EpiBasic

Simple statistical tools for epidemiological analyses

Developed by Svend Juul and Morten Frydenberg
Maintained by Stefan Hansen

Last update: August 2019

Introduction

EpiBasic is an easy-to-use tool for statistical analysis of tabular information, as you find it in epidemiological and clinical papers. It has been developed as a companion to the Danish textbook "Epidemiologi og evidens" (Juul et al 2017).

EpiBasic requires that Microsoft Excel is installed on your computer. You can write in the yellow cells only; all other cells are write-protected to prevent you from overwriting the formulas. When you open EpiBasic for the first time, you will see a textbook example filled in at each page, for illustrational purposes. 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 so make sure you have the latest version, which is available for download at:

https://ph.medarbejdere.au.dk/undervisning/software

If you have suggestions for improvement or believe you discovered an error, do not hesitate to mail me at stefanh@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.

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.

D Means: Descriptive statistics for means and standard deviations

From the mean, standard deviation (SD) and number of observations (n), calculate standard error (SE), 95%-prediction interval, 95%-confidence interval and simple tests for the mean based on the normal distribution. The confidence interval and tests are approximate and apply to any distribution with a sufficient number of observations. The prediction interval requires, however, that data follow (approximately) a normal distribution.

The sheet also displays exact confidence intervals and simple tests for the mean based on a t-distribution and exact confidence intervals and tests for the standard deviation based on a χ^2-distribution. All of these require data to follow (approximately) a normal distribution.

See Juul et al (2017; §2.3; §A.3), Bland (2000; §8.1-8.3; §10.1), Kirkwood & Sterne (2003; §6).

D Props: Descriptive statistics for proportions and odds

Calculate the proportion ($p = a/n$) and the odds = $p/(1–p)$ from a events among n observations, using the binomial distribution. The sheet displays 95%-confidence intervals for the proportion and odds (see Juul et al (2017; §2.4)) and allows you to make simple tests based on a normal distribution. 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 et al (2017; §2.5) , Kirkwood & Sterne (2003; §22). The simple test is based on a normal distribution. 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 two means

From mean, standard deviation and number of observations for two samples, examine the null hypothesis of no difference between the two means by calculating an approximate 95%-confidence interval of the difference and perform an approximate z-test, see Kirkwood & Sterne (2003; §7.3), Bland (2000; §9.7), Juul et al (2017; §3.1). This apply to any distribution with a sufficient number of observations.

The following exact tests are also performed and require that data follow (approximately) a normal distribution:

- Test for equal SDs by a variance ratio test; see Bland (2000; §10.8).
- A t-test for no difference in means assuming equal SDs; see Bland (2000; §10.3), Kirkwood & Sterne (2003; §7.4).
- A t-test for no difference in means assuming unequal SDs; see Dupont (2002; §1.4.13)

C Props: Risk/proportion difference, relative risk/proportion and odds ratio

This page lets you estimate risk/proportion difference (RD), relative risk/proportion (RR) and odds ratio (OR) from a 2×2 table. SE and confidence interval for RR and OR are calculated after log transformation. See Juul et al (2017; §3.3; §54.3), 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 and difference

Compare two incidence rates by their ratio and difference. SE and confidence interval for IRR are calculated after log transformation. See Juul wt al (2012; §3.4), 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).

Tables: Chi-square tests

Calculate the χ^2 statistic for contingency tables with up to 10 rows and 5 columns; see Juul et al(2017; §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 et al(2017; §A.5), Kirkwood & Sterne (2003; §17.5), Bland (2000; §13.8).

For the tests to be valid, all expected cell counts must be greater than or equal to 5. In the bottom of the page, the number of expected cell counts less than 5 is given.

Str MD: Stratified analysis for a mean difference

Perform a stratified analysis for a mean difference using the inverse variance ($1/SE^2$) as weight.

This page also compares any two strata to assess effect modification.

Str RD/RR/OR: Stratified analysis for a risk difference/relative risk/odds ratio

Perform stratified analysis, using the inverse variance ($1/SE^2$) as weight based on 2×2 tables. See Juul et al(2017; §7.3).

For an odds ratio, a Mantel-Haenszel analysis - see Str OR (MH) - is more robust with few observations.

This page also compares any two strata to assess effect modification.

Str IRR: Stratified analysis (incidence rate ratio)

Perform stratified analysis, using the inverse variance ($1/SE^2$) as weight. See Juul et al(2017; §7.3).

This page also compares any two strata to assess effect modification.

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

Perform stratified analysis, using the Mantel-Haenszel method. See Juul et al (2017; §A.6), Kirkwood & Sterne (2003; §18). For details, see Breslow and Day (1980, §4.4).

Str Any 1: Stratified analysis for any measure/parameter

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 et al (2017; §7.3). 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: $E_{weighted} = \Sigma(w \cdot E)/\Sigma w$.
The Standard error of the weighted mean is $SE_{weighted} = \sqrt{(1/\Sigma w)}$.

This page also compares any two strata to assess effect modification.

Str Any 2: Stratified analysis for any measure with user-specified weights

Calculate the weighted average of up to 26 estimates with user-specified weights.

P-values: From test statistic to p-value.

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).

Sample Size: Determine the necessary sample size to obtain a given power.

Given a significance level and wanted power, it will calculate the necessary sample size to realize these when comparing either a dichotomous outcome or continuous outcome in two groups.

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. et al Epidemiologi og evidens. 3nd ed. Copenhagen: Munksgaard, 2017.

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