PHCM9795: Foundations of Biostatistics

Timothy Dobbins

Invalid Date

Table of contents

Table of contents i

Course introduction 1
- Course information . . . 1
- Units of credit . . . 1
- Course aim . . . 1
- Learning outcomes . . . 1

1 Sample size estimation 3
- Learning objectives . . . 3
- Readings . . . 3
- 1.1 Introduction . . . 3
  - Under and over-sized studies . . . 3
- 1.2 Sample size estimation for descriptive studies . . . 4
  - Worked Example . . . 4
- 1.3 Sample size estimation for analytic studies . . . 5
  - Factors to be considered . . . 5
  - Power and significance level . . . 5
- 1.4 Detecting the difference between two means . . . 6
  - Worked Example . . . 6
- 1.5 Detecting the difference between two proportions . . . 8
  - Worked Example . . . 8
- 1.6 Detecting an association using a relative risk . . . 9
  - Worked Example . . . 10
- 1.7 Detecting an association using an odds ratio . . . 11
  - Worked Example . . . 11
- 1.8 Factors that influence power . . . 12
  - Dropouts . . . 12
  - Unequal groups . . . 12
- 1.9 Limitations in sample size estimations . . . 13
- 1.10 Summary . . . 14

10 Stata resources **15**
Sample size calculation for two independent samples t-test 15
Sample size calculation for paired samples t-test 16
Sample size calculation for difference between two independent proportions 17
Sample size calculation with relative risk and odds ratio 19

10 R resources **21**
Sample size calculation for the independent samples t-test 21
Sample size calculation for the paired t-test 22
Sample size calculation for difference between two independent proportions 23
Sample size calculation with a relative risk 23
Sample size calculation with an odds ratio 24
Estimating sample sizes with unequal groups 25

10 Learning Activities **27**
Activity 10.1 27
Activity 10.2 27
Activity 10.3 27
Activity 10.4 28

References **29**

Course introduction

Welcome to PHCM9795 Foundations of Biostatistics.

This introductory course in biostatistics aims to provide students with core biostatistical skills to analyse and present quantitative data from different study types. These are essential skills required in your degree and throughout your career.

We hope you enjoy the course and will value your feedback and comment throughout the course.

Course information

Biostatistics is a foundational discipline needed for the analysis and interpretation of quantitative information and its application to population health policy and practice.

This course is central to becoming a population health practitioner as the concepts and techniques developed in the course are fundamental to your studies and practice in population health. In this course you will develop an understanding of, and skills in, the core concepts of biostatistics that are necessary for analysis and interpretation of population health data and health literature.

In designing this course, we provide a learning sequence that will allow you to obtain the required graduate capabilities identified for your program. This course is taught with an emphasis on formulating a hypothesis and quantifying the evidence in relation to a specific research question. You will have the opportunity to analyse data from different study types commonly seen in population health research.

The course will allow those of you who have covered some of this material in your undergraduate and other professional education to consolidate your knowledge and skills. Students exposed to biostatistics for the first time may find the course challenging at times. Based on student feedback, the key to success in this course is to devote time to it every week. We recommend that you spend an average of 10-15 hours per week on the course, including the time spent reading the course notes and readings, listening to lectures, and working through learning activities and completing your assessments. Please use the resources provided to assist you, including online support.

Units of credit

This course is a core course of the Master of Public Health, Master of Global Health and Master of Infectious Diseases Intelligence programs and associated dual degrees, comprising 6 units of credit towards the total required for completion of the study program. A value of 6 UOC requires a minimum of 150 hours work for the average student across the term.

Course aim

This course aims to provide students with the core biostatistical skills to apply appropriate statistical techniques to analyse and present population health data.

Learning outcomes

On successful completion of this course, you will be able to:

1. Summarise and visualise data using statistical software.
2. Demonstrate an understanding of statistical inference by interpreting p-values and confidence intervals.
3. Apply appropriate statistical tests for different types of variables given a research question, and interpret computer output of these tests appropriately.
4. Determine the appropriate sample size when planning a research study.
5. Present and interpret statistical findings appropriate for a population health audience.

Module 1

Sample size estimation

Learning objectives

By the end of this module you will be able to:

- Explain the issues involved in sample size estimation for epidemiological studies;
- Estimate sample sizes for descriptive and analytic studies;
- Compute the sample size needed for planned statistical tests;
- Adjust sample size calculations for factors that influence study power.

Readings

Kirkwood and Sterne (2001); Chapter 35. [UNSW Library Link]

Bland (2015); Chapter 18. [UNSW Library Link]

For interest: Woodward (2013); Chapter 8. [UNSW Library Link]

1.1 Introduction

Determining the appropriate sample size (the number of participants in a study) is one of the most critical issues when designing a research study. A common question when planning a project is "How many participants do I need?" The sample size needs to be large enough to ensure that the results can be generalised to the population and will be accurate, but small enough for the study question to be answered with the resources available. In general, the larger the sample size, the more precise the study results will be.

Unfortunately, estimating the sample size required for a study is not straightforward and the method used varies with the study design and the type of statistical test that will be conducted on the data collected. In the past, researchers calculated the sample size by hand using complicated mathematical formula. More recently, look-up tables have been created which has removed the need for hand calculations. Now, most researchers use computer programs where parameters relevant to the particular study design are entered and the sample size is automatically calculated. In this module, we will use an abbreviated look-up table to demonstrate the parameters that need to be considered when estimating sample sizes for a confidence interval and use Stata for all other sample size calculations. The look-up table allows you to see at a glance, the impact of different factors on the sample size estimation.

Under and over-sized studies

In health research, there are different implications for interpreting the results if the sample size is too small or too large.

An under-sized study is one which lacks the power to find an effect or association when, in truth, one exists. If the sample size is too small, an important difference between groups may not be statistically significant and so will not be detected by the study. In fact, it is considered unethical

to conduct a health study which is poorly designed so that it is not possible to detect an effect or association if it exists. Often, Ethics Committees request evidence of sample size calculations before a study is approved.

A classic paper by Freiman et al examined 71 randomised controlled trials which reported an absence of clinical effect between two treatments.(Freiman et al. 1978) Many of the trials were too small to show that a clinically important difference was statistically significant. If the sample size of an analytic study is too small, then only very limited conclusions can be drawn about the results.

In general, the larger the sample size the more precise the estimates will be. However, large sample sizes have their own effect on the interpretation of the results. An over-sized study is one in which a small difference between groups, which is not important in clinical or public health terms, is statistically significant. When the study sample is large, the null hypothesis could be rejected in error and research resources may be wasted. This type of study may be unethical due to the unnecessary enrolment of a large number of people.

1.2 Sample size estimation for descriptive studies

To estimate the sample size required for a descriptive study, we usually focus on specifying the width of the confidence interval around our primary estimate. For example, to estimate the sample size for a study designed to measure a prevalence we need to:

- nominate the expected prevalence based on other available evidence;
- nominate the required level of precision around the estimate. For this, the width of the 95% confidence interval (i.e. the distance equal to 1.96 $\times$ SE) is used.

Table 10.1 is an abbreviated look-up table that we can use to estimate the sample size for this type of study. Note that the sample size required to detect an expected population prevalence of 5% is the same as to detect a prevalence of 95%. Similarly 10% is equivalent to 90% etc. It is symmetric about 50%. From Table 1.1, you can see that the sample size required increases as the expected prevalence approaches 50% and as the precision increases (i.e. the required 95% CI becomes narrower).

