StatCalc, 2x2 tables, sample size calculation, chi square for trend

Introduction

StatCalc is an epidemiologic calculator that produces statistics from summary data. Three types of calculations are offered:

- Statistics from 2-by-2 to 2-by-9 tables. Both single and stratified 2-by-2 tables can be analyzed to produce odds ratios and risk ratios (relative risks) with confidence limits, chi square tests, Fisher exact tests, Mantel-Haenszel summary odds ratios and chi squares, and associated p-values.
- Sample size & power calculations include Population Survey, Cohort or Cross-Sectional, and Unmatched Case-Control Study.
- Chi-square for trend by the Mantel extension of the Mantel-Haenszel summary odds ratio and chi square. This tests for the presence of a trend in dose response or other case control studies where a series of increasing or decreasing exposures is being studied.

How to use StatCalc

1. From the Epi Info 7 main menu, select StatCalc.
2. Select one of the following calculations using the up and down arrow keys or click: **Sample Size & Power, Chi Square for Trend, Tables (2X2, 2Xn), Poisson (rare event vs. std.) or Binomial (proportion vs. std.)**.
3. If you select the Sample size and power calculation, the following calculation options appear: Population Survey, Cohort or Cross-Sectional, Unmatched Case-Control.
    - Use the tab key or return key to move around the different cells available for data entry
4. Enter data for each calculation type. Calculations are performed when you enter data.
5. To modify values already entered, use the **Tab** key or click on the cell and enter the new information.
    - Right click and select the Print option in order to print the results.
    - Right click and select the Save as Image option to save a screen shot of your results as an image file.

Analysis of single 2-by-2 tables

A study of bladder cancers cases used randomly selected controls to explore the history of artificial sweetener use in the two groups. Is there an association between sweetener use and bladder cancer?

	Bladder Cancer	
Sweetener	**Yes**	**No**
Ever Used	1293	2455
Never Used	1707	3321

From the StatCalc application menu, select **Tables (2 x 2 x n)**. Enter the data.

Analysis of stratified tables

What is the relationship between alcohol consumption and myocardial infarction (MI)? The following case-control study shows an apparent association between alcohol consumption and MI with an odds ratio of 2.26.

	MI	
Alcohol	**Yes**	**No**
Yes	71	52
No	29	48

Smoking is known to be associated with both MI and with alcohol consumption. Stratifying the data by smoking status creates two tables, one for smokers and one for non-smokers

Nonsmokers		
	MI	
Alcohol	**Yes**	**No**
Yes	8	16
No	22	44

Smokers		
	MI	
Alcohol	**Yes**	**No**
	63	36
	7	4

1. From the StatCalc application main page, select Tables (2 x 2 x n)
2. Enter the data from the above Nonsmokers sample table in the first Strata tab (Strata 1).
3. Click on Strata 2. Enter the data from the above Smokers sample table.

The odds ratio for each table is 1.0 and the Mantel summary odds ratio is 1.0. The crude odds ratio and the Mantel summary odds ratio are quite different (2.26 and 1.0), leading to the conclusion that smoking was a confounding factor and that there appears to be no association (odds ratio= 1.0) between alcohol and MI.

- Note that the odds ratios in the two strata are the same (1.0); there is no interaction or effect modification between smoking and alcohol.

- In other words, the effect of alcohol on MI is the same for smokers and nonsmokers. When the effect varies in the different strata (the odds ratios are different), interaction or effect modification is said to be present.

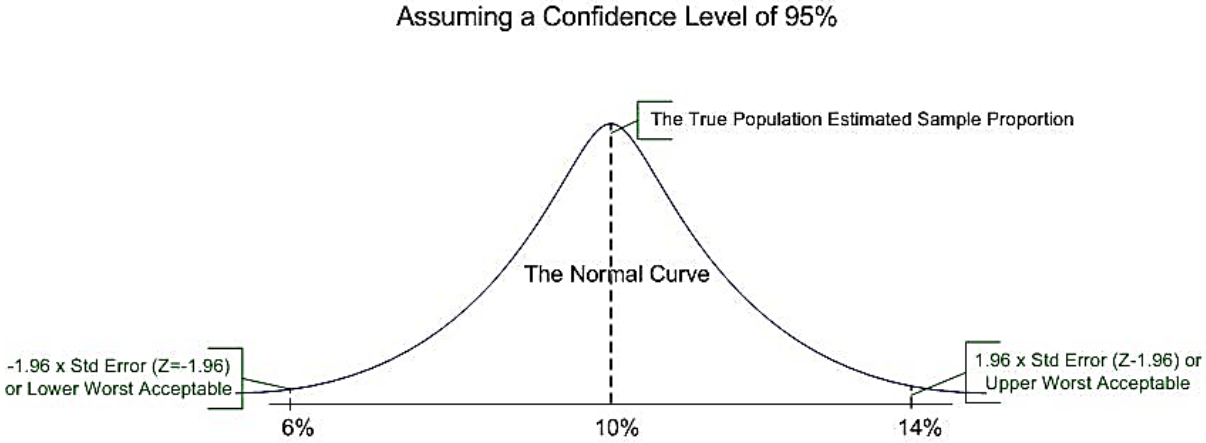
Sample size calculations

Population Survey or Description Study

Suppose you want to investigate whether the true prevalence of HIV in a population is 10%. A 95% confidence that the true proportion in the entire population will fall within the confidence interval calculated from the sample is desired.

