

Package ‘epibasix’

January 2, 2012

Version 1.1

Date 2009-05-13

Title Elementary Epidemiological Functions for a Graduate Epidemiology\{ }Biostatistics Course

Author Michael A Rotondi <mrotondi@uwo.ca>

Maintainer Michael A Rotondi <mrotondi@uwo.ca>

Depends R (>= 2.01)

Description This package contains elementary tools for analysis of common epidemiological problems, ranging from sample size estimation, through 2x2 contingency table analysis and basic measures of agreement (kappa, sensitivity/specificity). Appropriate print and summary statements are also written to facilitate interpretation wherever possible. This package is a work in progress, so any comments or suggestions would be appreciated. Source code is commented throughout to facilitate modification. The target audience includes graduate students in various epi/biostatistics courses.

License GPL (>= 2)

Repository CRAN

Date/Publication 2009-05-18 18:43:05

R topics documented:

corXY	2
diffDetect	3
epi2x2	4
epiKappa	6
epiTTest	7
mcNemar	9
n4means	10
n4props	12
sensSpec	13
univar	15

Index**17**

corXY

*Correlation of Two Vectors***Description**

This function displays the simple correlation of two vectors of equal length, as well as providing confidence limits and hypothesis tests.

Usage

```
corXY(X, Y, alpha=0.05, rho0 = 0, HA="not.equal", digits=3)
```

Arguments

X	A Vector of the same length as Y
Y	A Vector of the same length as X, This function requires the input of Vectors
alpha	The Type I error rate for Hypothesis Tests and Confidence Intervals
rho0	The Null Hypothesis for Hypothesis Tests
HA	The alternative hypothesis can be one of "less.than", "greater.than", or "not.equal"
digits	The number of digits to round results

Details

This function provides the required information, such as the Pearson correlation Hypothesis Tests and confidence intervals, while providing suitable detail in the and print statements for epidemiologists to understand the information at hand.

Value

rho	The Sample Pearson Correlation, as calculated in the cor function.
n	The sample size.
Test	The Test Statistic for the desired hypothesis test based on Fisher's Transformation.
p.Value	The p-value for the Hypothesis Test.
CIL	The lower bound of the constructed confidence interval for ρ , again based on Fisher's Z Transformation.
CIU	The Upper bound of the constructed confidence interval for ρ , again based on Fisher's Z Transformation.
alpha	The desired Type I Error Rate
rho0	The Null Hypothesis
HA	The supplied Alternative Hypothesis

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Casella G and Berger RL. Statistical Inference (2nd Ed.) Duxbury: New York, 2002. Koepsell TD and Weiss NS. Epidemiologic Methods. Oxford University Press: New York, 2003.

Examples

```
## Not run: Suppose we want to test whether two randomly generated normal vectors are uncorrelated
x <- rnorm(100);
y <- rnorm(100);
corXY(x,y);
```

diffDetect

Mean Difference Detetion Tool

Description

Provides Minimum Detectable Difference in Means Between Two Populations for fixed values of sigma and n. Useful for experimental design for randomized trials.

Usage

```
diffDetect(N,sigma,alpha=0.05, power=0.8, two.tailed=TRUE)
```

Arguments

N	A Vector (or single value) of fixed sample sizes.
sigma	A Vector (or single value) of fixed standard deviations sizes.
alpha	The desired Type I Error Rate
power	The desired level of power, recall power = 1 - Type II Error.
two.tailed	Logical, If TRUE calculations are based on a two-tailed Type I error, if FALSE, a one-sided calculation is performed.

Details

This function can be used as a tool for sensitivity analysis on the choice of population standard deviation. As is often the case, the sample size is fixed by practical considerations, such as cost or difficulty recruiting subjects. This simple tool may help determine whether it is worth performing an experiment that can only detect a given calculated difference between means.

Value

delta	A Matrix of minimum detectable differences for fixed values of n and sigma
N	A Vector (or single value) of specified sample sizes.
sigma	A Vector (or single value) of specified standard deviations sizes.
alpha	The desired Type I Error Rate
power	The desired level of power, recall power = 1 - Type II Error.
two.tailed	Logical, If TRUE calculations are based on a two-tailed Type I error, if FALSE, a one-sided calculation is performed.