Table 1.1: Sample size required to calculate a 95% confidence interval with a given precision

	Width of 95% confidence interval (precision)									
Prevalence	**1%**	**1.5%**	**2%**	**2.5%**	**3%**	**3.5%**	**4%**	**5%**	**10%**	**15%**
5% or 95%	1,825	812	457	292	203	149	115			
10% or 90%	3,458	1,537	865	554	385	283	217	139		
15% or 85%	4,899	2,177	1,225	784	545	400	307	196	49	
20% or 80%	6,147	2,732	1,537	984	683	502	385	246	62	28
25% or 75%	7,203	3,202	1,801	1,153	801	588	451	289	73	33

Worked Example

A descriptive cross-sectional study is designed to measure the prevalence of bronchitis in children age 0-2 years with a 95% CI of $\pm$ 4%. The prevalence is expected to be 20%. From the

table, a sample size of at least 385 will be required for the width of the 95% CI to be ± 4% (i.e. the reported precision of the summary statistic will be 20% (95% CI 16% to 24%)).

If the prevalence turns out to be higher than the researchers expected or if they decided that they wanted a narrower 95% CI (i.e. increase precision), a larger sample size would be required.

- What sample size would be required if the prevalence was 15% and the desired 95% CI was ± 3%?
- Answer: 545

1.3 Sample size estimation for analytic studies

Analytic study designs are used to compare characteristics between different groups in the population. The main study designs are analytic cross-sectional studies, case-control studies, cohort studies and randomised controlled trials. For analytic study designs, the outcome measure of interest can be a mean value, a proportion or a relative risk if a random sample has been enrolled. For case-control studies the most appropriate measure of association is an odds ratio.

Factors to be considered

The first important decision in estimating a required sample size for an analytic study is to select the type of statistical test that will be used to report or analyse the data. Each type of test is associated with a different method of sample size estimation.

Once the statistical method has been determined, the following issues need to be decided:

- Statistical power: the chance of finding a difference if one exists, e.g. 80%;
- Level of significance: the P value that will be considered significant, e.g. P<0.05;
- Minimum effect size of interest: the size of the difference between groups e.g. the difference in the proportion of parents who oppose immunisation in two different regions or the difference in mean values of blood pressure in two groups of people with different types of cardiac disease;
- Variability: the spread of the measurements, e.g. the expected standard deviation of the main outcome variable (if continuous), or the expected proportions;
- Resources: an estimate of the number of participants available and amount of funding to run the study.

In addition to deciding the level of power and probability that will be used, the difference between groups that is regarded as being important has to be estimated. The smallest difference between study groups that we want to detect is described as the minimum expected effect size. This is determined on the basis of clinical judgement, public health importance and expertise in the condition being researched, or may it be need to be determined from a pilot study or a literature review. The smaller the expected effect or association, the larger the sample size will need to obtain statistical significance. We also need some knowledge of how variable the measurement is expected to be. For this we often use the standard deviation for a continuous measure. As measurement variability increases, the sample size will need to increase in order to detect the expected difference between the groups. Above all, a study has to be practical in terms of the availability of a population from which to draw sufficient numbers for the study and in terms of the funds that are available to conduct the study.

Power and significance level

The power of a study, which was discussed in Module 4, is the chance of finding a statistically significant difference when one exists, i.e. the probability of correctly rejecting the null hypothesis. The relationship between the power of a study and statistical significance is shown in Table 1.2.

The power of a study is expressed as 1− β where β is the probability of a false negative (that is, the probability of a Type II error - incorrectly not rejecting the null hypothesis. In most research,

Table 1.2: Comparison of study result with the truth

Study result	**Truth**	
	Effect	No effect
Evidence	✓	α
No evidence	β	✓

power is generally set to 80% (a Type II error rate of 20%). However, in some studies, especially in some clinical trials where rigorous results are required, power is set to 90% (a Type II error rate of 10%).

The significance level, or α level, is the level at which the P value of a test is considered to be statistically significant. The α level is usually set at 5% indicating a probability of <0.05 will be regarded as statistically significant. Occasionally, especially if several outcome measures are being compared, the α level is set at 1% indicating a probability of <0.01 will be regarded as statistically significant.

The calculation of sample sizes for analytic studies are based on calculations that are somewhat tedious to compute by hand. Software packages are the standard method of calculating sample sizes for these types of study, and examples from both R and Stata will be provided.

1.4 Detecting the difference between two means

The test that is used to show that two mean values are significantly different from one another is the independent samples t-test (Module 5). The sample size needed for this test to have sufficient power can be calculated using R and Stata as shown in the Worked Example below.

Worked Example

There is a hypothesis that the use of the oral contraceptive (OC) pill in premenopausal women can increase systolic blood pressure. A study was planned to test this hypothesis using a two sided t-test. The investigators are interesting in detecting an increase of at least 5 mm Hg systolic blood pressure in the women using OC compared to the non-OC users with 90% power at a 5% significance level. A pilot study shows that the SD of systolic blood pressure in the target group is 25 mm Hg and the mean systolic blood pressure of non-OC user women is 90 mm Hg. What is the minimum number of women in each group that need to be recruited for the study to detect this difference?

Solution The effect size of interest is 5 mm Hg and the associated standard deviation is 25 mm Hg. For power of 90% and alpha of 5%, the sample size calculation using the Stata can be calculated using the `power twomeans` command:

Stata Output 10.1: Two independent samples t-test sample size calculation

```
. power twomeans 90 95, sd(25) power(0.9)

Performing iteration ...

Estimated sample sizes for a two-sample means test
t test assuming sd1 = sd2 = sd
Ho: m2 = m1 versus  Ha: m2 != m1

Study parameters:

        alpha =    0.0500
        power =    0.9000
```

```
        delta =    5.0000
           m1 =   90.0000
           m2 =   95.0000
           sd =   25.0000

Estimated sample sizes:

            N =     1,054
  N per group =       527
```

We can use the `epi.sscompc` function within the `epiR` package in R to calculate the sample size:

```
library(epiR)
library(pwr)

epi.sscompc(treat=90, control=95, n=NA, sigma=25, power=0.9)
```

```
$n.total
[1] 1052

$n.treat
[1] 526

$n.control
[1] 526

$power
[1] 0.9

$delta
[1] 5
```

Note that Stata and R provide slightly different estimated sample sizes. This difference is immaterial from a practical point of view, and highlights the importance of referencing which software package has been used when writing up results.

From the output, we can see that with 90% power we will need 526 or 527 participants in each group, i.e., 1052 or 1054 participants in total.

If the above were carried out by taking baseline measures of systolic blood pressure, and then again when the women were taking the OC pills, it would be a matched-pair study. Computing sample sizes for paired studies requires an estimate of the correlation between the paired obervations. If we do not have any estimates for this correlation, we can assume a value of 0. If the correlation is positive, a zero for correlation would give a more conservative estimate of sample size required (i.e. estimate a sample size larger than necessary). While a negative correlation would require a bigger sample size than a zero correlation, it is relatively uncommon to encounter negative correlations between pairs. Any discussions on the effect of correlation on sample size is beyond the scope of this course. Thus, we will always assume a correlation of zero between paired measurements in this course.

