EpiBasic

Simple statistical tools for epidemiological analysis

Svend Juul and Morten Frydenberg, September 2016

Introduction

EpiBasic is an easy-to-use tool for statistical analysis of tabular information, as you find it in epidemiological and clinical papers. It is developed as a companion to the Danish textbook "Epidemiologi og evidens" (Juul 2012).

EpiBasic requires that Microsoft Excel is installed on your computer. You can write in the <mark>yellow</mark> cells only; all other cells are write-protected to prevent you from overwriting the formula. When you open EpiBasic for the first time, you will see a textbook example filled in at each page, for illustration. To proceed, just delete all entries in the yellow fields, and fill in your own data.

Many analyses use log transformations; log transformed values are shown in *italic* typeface.

EpiBasic may be expanded or improved, and you should take a look now and then at http://www.ph.au.dk/uddannelse/software. We have taken great pains to ensure correctness by comparing to textbook examples and to the results produced by other software (Stata). If you have suggestions for improvement or believe you discovered an error, don't hesitate to mail one of us at sj@ph.au.dk or morten@ph.au.dk.

Start page

This page is just an index to the analysis pages; click a link to jump to the analysis page desired. Each analysis page has a link sending you back to the start page.

The name of most analysis pages starts with a prefix; the meaning is:
D: Descriptive statistics
C: Comparative statistics
Str: Stratified analysis
MH: Stratified analysis a.m. Mantel-Haenszel.

See a copy of the start page at the end of this document.

D means: Descriptive statistics for normally distributed data

From mean, Standard Deviation (SD) and number of observations, calculate Standard Error (SE), 95% prediction interval, and 95% confidence interval for the mean, provided data represent a normal distribution. The sheet also displays t-based intervals.

See Juul (2012; §2.3; §A.3), Bland (2000; §8.1-8.3; §10.1), Kirkwood & Sterne (2003; §6).

D prop: Descriptive statistics for proportions and odds

Calculate the proportion ($p = a/n$) from a events among n observations, using the binomial distribution. Also calculate the odds = $p/(1–p)$. 95% confidence interval for proportions and odds: see Juul (2012; §2.4). Note that the calculations for odds take place after log transformation.

The exact confidence interval for a proportion is calculated from the F-distribution; see Armitage et al (2002; §4.4). This interval is transformed directly to an exact confidence interval for odds.

D rates: Descriptive statistics for counts and rates

Calculate a rate ($r = a/t$) from a events among a number of subjects with a total time-at-risk of t, using the Poisson distribution. Calculate the confidence interval after log-transformation; see Juul (2012; §2.5) , Kirkwood & Sterne (2003; §22). To calculate the confidence interval for a count, just insert a time-at-risk of 1.

The exact confidence interval for a rate is calculated from the χ^2-distribution; see Armitage et al (2002; §5.2).

C means: Difference between means (z-test and t-test)

From mean, Standard Deviation and number of observations for two samples, examine the null hypothesis of no difference between means with:

- z-test, see Kirkwood & Sterne (2003; §7.3), Bland (2000; §9.7), Juul (2012; §4.1).
- Test for equal SDs by a variance ratio test; see Bland (2000; §10.8).
- t-test assuming equal SDs; see Bland (2000; §10.3), Kirkwood & Sterne (2003; §7.4).
- t-test assuming unequal SDs; see Dupont (2002; §1.4.13)

C risk: Risk ratio, risk difference, odds ratio

This page lets you estimate risk ratio (relative risk, RR), risk difference (RD) and odds ratio (OR) from a 2×2 table. SE and confidence interval for RR and OR are calculated after log transformation. See Juul (2012; §4.3; §5.3), Kirkwood & Sterne (2003; §16), Bland (2000; §9.8).

The page also displays exact confidence intervals for proportions (but not for the measures of association); see above on proportions and odds (descriptive statistics).

C rates: Incidence rate ratio (IRR) and difference (IRD)

Compare two incidence rates by their ratio and difference. SE and confidence interval for IRR are calculated after log transformation. See Juul (2012; §4.4), Kirkwood & Sterne (2003; §23).