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Matthews JNS. Introduction to Randomized Controlled Clinical Trials (2nd Ed.) Chapman & Hall: New York, 2006.

Examples

```
## Not run: Suppose, for financial considerations we can only enroll 100 people in a blood
pressure medication trial. What is the minimum difference we can detect between means if
sigma = 1, 5 or 10 mmHg, at standard levels?
## End(Not run)
n <- 100;
sigma <- c(1, 5, 10);
diffDetect(n,sigma);
```

epi2x2

Epidemiological 2x2 Contingency Table Analysis Tool

Description

This function analyzes 2x2 tables assuming either a case-control or cohort study. Information such as Pearson's chi-squared test, the odds ratio, risk difference and relative risk are computed, as well as confidence intervals.

Usage

```
epi2x2(X,alpha=0.05, digits=3)
```

Arguments

X	A 2x2 matrix in standard epidemiological format, that is, column one represents outcome present, column two outcome absent, while row one represents risk present and row two represents risk absent. This is crucial for correct computation of odds ratio and parameters.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Details

This function is similar to PROC FREQ in SAS, as it provides the comprehensive analysis of a 2x2 contingency table. Again, I must stress that the table must be entered in the appropriate format, or unsuitable estimates will result. In a case control study, cases should be entered as column one and controls as column two.

Value

X	The original input matrix.
Sy	Value for Pearson's Chi-squared statistic (with continuity correction).
Sy.p.value	P-value for the hypothesis test of no association.
Fisher.p.value	P-value for the hypothesis test of no association. (Using Fisher's Exact Test)
OR	Point Estimate of the odds ratio.
OR.CIL	Lower Confidence Limit for the odds ratio.
OR.CIU	Upper Confidence Limit for the odds ratio.
p1Co	Row One Risk (Cohort Study)
p2Co	Row Two Risk (Cohort Study)
rdCo	Risk difference (Cohort Study). That is $p1Co - p2Co$.
rdCo.CIL	Lower Confidence Limit for Risk Difference in a cohort study.
rdCo.CIU	Upper Confidence Limit for Risk Difference in a cohort study.
RR	Relative Risk (Cohort Study)
RR.CIL	Lower Confidence Limit for Relative Risk in a cohort study.
RR.CIU	Upper Confidence Limit for Relative Risk in a cohort study.
p1CC	Column One Risk (Case-Control Study)
p2CC	Column Two Risk (Case-Control Study)
rdCC	Risk difference (Case-Control Study). That is $p1CC - p2CC$.
rdCC.CIL	Lower Confidence Limit for Risk Difference in a case-control study.
rdCC.CIU	Upper Confidence Limit for Risk Difference in a case-control study.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Szklo M and Nieto FJ. Epidemiology: Beyond the Basics, Jones and Bartlett: Boston, 2007.

See Also

[mcNemar](#)

Examples

```
data <- cbind(c(100, 225), c(58, 45));
summary(epi2x2(data));
```

epiKappa	<i>Computation of the Kappa Statistic for Agreement Between Two Raters</i>
----------	--

Description

Computes the Kappa Statistic for agreement between Two Raters, performs Hypothesis tests and calculates Confidence Intervals.

Usage

```
epiKappa(C, alpha=0.05, k0=0.4, digits=3)
```

Arguments

C	An nxn classification matrix or matrix of proportions.
k0	The Null hypothesis, $kappa_0 = k_0$
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Details

The Kappa statistic is used to measure agreement between two raters. For simplicity, consider the case where each rater can classify an object as Type I, or Type II. Then, the diagonal elements of a 2x2 matrix are the agreeing elements, that is where both raters classify an object as Type I or Type II. The discordant observations are on the off-diagonal. Note that the alternative hypothesis is always greater then, as we are interested in whether kappa exceeds a certain threshold, such as 0.4, for Fair agreement.