We can compute the required sample size in Stata using the `power pairedmeans` command:

Output 10.2: Paired samples t-test sample size using Worked Example 10.2

```
. power pairedmeans 90 95, corr(0) power(0.9) sd(25)

Performing iteration ...
```

```
Estimated sample size for a two-sample paired-means test
Paired t test assuming sd1 = sd2 = sd
Ho: d = d0  versus  Ha: d != d0

Study parameters:

        alpha =    0.0500          ma1 =   90.0000
        power =    0.9000          ma2 =   95.0000
        delta =    0.1414           sd =   25.0000
           d0 =    0.0000         corr =    0.0000
           da =    5.0000
         sd_d =   35.3553

Estimated sample size:

            N =       528
```

As discussed in the R notes, calculating the sample size required for a paired t-test is a little more cumbersome in R. Here, only the output of the process is provided - refer to the R notes for detail on the code.

Output 10.2: Paired samples t-test sample size using Worked Example 10.2

```
     Paired t test power calculation

              n = 527.2954
              d = 0.1414214
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number of *pairs*
```

Assuming a correlation of 0 between the two sets of measurements, we can see that we will need 528 pairs of measurements to achieve a power of 90% (virtually the same as for an independent samples study).

1.5 Detecting the difference between two proportions

The statistical test for deciding if there is a significant difference between two independent proportions is a Pearson's chi-squared test (Module 7).

Other than the power and alpha required for the test, the expected prevalence or incidence rate of the outcome factor needs to be estimated for each of the two groups being compared, based on what is known from other studies or what is expected. Occasionally, we may not know the expected proportion in one of the groups, e.g. in a randomised control trial of a novel intervention. In the sample size calculation for such a study, we should instead justify the minimum expected difference between the proportions based on what is important from a clinical or public health perspective. Based on the minimum difference, we can then derive the expected proportion for both groups. Note that the smaller the difference, the larger the sample size required.

The sample size required in each group to observe a difference in two independent proportions can be calculated using the power twoproportions command in Stata.

Worked Example

If we expect that the prevalence of smoking in two comparison groups (e.g. males and females) will be 35% and 20%. The sample size required in each group to show that the prevalences are

significantly different at P<0.05 with 80% power is shown in Output 10.3.

Stata Output 10.3: Sample size calculation for two independent proportions

```
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
        power =    0.8000
        delta =   -0.1500  (difference)
           p1 =    0.3500
           p2 =    0.2000

Estimated sample sizes:

            N =       276
  N per group =       138
```

From Output 10.3, we see that we would need 138 males and 138 females (i.e. a total sample size of 276 participants).

What sample size would be required if the prevalence of smoking among men was 30%?

Answer = 294 men and 294 women would be needed.

[Command: `power twoproportions .3 .2, test(chi2)`]

To do the same problem using R, we use the `epi.sscohortc` function. We need to specify the risk of the the outcome each group (labelled group 0 and 1) and the desired power:

```
epi.sscohortc(irexp1=0.35, irexp0=0.2, n=NA, power=0.8)
```

```
$n.total
[1] 276

$n.exp1
[1] 138

$n.exp0
[1] 138

$power
[1] 0.8

$irr
[1] 1.75

$or
[1] 2.153846
```

1.6 Detecting an association using a relative risk

The relative risk is used to describe the association between an exposure and an outcome variable if the sample has been randomly selected from the population. This statistic is often used to describe the effect or association of an exposure in a cross-sectional or cohort study or the effect/association of a treatment in an randomised controlled trial. To estimate the sample

size required for the RR to have a statistically significant P value, i.e. to show a significant association, we need to define: - the size of the RR that is considered to be of clinical or public health importance; - the event rate (rate of outcome) among the group who are not exposed to the factor of interest (reference group); - the desired level of significance (usually 0.05); - the desired power of the study (usually 80% or 90%).

In general, a RR of 2.0 or greater is considered to be of public health importance, however, a smaller RR can be important when exposure is high. For example, there may be a relatively small risk of respiratory infection among young children with a parent who smokes (RR ~ 1.2). If 25% of children are exposed to smoking in their home, then the high exposure rate leads to a very large number of children who have preventable respiratory infections across the community.

Worked Example

A study is planned to investigate the effect of an environmental exposure on the incidence of a certain common disease. In the general (unexposed) population the incidence rate of the disease is 50% and it is assumed that the incidence rate would be 75% in the exposed population. Thus the relative risk of interest would be 1.5 (i.e. 0.75 / 0.50). We want to detect this effect with 90% power at a 5% level of significance.

We can use the Stata command `power twoproportions` as below:

Stata Output 10.4: Sample size calculation for relative risk

```
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
        power =    0.9000
        delta =    1.5000  (relative risk)
           p1 =    0.5000
           p2 =    0.7500
        rrisk =    1.5000

Estimated sample sizes:

            N =       154
  N per group =        77
```

From Output 10.4, we can see that for a control proportion of 0.5 and RR of 1.5, we need a total sample size of 154, that is 77 people would be needed in each of the exposure groups.

The `epiR` package does not have a function to estimate sample size and power directly for a relative risk, but we can use the `epi.sscohortc` function. To do this, we recognise that the assumed rate in the exposed group will equal the rate in the unexposed group multiplied by the relative risk.

Here we define the risk in the unexposed group as 0.5 and the desired relative risk to detect is 1.5. So we specify `irexp0 = 0.5` and `irexp1 = 1.5 * 0.5`:

```
epi.sscohortc(irexp1=1.5*0.5, irexp0=0.5, n=NA, power=0.9)
```

```
$n.total
[1] 154

$n.exp1
[1] 77
```

```
$n.exp0
[1] 77

$power
[1] 0.9

$irr
[1] 1.5

$or
[1] 3
```

1.7 Detecting an association using an odds ratio

If we are designing a case-control study, the appropriate measure of effect is an odds ratio. The method for estimating the sample size required to detect an odds ratio of interest is slightly different to that for the relative risk. However, the same parameters are required for the estimation:

- the minimum OR to be considered clinically important;
- the proportion of exposed among the control group;
- the desired level of significance (usually 0.05);
- the desired power of the study (usually 80% or 90%).

Worked Example

A case-control study is designed to examine an association between an exposure and outcome factor. Existing literature shows that 30% of the controls are expected to be exposed. We want to detect a minimum OR of 2.0 with 90% power and 5% level of significance.

```
. power twoproportions .3, test(chi2) oratio(2) power(0.9)
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

Study parameters:

        alpha =    0.0500
        power =    0.9000
        delta =    2.0000  (odds ratio)
           p1 =    0.3000
           p2 =    0.4615
   odds ratio =    2.0000

Estimated sample sizes:

            N =       376
  N per group =       188
```

We can use the `epi.sscc` function in R to calculate a sample size based on an odds ratio in a case-control study:

```
epi.sscc(OR=2, p0=0.3, n=NA, power=0.9)
```

```
$n.total
[1] 376
```

```
$n.case
[1] 188

$n.control
[1] 188

$power
[1] 0.9

$OR
[1] 2
```

We find that 188 controls and 188 cases are required i.e. a total of 376 participants.

This sample size would be smaller if we increased the effect size (OR) or reduced the study power to 80%. You could try this in either Stata or R (answer: 141 per group if power is reduced to 80%).

1.8 Factors that influence power

Dropouts

It is common to increase estimated sample sizes to allow for drop-outs or non-response. To account for drop-outs, the estimated sample size can be divided by (1 minus the dropout rate). Consider the following case:

- n-completed: the number who will complete the study (i.e. n after drop-out)
- n-recruited: the number who should be recruited (i.e. n before drop-out)
- d: drop-out rate (as a proportion - i.e. a number between 0 and 1)

Then n-completed = n-recruited × (1 - d)

Re-arranging this formula gives: n-recruited = n-completed ÷ (1 - d).

