re-enter the program with the preliminary data.

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

```
$?
1                        #          Another calculation
```


Chapter 10

Proportion response: conservative vs. expensive treatment

10.1 Explanation of the investigation

Most two-sample trials are designed to show that two groups, treatments, etc., are different. This trial described here intended to demonstrate that the two treatments are almost equally effective. Identity of effectiveness cannot be demonstrated, so the most that can be accomplished is to show that, with a large probability, the two treatments do not differ much.

The paradigm for this type of demonstration is treatment for a disease; response to the treatment is considered good. The outcome measured is the proportion of patients that responds to each treatment. The expensive treatment is also the current treatment; it is desired to show that the response rate to a proposed conservative treatment is not much less (worse) than that of the expensive treatment.

The high price exacted by the expensive treatment is frequently patient morbidity, although monetary cost is usually also involved. The conservative treatment is less drastic than the expensive current treatment, it produces less morbidity, and it usually has a smaller dollar cost. However, the conservative treatment will be adopted only if its response rate is not much worse than that of the current regimen. An example of the two types of treatment involves breast cancer and has received some public attention. The expensive treatment is radical mastectomy, and the conservative alternative is a lumpectomy plus radiation to local lymph nodes. The important problem with the expensive

treatment is disfigurement.

The calculated outcome of the trial will be a (one-sided) confidence interval on the proportion of responders to the expensive treatment minus the proportion of responders to the conservative treatment. The degree of confidence to be used must be specified, and it is a probability near one. Values of 0.9, 0.95, and 0.99 are popular. The trial is designed to make this confidence interval smaller than an amount specified as the critical difference. The value of the critical difference varies with the particular application, but it is a small number – 0.1 is a popular value.

The same sort of calculation arises in cases in which it is desired to have a small outcome probability. For example, a new and less expensive manufacturing process is to be shown not to produce a much greater defect rate than the current process. This routine can be used in these cases by considering, for example, the proportion of non-defective items rather than the proportion of defective items.

SIDES: The procedure is one-sided.

10.2 Parameters of the procedure

Proportion of responders to the conservative treatment: The assumed true proportion.

Proportion of responders to the expensive treatment: The assumed true proportion.

Sample size of conservative group: The number of patients to be given the conservative treatment.

Sample size of expensive group: The number of patients to be given the expensive treatment.

Critical difference: The upper bound on the proportion of responders to the expensive treatment minus the proportion of responders to the conservative treatment.

Confidence level: The degree of confidence used to construct the confidence interval on the difference in proportions.

Probability of demonstration: The probability that the confidence limit constructed will be bounded by the critical difference. This value is analogous to power, so 0.8 is a popular value.

Equal sample size: If this option is chosen for the calculation, the sample size required in both groups for the procedure to achieve the specified probability of demonstration.

10.3 Method

The normal approximation to the binomial distribution is used.

Reference: Makuch, R. and Simon, R.: "Sample Size Requirements for Evaluating a Conservative Therapy," *Cancer Treatment Report* 62: pp. 1037-1040, 1978.

10.4 Examples

If both the conservative and expensive treatment produce a response rate of 70%, how many patients must be in each group to produce an 80% probability that a 95% confidence interval on the difference is no larger than 0.1?

If in the same problem, the largest feasible sample size is 100 patients per group, what must the response rate for the conservative treatment be?

```
===== Begin Echo of Computer Session =====
```

```
***** Proportion response - conservative vs. expensive treatment *****
              (normal approximation to the binomial)
```

```
Parameter Specification: -----
```

```
Enter:  1: Prop. response to conservative treatment
        2: Prop. response to expensive treatment
        3: Sample size of conservative group
        4: Sample size of expensive group
        5: Critical difference
        6: Confidence level
        7: Probability of demonstration
```

```
Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and
4 will calculate equal sample sizes for both groups. Enter a "t" if a table
of values for this parameter is to be specified. Entering "t"s in positions
3 and 4 will allow a table of equal sample sizes. Enter the numeric value
of the parameter otherwise. Entering zeros for all items skips the
calculation.
```

```
$?
```

```
0.7 0.7 ? ? 0.1 0.95 0.8      #      Proportion responders ex. #1
```

```
***** Proportion response - conservative vs. expensive *****
```

These calculations are for a one sided test.

Prop. response to conservative treatment 0.700

Prop. response to expensive treatment 0.700

Critical difference 0.100

Confidence level 0.950

Probability of demonstration 0.800

The Equal sample size for both groups is calculated to be 259.667

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

(1) To perform another calculation of the same sort

(2) To return to higher level menu

(3) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Prop. response to conservative treatment

2: Prop. response to expensive treatment

3: Sample size of conservative group

4: Sample size of expensive group

5: Critical difference

6: Confidence level

7: Probability of demonstration

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the

Chapter 11

Two-sample binomial with historical control group

11.1 The test

The null hypothesis is that the probability of an event in each trial is the same in both of the groups. One group is a historical control that has been analyzed before the planning of the current comparative trial, and the proportion of events in this group is known.

The test for which these calculations are performed is the χ^2 test comparing two binomial probabilities.

SIDES: Both one- and two-sided alternatives are available.

11.2 Parameters of the test

Probability of control group: The probability of an event in each trial of the historic control group.

Probability of experimental group: The probability, assuming the alternative hypothesis, of an event in each trial of the experimental group.

Number in control group: The number of individual trials of the control group.

Number in experimental group: The number of individual trials of the experimental group.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

11.3 Method

The normal approximation to the arcsin transformation of the binomial distribution is used.

Reference: Makuch, RW and Simon, RM.: "Sample Size Considerations for Non-Randomized Comparative Studies," *J. Chron. Dis.* 33: pp. 175-181, 1980.

11.4 Examples

The examples shown are the same sample size calculations that were performed for the two-sample binomial. In the two-sample binomial, the sample sizes were taken to be equal. In these examples, the historical control group is taken to have the lower probability of the two, and the sample size of the historic group is that needed in both groups for the two-sample binomial. The difference between the sample size for the historic and experimental group is thus an indication of the savings due to having a historical control.

The calculations shown determine the sample size necessary to detect a difference in probability of 0.1 for varying values of the lower probability. As is well known, the sample size needed is maximal when the two probabilities straddle 0.5 equally.

===== Begin Echo of Computer Session =====

***** Two-sample binomial with historical control group *****

Problem Initialization: -----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1	#	1-sided tests
---	---	---------------

Parameter Specification: -----

Enter: 1: Probability of an event in the control group

2: Probability of an event in the experimental group

- 3: Number of subjects in the control group
- 4: Number of subjects in the experimental group
- 5: Significance Level
- 6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?
 0.1 0.2 153.5 ? 0.05 0.8 # 2-sx bin. w/hist. ctrl. ex. #1

***** Two-sample binomial with historical control group *****

These calculations are for a one sided test.

Probability of an event in the control group 0.100

Probability of an event in the experimental group 0.200

Number of subjects in the control group 153.500

Significance Level 0.050

Power 0.800

The Number of subjects in the experimental group is calculated to be 113.254

Press Return / Enter to continue:

\$?
 # Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Probability of an event in the control group
- 2: Probability of an event in the experimental group
- 3: Number of subjects in the control group
- 4: Number of subjects in the experimental group
- 5: Significance Level
- 6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.25 0.35 258 ? 0.05 0.8 # 2-sx bin. w/hist. ctrl. ex. #2

***** Two-sample binomial with historical control group *****

These calculations are for a one sided test.

Probability of an event in the control group 0.250

Probability of an event in the experimental group 0.350

Number of subjects in the control group 258.000

Significance Level 0.050

Power 0.800

The Number of subjects in the experimental group is calculated to be 188.906

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Probability of an event in the control group
 - 2: Probability of an event in the experimental group
 - 3: Number of subjects in the control group
 - 4: Number of subjects in the experimental group
 - 5: Significance Level
 - 6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.45 0.55 308 ? 0.05 0.8 # 2-sx bin. w/hist. ctrl. ex. #3

***** Two-sample binomial with historical control group *****

These calculations are for a one sided test.

Probability of an event in the control group 0.450

Probability of an event in the experimental group 0.550

Number of subjects in the control group 308.000

Significance Level 0.050

Power 0.800

The Number of subjects in the experimental group is calculated to be 226.295

=====
 ===== End Echo of Computer Session =====
 =====

Chapter 12

k-sample binomial

12.1 The test

The null hypothesis is that the probability of an event in each trial is the same in all k of the groups.

The test for which these calculations are performed is the χ^2 test comparing binomial probabilities, *i.e.*, the χ^2 test on a $2 \times k$ table.

SIDES: The χ^2 test is multiple sided.

12.2 Situations for which calculations are provided

This routine differs from others in DSTPLAN in that it begins with questions that establish the testing framework for the calculation before listing the parameters that can be calculated. The computable parameters vary somewhat with this framework.

The first question requests the number of groups, k .

The second question asks whether all groups are the same size. Sample sizes for the test can be calculated regardless of the answer. If all groups are the same size, then the common size will be computed. If the groups are not all the same size, then the size of each group will be requested. This size is used only to establish the proportion of the total sample belonging to each group. The computed sample size of each group necessary to achieve the specified power will be printed for the user; the sample sizes so calculated will be proportional to the original sizes specified.

The ability to handle unequal sized groups is convenient for observational studies in which the investigator has no control over the proportion in each group. For example, a study will investigate the proportion of smokers in groups determined by economic status. This will be done by sampling some human population – for example all those living in a particular town. In this scheme, the investigator has no control over the proportion in each economic status; these proportions will emerge from the population sample. However, an initial estimate of these proportions is needed to perform the sample size calculation.

The third question asks whether the probability is the same for all but one of the groups; this refers to the probabilities at the alternative hypothesis value at which power is calculated. Only if these probabilities are the same can the probability of the remaining group be calculated so as to yield a specified power. (If the other probabilities differ, there are so many such answers that the experimenter would find a listing of them more confusing than helpful.) When the value of the different probability is calculated to yield a specified power, there are two answers – one is less than the common probabilities, the other is greater. Sometimes one of these values is calculated to be negative or greater than 1.0. These spurious values should be ignored.

12.3 Parameters of the problem

Group k probability: Available only in the case that the first $k - 1$ groups have the same probability; in this case it is the probability of an event in each trial of the k th group that will yield the specified power. Two answers are calculated, one below and one above the value of the common probability.

Number in each group: If the groups are of unequal size, the user will be asked for the number in each group and the answer will be proportional to the numbers so specified. This allows calculation of the factor by which a pilot study must be expanded to provide a specified power.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

12.4 Method

The non-central χ^2 is used for these calculations.

Reference: Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, Academic Press, New York, NY. (1977) Pages 215-272.

12.5 Examples

Three treatments for cancer are being tested for response rate. If the true response rate is 20% for two of them, how large must the other rate be in order to show that the rates are different with probability 0.8 at a 5% significance level when 100 patients are placed on each treatment?

A study is investigating whether the proportion of individuals who smoke differs by economic status. There are three categories of status, low which constitutes 20% of the population, middle which constitutes 60% of the population, and high which completes the final 20% of the population. If the proportion of smokers is 0.5, 0.3, and 0.1 respectively for the economic categories, how large a sample is required to show the effect at the 5% level with a power of 80%?

===== Begin Echo of Computer Session =====

***** K-sample binomial *****

Problem Initialization:-----

Enter number of groups:

\$?

3 # 3 groups

Are all of the 3 groups of equal size? (y/n)

\$?

y # 3 groups are of equal size

Do all but one of the 3 groups have the same probability? (y/n)

\$?

y # All but one have same prob.

Parameter Specification: -----

Enter: 1: Group 3 probability

2: Number in each group

3: Significance

4: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

? 100 0.05 0.8 # K-sample (K=3) bin. ex. #1

Enter the single probability of the first 2 groups.
\$?
0.2 # Sing. prb. (1st 2 grp.s)

***** K-sample binomial *****

All the groups are of equal size
Probability of the first 2 groups 0.200
Number in each group 100.000
Significance 0.050
Power 0.800

The Group k probabilities are calculated to be 0.063
and 0.366

Choice of Next Action: -----

- Enter:
- (1) To perform another calculation of the same sort
 - (2) To change the setup for this type of problem
 - (3) To return to higher level menu
 - (4) To quit DSTPLAN

Select one of the above:

\$?
2 # Change setup

Problem Initialization:-----

Enter number of groups:

\$?
3 # 3 groups

Are all of the 3 groups of equal size? (y/n)

\$?
n # 3 groups are NOT of eq. size

Do all but one of the 3 groups have the same probability? (y/n)
 \$?
 n # All but one NOT same prob.