Value

kappa	The computation of the kappa statistic.
seh	The standard error computed under H0
seC	The standard error as computed for Confidence Intervals
CIL	Lower Confidence Limit for κ
CIU	Upper Confidence Limit for κ
Z	Hypothesis Test Statistic, $\kappa = K0 = K0$ vs. $\kappa > K0$
p.value	P-Value for hypothesis test
Data	Returns the original matrix of agreement.
k0	The Null hypothesis, kappa = k0
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

- Szklo M and Nieto FJ. Epidemiology: Beyond the Basics, Jones and Bartlett: Boston, 2007.
- Fleiss J. Statistical Methods for Rates and Proportions, 2nd ed. New York: John Wiley and Sons; 1981.

See Also

[sensSpec](#)

Examples

```
X <- cbind(c(28,5), c(4,61));
summary(epiKappa(X, alpha=0.05, k0 = 0.6));
```

epiTTest

Epidemiological T-Test Function

Description

This function computes the standard two sample T-Test, as well as performing hypothesis tests and computing confidence intervals for the equality of both population means.

Usage

```
epiTTest(X,Y, alpha=0.05, pooled=FALSE, digits=3)
```

Arguments

X	A vector of observed values of a continuous random variable.
Y	A vector of observed values of a continuous random variable.
alpha	The desired Type I Error Rate for Confidence Intervals
pooled	Logical: If TRUE, a pooled estimate of the variance is used. That is, the variance is assumed to be equal in both groups. If FALSE, the Satterthwaite estimate of the variance is used.
digits	Number of Digits to round calculations

Details

This function performs the simple two-sample T-Test, while providing detailed information regarding the analysis and summary information for both groups. Note that this function requires the input of two vectors, so if the data is stored in a matrix, it must be separated into two distinct vectors, X and Y.

Value

nx	The number of observations in X.
ny	The number of observations in Y.
mean.x	The sample mean of X.
mean.y	The sample mean of Y.
s.x	The standard deviation of X.
s.y	The standard deviation of Y.
d	The difference between sample means, that is, mean.x - mean.y.
s2p	The pooled variance, when applicable.
df	The degrees of freedom for the test.
TStat	The test statistic for the null hypothesis $\mu_X - \mu_Y = 0$.
p.value	The P-value for the test statistic for $\mu_X - \mu_Y = 0$.
CIL	The lower bound of the constructed confidence interval for $\mu_X - \mu_Y = 0$.
CIU	The upper bound of the constructed confidence interval for $\mu_X - \mu_Y = 0$.
pooled	Logical: as above for assuming variances are equal.
alpha	The desired Type I Error Rate for Confidence Intervals

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

- Casella G and Berger RL. Statistical Inference (2nd Ed.) Duxbury: New York, 2002.
 Szklo M and Nieto FJ. Epidemiology: Beyond the Basics, Jones and Bartlett: Boston, 2007.

Examples

```
X <- rnorm(100,10,1);
Y <- rnorm(100);
summary(epiTTTest(X,Y, pooled = FALSE));
```

mcNemar

Pair-Matched Analysis Tool

Description

This function performs elementary pair-matched analysis using McNemar's test and computing risk differences.

Usage

```
mcNemar(X, alpha= 0.05, force=FALSE, digits=3)
```

Arguments

X	A 2x2 matrix, with disease status (Yes/No) for the exposed individual in the columns and disease status (Yes/No) for the control individuals in the rows. Note that for a matched-pair analysis, each entry corresponds to a pair of subjects.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
force	Logical: McNemar's test is typically valid when the number of discordant pairs exceeds 30. The function may be forced to work, without regards to this concern with FORCE=TRUE.
digits	Number of Digits to round calculations

Details

McNemar's OR is computed as b/c . While standard errors are computed using a transformation. The risk difference is computed as $(b-c)/n$. Note that this technique can be used for cohort studies as well as matched trials.