Unequal groups

Many factors that come into play in a study can reduce the estimated power of a study. In clinical trials, it is not unusual for recruitment goals to be much harder to achieve than expected and therefore for the target sample size to be impossible to realise within the timeframe planned for recruitment.

In case-control studies, the number of potential case participants available may be limited but study power can be maintained by enrolling a greater number of controls than cases. Or in an experimental study, more participants may be randomised to the new treatment group to test its effects accurately when much is known about the effect of standard care and a more precise estimate of the new treatment effect is required.

However, there is a trade-off between increasing the ratio of group size and the total number that needs to be enrolled. Consider Worked Example @ref(wex10-5): selecting an equal number of controls and cases would require 188 cases and 188 controls, a total of 376 participants.

We may want to reduce the number of cases required, by selecting 2 controls for every case. When performing sample size calculations with unequal groups, Stata refers to cases as N2, and controls as N1. Selecting 2 controls (N1) per case (N2) (corresponding to a ratio of N2/N1 0.5 in Stata) would require 140 cases and 280 controls, a total of 420 participants. We can extend this example and investigate the impact of changing the ratio of controls per case.

This can be visualised graphically, as in Figure 1.1.

We can see that the number of cases required drops off if we go from 1 to 2 controls per case, and again from 2 to 3 controls per case. Once we go from 3 to 4 controls per case, we only

Table 1.3: Increasing the number of controls per case

Controls per case	Stata's allocation ratio (N2/N1)	Number of cases required	Number of controls required	Total participants required
1	1	188	188	376
2	0.5	140	280	420
3	0.3333	124	371	495
4	0.25	116	462	578
5	0.2	111	553	664
6	0.1666	108	644	752
7	0.1429	105	734	839
8	0.125	104	825	929
9	0.1111	102	916	1,018
10	0.1	101	1,006	1,107

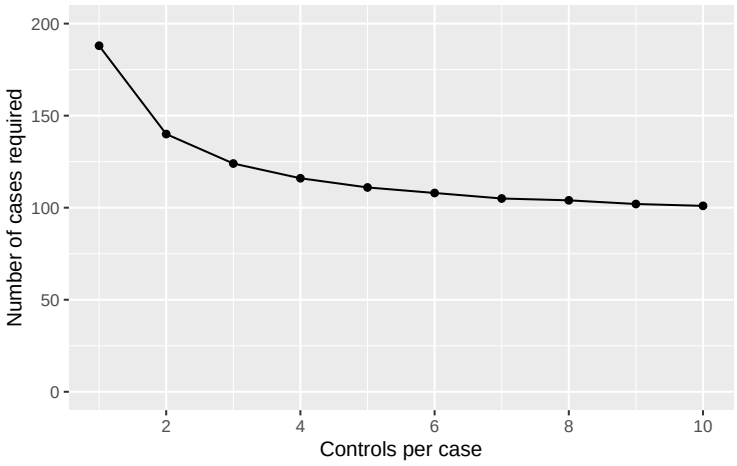


Figure 1.1: Increasing the number of controls per case

reduce the number of cases by 8 (124 vs 116 cases), but at an increase of 91 (371 vs 462) controls. Clearly, this reduction in cases is not offset by the extra controls required.

In Stata, nratio is experimental group / control group. I'm not sure the ratio in R is correct!

	Stata	R
Difference in means	n-exposed / n-controls	n-exposed / n-unexposed
Difference in proportions	n-exposed / n-controls	n-exposed / n-unexposed
Relative risk	n-exposed / n-controls	n-exposed / n-unexposed
Odds ratio	n-cases / n-controls	n-controls / n-cases

1.9 Limitations in sample size estimations

In this module we have seen how to use Stata for estimating the sample size requirement of a study given the statistical test that will be used and the expected characteristics of the sample.

However, once a study is underway, it is not unusual for sample size to be compromised by the lack of research resources, difficulties in recruiting participants or, in a clinical trial, participants wanting to change groups when information about the new experimental treatment rapidly becomes available in the press or on the internet.

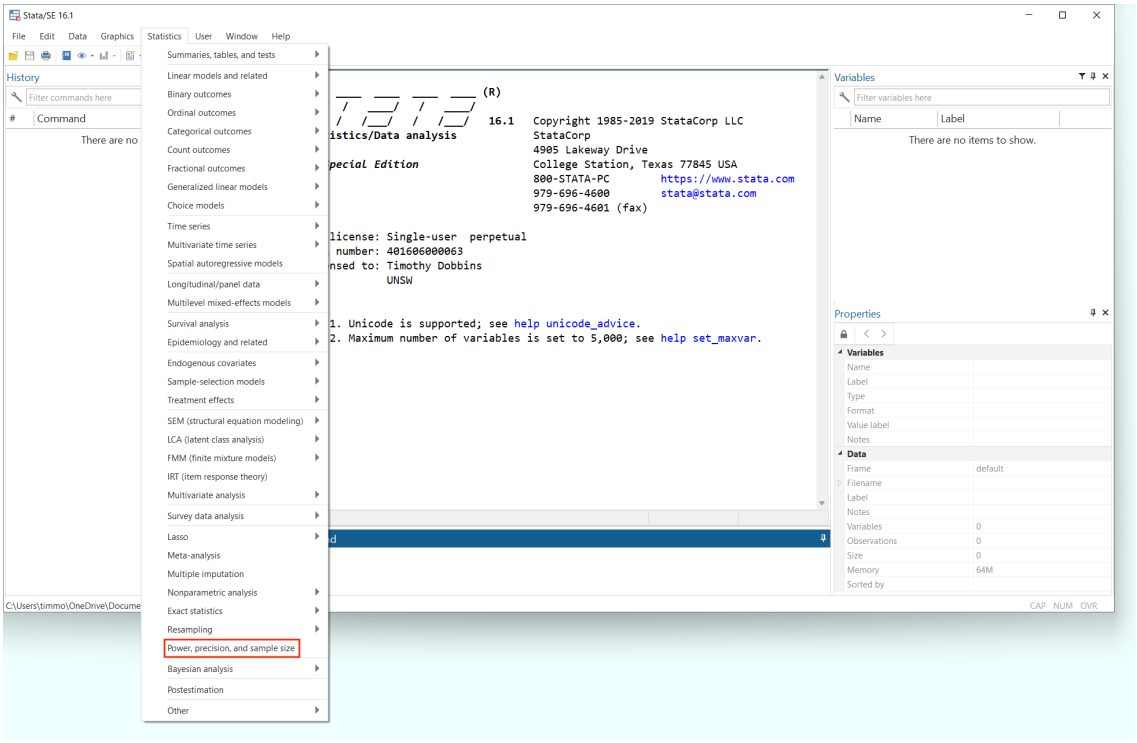
One approach that is increasingly being used is to conduct a blinded interim analysis say when 50% of the total data that are planned have been collected. In this, a statistician external to the research team who is blinded to the interpretation of the group code is asked to measure the effect size in the data with the sole aim of validating the sample size requirement. It is rarely a good idea to use an interim analysis to reduce the planned sample size and terminate a trial early because the larger the sample size, the greater the precision with which the treatment effect is estimated. However, interim analyses are useful for deciding whether the sample size needs to be increased in order to answer the study question and avoid a Type II error.