Parameter Specification: -----

Do you want to solve for group sizes? (y/n)
 \$?
 y # DO solve for group sizes

Enter: 1: Significance
 2: Power

Enter a 't' if a table of values for this parameter is to be specified
 (only one 't' entry allowed). Enter the numeric value of the parameter
 otherwise. Entering zeros for all items skips the calculation and allows a
 choice of the next action.

\$?
 0.05 0.8 # K-sx (K=3) bin. ex. #2
 Enter the probabilities of the 3 groups.

\$?
 0.5 0.3 0.1 # Prob.s of the 3 grps.
 Enter the proportional number of observations in the 3 groups.

\$?
 20 60 20 # The prop. num. obs.

***** K-sample binomial *****

Probability of group 1 0.500
 Probability of group 2 0.300
 Probability of group 3 0.100
 Proportional number of obs. in group 1 20.000
 Proportional number of obs. in group 2 60.000
 Proportional number of obs. in group 3 20.000
 Significance 0.050

Chapter 13

Case-control

13.1 The test

The effect of a purported risk factor on a rare disease is studied by the following strategy. A number of cases (those who have the disease) and controls (who don't have the disease) are examined for the risk factor. The proportion of cases with the risk factor is compared to the proportion of controls using the two by two χ^2 test. If the risk factor increases or decreases the probability of having the disease, then these proportions should differ.

In practice, the number of cases studied is usually the same as the number of controls. Furthermore, the controls are usually matched to the cases in some fashion, *e.g.*, age, sex, other known risk factors. In spite of the matching of controls to cases, the analysis is unpaired – *i.e.*, the proportion with the risk factor in the case and control groups are compared; the number of pairs, in which the case has the risk factor and the control doesn't have the risk factor, do not enter into the test. If a matched analysis is planned, these calculations shouldn't be used; use the matched pair calculations instead.

The planning is performed in terms of two population parameters. One is the frequency of the risk factor in the whole population to be studied, and the other is the relative risk of the disease for those with the risk factor as compared to those without the risk factor. The frequency of the risk factor is the proportion of those who have the risk factor in the total population. The relative risk is the proportion of those who have the disease in the risk factor bearing population divided by the proportion of those who have the disease in the non risk factor bearing population. Using standard notation for conditional probability, the frequency of the risk factor is:

$$F = P(R)$$

and the relative risk is:

$$RR = P(D|R)/P(D|\neg R)$$

For planning purposes, it is assumed that the proportion of those with the risk factor in the non disease bearing population is the same as the proportion of those with the risk factor in the total population, *i.e.*, the frequency. This is true assuming the null hypothesis that the relative risk is 1. Even assuming that the null hypothesis is false this approximation is good if the disease is rare, say, occurring in less than 10% of the total population. Precisely, the proportion of those with the risk factor in the non disease bearing population is the frequency multiplied by the ratio of one minus the proportion of those who have the disease in the risk factor bearing population to one minus the proportion of those with the disease in the total population. Symbolically:

$$P(R|\neg D) = \frac{1 - P(D|R)}{1 - P(D)} F$$

Should the proportion of those with the disease in the whole population not be small, this proportion can be used to calculate the two probabilities being compared and the two-sample binomial calculations used.

SIDES: Both one- and two-sided alternatives are available.

13.2 Parameters of the test

Frequency of risk factor: The proportion of those who have the risk factor in the entire population studied. In most cases, there is an interval of frequencies that will yield the desired power for a specified relative risk. The endpoints of the interval are printed by DSTPLAN if the frequency is chosen for calculation.

Relative risk: The ratio of the probability of disease given the risk factor to the probability of disease given no risk factor. If the relative risk is chosen for calculation, two values will be computed, one that is less than 1.0 and one that is greater than 1.0; these values and values further from one are those that can be shown to be different from 1.0 with the power specified. Should either risk print as a negative number, it indicates that the value does not exist.

Number of cases: The number of cases, *i.e.*, those with the disease, sampled from the population. Usually, the number of cases is the same as the number of controls, *i.e.*, those without the disease.

Number of controls: The number of controls, *i.e.*, those without the disease, sampled from the population. Usually, the number of controls is the same as the number of cases, *i.e.*, those with the disease.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

Equal number of cases and controls: When chosen for calculation, computes the equal number of cases and controls needed to yield the specified power.

13.3 Method

The formulae from the reference cited are used to obtain the probabilities being compared. The normal approximation of the arcsin transformation of the binomial distribution is used to compute power (instead of the direct normal approximation of the reference).

The mathematical form of power as a function of frequency is complicated. Hence a scan over frequency values from 0 to 1 in steps of 0.0001 is performed in finding the interval of values yielding at least the specified power. When the power crosses the desired level between two scanned values, linear interpolation is used. Although several examinations of power as a function of frequency seem to show that the set of acceptable values is an interval on the real line, this is not guaranteed. Consequently, the number of times that the desired power value is crossed during the scan is counted. Should the number be greater than two, the set of allowable values forms two or more disjoint intervals. A warning message to this effect is then printed.

Reference: Schlesselman, JJ.: "Sample Size Requirements in Cohort and Case-Control Studies of Disease," *American Journal of Epidemiology* 99(6): pp. 381-384, 1974.

13.4 Examples

The effect of heavy coffee consumption on heart disease is to be investigated. A number of patients hospitalized with heart disease and an equal number of controls matched by sex, age, and residence will be studied. The proportion of heavy coffee consumption will be compared for the two groups. A one-sided test (that coffee consumption increases the risk of heart disease) will be used at a significance level of 0.05.

If the population frequency of heavy coffee consumption is 0.3 and it is desired to have a power of 0.8 of detecting a relative risk of 2, how many cases and controls must be in the study? For what range of frequency of heavy coffee consumers in the population will this study have the power of 0.8? If the study must be limited to 100 cases and 100 controls and if the frequency of heavy coffee consumption is 0.3, what relative risk is detectable at a power of 0.8?

Suppose the study were investigating a factor that is supposed to be protective against heart disease, such as the consumption of fish and fish products. Perform the same calculations using a relative risk of one-half instead of 2.0.

===== Begin Echo of Computer Session =====

***** Case-control (arcsin approximation) *****

Problem Initialization: -----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Frequency of risk factor

2: Relative risk

3: Number of cases

4: Number of controls

5: Significance

6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

0.3 2 ? ? 0.05 0.8 # Case-control ex. #1

***** Case-control (arcsin approximation) *****

These calculations are for a one sided test.

Frequency of risk factor 0.300

Relative risk 2.000

Significance 0.050

Power 0.800

The Equal number of cases and controls is calculated to be 110.499

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Frequency of risk factor
 - 2: Relative risk
 - 3: Number of cases
 - 4: Number of controls
 - 5: Significance
 - 6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

? 2 111 111 0.05 0.8 # Case-control ex. #2

***** Case-control (arcsin approximation) *****

These calculations are for a one sided test.

Relative risk 2.000

Number of cases 111.000

Number of controls 111.000

Significance 0.050

Power 0.800

The limits on the frequency interval are 0.296 and 0.543

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Frequency of risk factor
 - 2: Relative risk
 - 3: Number of cases
 - 4: Number of controls
 - 5: Significance
 - 6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

0.3 ? 100 100 0.05 0.8 # Case-control ex. #3

***** Case-control (arcsin approximation) *****

These calculations are for a one sided test.

Frequency of risk factor 0.300
Number of cases 100.000
Number of controls 100.000
Significance 0.050
Power 0.800

The relative risk less than 1 is 0.426,
greater than 1 is 2.070

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Frequency of risk factor
2: Relative risk
3: Number of cases
4: Number of controls
5: Significance
6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions

3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

0.3 t ? ? 0.05 0.8 # Case-control ex. #4

You are specifying a list for values of Relative risk

The list currently contains 0 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values
(8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

2 # The number of values

Enter 2 values

\$?

2 0.5 # The 2 values

The list currently contains 2 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values
(8) QUIT

\$?

8 # QUIT

***** Case-control (arcsin approximation) *****

These calculations are for a one sided test.

Frequency of risk factor 0.300

Significance 0.050

Power 0.800

Relative risk	Equal number of cases and controls
2.000	110.499
0.500	144.819

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Frequency of risk factor
2: Relative risk
3: Number of cases
4: Number of controls
5: Significance
6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

? t 145 145 0.05 0.8 # Case control example #5

You are specifying a list for values of Relative risk
The list currently contains 0 values.
The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

2 # The number of values

Enter 2 values

\$?

2 0.5 # The 2 values

The list currently contains 2 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

8 # QUIT

***** Case-control (arcsin approximation) *****

These calculations are for a one sided test.

Number of cases 145.000

Number of controls 145.000

Significance 0.050

Power 0.800

Relative risk	Lower limit on frequency interval	Upper limit on frequency interval
2.000	0.176	0.700
0.500	0.300	0.824

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Frequency of risk factor

2: Relative risk

3: Number of cases

4: Number of controls

5: Significance

6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

0.3 ? 100 100 0.05 0.8 # Case-control ex. #6 (#3)

***** Case-control (arcsin approximation) *****

Chapter 14

Matched case-control

14.1 The test

As in the unmatched case-control analysis, cases and controls are matched according to age, sex, or other known risk factors. However, instead of comparing the proportion of cases and controls with the purported risk factor, this test examines the number of discordant pairs in which the case is exposed (denoted by O). If the risk factor increases the probability of having the disease, then O is binomially distributed with proportion $R/(R + 1)$, where R is the relative risk for cases versus controls. If the risk factor has no effect on the probability of having the disease, then O is binomially distributed with proportion $1/2$.

SIDES: Only a one-sided alternative is available.

14.2 Parameters of the test

Frequency of risk factor: The proportion of the population studied that has the risk factor. In most cases, there is an interval of frequencies that will yield the desired power for a specified relative risk. The endpoints of the interval are printed by DSTPLAN if the frequency is chosen for calculation.

Relative risk: The probability of disease given the risk factor divided by the probability of disease given no risk factor. If the relative risk is chosen for calculation, two values will be computed, one that is less than 1.0 and one that is greater than 1.0; these values and values further from one are those that can be shown to be different from 1.0 with the power specified. Should either risk print as a negative number, it indicates that the value does not exist.

Number of pairs of cases and controls: The number of pairs of cases

(those with the disease) and controls (those without the disease) sampled from the population.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

14.3 Method

This test reduces to a one-sample binomial with p_s equal to $1/2$ and p_a equal to $R/(R+1)$, with sample size determined from the frequency of the risk factor and the number of discordant pairs in which the case is exposed (O). In the binomial one-sample case, critical values bounding the desired significance level are usually printed with the associated significance and power values whenever either value is changed from that requested. In this case, sometimes only the values associated with the critical value lower than the requested significance level are given, and are printed under the labels “new significance level” and “new power”.

The mathematical form of power as a function of frequency is complicated. Hence a scan over frequency values from 0 to 1 in steps of 0.0001 is performed in finding the interval of values yielding at least the specified power. Although several examinations of power as a function of frequency seem to show that the set of acceptable values is an interval on the real line, this is not guaranteed. Consequently, the number of times that the desired power value is crossed during the scan is counted. Should the number be greater than two, the set of allowable values forms two or more disjoint intervals. A warning message to this effect is then printed.

Reference: Breslow, N.E. and Day, N.E.: *Statistical Methods in Cancer Research: Volume II - The Analysis of Cohort Studies*, International Agency for Research on Cancer, Lyon, France, 1987.

14.4 Examples

The effect of heavy coffee consumption on heart disease is to be investigated. A number of patients hospitalized with heart disease and an equal number of controls matched by sex, age, and residence will be studied. A one-sided test (that coffee consumption increases the risk of heart disease) will be used at a significance level of 0.05.

If the population frequency of heavy coffee consumption is 0.3 and it is

desired to have a power of 0.8 of detecting a relative risk of 2, how many pairs of cases and controls must be in the study? For what range of frequency of heavy coffee consumers in the population will this study have the power of 0.8? If the study must be limited to 100 pairs of cases and controls and if the frequency of heavy coffee consumption is 0.3, what relative risk is detectable at a power of 0.8?