Value

X	The original input matrix.
ORMc	McNemar's Odds Ratio
ORMC.CIL	Lower Confidence Limit for McNemar's OR
ORMC.CIU	Upper Confidence Limit for McNemar's OR
rd	Point Estimate of the risk difference
rd.CIL	Lower Confidence Limit for the risk difference
rd.CIU	Upper Confidence Limit for the risk difference

XMc	Value for McNemar's Chi-squared statistic
.	
XMc.p.Value	P-value for the hypothesis test of no association.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Szklo M and Nieto FJ. Epidemiology: Beyond the Basics, Jones and Bartlett: Boston, 2007.

See Also

[epi2x2](#)

Examples

```
## Not run: Data for matched-cohort study, comparing smokers to non-smokers for the presence
of lung cancer.
## End(Not run)
X <- cbind(c(15,5), c(19,61));
summary(mcnemar(X, alpha=0.05, force=TRUE));
```

n4means	<i>Number of Subjects Required for a Randomized Trial with a Continuous Outcome</i>
---------	---

Description

This function provides detailed sample size estimation information to determine the number of subjects that must be enrolled in a randomized trial with a continuous outcome.

Usage

```
n4means(delta, sigma, alpha=0.05, power=0.8, AR=1, two.tailed=TRUE, digits=3)
```

Arguments

delta	The minimum detectable difference between population means.
sigma	The standard error of the outcome.
AR	The Allocation Ratio: One implies an equal number of subjects per treatment and control group (maximum efficiency), > 1, implies more subjects will be enrolled in the control group (e.g. in the case of costly intervention), < 1 implies more in the treatment group (rarely used).

alpha	The desired Type I Error Rate
power	The desired level of power, recall power = 1 - Type II Error.
two.tailed	Logical, If TRUE calculations are based on a two-tailed Type I error, if FALSE, a one-sided calculation is performed.
digits	Number of Digits to round calculations

Details

This function provides detailed information, similar to PROC POWER in SAS, but with less functionality and more concise output. It is used for sample size estimation in a randomized trial where the outcome is continuous, such as blood pressure, or weight.

Value

nE	The minimum number of subjects required in the Experimental group.
nC	The minimum number of subjects required in the Control group.
delta	The minimum detectable difference between population means.
sigma	The standard error of the outcome.
alpha	The desired Type I Error Rate
power	The desired level of power, recall power = 1 - Type II Error.
AR	The Allocation Ratio

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Matthews JNS. Introduction to Randomized Controlled Clinical Trials (2nd Ed.) Chapman & Hall: New York, 2006.

See Also

[n4props](#)

Examples

```
## Not run: Suppose we wish to test whether a blood pressure medication reduces diastolic blood
## pressure by 10 mm Hg, at standard significance and power, assume the standard deviation is 10 mm Hg.
## End(Not run)
n4means(delta=10, sigma=10, alpha=0.05, power=0.80);
```

n4props	<i>Number of Subjects Required for a Randomized Trial with Binary Outcomes</i>
---------	--

Description

This function provides detailed sample size estimation information to determine the number of subjects that must be enrolled in a randomized trial with a binary outcome.

Usage

```
n4props(pe, pc, alpha=0.05, power = 0.80, AR=1, two.tailed=TRUE, digits=3)
```

Arguments

pe	The anticipated proportion of individuals in the experimental group with the outcome.
pc	The anticipated proportion of individuals in the control group with the outcome.
AR	The Allocation Ratio: One implies an equal number of subjects per treatment and control group (maximum efficiency), > 1, implies more subjects will be enrolled in the control group (e.g. in the case of costly intervention), < 1 implies more in the treatment group (rarely used).
alpha	The desired Type I Error Rate
power	The desired level of power, recall power = 1 - Type II Error.
two.tailed	Logical, If TRUE calculations are based on a two-tailed Type I error, if FALSE, a one-sided calculation is performed.
digits	Number of Digits to round calculations

Details

This function provides detailed information, similar to PROC POWER in SAS, but with less functionality and more concise output. It is used for sample size estimation in a randomized trial where the response is binary. A simple example may include whether an individual dies from a heart attack. In epidemiological terms, pe and pc can be thought of as the expected prevalence of the outcome in the experimental and control group.