1.10 Summary

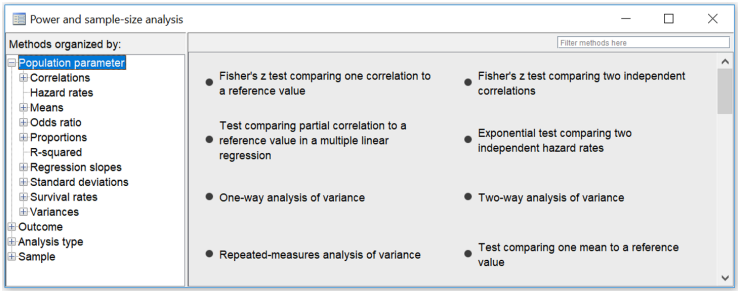
In this module we have discussed the importance of conducting a clinical or epidemiological study with enough participants so that an effect or association can be identified if it exists (i.e. study power), and how this has to be balanced by the need to not enrol more participants than necessary because of resource issues. We have looked at the parameters that need to be considered when estimating the sample size for different studies and have used a look-up table to estimate required sample size for a prevalence study and Stata to estimate appropriate sample sizes in epidemiological research under the most straightforward situations. The common requirement in all the situations is that the researchers need to specify the minimum effect measure (e.g. difference in means, OR, RR etc) they want to detect with a given probability (usually 80% to 90%) at a certain level of significance (usually $P<0.05$). The ultimate decision on the sample size depends on a compromise among different objectives such as power, minimum effect size, and available resources. To make the final decision, it is helpful to do some trial calculations using revised power and the minimum detectable effect measure.

10 Stata resources

The **Power, precision and sample size** menu item is located at the very bottom of the **Statistics** menu. **You may need to scroll down the menu to locate the item.**

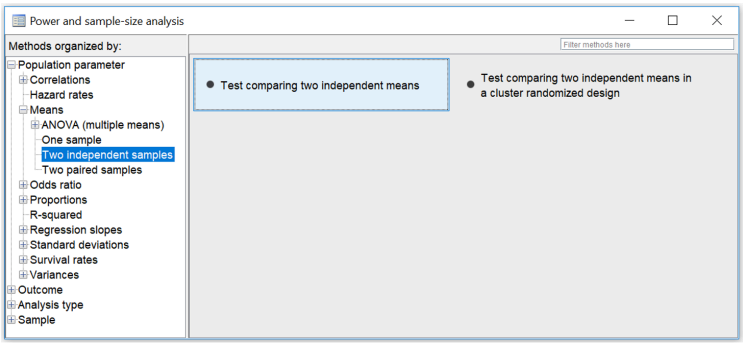


The **Power, precision and sample-size** analysis dialog box appears as shown below.

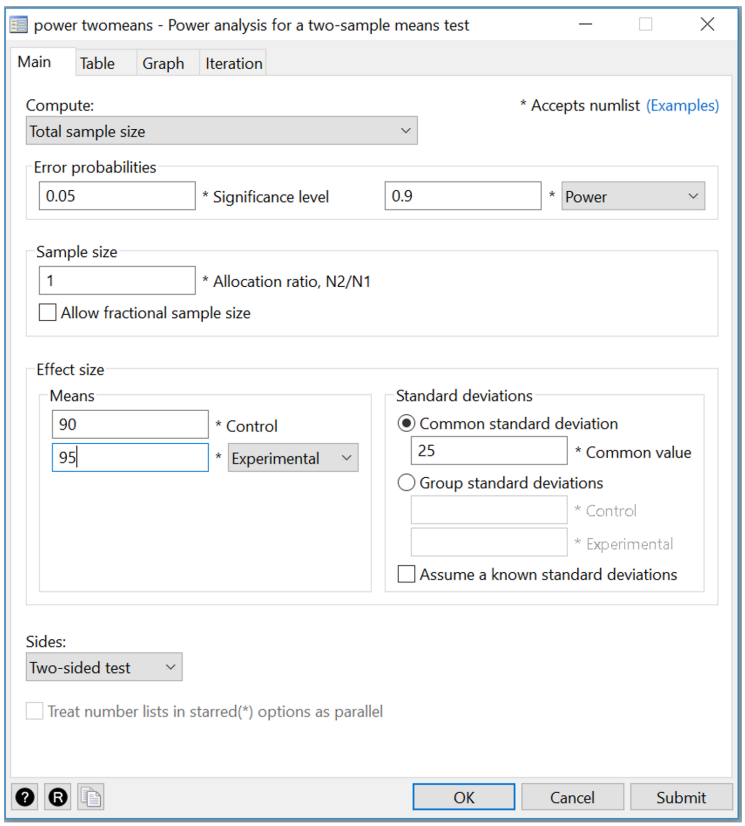


Sample size calculation for two independent samples t-test

To do the problem discussed in Worked Example 10.2, you can expand the **Means** item under **Population parameter** on the left-hand-side of the **Power and sample-size analysis** dialog box, then choose **Two independent** samples. On the right-hand-side of the dialog box, choose **Test comparing two independent means**.



The **power twomeans** dialog box will appear. Based on the information in Worked Example 10.2, change **Power** to `0.9` and enter `25` as the **Common standard deviation**. For the means, we can assume that **Control** mean systolic blood pressure is `90` mmHg and the the **Experimental** mean systolic blood pressure is 5 mmHg higher at `95` mmHg. We use equal numbers in each group (allocation ratio of `1` which is the default) because that would give us the smallest total sample size required.

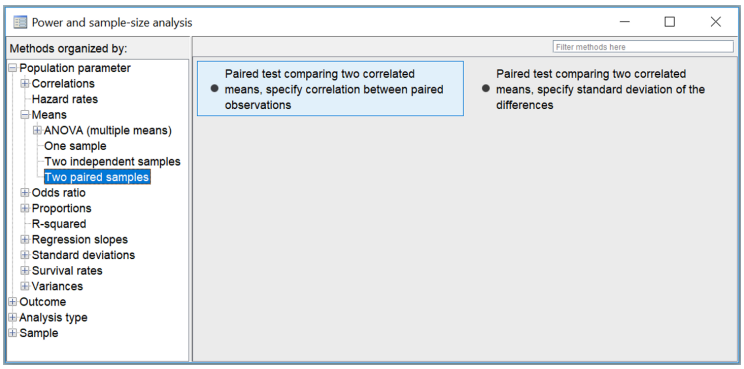


Click **OK** or **Submit** to produce Output 10.1.