Suppose the study were investigating a factor that is supposed to be protective against heart disease, such as the consumption of fish and fish products. Perform the same calculations using a relative risk of one-half instead of 2.0.

===== Begin Echo of Computer Session =====

***** Matched Case-control *****

This routine does calculations only for one-sided tests

Parameter Specification: -----

Enter: 1: Frequency of risk factor
2: Relative risk
3: Number of pairs of cases and controls
4: Significance
5: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.3 2 ? 0.05 0.8 # Matched case-control ex. #1

***** Matched Case Control *****

Frequency of risk factor 0.300

Relative risk 2.000

Requested Significance 0.050

Requested Power 0.800

The Number of pairs of cases and controls is calculated to be 119.683
The new significance level is 0.043
The new power is 0.812
Press Return / Enter to continue:

\$?

Done seeing answer

Bracketing values:

Significance	Power
0.0435	0.8120
0.0740	0.8762

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Frequency of risk factor
 - 2: Relative risk
 - 3: Number of pairs of cases and controls
 - 4: Significance
 - 5: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

? 2 120 0.05 0.8 # Matched case-control ex. #2

***** Matched Case Control *****

Relative risk 2.000

Number of pairs of cases and controls 120.000

Requested Significance 0.050

Requested Power 0.800

The limits on the frequency interval are 0.298 and 0.541
The set of frequencies is not a single interval. Low and high
values of frequency reported.

The endpoints of the largest single interval within the set of
frequencies are 0.333 and 0.500

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Frequency of risk factor
2: Relative risk
3: Number of pairs of cases and controls
4: Significance
5: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values
for this parameter is to be specified (only one "t" entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.

\$?

0.3 ? 100 0.05 0.8 # Matched case-control ex. #3

***** Matched Case Control *****

Frequency of risk factor 0.300
Number of pairs of cases and controls 100.000
Requested Significance 0.050
Requested Power 0.800

The new significance level is 0.030
The relative risk less than 1 is 0.412,
greater than 1 is 2.111

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Frequency of risk factor
2: Relative risk
3: Number of pairs of cases and controls
4: Significance
5: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.3 t ? 0.05 0.8 # Matched case-control ex. #4

You are specifying a list for values of Relative risk

The list currently contains 0 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

2 # The number of values

Enter 2 values

\$?

2 0.5 # The 2 values

The list currently contains 2 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

8 # QUIT

***** Matched Case Control *****

Frequency of risk factor 0.300

Requested Significance 0.050

Requested Power 0.800

Relative risk	Number of pairs of cases and controls	New Significance level	New Power
2.000	119.683	0.043	0.812
0.500	156.508	0.043	0.812

Press Return / Enter to continue:

\$?

Done seeing answer

Bracketing values:

Relative Risk	Significance	Power
2.0000	0.0435	0.8120
	0.0740	0.8762
0.5000	0.0435	0.8120
	0.0740	0.8762

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Frequency of risk factor
 - 2: Relative risk
 - 3: Number of pairs of cases and controls
 - 4: Significance
 - 5: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

? 0.5 157 0.05 0.8 # Matched case-control ex. #5

***** Matched Case Control *****

Relative risk 0.500

Number of pairs of cases and controls 157.000

Requested Significance 0.050

Requested Power 0.800

The limits on the frequency interval are 0.299 and 0.824
The set of frequencies is not a single interval. Low and high values of frequency reported.

The endpoints of the largest single interval within the set of frequencies are 0.313 and 0.815

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Frequency of risk factor

- 2: Relative risk
- 3: Number of pairs of cases and controls
- 4: Significance
- 5: Power

Enter a "?" for the item to be calculated. Enter a "t" if a table of values for this parameter is to be specified (only one "t" entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.3 ? 100 0.05 0.8 # Matched case-control ex. #6

***** Matched Case Control *****

Frequency of risk factor 0.300

Number of pairs of cases and controls 100.000

Requested Significance 0.050

Requested Power 0.800

The new significance level is 0.030

The relative risk less than 1 is 0.412,
greater than 1 is 2.111

=====
===== End Echo of Computer Session =====

Chapter 15

Sample size correction for loss of subjects over time

15.1 The calculation

Unlike the other components of DSTPLAN, this calculation is not directly related to a power calculation. Instead, it corrects an original sample size to account for a constant loss rate over time.

As an example, suppose the experimenter is performing a five year study and desires to have 100 subjects left at the end of the study. From previous experience, it is estimated that 10% of the subjects will be lost per year. How many subjects must initially be in the study to provide a 90% probability that at least 100 will remain to the end?

15.2 Parameters

Mean proportion dropouts per unit time (D): This is the proportion of subjects that are lost in each time period. This is a number between 0 and 1; in the example above, this number is 0.1.

Initial number in study (I): Self explanatory.

Number at end of study (E): The number remaining after the losses.

Length of study (T): The time the study will last in the same time units used for the mean proportion of dropouts per unit time; in the example, the units are years and this number is 5.

Probability (P): The probability that the number at the end of the study

will be at least E; in the example, this number is 0.9.

15.3 Method

The loss is a Poisson process which implies that an individual is equally likely to be lost at any time during the study, and that all individuals are equally likely to be lost. In other words, the probability of being lost is the same for all individuals and the same for all points in time during the study.

The probability of a single individual remaining in the study after T time units is:

$$P_s = (1 - D)^T$$

The probability of E or more subjects remaining at the end of the study is the sum of the binomial probabilities of exactly E “successes” to exactly I “successes” where there are I total trials and the probability of success in any single trial is P_s .

15.4 Example

```
===== Begin Echo of Computer Session =====
```

```
***** Poisson Loss Correction *****
```

```
Parameter Specification: -----
```

```
Enter:  1: Mean proportion dropouts per unit time
        2: Initial number in study
        3: Number at end of study
        4: Length of study
        5: Prob. that # at end of study >= specification
```

```
Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.
```

```
$?
```

```
0.1 ? 100 5 0.9          #          Samp. size corr. for loss ex. #1
```

```
***** Poisson Loss Correction *****
```

Mean proportion dropouts per unit time 0.100
Number at end of study 100.000
Length of study 5.000
Prob. that # at end of study >= specification 0.900
The Initial number in study is calculated to be 182.968
===== End Echo of Computer Session =====

Chapter 16

One-sample Poisson

16.1 The test

The null hypothesis is that the rate at which events occur is a specified value, R_s .

The test associated with these calculations is the exact test for the one-sample Poisson distribution.

SIDES: Only a one-sided alternative is available.

16.2 Parameters of the test

Null hypothesis rate: This is the rate, R_s , of the number of events per unit time according to the null hypothesis. This rate is used to obtain the critical region of the test.

Alternative hypothesis rate: The rate, R_a , of the number of events per unit time according to the alternative hypothesis. This rate is used to calculate the power of the test.

Time: The amount of time the Poisson process is observed.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

16.3 Units of time

The two rates and time must be entered with respect to the the same time units, *e.g.*, days, weeks, years, or totally arbitrary user defined units. The time units should be chosen so that the rates are reasonably sized numbers; in particular, the rates must be within the range of numbers handled by DSTPLAN.

If Poisson distributed observations are being tested rather than a Poisson process in time, then the rate is the mean of the observation and the time is the number of such observations that are taken.

In the Poisson approximation to the binomial distribution, the mean of the Poisson distribution is the probability of an event multiplied by the number of such events. Scaling, for example by using the number of events per thousand trials, may be necessary in this case.

16.4 Method

The critical region is calculated as the high or low number of events (high if R_a is greater than R_s , low if R_a is less than R_s) such that the probability of the number of events falling in the critical region is less than or equal to the significance level and as close to its value as possible. The power is calculated as the probability of the number of events falling in the critical region if the alternative hypothesis is correct.

16.5 Comments

The outcome of a trial is an integral number of events. The critical region of a test is consequently bounded by an integer, so there are only a fixed number of significance levels available and a user-specified significance level is thus usually not exactly attainable. DSTPLAN deals with this problem by choosing the critical region as large as possible under the constraint that the specified significance value may not be exceeded. This choice frequently causes the significance level to be less than that specified. For similar reasons, the power may be increased over that requested. To give the user some insight, the critical values bounding the desired significance level are printed with the associated significance and power values whenever either value is changed from that requested.

The inability to exactly match requested significance levels is the reason that two-sided tests are not considered in the one-sample Poisson case. Generally, in two-sided tests, when the null hypothesis is true, half of the probability of an outcome in the critical region is in the top portion of this region and

half in the bottom. Because significance levels are not generally attained, this condition cannot be exactly achieved in the one-sample binomial case. The user is responsible for choosing the suitable approximation by considering the two sides of the test separately.

16.6 Examples

A publisher wants to show that the error rate in a book is less than one error per ten pages by a careful examination of 100 pages. How low must the true error rate be in order to have an 80% chance of demonstrating this using a one-sided test with significance level 0.05?

A drug is being tested for mutagenicity by administering it to a cell culture and counting the resulting number of mutants. If the spontaneous mutation rate is one per ten million cells and it is desired to have an 80% chance of detecting a doubling of this rate to two per ten million using a one-sided test with significance level 0.05, how many cells must be exposed? (Note: the raw numbers are exceedingly small so we change units to one million cells; the rates then become 0.1 and 0.2. The answer is also in units of millions of cells.)

===== Begin Echo of Computer Session =====

***** One-sample Poisson (exact method) *****

This routine does calculations only for one-sided tests

Problem Specification: -----

Enter: 1: Null hypothesis rate
2: Alternative hypothesis rate
3: Time of study
4: Significance level
5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.1 ? 100 0.05 0.8 # One-sample Poisson ex. #1

Is the null hypothesis rate less than
the alternative hypothesis rate? (y/n)

\$?
n # Null rate NOT < alt. rate

***** One-sample Poisson (exact method) *****

Null hypothesis rate 0.100

Time of study 100.000

Significance level 0.050

Power 0.800

The Alternative hypothesis rate is calculated to be 0.031

The new significance level is 0.029

The new power is 0.800

Press Return / Enter to continue:

\$?
Done seeing answer

Bracketing values:		
Critical value	Significance	Power
4.0	0.0293	0.8000
5.0	0.0671	0.9068

Choice of Next Action: -----

Enter:

(1) To perform another calculation of the same sort

(2) To return to higher level menu

(3) To quit DSTPLAN

Select one of the above:

\$?
1 # Another calculation

Problem Specification: -----

Enter: 1: Null hypothesis rate

2: Alternative hypothesis rate

- 3: Time of study
- 4: Significance level
- 5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.1 0.2 ? 0.05 0.8 # One-sample Poisson ex. #2

***** One-sample Poisson (exact method) *****

Null hypothesis rate 0.100

Alternative hypothesis rate 0.200

Significance level 0.050

Power 0.800

The Time of study is calculated to be 90.625

The new significance level is 0.044

The new power is 0.800

Press Return / Enter to continue:

\$?

Done seeing answer

Bracketing values:

Critical value	Significance	Power
15.0	0.0435	0.8000
14.0	0.0770	0.8637

===== End Echo of Computer Session =====

Chapter 17

Two-sample Poisson distribution

17.1 The test

The null hypothesis is that the rate (number of events per unit time) of two Poisson processes is the same.

The test associated with these calculations is the F test for comparing two exponential distributions. That is, the mean time to an event will be computed for each process by dividing the total time on trial of the process by the number of events. The equality of the means will be tested as exponential distributions.

SIDES: Both one- and two-sided alternatives are available.

17.2 Parameters of the test

Group 1 rate: The rate of the first process under the alternative hypothesis. If this parameter is calculated, it will be less than the group 2 rate.

Group 2 rate: The rate of the second process under the alternative hypothesis. If this parameter is calculated, it will be greater than the group 1 rate.

Time of observation of process 1: The amount of time that the first process is observed during the trial.

Time of observation of process 2: The amount of time that the second process is observed during the trial.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

Observation time for equal numbers of events: The observation time will be calculated for each process so that the expected number of events for the two processes is the same, and so that the specified power will be obtained. These times yield the shortest overall observation time for the trial that will produce the desired power.

Equal observation times: Equal observation times for each process that will yield the desired power.

17.3 Units of time

The two rates and times must be entered with respect to the same time units, *e.g.*, days, weeks, years, or totally arbitrary user defined units. The time units should be chosen so that the rates are reasonably sized numbers; in particular, the rates must be within the range of numbers handled by DSTPLAN.