The page also displays exact confidence intervals for rates (but not for the measures of association); see above on rates (descriptive statistics).

Str any: Stratified analysis for any measure of association

You can perform a stratified analysis with any measure of association from an estimate and its standard error, e.g. for a meta-analysis (there are more advanced tools for this). See Juul (2012; §8.4). The principle is:

The contribution from each stratum (the weight, w) is proportional to its inverse variance:
$$w = 1/SE^2.$$
The weighted mean of the estimates (E) is: $E_{weighted} = \Sigma(w\ E)/\Sigma w$.
The Standard error of the weighted mean is $SE_{weighted} = \sqrt{(1/\Sigma w)}$.

This page also compares the first two strata to assess effect modification.

Str RR: Stratified analysis (relative risk)

Perform stratified analysis, using the inverse variance $(1/SE^2)$ as weight. See Juul (2012; §8.4).

Str OR: Stratified analysis (odds ratio)

Perform stratified analysis, using the inverse variance $(1/SE^2)$ as weight. See Juul (2012; §8.4).

A Mantel-Haenszel analysis (see MH OR) is more robust with few observations.

Str IRR: Stratified analysis (incidence rate ratio)

Perform stratified analysis, using the inverse variance $(1/SE^2)$ as weight. See Juul (2012; §8.4).

MH OR: Stratified analysis (odds ratio, Mantel-Haenszel)

Perform stratified analysis, using the Mantel-Haenszel method. See Juul (2012; §A.6), Kirkwood & Sterne (2003; §18). For details, see Breslow and Day (1980, §4.4).

Chisq: Chi-square test

Calculate the χ^2 statistic for contingency tables with up to 10 rows and 5 columns; see Juul (2012; §A.4), Kirkwood & Sterne (2003; §17), Bland (2000; §13.1).

If the rows and columns represent ordinal (rank) scales – or if one of them is dichotomous – the trend test with one degree of freedom is more sensitive than the standard χ^2 test. You may change the trend scores used for the analysis. See Juul (2012; §A.5), Kirkwood & Sterne (2003; §17.5), Bland (2000; §13.8).

Tables: Statistical tables

This page lets you calculate P-values for the normal (z) distribution, t distribution and χ^2 distribution. One-sided P-values are not displayed.

You may also find, for example, the t-value with 20 degrees of freedom, corresponding to a P-value of 5% (or a 95% confidence interval).

References

Armitage P, Berry G, Matthews JNS. Statistical methods in medical research, 4th ed. Oxford: Blackwell Science, 2002

Bland M. An introduction to medical statistics. Oxford: Oxford University Press, 2000.

Breslow BE, Day NE. Statistical methods in cancer research, vol. 1. Lyon: International Agency for Research on Cancer, 1980.

Dupont WD. Statistical modeling for medical researchers. Cambridge: Cambridge University Press, 2002.

Juul S. Epidemiologi og evidens. 2nd ed. Copenhagen: Munksgaard, 2012.

Kirkwood BR, Sterne JAC. Essential medical statistics. Malden, MA: Blackwell Science, 2003.

EpiBasic version 3.0
Svend Juul og Morten Frydenberg 2013

The sheets are write-protected, meaning that you cannot by mistake delete formula etc.
You are allowed to write in the [yellow] cells only.

Each analysis page name has a prefix; the meaning is:
D: Descriptive statistics
C: Comparative statistics
Str: Stratified analysis
MH: Stratified analysis a.m. Mantel-Haenszel

	Descriptive statistics	Simple comparison	Stratified analysis	Stratified (Mantel-Haenszel)
Means	D means	C means	Str any	
Proportions	D prop	C risk	Str RR	
Odds	D prop	C risk	Str OR	MH OR
Rates	D rates	C rates	Str IRR	

Weighted average	Weighted

Chi-squared test	Chisq

Statistical tables	Tables

In many sheets you find cells with *italic* typeface.
They contain intermediate calculations on a log scale.

EpiBasic may be downloaded from: http://ph.au.dk/uddannelse/software
You download two files:
EpiBasic.xlsx   (this workbook)
EpiBasic.pdf   (help and documentation)