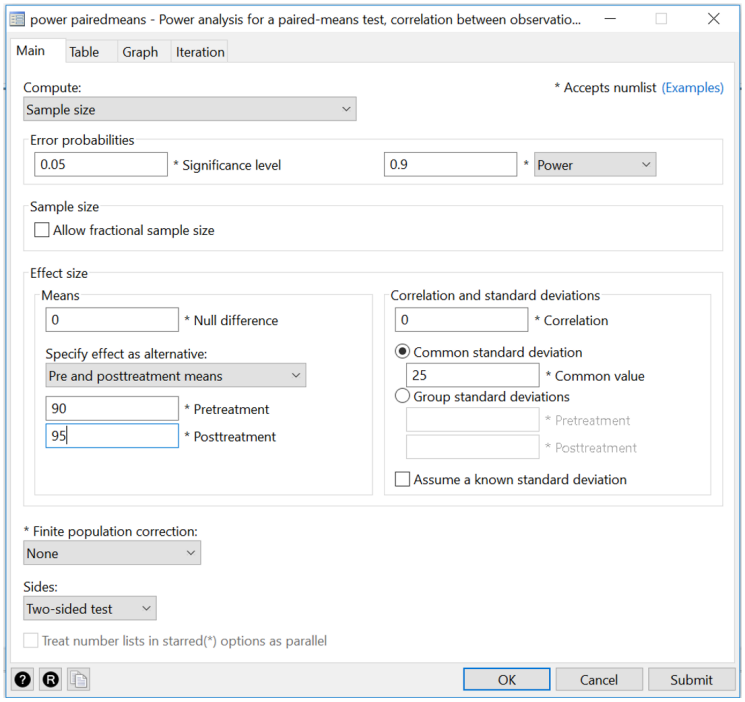
[Command: `power twomeans 90 95, sd(25) power(0.9)`]

Sample size calculation for paired samples t-test

For a paired sample t-test, go back to the **Power and sample-size analysis** dialog box and click on **Two paired samples** under **Means** on the left-hand-side, then choose **Paired test comparing two correlated means, specify correlation between paired observations** as shown below.



In the **power pairedmeans** dialog box, change **Power** to `0.9` and enter `25` as the **Common standard deviation** as shown below. For the means, we can assume that **Control** mean systolic blood pressure is 90 mmHg and the the **Experimental** mean systolic blood pressure is 5 mmHg higher at `95` mmHg as we had for the independent samples.

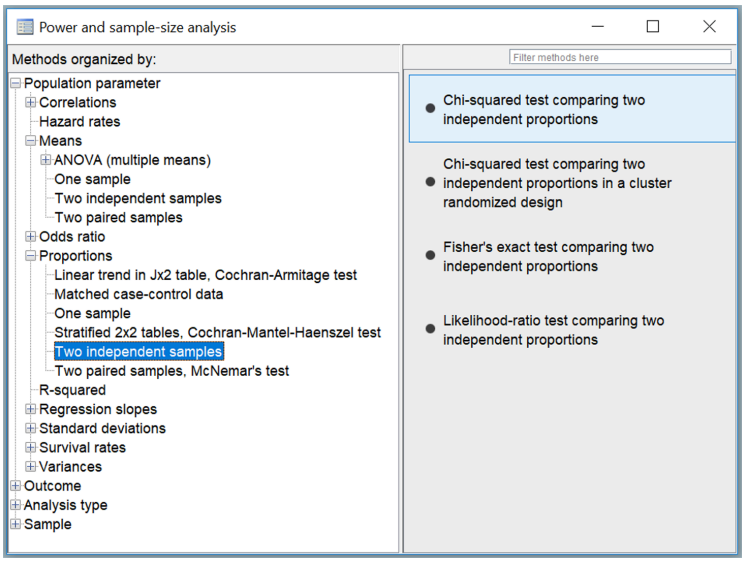


Click **OK** or **Submit** to produce Output 10.2.

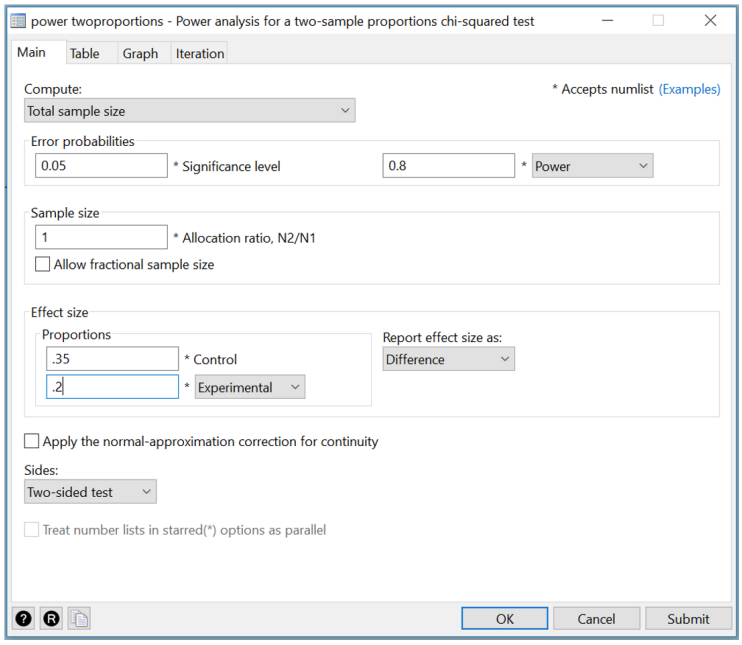
[Command: `power pairedmeans 90 95, corr(0) power(0.9) sd(25)`]

Sample size calculation for difference between two independent proportions

To do the problem discussed in Worked Example 10.3, you can expand the **Proportions** item under **Population parameter** on the left-hand-side of the Power and sample-size analysis dialog box, then choose **Two independent samples**. On the right-hand-side of the dialog box, choose **Chi-squared test comparing two independent proportions**.



The **power twoproportions** dialog box will appear. Based on the information in Worked Example 10.3, check the **Power** is `0.8` and **Significance level** is `0.05` (these are the default values), then enter `0.35` and `0.2` as the **Proportions** as shown below. As with the two independent samples example, we can assume equal numbers in each group (**Allocation ratio** of `1` which is the default).

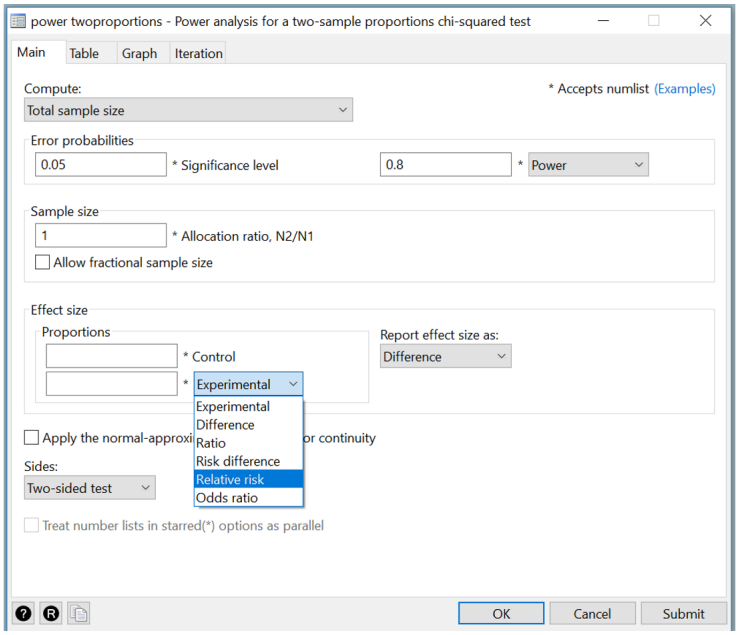


Click **OK** or **Submit** to obtain Output 10.3.