If Poisson distributed observations are being tested rather than Poisson processes in time, then the rates are the means of the respective observations and the times are the respective number of such observations that are taken.

In the Poisson approximation to the binomial distribution, the mean of the Poisson distribution is the probability of an event multiplied by the number of such events. Scaling, for example by using the number of events per thousand trials, may be necessary in this case.

17.4 Method

The observation time of each process is sufficient to produce an expected number of events equal to the required number plus one-half. The required number is calculated using the method described for the two-sample exponential distribution.

Reference: Cox, DR.: "Some Simple Approximate Tests for Poisson Variates," *Biometrika*, 30: pp. 354-360, 1953.

17.5 Examples

A baker wants to show that his raisin bread contains more raisins than does that of the competition by counting the number in 20 loaves of his bread and 20 loaves of the competitor's bread. If the mean number of raisins per loaf in

his bread is 250, what must the mean number be for the competition in order to show the difference with probability 0.8 using a one-sided test with significance level 0.05?

A drug is tested for mutagenicity by administering it to a cell culture and counting the number of mutants in that culture and a control. The control mutation rate is known to be one per ten million cells and it is desired to have an 80% chance of detecting a doubling of this rate to two per ten million using a one-sided test with significance level 0.05. If the same expected number of events are to be obtained, how many cells must be in each of the exposed and control cultures? If the two cultures are to be the same size, what must this size be? (Note: the raw rates are exceedingly small, so the units are changed to millions of cells; the rates then become 0.1 and 0.2. The answers are also in millions of cells.)

===== Begin Echo of Computer Session =====

***** Two-sample Poisson (Exact method) *****

Problem Initialization: -----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Group 1 rate
2: Group 2 rate
3: Process 1 observation time
4: Process 2 observation time
5: Significance level
6: Power

Enter a '?' for the item to be calculated. Entering a '?' in positions 3 and 4 will calculate either equal observation times or observation times for equal numbers of events in both groups. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

? 250 20 20 0.05 0.8 # Two-sample Poisson ex. #1

***** Two-sample Poisson (Exact method) *****

These calculations are for a one sided test.

Group 2 rate 250.000

Process 1 observation time 20.000

Process 2 observation time 20.000

Significance level 0.050

Power 0.800

The Group 1 rate is calculated to be 237.720

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Group 1 rate
 - 2: Group 2 rate
 - 3: Process 1 observation time
 - 4: Process 2 observation time
 - 5: Significance level
 - 6: Power

Enter a '?' for the item to be calculated. Entering a '?' in positions 3 and 4 will calculate either equal observation times or observation times

for equal numbers of events in both groups. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?
0.1 0.2 ? ? 0.05 0.8 # Two-sample Poisson ex. #2

Special Case: -----

Enter:

- (1) To calculate equal observation times
- (2) To calculate the observation times for equal number of events

\$?
2 # Obs. times for eq. num. events

***** Two-sample Poisson (Exact method) *****

These calculations are for a one sided test.

Group 1 rate 0.100

Group 2 rate 0.200

Significance level 0.050

Power 0.800

The process 1 observation time is 266.528.
The process 2 observation time is 133.264 for equal numbers of events.

Press Return / Enter to continue:

\$?
Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu

(4) To quit DSTPLAN
Select one of the above:
\$?
1 # Another calculation

Parameter Specification: -----

- Enter: 1: Group 1 rate
2: Group 2 rate
3: Process 1 observation time
4: Process 2 observation time
5: Significance level
6: Power

Enter a '?' for the item to be calculated. Entering a '?' in positions 3 and 4 will calculate either equal observation times or observation times for equal numbers of events in both groups. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?
0.1 0.2 ? ? 0.05 0.8 # Two-sample Poisson ex. #3

Special Case: -----

- Enter:
(1) To calculate equal observation times
(2) To calculate the observation times for equal number of events

\$?
1 # Eq. obs. times

***** Two-sample Poisson (Exact method) *****

These calculations are for a one sided test.

Group 1 rate 0.100
Group 2 rate 0.200
Significance level 0.050
Power 0.800

Chapter 18

One-sample normal distribution

18.1 The test

The null hypothesis is that the mean of the normal distribution is a specified number, M_S .

The test associated with these calculations is the one-sample t-test.

The standard deviation of the normal distribution will be estimated from the data; it is not assumed to be known.

SIDES: Both one and two-sided alternatives are available.

18.2 Parameters of the test

Null hypothesis mean: The mean of the normal distribution according to the null hypothesis.

Alternative mean: The mean of the normal distribution according to the alternative hypothesis.

Standard deviation: The standard deviation of the normal distribution.

Sample size: The number of observations in the single sample.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

18.3 Method

Reference: Mace, Arthur E.: *Sample-Size Determination*, Robert E. Krieger Publishing Company, Huntington, NY. (1974) Pages 74-77.

18.4 Examples

A drug is studied to see if it raises intelligence by administering it to a number of subjects and then measuring their scores on a standardized test which is known to produce a mean value of 100 with a standard deviation of 10. If an increase to a true mean of 102 is to be detected with a probability of 0.8 with a one-sided test at a 0.05 significance level, what sample size is required? For a two-sided test?

What is the power of these two tests if the mean is only increased to 101?

```
===== Begin Echo of Computer Session =====
```

```
***** One-sample normal (exact method) *****
```

```
Problem Initialization:-----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
1 # 1-sided tests
```

```
Parameter Specification: -----
```

```
Enter: 1: Null hypothesis mean
```

```
2: Alternative hypothesis mean
```

```
3: Standard deviation
```

```
4: Sample size
```

```
5: Significance level
```

```
6: Power
```

```
Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.
```

```
$?
```

```
100 102 10 ? 0.05 0.8 # One-sample normal ex. #4
```

***** One-sample normal (exact method) *****

These calculations are for a one sided test.

Null hypothesis mean 100.000

Alternative hypothesis mean 102.000

Standard deviation 10.000

Significance level 0.050

Power 0.800

The Sample size is calculated to be 155.926

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

2 # To change setup for this problem

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

2 # 2-sided tests

Parameter Specification: -----

- Enter:
- 1: Null hypothesis mean
 - 2: Alternative hypothesis mean
 - 3: Standard deviation
 - 4: Sample size
 - 5: Significance level

6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

100 102 10 ? 0.05 0.8 # One-sample normal ex. #5

***** One-sample normal (exact method) *****

These calculations are for a two sided test.

Null hypothesis mean 100.000

Alternative hypothesis mean 102.000

Standard deviation 10.000

Significance level 0.050

Power 0.800

The Sample size is calculated to be 198.151

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

2 # To change setup for this problem

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

- Enter: 1: Null hypothesis mean
2: Alternative hypothesis mean
3: Standard deviation
4: Sample size
5: Significance level
6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

100 101 10 156 0.05 ? # One-sample normal ex. #6

***** One-sample normal (exact method) *****

These calculations are for a one sided test.

Null hypothesis mean 100.000

Alternative hypothesis mean 101.000

Standard deviation 10.000

Sample size 156.000

Significance level 0.050

The Power is calculated to be 0.344

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

2 # To change setup for this problem

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

2 # 2-sided tests

Parameter Specification: -----

- Enter: 1: Null hypothesis mean
- 2: Alternative hypothesis mean
- 3: Standard deviation
- 4: Sample size
- 5: Significance level
- 6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

100 101 10 199 0.05 ? # One-sample normal ex. #7

***** One-sample normal (exact method) *****

These calculations are for a two sided test.

Null hypothesis mean 100.000

Alternative hypothesis mean 101.000

Standard deviation 10.000

Sample size 199.000

Significance level 0.050

The Power is calculated to be 0.289

=====
===== End Echo of Computer Session =====

Chapter 19

Two-sample normal distribution

19.1 The test

The null hypothesis is that the mean of the two normally distributed groups is the same.

The test associated with these calculations is the two-sample t-test.

The common standard deviation of the normal distributions of the groups will be estimated from the data; it is not assumed to be known.

SIDES: Both one and two-sided alternatives are available.

19.2 Parameters of the test

Group 1 mean: The mean of the first group under the alternative hypothesis. If this parameter is calculated, it will be less than the group 2 mean.

Group 2 mean: The mean of the second group under the alternative hypothesis. If this parameter is calculated, it will be greater than the group 1 mean.

Number in group 1: The number of observations in group 1.

Number in group 2: The number of observations in group 2.

Standard deviation of both groups: The common standard deviation of both groups.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

Number for equal sample size: The number in each group required to yield the specified power.

19.3 Method

Reference: Mace, Arthur E.: *Sample-Size Determination*, Robert E. Krieger Publishing Company, Huntington, NY. (1974) Pages 77-80.

19.4 Examples

A drug is studied to see if it raises intelligence by administering it to a number of subjects and then measuring their scores on a standardized test. An equal number of controls are given a placebo and also take the standardized test. If a change from a mean test score of 100 to a treated mean of 102 is to be detected with probability 0.8, how many subjects must there be in the control and treatment groups? The standard deviation of the test is known to be 10, although the sample standard deviation will be used in the test.

Both 1- and 2-sided tests at a significance level of 0.05 are to be examined. For both tests so constructed, what is the probability of detecting a change from 100 to 101?

===== Begin Echo of Computer Session =====

***** Two-sample normal with equal variances *****
(exact method)

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Group 1 mean

2: Group 2 mean

3: Number in group 1

- 4: Number in group 2
- 5: Standard deviation of both groups
- 6: Significance level
- 7: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

```
$?
100 102 ? ? 10 0.05 0.8      #      Two-sample normal eq. var ex. #1
```

***** Two-sample normal with equal variances (exact method) *****

These calculations are for a one sided test.

```
Group 1 mean ..... 100.000
Group 2 mean ..... 102.000
Standard deviation of both groups ..... 10.000
Significance level ..... 0.050
Power ..... 0.800
```

The Equal sample size for both groups is calculated to be 309.806
Press Return / Enter to continue:

```
$?
#      Done seeing answer
Choice of Next Action: -----
```

- Enter:
- (1) To perform another calculation of the same sort
 - (2) To change the setup for this type of problem
 - (3) To return to higher level menu
 - (4) To quit DSTPLAN

Select one of the above:

```
$?
```

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Group 1 mean
- 2: Group 2 mean
- 3: Number in group 1
- 4: Number in group 2
- 5: Standard deviation of both groups
- 6: Significance level
- 7: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

100 101 310 310 10 0.05 ? # Two-sample normal eq. var ex. #2

***** Two-sample normal with equal variances (exact method) *****

These calculations are for a one sided test.

Group 1 mean 100.000

Group 2 mean 101.000

Number in group 1 310.000

Number in group 2 310.000

Standard deviation of both groups 10.000

Significance level 0.050

The Power is calculated to be 0.344

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

2 # To change setup for this problem

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

2 # 2-sided tests

Parameter Specification: -----

- Enter:
- 1: Group 1 mean
 - 2: Group 2 mean
 - 3: Number in group 1
 - 4: Number in group 2
 - 5: Standard deviation of both groups
 - 6: Significance level
 - 7: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

100 102 ? ? 10 0.05 0.8 # Two-sample normal eq. var ex. #3

***** Two-sample normal with equal variances (exact method) *****

These calculations are for a two sided test.

Group 1 mean 100.000

Group 2 mean 102.000

Standard deviation of both groups 10.000

Significance level 0.050

Power 0.800

The Equal sample size for both groups is calculated to be 393.407

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Group 1 mean
 - 2: Group 2 mean
 - 3: Number in group 1
 - 4: Number in group 2
 - 5: Standard deviation of both groups
 - 6: Significance level
 - 7: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

100 101 394 394 10 0.05 ? # Two-sample normal eq. var ex. #4

***** Two-sample normal with equal variances (exact method) *****

These calculations are for a two sided test.

Group 1 mean 100.000

Group 2 mean 101.000

Number in group 1 394.000

Number in group 2 394.000

Standard deviation of both groups 10.000

Significance level 0.050

The Power is calculated to be 0.288

===== End Echo of Computer Session =====

Chapter 20

Two-sample normal with unequal variances

20.1 The distribution and test

The assumption of this test is that the two independent samples arise from normally distributed populations whose variances and hence standard deviations differ.

The null hypothesis is that the mean of the two groups is the same.