1. From the StatCalc application main page, select **Sample size & power > Population survey**.
2. Enter the Population Size of 5,000.
3. Enter the Expected Frequency of 10%.
4. Enter the Confidence Limits as 4%, i.e. the true mean of 10% will fall between 10% ± 4% (6% and 14%) in 95% of the proportions obtained from repeated samples.

   Pictorially, the equation works with the following values, if confidence is 95%.



- Cluster sampling can be used to make the study easier or cheaper to perform.
- The effect on sample size depends on the number of clusters and the variance between and within clusters expressed as a constant called "Design Effect", often between 1.0 and 2.0.
- The sample sizes for simple random samples are multiplied by the Design Effect to obtain the sample size for the clustered sample.
- Sample sizes must be adjusted upward to allow for nonresponse and other factors that decrease the yield of usable responses. If you expect 10% nonparticipation and 5% non-response to any given question, then the sample size must be multiplied by 100% /85% or 1.18.

Cohort or cross-sectional

Example: A Cross-Sectional Study

In a foodborne outbreak, there were 2397 persons at a banquet that was followed by illness. You wish to conduct a cross-sectional study to test for an association between illness and food consumption. With the cook's help, you estimate that about 40% of the crowd had the suspect potato salad and the rest had the green garden salad. You guesstimate that the ratio of unexposed to exposed is 60:40. As a rough guess we choose 5% as the case rate among the unexposed.

What number of persons will be needed to detect and odds ratio of 2.0 or greater?

1. From the StatCalc application main page, select **Sample size & power > Cohort or cross-sectional**.
2. Select the two-sided confidence level as **95**%.
3. Enter the Power as **80**%.
4. Enter the Ratio (Unexposed : Exposed) as **1.5** (i.e. 60/40)
5. Enter the % outcome in unexposed group as **5**%.
6. Enter the Odds ratio of **2.0**. The Risk ratio and % outcome in exposed group fields will automatically populate.

Example: A Cohort (Exposure-based) Study

In a vaccine trial, subjects are to be randomly allocated to receive a vaccine or placebo. The normal rate of disease in the unimmunized is 10%. What sample size will be required to show a two-fold reduction in risk (rate of 5%) associated with immunization? Use a confidence level of

95%, power of 80%, and the rates mentioned, with a 1:1 ratio of immunized and unimmunized subjects.

1. From the StatCalc application main page, select **Sample size & power > Cohort or cross-sectional**.
2. Select the two-sided confidence level as **95**%.
3. Enter the Power as **80**%.
4. Enter the Ratio (Unexposed : Exposed) as **1**
5. Enter the % outcome in unexposed group as **10**%.
6. Enter the % outcome in exposed group fields as **5**%.

Unmatched Case Control Study

Twelve infants with staphylococcal infections in a hospital outbreak are to be compared with uninfected infants in the same nursery in a search for risk factors. We do not know the expected proportion of infants in either group having risk factors such as exposure to a particular nurse or kind of skin treatment, but to study the worst case, we choose 50% as the frequency in the controls and calculate sample size for a variety of odds ratios.

1. From the StatCalc application main page, select **Sample size & power > Unmatched case-control**.
2. Select the two-sided confidence level as **95**%.
3. Enter the Power as **80**%.
4. Enter the Ratio of controls to cases as **6**
5. Enter the Percent of controls exposed as **50**%.
6. Enter the Odds ratio as **10**.
    - What happens when you vary the ratio of controls to cases and the odds ratio?
    - Only a very striking odds ratio can be detected with such a small number of cases, or that failure to demonstrate significant results does not rule out the presence of risk factors.

Chi-square for trend

Often used for dose-response studies or to test trends with any ordered variable. This example uses data from a case-control study of women with and without myocardial infarction (MI) at three different levels of cigarette smoking, measured as cigarettes per day.

We assign a score to categorize the number of cigarettes smoked as an ordered variable. The outcome of myocardial infarction is dichotomous (MI vs. control)

Cig/day	Score	Myocardial Infarction	Control
None	0	38	751
1-24	1	82	617
>24	2	126	316

From the StatCalc application main page, select **Chi Square for Trend**. Enter the data.