[Command: `power twoproportions .35 .2, test(chi2)`]

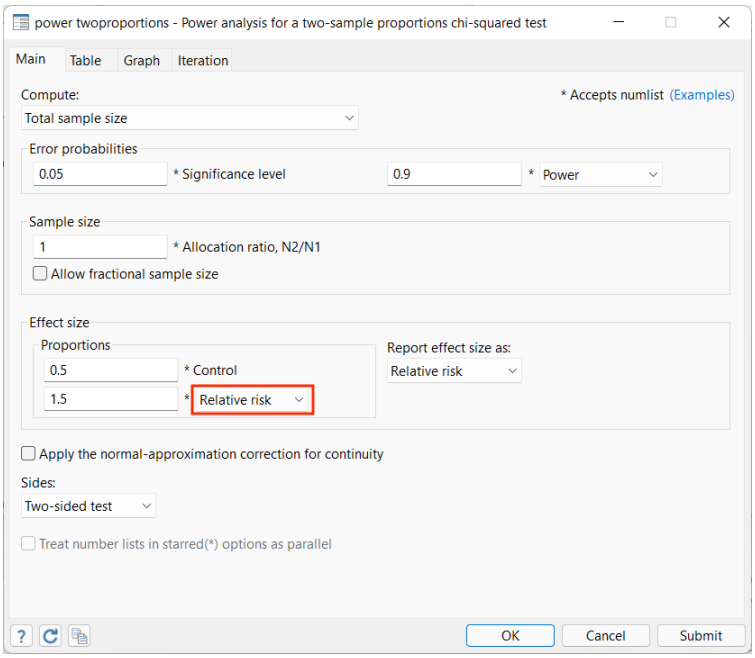
Note: It doesn't matter if you swap the proportions for the **Control** and **Experimental** group, i.e. the command `power twoproportions .2 .35 , test(chi2)` gives the same results.

If you had difference in proportion, relative risk or odds ratio for the sample size calculation, you can choose them from the drop-down list under Experimental as shown below.



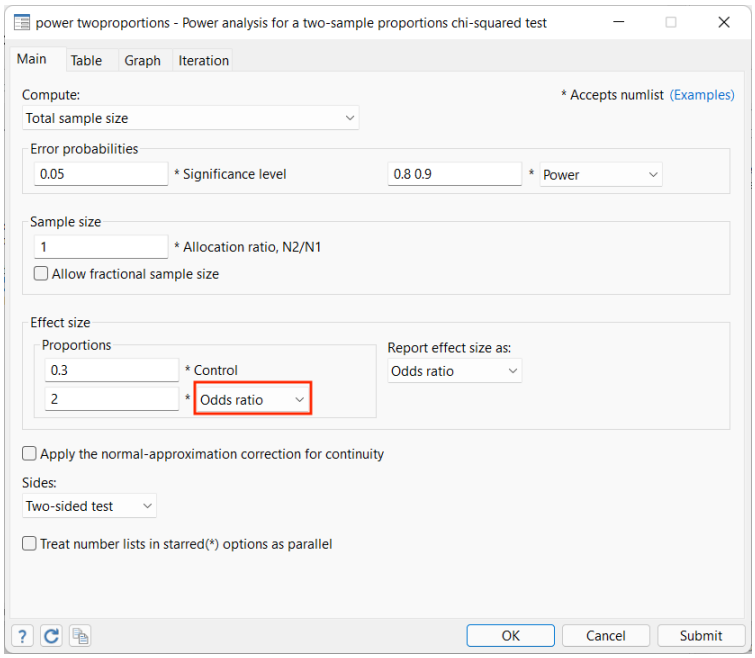
Sample size calculation with relative risk and odds ratio

Using information from Worked Example 10.4, we change **Power** to `0.9`, enter `0.5` as the **Control Proportion**. As we are working with a relative risk, choose **Relative risk** as the effect size, and enter `1.5` as shown below.



Click **OK** or **Submit** to obtain Output 10.4.

Now we calculate the sample size for Worked Example 10.5. Enter `0.3` as the **Control Proportion**, choose **Odds ratio** as the estimate from the dropdown list and enter `2` as the **Odds ratio.** In this example, you can also try entering two values for **Power**: `0.9` and `0.8` as shown below.



Click **OK** or **Submit** to obtain the output below.

```
Estimated sample sizes for a two-sample proportions test
Pearson's chi-squared test
Ho: p2 = p1  versus  Ha: p2 != p1

+---------------------------------------------------------------------------+
|   alpha    power       N      N1      N2   delta      p1      p2  oratio |
|---------------------------------------------------------------------------|
|     .05       .8     282     141     141       2      .3   .4615       2 |
|     .05       .9     376     188     188       2      .3   .4615       2 |
+---------------------------------------------------------------------------+
```

10 R resources

Many power and sample size procedures are available in the `epiR` package. We will also use one function from the `pwr` package.

```
# If not yet installed, submit the following:
# install.packages("epiR")
# install.packages("pwr")
library(epiR)
library(pwr)
```

We will use three functions from the `epiR` package in this module:

- `epi.sscompc` to estimate the sample size to compare continuous outcomes
- `epi.sscohortc` to estimate the sample size to compare two independent proportions from a cohort or cross-sectional study
- `epi.sscc` to estimate the sample size to compare two independent proportions from a case-control study

We will use one function from the `pwr` package:

- `pwr.t.test` estimate the sample size to compare means from a paired study

Sample size calculation for the independent samples t-test

To do the problem discussed in Worked Example 10.2, we use the `epi.sscompc` function:

```
epi.sscompc(treat, control, n, sigma, power,
   r = 1, design = 1, sided.test = 2, nfractional = FALSE, conf.level = 0.95)
```

The first line contains parameters that we usually specify, with the second line usually left as the defaults. We must define the expected mean in the treatment and control groups, and the standard deviation of the measure. We specify one of `n` or `power` to be the measure to estimate, by specifying the unknown value as being equal to R's missing value, `NA`.

For example, to calculate the required sample size in Worked Example 10.2, we specify:

- the assumed mean in the experimental, or treatment, group: 90mmHg
- the assumed mean in the control group: 95mmHg
- the standard deviation of blood pressure: 25mmHg
- the required power, 0.9 (representing 90%)

The values on the second line of the function are defined by default, and we can leave these as default.

Putting this all together, and specifying the sample size as unknown:

```
epi.sscompc(treat=90, control=95, n=NA, sigma=25, power=0.9)
```

```
$n.total
[1] 1052

$n.treat
[1] 526

$n.control
[1] 526

$power
[1] 0.9

$delta
[1] 5
```

The results indicate that we need 526 participants in each group, or 1052 in total. Note that these numbers are slightly different from the Stata estimates (527 in each group).

We can define whether we want unequal numbers in each group by specifying `r`: the number in the treatment group divided by the number in the control group.

Sample size calculation for the paired t-test

Calculating the sample size required for a paired t-test is a little more cumbersome. We can use the following code to specify:

- `m1`: the mean of the first paired observations
- `m2`: the mean of the second paired observations
- `s_group`: the common standard deviation
- `corr`: the assumed correlation between the paired observations (conservatively set to 0)

The code below then uses the `pwr.t.test` function within the `pwr` library to estimate the number of pairs. For example, to replicate Output 10.2:

```
m1 <- 90
m2 <- 95
s_group <- 25
corr <- 0

s_paired <- sqrt(2 * s_group^2 - 2*corr*s_group^2)

d <- ((m1 - m2)/s_paired)

pwr.t.test(d=d, power=0.9, type="paired")
```

```
     Paired t test power calculation

              n = 527.2954
              d = 0.1414214
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number of *pairs*
```

As per the Stata calculations, we require 528 pairs of observations (noting that **sample sizes are always rounded up**).