The test used to analyze the data will be the t-test using Welch's approximation for the degrees of freedom.

SIDES: Both one- and two-sided alternatives are available.

20.2 Parameters of the test

Group 1 mean: The mean of the first population under the alternative hypothesis. If this parameter is calculated, it will be less than the group 2 mean.

Group 2 mean: The mean of the second population under the alternative hypothesis. If this parameter is calculated, it will be greater than the group 1 mean.

Standard deviation of group 1: The standard deviation of the first population.

Standard deviation of group 2: The standard deviation of the second population.

Number in group 1: The number of observations in group 1.

Number in group 2: The number of observations in group 2.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

NB: If sample sizes are calculated, the program provides an option of whether to use efficient allocation of subjects to groups or whether to use equal sample sizes in each group. The efficient allocation uses a number of subjects proportional to the population standard deviation in each group. This allocation minimizes the standard error of the difference in means. See the comments at the end of the example.

20.3 Method

The t-test with Welch's correction to the degrees of freedom for unequal variances is used.

20.4 Example

Given two populations, normally distributed with means of 10 and 20 and standard deviations of 10 and 20 respectively, what sample sizes for both groups are required to produce probabilities of 0.8, 0.9, 0.95, and 0.99 of detecting that there is a difference in the means, using a two-sided test with a significance of 0.05? Perform the calculation using both efficient subject allocation and equal allocation to the two groups.

```
===== Begin Echo of Computer Session =====
```

```
***** Two-sample normal with unequal variances *****  
          (Welch's Approximation)
```

```
Problem Initialization:-----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
2                #          2-sided tests
```

Parameter Specification: -----

- Enter: 1: Group 1 mean
- 2: Group 2 mean
- 3: Number in group 1
- 4: Number in group 2
- 5: Standard deviation of group 1
- 6: Standard deviation of group 2
- 7: Significance level
- 8: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a 't' if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?

10 20 ? ? 10 20 0.05 t # 2-samp. norm. uneq. vars ex.#1

You are specifying a list for values of Power

The list currently contains 0 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

4 # The number of values

Enter 4 values

\$?

0.8 0.9 0.95 0.99 # The 4 values

The list currently contains 4 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified
 (2) ADD values equally spaced between bounds
 (3) ADD values as (2) with logarithmic spacing
 (4) PRINT list
 (5) DELETE a single element of list
 (6) DELETE consecutive elements of list
 (7) SORT list in ascending order and eliminate duplicate values
 (8) QUIT

\$?

8 # QUIT

Enter (1) for equal sample size calculation
 (2) for efficient allocation of sample sizes calculation

\$?

1 # eq. samp. size calc.

***** Two-sample normal with unequal variances (Welch's ap) *****

These calculations are for a two sided test.

Group 1 mean 10.000

Group 2 mean 20.000

Standard deviation of group 1 10.000

Standard deviation of group 2 20.000

Significance level 0.050

Power	Equal sample size for both groups
0.800	40.581
0.900	53.868
0.950	66.302
0.990	93.186

Press Return / Enter to continue:

\$?

Done seeing answer

Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values
(8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

4 # The number of values

Enter 4 values

\$?

0.8 0.9 0.95 0.99 # The 4 values

The list currently contains 4 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values
(8) QUIT

\$?

8 # QUIT

Enter (1) for equal sample size calculation

(2) for efficient allocation of sample sizes calculation

\$?

2 # eff. alloc. of samp. size

***** Two-sample normal with unequal variances (Welch's ap) *****

These calculations are for a two sided test.

Group 1 mean 10.000
 Group 2 mean 20.000
 Standard deviation of group 1 10.000
 Standard deviation of group 2 20.000
 Significance level 0.050

Power	Number in group 1	Number in group 2
0.800	24.333	48.667
0.900	32.333	64.667
0.950	39.667	79.333
0.990	56.000	112.000

=====
 ===== End Echo of Computer Session =====
 =====

The efficient allocation produced a smaller total sample size than the equal allocation. In other words, the number of subjects in both arms combined was less for efficient allocation studies than for equal allocation studies. Here is a table of the total sample sizes calculated for both kinds of allocation at the four different power levels:

Power	Efficient	Equal
0.800	74	82
0.900	98	108
0.950	120	134
0.990	168	188

Chapter 21

Two-sample log-normal

21.1 The distribution and test

The log-normal is the distribution of a set of values whose logarithms are normally distributed. The distribution is recognized as a reasonable approximation to data by two properties. (1) All values are positive and (2) the distribution is skewed to the right. The second property indicates that there are more observations far above the peak of a histogram than there are far below the peak.

The calculations performed by this routine assume that the variances of the logarithms of the values in the two groups are the same; in terms of raw (observed) values this implies a common coefficient of variation. The coefficient of variation is the standard deviation divided by the mean.

The null hypothesis is that the mean of the two groups is the same.

The test used to analyze the data will be the t-test applied to the logarithms of the observed data.

SIDES: Both one- and two-sided alternatives are available.

21.2 Parameters of the test

Group 1 mean (raw values): The mean of the first group (in observed, not logarithmically transformed values) under the alternative hypothesis. If this parameter is calculated, it will be less than the group 2 mean.

Group 2 mean (raw values): The mean of the second group (in observed, not logarithmically transformed values) under the alternative hypothesis. If this parameter is calculated, it will be greater than the group 1 mean.

Common coefficient of variation (raw values): The standard deviation of either group divided by the mean of the group, where the standard deviation and the mean are given in observed, not logarithmically transformed values. The values for the coefficient of variation of the two groups are assumed to be the same.

Number in group 1: The number of observations in group 1.

Number in group 2: The number of observations in group 2.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

21.3 Method

The means and variances are transformed back and forth between raw and logarithmic units as needed. The two-sample normal routine is used for calculations.

21.4 Example

Given two log normal means of 10 and 20 and a coefficient of variation of 1, what equal sample size for both groups is required to produce probabilities of 0.8, 0.9, 0.95, and 0.99 of detecting that there is a difference in the means, using a two-sided test with a significance of 0.05?

```
===== Begin Echo of Computer Session =====
```

```
***** Two-sample log-normal *****
```

```
Problem Initialization:-----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
2                #                2-sided tests
```

```
Parameter Specification: -----
```

```
Enter:  1: Group 1 mean - raw values
```

```
        2: Group 2 mean - raw values
```

- 3: Number in group 1
- 4: Number in group 2
- 5: Coefficient of Variation - raw values
- 6: Significance level
- 7: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation.

```
$?
10 20 ? ? 1 0.05 t          #          Two-sample log-normal ex. #1
You are specifying a list for values of Power
The list currently contains 0 values.
The maximum number allowed is 20.
```

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

```
$?
1          #          ADD values indiv. spec.
Enter the number of values that will be entered
```

```
$?
4          #          The number of values
Enter 4 values
```

```
$?
0.8 0.9 0.95 0.99          #          The 4 values
The list currently contains 4 values.
The maximum number allowed is 20.
```

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list

- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

8 # QUIT

***** Two-sample log-normal *****

These calculations are for a two sided test.

Group 1 mean - raw values 10.000

Group 2 mean - raw values 20.000

Coefficient of Variation - raw values 1.000

Significance level 0.050

Power	Equal sample size for both groups
0.800	23.647
0.900	31.310
0.950	38.483
0.990	53.994

===== End Echo of Computer Session =====

Chapter 22

One-sample exponential distribution

22.1 The test

The null hypothesis is that the mean of the exponential distribution is a specified quantity, M_s .

The test associated with these calculations is the exact test.

SIDES: Both one- and two-sided alternatives are available.

22.2 Parameters of the test

Null hypothesis mean: The mean of the exponential distribution according to the null hypothesis.

Alternative mean: The mean of the exponential distribution according to the alternative hypothesis.

Sample size: The number of observations in the single sample.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

22.3 Method

Reference: Mace, Arthur E.: *Sample-Size Determination*, Robert E. Krieger Publishing Company, Huntington, NY. (1974) Pages 107-110.

22.4 Examples

A computer manufacturer wants to demonstrate that a machine has a failure rate of less than one failure per thousand hours. The time to failure of the machine is distributed exponentially, a one-sided test at a significance level of 0.05 will be used, 10 machines will be tested until each fails, and an 80% chance of demonstration is desired. What must the true mean time to failure be to meet these specifications? If the true time to failure is 1500 hours, how many machines must be tested to failure to show that the mean time is greater than 1000?

===== Begin Echo of Computer Session =====

***** One-sample Exponential (Exact method) *****

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Null hypothesis mean
2: Alternative hypothesis mean
3: Sample size
4: Significance level
5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

1000 ? 10 0.05 0.8 # One-sample exponential ex. #1

Special Case: -----

Is null hypothesis mean less than alternative hypothesis mean? (y/n)
\$?
y # null mean < alt. mean

***** One-sample Exponential (Exact method) *****

These calculations are for a one sided test.

Null hypothesis mean 1000.000

Sample size 10.000

Significance level 0.050

Power 0.800

The Alternative hypothesis mean is calculated to be 2154.581

Press Return / Enter to continue:

\$?
Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?
1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Null hypothesis mean
 - 2: Alternative hypothesis mean
 - 3: Sample size
 - 4: Significance level
 - 5: Power

Chapter 23

Two-sample exponential distribution

23.1 The test

The null hypothesis is that the mean of the two exponentially distributed groups is the same.

The test associated with these calculations is the F test for comparing two exponential distributions.

SIDES: Both one- and two-sided alternatives are available.

23.2 Parameters of the test

Group 1 mean: The mean of the first group under the alternative hypothesis. If this parameter is calculated, it will be less than the group 2 mean.

Group 2 mean: The mean of the second group under the alternative hypothesis. If this parameter is calculated, it will be greater than the group 1 mean.

Number in group 1: The number of observations in group 1.

Number in group 2: The number of observations in group 2.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

Number for equal sample size: The equal number in each group required

to yield the specified power.

23.3 Method

Reference: Mace, Arthur E.: *Sample-Size Determination*, Robert E. Krieger Publishing Company, Huntington, NY. (1974) Pages 110-113.

23.4 Examples

A computer manufacturer believes that a change in design has improved the mean time to failure from 1000 hours to 2000 hours. He wants to show that the new machine is improved by testing an equal number of old and new machines until they fail. A one-sided test at significance level 0.05 will be used, the desired power is 0.8. How many machines must be tested?

Show that the answer is invariant to scale by using 1 and 2 instead of 1000 and 2000.

```
===== Begin Echo of Computer Session =====
```

```
***** Two-sample Exponential (Exact method) *****
```

```
Problem Initialization:-----
```

```
Enter 1 or 2 for 1- or 2-sided test:
```

```
$?
```

```
1 # 1-sided tests
```

```
Parameter Specification: -----
```

```
Enter: 1: Group 1 mean
```

```
2: Group 2 mean
```

```
3: Group 1 sample size
```

```
4: Group 2 sample size
```

```
5: Significance level
```

```
6: Power
```

```
Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a "t" if a table of values for this parameter is to be specified. Entering "t"s in positions 3 and 4 will allow a table of equal sample sizes. Enter the numeric value
```

of the parameter otherwise. Entering zeros for all items skips the calculation.

\$?
1000 2000 ? ? 0.05 0.8 # Two-sample exponential ex. #1

***** Two-sample Exponential (Exact method) *****

These calculations are for a one sided test.

Group 1 mean 1000.000

Group 2 mean 2000.000

Significance level 0.050

Power 0.800

The Equal sample size for each group is calculated to be 26.153

Press Return / Enter to continue:

\$?
Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?
1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Group 1 mean
 - 2: Group 2 mean
 - 3: Group 1 sample size
 - 4: Group 2 sample size
 - 5: Significance level
 - 6: Power

Enter a "?" for the item to be calculated. Entering "?"s in positions 3 and

Chapter 24

One-sample clinical trial

24.1 Description

Patients enter a study according to a Poisson process with a constant accrual rate. Accrual continues for a fixed time; there is also a follow-up time for the study in which entered patients are followed, but no new patients are admitted. Survival of all patients is exponential.

For this explication, the event of interest is considered to be death, but any event could be used so long as it occurs at most once.

Some particular unit of time must be used for the specification and interpretation of survival times, accrual rate, accrual time, and follow-up time. While the choice of time unit is completely arbitrary (seconds, days, centuries or a totally capricious interval such as 6.23 months), the same time unit must be used for all five of these parameters. For example, if the chosen time unit is weeks, then survival times, accrual time and follow-up time will all be in weeks, while accrual rate will be in numbers per week.