Sample size calculation for difference between two independent proportions

To do the problem discussed in Worked Example 10.3, we use the `epi.sscohortc` function:

```
epi.sscohortc(irexp1, irexp0, pexp = NA, n = NA, power = 0.80,
   r = 1, N, design = 1, sided.test = 2, finite.correction = FALSE,
   nfractional = FALSE, conf.level = 0.95)
```

We can enter:

- irexp1: the assumed risk of the outcome in the exposed group: here 0.35
- irexp0: the assumed risk of the outcome in the unexposed group: here 0.2
- n: the total sample size, to be determined
- power: the required power: here 0.8 (representing 80%)

```
epi.sscohortc(irexp1=0.35, irexp0=0.2, n=NA, power=0.8)
```

```
$n.total
[1] 276

$n.exp1
[1] 138

$n.exp0
[1] 138

$power
[1] 0.8

$irr
[1] 1.75

$or
[1] 2.153846
```

Note: It doesn't matter if you swap the proportions for the **exposed** and **unexposed** groups, i.e. the command `epi.sscohortc(irexp1=0.2, irexp0=0.35, n=NA, power=0.8)` gives the same results.

Sample size calculation with a relative risk

The `epiR` package does not have a function to estimate sample size and power directly for a relative risk, but we can use the `epi.sscohortc` function. To do this, we recognise that the assumed rate in the exposed group will equal the rate in the unexposed group multiplied by the relative risk.

Here we will replicate Output 10.4, where p0=0.5 and the desired relative risk to detect is 1.5. So we specify `irexp0 = 0.5` and `irexp1 = 1.5 * 0.5`:

```
epi.sscohortc(irexp1=1.5*0.5, irexp0=0.5, n=NA, power=0.9)
```

```
$n.total
[1] 154

$n.exp1
[1] 77

$n.exp0
[1] 77

$power
[1] 0.9

$irr
[1] 1.5

$or
[1] 3
```

Hence we require 77 participants in each group, or 154 participants in total.

Sample size calculation with an odds ratio

We can use the `epi.sscc` function to calculate a sample size based on an odds ratio in a case-control study:

```
epi.sscc(OR, p1 = NA, p0, n, power, r = 1,
   phi.coef = 0, design = 1, sided.test = 2, nfractional = FALSE,
   conf.level = 0.95, method = "unmatched", fleiss = FALSE)
```

Using information from Worked Example 10.4, we specify:

- OR: the odds ratio to be detected, here 1.5
- p0: the proportion of the outcome in the controls, here 0.5
- n: the sample size, here to be calculated
- power: the required study power, here 0.9

```
epi.sscc(OR=1.5, p0=0.5, n=NA, power=0.9)
```

```
$n.total
[1] 1038

$n.case
[1] 519

$n.control
[1] 519

$power
[1] 0.9

$OR
[1] 1.5
```

Now we calculate the sample size for Worked Example 10.5:

```
  epi.sscc(OR=2, p0=0.3, n=NA, power=0.9)
```

```
$n.total
[1] 376

$n.case
[1] 188

$n.control
[1] 188

$power
[1] 0.9

$OR
[1] 2
```

Here we see that we require a total of 376 participants to detect an odds ratio of 2.0 with 90% power;

```
  epi.sscc(OR=2, p0=0.3, n=NA, power=0.8)
```

```
$n.total
[1] 282

$n.case
[1] 141

$n.control
[1] 141

$power
[1] 0.8

$OR
[1] 2
```

or a total of 282 participants to detect an odds ratio of 2.0 with 80% power.

Estimating sample sizes with unequal groups

To change the allocation ratio in any of these study type, we can specify `r`. Note that the definition of `r` differs slightly depending on the function used, so it pays to check the help-file for each function. In summary:

- for `epi.sscompc()`, `r` is the number in the treatment group divided by the number in the control group
- for `epi.sscohortc()`, `r` is the number in the exposed group divided by the number in the unexposed group
- for `epi.sscc()`, `r` is the number in the control group divided by the number in the case group

10 Learning Activities

Activity 10.1

We are planning a study to measure the prevalence of a relatively rare condition (say approximately 5%) in children age 0-5 years in a remote community.

a) What type of study would need to be conducted?
b) Use the correct sample size table included in your notes to determine how many children would need to be enrolled for the confidence interval to be
    i. 2%
    ii. 4% around the prevalence?

What would the resulting prevalence estimates and 95% CIs be?

Activity 10.2

We are planning an experimental study to test the use of a new drug to alleviate the symptoms of the common cold compared to the use of Vitamin C. Participants will be randomised to receive the new experimental drug or to receive Vitamin C. How many participants will be required in each group (power = 80%, level of significance = 5%).

a) If the resolution of symptoms is 10% in the control group and 40% in the new treatment group?
b) How large will the sample size need to be if we decide to recruit two control participants to every intervention group participant?
c) If we decide to retain a 1:1 ratio of participants in the intervention and controls groups but the resolution of symptoms is 20% in the control group and 40% in the new treatment group?
d) How many participants would we need to recruit (calculated in c) if a pilot study shows that 15% of people find the new treatment unpalatable and therefore do not take it?

Activity 10.3

In a case-control study, we plan to recruit adult males who have been exposed to fumes from an industrial stack near their home and a sample of population controls in whom we expect that 20% may also have been exposed to similar fumes through their place of residence or their work. We want to show that an odds ratio of 2.5 for having respiratory symptoms associated with exposure to fumes is statistically significant.

a) What statistical test will be needed to measure the association between exposure and outcome?
b) How large will the sample size need to be to show that the OR of 2.5 is statistically significant at $P < 0.05$ with 90% power if we want to recruit equal number of cases and controls?
c) What would be the required sample size (calculated in b) if the minimum detectable OR were 1.5?
d) If there are problems recruiting cases to detect an OR of 1.5 (as calculated in c), what would the sample size need to be if the ratio of cases to controls was increased to 1:3?

Activity 10.4

In the above study to measure the effects of exposure to fumes from an industrial stack, we also want to know if the stack has an effect on lung function which can be measured as forced vital capacity in 1 minute (FEV1). This measurement is normally distributed in the population.

a) If the research question is changed to wanting to show that the mean FEV1 in the exposed group is lower than the mean FEV1 in the control group what statistical test will now be required?
b) Population statistics show that the mean FEV1 and its SD in the general population for males are 4.40 L (SD=1.25) which can be expected in the control group.

We expect that the mean FEV1 in the cases may be 4.0 L. How many participants will be needed to show that this mean value is significantly different from the control group with $P < 0.05$ with an 80% power if we want to recruit equal number in each group?

c) How much larger will the sample size need to be if the mean FEV1 in the cases is 4.20 L?

References

Bland, Martin. 2015. *An Introduction to Medical Statistics*. 4th ed. Oxford, New York: Oxford University Press.

Freiman, Jennie A., Thomas C. Chalmers, Harry Smith, and Roy R. Kuebler. 1978. "The Importance of Beta, the Type II Error and Sample Size in the Design and Interpretation of the Randomized Control Trial." *New England Journal of Medicine* 299 (13): 690–94. https://doi.org/10.1056/NEJM197809282991304.

Kirkwood, Betty, and Jonathan Sterne. 2001. *Essentials of Medical Statistics*. 2nd ed. Malden, Mass: Wiley-Blackwell.

Woodward, Mark. 2013. *Epidemiology: Study Design and Data Analysis*. 3rd ed. Chapman and Hall/CRC.