24.2 The test

The null hypothesis is that the parameter of the exponential survival distribution is a specified value. (See discussion below.)

The test associated with these calculations is the exact test.

24.3 Parameters of the test

Survival time under the null hypothesis: The parameter of the exponential survival distribution under the null hypothesis. The parameter is expressed in user chosen arbitrary units of time. (See discussion below.)

Survival time under the alternative hypothesis: The parameter of the exponential survival distribution under the alternative hypothesis. The parameter is expressed in user chosen arbitrary units of time. (See discussion below.)

Accrual rate: The average number of patients entering the study per user chosen arbitrary unit of time.

Accrual time: The amount of time, in user chosen arbitrary units, that the study will be open to patient accrual.

Follow-up time: The amount of time, in user chosen arbitrary units, that the patients who entered during the accrual time will be followed after the accrual period has ended. In many studies, this follow-up time is zero.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

24.4 Choices for expressing the parameter of the exponential survival distribution

There are several equivalent methods for specifying the parameter of an exponential survival distribution, and all are encountered in practice. The default method is the median survival time. However, before the parameter to be calculated is selected, DSTPLAN asks whether the method of specifying the parameter of the survival distribution is to be changed. The choices are as follows:

Median (default): This is the time, in user chosen arbitrary time units, at which the cumulative proportion surviving is 50%.

Mean: This is the mean, in user chosen arbitrary time units, of the exponential survival distribution.

Proportion surviving at a specified time: This is the proportion surviving at a user specified time. The proportion is a number between 0 and 1, while the time is given in user chosen arbitrary time units. If this choice is selected, the user must also specify the time at which the specified proportion survives.

If a parameter of a survival distribution is to be calculated, the mean and median are both reported. Should a time be specified for the proportion surviving, the calculated proportion surviving at this time will also be reported.

24.5 Example

A study is conducted under the null hypothesis that median survival for patients receiving a particular treatment is 2 years versus the alternative that median survival is 3 years in a one-sided test. Patients are to be accrued for 3 years and there is to be no follow-up period after the 3 year accrual period has ended. What must the accrual rate be for the test to have an 80% chance of showing the alternative is true at a significance level of 0.05?

What must the accrual rate be in the above study if the patients are to be followed for an additional 2 years after the 3 year accrual period has ended?

If the accrual rate were 20 patients per year, how long would the study with no follow-up period have to accrue patients in order to achieve the same power and significance level?

===== Begin Echo of Computer Session =====

***** One-Sample Clinical Trial *****

Problem Initialization:-----

Do you want to change the method for specifying the parameters of the exponential survival distributions? (y/n)

(The current method is median survival time.)

\$?

n # DON'T change spec. of exp. surv.

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Median survival time under null hypothesis
2: Median survival time under alternative hypothesis
3: Accrual rate
4: Accrual time
5: Follow-up time

6: Significance level

7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

2 3 ? 3 0 0.05 0.8 # 1-sample clinical trial ex. #1

***** One Sample Randomized Clinical Trial *****

These calculations are for a one sided test.

Median survival time under null hypothesis 2.000

Median survival time under alternative hypothesis 3.000

Accrual time 3.000

Follow-up time 0.000

Significance level 0.050

Power 0.800

The Accrual rate is calculated to be 44.354

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Median survival time under null hypothesis
2: Median survival time under alternative hypothesis
3: Accrual rate
4: Accrual time
5: Follow-up time
6: Significance level
7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

2 3 ? 3 2 0.05 0.8 # 1-sample clinical trial ex. #2

***** One Sample Randomized Clinical Trial *****

These calculations are for a one sided test.

Median survival time under null hypothesis 2.000

Median survival time under alternative hypothesis 3.000

Accrual time 3.000

Follow-up time 2.000

Significance level 0.050

Power 0.800

The Accrual rate is calculated to be 22.653

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem

(3) To return to higher level menu

(4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter: 1: Median survival time under null hypothesis
- 2: Median survival time under alternative hypothesis
- 3: Accrual rate
- 4: Accrual time
- 5: Follow-up time
- 6: Significance level
- 7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

2 3 20 ? 0 0.05 0.8 # 1-sample clinical trial ex. #3

***** One Sample Randomized Clinical Trial *****

These calculations are for a one sided test.

Median survival time under null hypothesis 2.000

Median survival time under alternative hypothesis 3.000

Accrual rate 20.000

Follow-up time 0.000

Significance level 0.050

Power 0.800

The Accrual time is calculated to be 4.731

===== End Echo of Computer Session =====

24.6 The randomized clinical trial (exponential survival)

24.7 Description

In a randomized clinical trial, patients entering the study are assigned randomly with equal probability to one of two groups, which typically determine the treatment received. Patients accrue to the study according to a Poisson process, *i.e.*, the probability of a new patient's arriving in any time interval is the same as during any other time period of equal length. The average number of patients entering the study per unit time is the accrual rate. When a patient enters the study, the distribution of his or her survival is distributed exponentially. The survival time depends systematically only on the patient's group membership.

The study continues for a fixed period of time. For the duration of the accrual time, patients enter the study as they arrive. At the end of the accrual time no new patients are accepted on the study, but those already included (who are still alive) are followed for an additional fixed follow-up period.

The data are analyzed by the F-test for the equality of means of two exponential samples or equivalently by the log-rank test.

For this explication, the event of interest is considered to be death, but any event could be used so long as it occurs at most once. For example, in comparing two treatments for infertility in women, one might well look at the time interval from treatment to the first live birth.

Some particular unit of time must be used for the specification and interpretation of survival times, accrual rate, accrual time, and follow-up time. While the choice of time unit is completely arbitrary (seconds, days, centuries or a totally capricious interval such as 6.23 months), the same time unit must be used for all five of these parameters. For example, if the chosen time unit is weeks, then survival times, accrual time and follow-up time will all be in weeks, while accrual rate will be in numbers per week.

24.8 The test

Patients enter a study according to a Poisson process with a constant accrual rate. Accrual continues for a fixed time; there is also a follow-up time for the study in which entered patients are followed, but no new patients are admitted.

Survival of all patients is exponential, but the parameter of this exponential may differ depending on the group.

The null hypothesis is that the survival distribution of the two groups is the same, *i.e.*, the exponential distributions are the same for the two groups.

The test associated with these calculations is the F test for equality of two exponential distributions or equivalently the log-rank test.

24.9 Parameters of the test

Survival time of group 1: The parameter of the exponential survival distribution under the alternate hypothesis for those patients in group 1. The parameter is expressed in user chosen arbitrary units of time. (See discussion below.)

Survival time of group 2: The parameter of the exponential survival distribution under the alternate hypothesis for those patients in group 2. The parameter is expressed in user chosen arbitrary units of time. (See discussion below.)

Accrual rate: The average number of patients entering the study per user chosen arbitrary unit of time. The accrual rate includes both arms of the study, as those entering will be randomized.

Accrual time: The amount of time, in user chosen arbitrary units, that the study will be open to patient accrual.

Follow-up time: The amount of time, in user chosen arbitrary units, that the patients who entered during the accrual time will be followed after the accrual period has ended. In many studies, this follow-up time is zero.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

24.10 Choices for expressing the parameter of the exponential survival distribution

There are several equivalent methods for specifying the parameter of an exponential survival distribution, and all are encountered in practice. The default method is the median survival time. However, before the parameter to be calculated is selected, DSTPLAN asks whether the method of specifying the

parameter of the survival distribution is to be changed. The choices are as follows:

Median (default): This is the time, in user chosen arbitrary time units, at which the cumulative proportion surviving is 50%.

Mean: This is the mean, in user chosen arbitrary time units, of the exponential survival distribution.

Proportion surviving at a specified time: This is the proportion surviving at a user specified time. The proportion is a number between 0 and 1, while the time is given in user chosen arbitrary time units. If this choice is selected, the user must also specify the time at which the specified proportion survives.

If a parameter of a survival distribution is to be calculated, the mean and median are both reported. Should a time be specified for the proportion surviving, the calculated proportion at this time will also be reported.

24.11 Method

Reference: George, SL and Desu, MM: "Planning the Size and Duration of a Clinical Trial Studying the Time to Some Critical Event," *Journal of Chronic Disease* 27:15-24, 1974.

24.12 Example

Current treatment for a form of cancer yields a 2 year median survival time; a new treatment is thought to improve this figure. If 30 patients per year can be randomized to the old and new treatment, and the study will accrue patients for the 3 years that it will last, what must the median survival for the new treatment be to have an 80% chance of showing its superiority at a significance level of 0.05?

What must the median survival for the new treatment be in the above study if the patients are to be followed for an additional 2 years after the 3 year accrual period has ended?

With no followup period after accrual, how long must the study last to demonstrate an improvement when the new treatment doubles the median survival time to 4 years?

In this example, the default method of specifying the parameter of the survival distribution, the median survival time, will be used.

===== Begin Echo of Computer Session =====

***** Randomized clinical trial *****
(method of George and Desu)

Problem Initialization:-----

Do you want to change the method for specifying the
parameters of the exponential survival distributions? (y/n)
(The current method is median survival time.)

\$?

n # DON'T change spec. of exp. surv.

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Median survival time of group 1
2: Median survival time of group 2
3: Accrual rate
4: Accrual time
5: Follow-up time
6: Significance level
7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.

\$?

2 ? 30 3 t 0.05 0.8 # Rand. clin. tr. (exp.) ex. #1

You are specifying a list for values of Follow-up time

The list currently contains 0 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values

(8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

2 # The number of values

Enter 2 values

\$?

0 2 # The 2 values

The list currently contains 2 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified

(2) ADD values equally spaced between bounds

(3) ADD values as (2) with logarithmic spacing

(4) PRINT list

(5) DELETE a single element of list

(6) DELETE consecutive elements of list

(7) SORT list in ascending order and eliminate duplicate values

(8) QUIT

\$?

8 # QUIT

***** Randomized clinical trial *****

These calculations are for a one sided test.

Median survival time of group 1 2.000

Accrual rate 30.000

Accrual time 3.000

Significance level 0.050

Power 0.800

Follow-up time	Median survival time of group 2
0.000	6.786

2.000

4.172

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Median survival time of group 1
 - 2: Median survival time of group 2
 - 3: Accrual rate
 - 4: Accrual time
 - 5: Follow-up time
 - 6: Significance level
 - 7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

2 4 30 ? 0 0.05 0.8 # Rand. clin. tr. (exp.) ex. #2

***** Randomized clinical trial *****

These calculations are for a one sided test.

Median survival time of group 1 2.000

Median survival time of group 2 4.000

\$?
3 # time prop. surv. will be spec.

Enter 1 or 2 for 1- or 2-sided test:

\$?
1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Proportion surviving in group 1 at time 3.000
2: Proportion surviving in group 2 at time 3.000
3: Accrual rate
4: Accrual time
5: Follow-up time
6: Significance level
7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?
0.352 ? 30 3 t 0.05 0.8 # Rand. clin. tr. (exp.) ex. #3

You are specifying a list for values of Follow-up time

The list currently contains 0 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values
(8) QUIT

\$?
1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?
2 # The number of values

Enter 2 values

\$?
0 2 # The 2 values

The list currently contains 2 values.
The maximum number allowed is 20.

- Enter (1) ADD values individually specified
(2) ADD values equally spaced between bounds
(3) ADD values as (2) with logarithmic spacing
(4) PRINT list
(5) DELETE a single element of list
(6) DELETE consecutive elements of list
(7) SORT list in ascending order and eliminate duplicate values
(8) QUIT

\$?

8 # QUIT

***** Randomized clinical trial *****

These calculations are for a one sided test.

Proportion surviving in group 1 at time 3.000 0.352
Accrual rate 30.000
Accrual time 3.000
Significance level 0.050
Power 0.800

Follow-up time	Proportion surviving in group 2 at time 3.000
0.000	0.734
2.000	0.606

Press Return / Enter to continue:

\$?

Done seeing answer
Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Proportion surviving in group 1 at time 3.000
 - 2: Proportion surviving in group 2 at time 3.000
 - 3: Accrual rate
 - 4: Accrual time
 - 5: Follow-up time
 - 6: Significance level
 - 7: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.352 0.594 30 ? 0 0.05 0.8 # Rand. clin. tr. (exp.) ex. #4

***** Randomized clinical trial *****

These calculations are for a one sided test.

Proportion surviving in group 1 at time 3.000 0.352

Proportion surviving in group 2 at time 3.000 0.594

Accrual rate 30.000

Follow-up time 0.000

Significance level 0.050

Power 0.800

Chapter 25

Exponential survival: comparison with historic control

25.1 Description

In this trial, a comparison of the exponentially distributed survival of an experimental group will be made to that of a historic control. The survival experience of the historic control group has been analyzed by the time that this study is planned.

Patients accrue to the study according to a Poisson process, *i. e.*, the probability of a new patient's arriving in any time interval is the same as during any other time period of equal length. The average number of patients entering the study per unit time is the accrual rate. Some fraction of entering patients may be allocated to the control group, although usually all receive the experimental treatment. Allocation to the control group is necessary only when this group is too small to permit detection of the expected difference in survival.

The duration of the study consists of an accrual period and an optional follow-up period. During the accrual time, patients enter the study as they arrive. After the accrual period no new patients are accepted on the study, but those already included are followed for the follow-up period. Patients in the historic control group may or may not be followed during the planned study. DSTPLAN has an option for performing calculations in either case.

The data are analyzed by the F-test for the equality of means of two exponential samples, or asymptotically equivalently by the log rank test.

Some particular unit of time must be used for the specification and interpretation of survival times, accrual rate, accrual time, and follow-up time. While the choice of time unit is completely arbitrary (seconds, days, centuries or a totally capricious interval such as 6.23 months), the same time unit must be used for all five of these parameters. For example, if the chosen time unit is weeks, then survival times, accrual time and follow-up time will all be in weeks, while accrual rate will be in numbers per week.

Note: The statistical test associated with the trial is inherently one-sided, *i.e.*, we test only for the possibility that the patients on the experimental treatment live longer than those on the control treatment. If specifications of survival times are given that reverse this direction, then the program complains and refuses to provide an answer.

25.2 Information required to plan the trial

Survival time of the historic control group: The parameter of the exponential survival distribution for those patients in the historic control group. The parameter is expressed in user chosen arbitrary units of time. (See the discussion below for alternative forms in which this information can be entered.)

Survival time of the group receiving the experimental treatment: The parameter of the exponential survival distribution under the alternate hypothesis for those patients in the experimental group. The parameter is expressed in user chosen arbitrary units of time. (See the discussion below for alternative forms in which this information can be entered.)

Whether the historic control group will be followed during the trial: Following those still alive in the control group is a cheap way of increasing the power or decreasing the duration of the trial. However, in some cases, investigators will not want to review charts or collect this information.

The proportion of entering patients that will be assigned to the control group: Usually this is zero unless the control group is too small to allow detection of the anticipated difference in survival. If accrual time is calculated, so is the optimal proportion to assign to the control group to minimize the accrual time.

Accrual rate: The average number of patients entering the study per user chosen arbitrary unit of time. If new patients are to be assigned to the control group, the accrual rate includes both arms of the study, as those entering will be randomized.

Accrual time: The amount of time, in user chosen arbitrary units, that the study will be open to patient accrual.

Follow-up time: The amount of time, in user chosen arbitrary units, that the patients will be followed after the accrual period has ended. In many studies, this follow-up time is zero.

Significance: The probability of falsely rejecting the null hypothesis of equality of survival for the control and experimental groups when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

25.3 Choices for expressing the parameter of the exponential survival distribution

Several methods for specifying the parameter of an exponential survival distribution exist and all are encountered in practice. Consequently, DSTPLAN offers several of these methods. The choices are as follows:

Total observation time and number dead (default for the historic control group): This method is only available for the historic control group since the required data are only available for the historic control group. The observation time for each subject is the time, in user chosen arbitrary time units, from entry on the study until last followup for those who are alive; it is the time, in user chosen arbitrary time units, from entry on the study until death for those who are dead. The total observation time is the sum over all of the subjects of each individual's observation time. This is the preferred method for entering the survival parameter for the historic control group since it allows for maximum likelihood estimation of the survival parameter.

Median (default for the experimental group): This is the time, in user chosen arbitrary time units, at which the cumulative proportion surviving is 50%.

Mean: This is the mean, in user chosen arbitrary time units, of the exponential survival distribution. This option is not provided for the historic control group because most survival analysis programs do not estimate the mean.

Proportion surviving at a specified time: This is the proportion surviving at a user specified time. The proportion is a number between 0 and 1, while the time is given in user chosen arbitrary time units. If this choice is selected, the user must also specify the time at which the specified proportion survives. This option is provided for those cases in which the control group may not have reached a survival of 50%.

If a parameter of a survival distribution is to be calculated, the mean and median are both reported. Should a time for the proportion surviving be specified, the calculated proportion at this time will also be reported.

25.4 Method

Reference: Dixon, DO and Simon, R: "Sample Size Considerations for Studies Comparing Survival Curves using Historical Controls," *J Clin Epidemiol*, Vol 41, No 12, pp 1209-1213.

25.5 Example

Fifty patients have been treated and their median survival time is twelve months. At the time of analysis, half are dead and half are alive. A new treatment is thought to double the survival time. A planned study can accrue 5 patients per month; in the study, all patients (including the controls) will be followed for twelve months after the accrual period. How long must the accrual period for the study be to achieve an 80% probability of detecting the difference using a one-sided test with significance level 0.05?

Note: The example is contrived to be one of the cases covered in Table 3 of the reference. The total observation time of the control group was calculated by noting that the estimate of the hazard rate of an exponential is the number dead divided by the total observation time. On the other hand, the hazard rate is fixed by the median survival time. Equating the two yields a total observation time of 433 months. The table in Dixon and Simon shows that 45 patients must be entered in the study; this corresponds to the accrual rate (5 patients per month) times the accrual time (9 months) of this example.

===== Begin Echo of Computer Session =====

Exponential Survival - Comparison with a Historic Control
(method of Dixon and Simon)

Problem Initialization:-----

Do you want to change the method for specifying the parameter
of the control group survival distribution? (y/n)
(current method is the total observed time and the number dead)

\$?

n # DON'T chng spec. ctrl. gp. surv.

Do you want to change the method for specifying the parameter
of the experimental group survival distribution? (y/n)
(current method is the median survival time)

\$?
n # DON'T chng spec. expr. gp. surv.

Parameter Specification: -----

Will the patients in the control group be followed
while patients are accrued for the study? (y/n)

\$?
y # ctrl. gp. foll. while pat.s acc.

Enter the proportion of entering patients assigned to the control group.
(Note: Choosing to calculate the accrual time will cause the program to
also calculate the optimal proportion, optimal being that proportion which
yields the minimal accrual time.)

\$?
0.0 # prop. ent. pat.s asgnd. to ctrl.

Enter: 1: Median survival time experimental group

2: Accrual rate

3: Accrual time

4: Follow-up time

5: Significance level

6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.

\$?
24 5 ? 12 0.05 0.8 # Comp. w/ hist. ctrl. ex. #1

Enter number dead and number alive in control group

\$?
25 25 # no. dead & alive (ctrl. gp.)

Enter the total observed time in control group.

\$?
433 # tot. obs. time in ctrl. gp.

***** Exponential Survival - Control *****

Median survival time experimental group 24.000

```

Accrual rate ..... 5.000

Follow-up time ..... 12.000

Significance level ..... 0.050

Power ..... 0.800

Prop. assigned to control group ..... 0.000

Number dead in control group ..... 25.000

Number alive in control group ..... 25.000

Total observed time in control group ..... 433.000

The Accrual time is calculated to be 8.996
The optimal proportion to assign to the control group is 0.000
It yields an accrual time of 8.996

===== End Echo of Computer Session =====

```

25.6 Another example

This is taken directly from Dixon and Simon and demonstrates the calculation of the optimal fraction of new patients to assign to the control group to minimize the accrual time of the study:

Suppose the historical controls consist of 50 patients, 10 of whom have failed and whose median survival was 24 months. We show in Table 2 that 241 patients are required to detect $\delta = 1.5$ if new patients are accrued to the experimental treatment only, at a rate of three per month. By allocating approximately one-quarter of the new patients to the control group to supplement the historical data, the accrual requirements can be reduced to 181 new patients, a savings not only of new patients, but also of time: the accrual period is shortened by 20 months.

Notes: We illustrate the option to specify the survival parameter of the control group by providing the proportion surviving (80% from the example) at some time. The time, 7.75 months, was calculated as the time at which there

would be 80% survival in an exponential failure process with median time to failure of 24 months.

In actual use of the calculation, the time and survival figures would be obtained from an analysis of the control data. Also note that the significance is specified to be 0.05 while the power is specified to be 0.8.

===== Begin Echo of Computer Session =====

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

2 # To change setup for this problem

Problem Initialization:-----

Do you want to change the method for specifying the parameter of the control group survival distribution? (y/n)
(current method is the total observed time and the number dead)

\$?

y # DO chng spec. of ctrl. gp. surv.

Enter (1) to use the total observed time and number dead
(2) to use the median survival time
(3) to use the proportion surviving to some time

\$?

3 # use prop. surv. to some time

Enter time at which proportion surviving will be specified:

\$?

7.75 # time prop. surv. specified

Do you want to change the method for specifying the parameter

of the experimental group survival distribution? (y/n)
(current method is the median survival time)

\$?

n # DON'T chng spec. expr. gp. surv.

Parameter Specification: -----

Will the patients in the control group be followed
while patients are accrued for the study? (y/n)

\$?

y # ctrl. gp. foll. while pat.s acc.

Enter the proportion of entering patients assigned to the control group.
(Note: Choosing to calculate the accrual time will cause the program to
also calculate the optimal proportion, optimal being that proportion which
yields the minimal accrual time.)

\$?

0.0 # prop. ent. pat.s asgnd. to ctrl.

Enter: 1: Median survival time experimental group

2: Accrual rate

3: Accrual time

4: Follow-up time

5: Significance level

6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.

\$?

36 3 ? 0 0.05 0.8 # Comp. w/ hist. ctrl. ex. #2

Enter number dead and number alive in control group

\$?

10 40 # no. dead & alive (ctrl. gp.)

Enter the control proportion surviving to time 7.75

\$?

0.8 # ctrl. prop. surv. to T 7.75

***** Exponential Survival - Control *****

Median survival time experimental group 36.000

Chapter 26

Two hazard rates depending on time

26.1 Description

It is common in medical research for a failure rate to either increase or decrease with the time after diagnosis. DSTPLAN provides sample size and related calculations for an approximation to this case. Survival is exponential with a hazard h_1 up to time, T_B , following diagnosis; after T_B , survival is exponential with a hazard h_2 . (The hazard is the mean number of deaths per unit time at risk; the mean time to failure is the reciprocal of the hazard.) There are two groups and the proportional hazards model applies, *i.e.*, the break time, T_B , is the same for both groups and the hazards, h_1 and h_2 , of group one are a constant multiple of those for group two.

Patients accrue to the study according to a Poisson process, *i.e.*, the probability of a new patient's arriving in any time interval is the same as during any other time period of equal length. The average number of patients entering the study per unit time is the accrual rate.

The study consists of an accrual period and (optionally) a follow-up period. During the accrual time, patients enter the study as they arrive. After the accrual period no new patients are accepted on the study, but those already included are followed for the follow-up period.

The survival experiences of the two groups are compared by the log rank test.

A common unit of time must be used for the parameters of the model for survival, accrual rate, accrual time, and follow-up time. The choice of the time

unit is completely arbitrary (seconds, days, centuries or a totally capricious interval such as 6.23 months).

26.2 Information required to plan the trial

The model for survival: DSTPLAN gives three means of specifying the model for a single group:

1. The time of the break in hazard rates, T_B , and the two hazard rates. This can be used when either the hazard rates prior to and after T_B are known or equivalently when the two mean times to failure are at hand since the mean time to failure is the reciprocal of the hazard rate.
2. The time of the break in hazard rates, T_B , and the proportion surviving at two times. One or both of these times must be greater than T_B .
3. The proportion surviving at three different times. One time must be less than T_B and two must be greater than T_B .

Options (2) and (3) are provided so that the model can be specified from information from a survival curve. DSTPLAN checks for illegal models in which either T_B or one of the hazards would be negative.

Whether the specified model applies to the group with the high or low survival rate: Generally the model will apply to the group with the low survival rate since this is current experience which it is sought to improve. Remember that survival and hazard are inverse of each other. A low survival rate corresponds to a high hazard and conversely.

Whether the test will be one- or two-sided: Usually, the test is one-sided as one is only interested in improvements in survival.

Ratio of high to low hazard at any time: This ratio is the same before and after T_B .

Accrual rate: The average number of patients entering the study per user chosen arbitrary unit of time.

Accrual time: The amount of time in user chosen arbitrary units that the study will be open to patient accrual.

Follow-up time: The amount of time in user chosen arbitrary units that the patients who entered during the accrual time will be followed after the accrual period has ended. In many studies, the follow-up time is zero.

Significance: The probability of falsely rejecting the null hypothesis of equality of survival for the two groups when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

26.3 Method

Reference: Schoenfeld, D: "The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions," *Biometrika* (1981) Vol 68, pp. 316-319.

26.4 Example

Current experience with recurrence of a certain type of cancer shows a lessening rate with time. It is thought that the experience can reasonably be approximated by a hazard rate of two (mean time to recurrence of one-half) in units of years for the first 2.5 years after diagnosis and a hazard of one thereafter. If 20 patients per year for five years can be accrued and if no follow-up time is planned after the accrual period, what improvement must a new treatment yield to have an 80% chance of showing the reality of the improvement using a one-sided significance level of 0.05?

(We see from the session echoed below that the answer is 1.7, *i.e.*, the mean survival time must improve by a factor of 1.7 to be detectable. For the same problem with exponential survival, if the hazard is 1, then the detectable difference is a factor of 1.8; if the hazard is 2, it is 1.7. These answers can be obtained either from this routine setting the two hazards to be the same or from the randomized clinical trial routine. The two answers will differ slightly, as somewhat different methods are used.)

===== Begin Echo of Computer Session =====

***** Double-Exponential Model *****

Problem Initialization: -----

How do you want to specify the model?

- (1) Time of break, TB, and two hazard rates
- (2) Time of break and proportion surviving at two times, at least one of which is greater than TB
- (3) Proportion surviving at three different times, one less than TB and two greater than TB

\$?

1 # Model spec. by TB & 2 haz. rates

A model must be specified. Enter H if this model applies to the high survival rate group or L if it applies to the low survival rate group.

\$?

1 # Low survival rate group model

Enter the time of break ,TB, the time at which the failure rate changes.

\$?

2.5 # T(ime)B(reak) (fail rate chngs.)

Enter the hazard rates for the model:

\$?

2 1 # haz. rates for the model

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Ratio of high to low hazard

2: Accrual rate

3: Accrual time

4: Follow-up time

5: Significance

6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

? 20 5 0 0.05 0.8 # Rnd. cln. (2 diff. haz.) ex #1

***** Exponential Survival Double-Exponential Model *****

This model applies to the low survival rate group. These calculations are for a one sided test.

Time to break 2.500

Hazard rate 2.000

Hazard rate 1.000
Accrual rate 20.000
Accrual time 5.000
Follow-up time 0.000
Significance 0.050
Power 0.800

The Ratio of high to low hazard is calculated to be 1.709
Choice of Next Action: -----

Enter:

- (1) To perform another calculation with the same model
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one if the above:

\$?

1 # Another calculation

Parameter Specification: -----

Enter: 1: Ratio of high to low hazard
2: Accrual rate
3: Accrual time
4: Follow-up time
5: Significance
6: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

1.7 t ? 0 0.05 0.8 # Rnd. cln. (2 diff. haz.) ex #2

You are specifying a list of values for Accrual rate

The list currently contains 0 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified
 (2) ADD values equally spaced between bounds
 (3) ADD values as (2) with logarithmic spacing
 (4) PRINT list
 (5) DELETE a single element of list
 (6) DELETE consecutive elements of list
 (7) SORT list in ascending order and eliminate duplicate values
 (8) QUIT

\$?

1 # ADD values indiv. spec.

Enter the number of values that will be entered

\$?

2 # The number of values

Enter 2 values

\$?

10 5 # The 2 values

The list currently contains 2 values.

The maximum number allowed is 20.

Enter (1) ADD values individually specified
 (2) ADD values equally spaced between bounds
 (3) ADD values as (2) with logarithmic spacing
 (4) PRINT list
 (5) DELETE a single element of list
 (6) DELETE consecutive elements of list
 (7) SORT list in ascending order and eliminate duplicate values
 (8) QUIT

\$?

8 # QUIT

***** Exponential Survival Double-Exponential Model *****

This model applies to the low survival rate group. These calculations are for a one sided test.

Time to break 2.500

Hazard rate 2.000

Hazard rate 1.000

Chapter 27

One-sample correlation

27.1 The test

The null hypothesis is that the correlation coefficient between two observed variables is a specified value, C_s .

The test associated with these calculations uses the approximate normality of the transformed sample correlation coefficient (see Method).

SIDES: Both one- and two-sided alternatives are available.

27.2 Parameters of the test

Null hypothesis correlation: This is the value, C_s , of the correlation coefficient between the variables under the null hypothesis. This value is used to determine the critical region.

Alternative hypothesis correlation: The value, C_a , of the correlation coefficient between the variables under the alternative hypothesis. This value is used to determine the power of the test.

Sample Size: The number of trials or observations in the single sample.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

27.3 Method

Uses the approximate normality of the transform, Z , of the correlation coefficient, C :

$$Y = \frac{1}{2} \log \left(\frac{1+C}{1-C} \right)$$

$$Z = Y - \frac{3 * Y + C}{4N}$$

where N is the sample size.

Reference: Kendall, MG and Stuart, A: *The Advanced Theory of Statistics*, Harner Publishing Company, New York, NY. (1963). Volume I, page 391, formula (16.81).

27.4 Comment

In the vast majority of cases the trial is designed to show that the correlation is non-zero, *i.e.*, there is an association between the variables. In these cases, the null hypothesis value of the correlation is zero.

A one-sided test is somewhat more common than a two-sided test because the nature of the association, if any, is usually known a priori. For example, more nutrition administered intravenously should decrease weight loss in patients.

27.5 Examples

What true correlation produces an 80% chance of being distinguished from zero using a one-sided test of a sample of size 10 with a 0.05 significance level? What are the corresponding true correlations for sample sizes of 50 and 100?

How small must a true correlation be to have an 80% chance of showing that it is less than 0.5 using a one-sided test with significance level of 0.05 and a sample size of 50?

```
===== Begin Echo of Computer Session =====
```

```
***** One-sample Correlation *****
```

```
Problem Initialization:-----
```

```

Enter 1 or 2 for 1- or 2-sided test:
$?
1          #          1-sided tests

Parameter Specification: -----

Enter:  1: Null hypothesis correlation
        2: Alternative hypothesis correlation
        3: Sample size
        4: Significance level
        5: Power
Enter a '?' for the item to be calculated. Enter a 't' if a table of values
for this parameter is to be specified (only one 't' entry allowed). Enter
the numeric value of the parameter otherwise. Entering zeros for all items
skips the calculation and allows a choice of the next action.
$?
0 ? t 0.05 0.8          #          One-sample corr. coeff. ex. #1

Special Case: -----

Is the null hypothesis correlation less than
the alternative hypothesis correlation? (y/n)
$?
y          #          null corr. < alt. corr.

You are specifying a list for values of Sample size
The list currently contains 0 values.
The maximum number allowed is 20.

Enter (1) ADD values individually specified
      (2) ADD values equally spaced between bounds
      (3) ADD values as (2) with logarithmic spacing
      (4) PRINT list
      (5) DELETE a single element of list
      (6) DELETE consecutive elements of list
      (7) SORT list in ascending order and eliminate duplicate values
      (8) QUIT
$?
1          #          ADD values indiv. spec.
Enter the number of values that will be entered
$?

```

3 # The number of values

Enter 3 values

\$?

10 50 100 # The 3 values

The list currently contains 3 values.

The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

8 # QUIT

***** One-sample Correlation *****

These calculations are for a one sided test.

Null hypothesis correlation 0.000

Significance level 0.050

Power 0.800

Sample size	Alternative hypothesis correlation
10.000	0.743
50.000	0.350
100.000	0.248

Press Return / Enter to continue:

\$?

Done seeing answer

Choice of Next Action: -----

Enter:

- (1) To perform another calculation of the same sort
- (2) To change the setup for this type of problem
- (3) To return to higher level menu
- (4) To quit DSTPLAN

Select one of the above:

\$?

1 # Another calculation

Parameter Specification: -----

- Enter:
- 1: Null hypothesis correlation
 - 2: Alternative hypothesis correlation
 - 3: Sample size
 - 4: Significance level
 - 5: Power

Enter a '?' for the item to be calculated. Enter a 't' if a table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

0.5 ? 50 0.05 0.8 # One-sample corr. coeff. ex. #2

Special Case: -----

Is the null hypothesis correlation less than the alternative hypothesis correlation? (y/n)

\$?

n # null corr. < alt. corr.

***** One-sample Correlation *****

These calculations are for a one sided test.

Null hypothesis correlation 0.500

Sample size 50.000

Significance level 0.050

Power 0.800

Chapter 28

Two-sample correlation

28.1 The test

The null hypothesis is that the correlation coefficient between two observed variables is the same in both of the groups.

The test associated with these calculations uses the approximate normality of the difference in transformed sample correlation coefficients (see Method).

SIDES: Both one- and two-sided alternatives are available.

28.2 Parameters of the test

Group 1 correlation: The value of the correlation coefficient, C_1 , between the variables in group 1.

Group 2 correlation: The value of the correlation coefficient in group 2, C_2 , between the same variables used to obtain the group 1 correlation. This value may be greater than or less than the value of C_1 .

Sample Size of group 1: The number of trials or observations in group 1.

Sample Size of group 2: The number of trials or observations in group 2.

Significance: The probability of falsely rejecting the null hypothesis when it is true.

Power: The probability of correctly rejecting the null hypothesis when the alternative is true.

Number for equal sample size: The equal number in each group necessary to achieve the specified power for the given values of the other parameters.

28.3 Method

Uses the approximate normality of the difference of transforms, $Z_1 - Z_2$, of the correlation coefficients, C_1 and C_2 :

$$Y_i = \frac{1}{2} \log \left(\frac{1 + C_i}{1 - C_i} \right)$$

$$Z_i = Y_i - \frac{3 * Y_i + C_i}{4N_i}$$

where C_i is the correlation coefficient of the i th group, N_i is the sample size of the i th group, and $i = 1$ or 2 .

28.4 Examples

The calculations shown determine the equal sample size in each group necessary to detect a difference in correlation coefficients in a one-sided test with a power of 0.8 and a significance of 0.05. The correlation coefficients considered are 0.3, 0.4, 0.5, 0.6, and 0.65 for the lower correlation and 0.7 for the higher correlation.

===== Begin Echo of Computer Session =====

***** Two-sample Correlation *****

Problem Initialization:-----

Enter 1 or 2 for 1- or 2-sided test:

\$?

1 # 1-sided tests

Parameter Specification: -----

Enter: 1: Group 1 correlation
2: Group 2 correlation
3: Sample size of group 1
4: Sample size of group 2
5: Significance level
6: Power

Enter a '?' for the item to be calculated. Entering a '?' in positions 3 and 4 will calculate equal sample sizes for both groups. Enter a 't' if a

table of values for this parameter is to be specified (only one 't' entry allowed). Enter the numeric value of the parameter otherwise. Entering zeros for all items skips the calculation and allows a choice of the next action.

\$?

t 0.7 ? ? 0.05 0.8 # Two-sample corr. coeff. ex. #1
You are specifying a list for values of Group 1 correlation
The list currently contains 0 values.
The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

1 # ADD values indiv. spec.
Enter the number of values that will be entered

\$?

5 # The number of values
Enter 5 values

\$?

0.3 0.4 0.5 0.6 0.65 # The 5 values

The list currently contains 5 values.
The maximum number allowed is 20.

- Enter (1) ADD values individually specified
- (2) ADD values equally spaced between bounds
- (3) ADD values as (2) with logarithmic spacing
- (4) PRINT list
- (5) DELETE a single element of list
- (6) DELETE consecutive elements of list
- (7) SORT list in ascending order and eliminate duplicate values
- (8) QUIT

\$?

8 # QUIT

***** Two-sample Correlation *****