Value

nE	The minimum number of subjects required in the Experimental group.
nC	The minimum number of subjects required in the Control group.
pe	The anticipated proportion of individuals in the experimental group with the outcome.
pc	The anticipated proportion of individuals in the control group with the outcome.
alpha	The desired Type I Error Rate
power	The desired level of power, recall power = 1 - Type II Error.
AR	The Allocation Ratio

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Matthews JNS. Introduction to Randomized Controlled Clinical Trials (2nd Ed.) Chapman & Hall: New York, 2006.

See Also

[n4means](#)

Examples

```
## Not run: Suppose a new drug is thought to reduce heart attack mortality from 0.10 to 0.03. Calculate the required
n4props(0.03, 0.10, AR=1, alpha=0.05, power=0.80);
```

sensSpec

Sensitivity and Specificity Analysis of a 2x2 Matrix

Description

This function provides detailed information regarding the comparison of two competing methods, for example self-report and gold-standard treatment through a sensitivity/specificity analysis.

Usage

```
sensSpec(X, alpha=0.05, CL=TRUE, digits=3)
```

Arguments

X	A 2x2 matrix, with Gold Standard Class A and B in the columns and Comparison Method A and B in the rows.
CL	Logical: If TRUE, Confidence Intervals are calculated and displayed in summary method.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Details

This function is designed to calculate Sensitivity, Specificity, Youden's J and Percent Agreement. These tools are used to assess the validity of a new instrument or self-report against the current gold standard. In general, self-report is less expensive, but may be subject to information bias. Computational formulae can be found in the reference.

Value

X	The original input matrix.
sens	The point estimate of sensitivity
spec	The point estimate of specificity
PA	The point estimate of Percent Agreement
YoudenJ	The point estimate of Youden's J
sens.s	The standard deviation of sensitivity
spec.s	The standard deviation of specificity
PA.s	The standard deviation of Percent Agreement
YoudenJ.s	The standard deviation of Youden's J
sens.CIL	The lower bound of the constructed confidence interval for true sensitivity.
sens.CIU	The upper bound of the constructed confidence interval for true sensitivity
spec.CIL	The lower bound of the constructed confidence interval for true specificity.
spec.CIU	The upper bound of the constructed confidence interval for true specificity.
PA.CIL	The lower bound of the constructed confidence interval for Percent Agreement.
PA.CIU	The upper bound of the constructed confidence interval for Percent Agreement.
YoudenJ.CIL	The lower bound of the constructed confidence interval for Youden's J.
YoudenJ.CIU	The upper bound of the constructed confidence interval for Youden's J.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
digits	Number of Digits to round calculations

Note

All confidence limits rely on simple asymptotic theory, as such, confidence limits may lie outside of [0,1]. A more accurate method is available in the `twoby2` function of the `Epi` package, which employs a logit transformation.

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Szklo M and Nieto FJ. *Epidemiology: Beyond the Basics*, Jones and Bartlett: Boston, 2007.

See Also

[kappa](#)

Examples

```
## Not run: From Szklo and Nieto, p. 315
dat <- cbind(c(18,1), c(19,11));
summary(sensSpec(dat));
```

 univar

Univariate Analysis of a Single Variable

Description

This function provides detailed univariate analysis for a single variable. Values include the sample mean, median, standard deviation and range, as well as tools for hypothesis tests and confidence intervals.

Usage

```
univar(X, alpha=0.05, mu0 = 0, shapiro=FALSE, digits=3)
```

Arguments

X	A Vector of observed values from a continuous distribution
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
mu0	The null hypothesis for the true population mean.
shapiro	Logical: TRUE returns the Shapiro-Wilks Test for normality, this portion calls the shapiro.test function.
digits	Number of Digits to round calculations

Details

This function provides a thorough summary of information within a vector. It conveniently calculates useful statistics at the call of a single command. Furthermore, it provides methods to test the hypothesis/construct confidence intervals for the true population mean.

Value

n	Number of Observations Used
mean	The sample mean of the observations in X.
median	The sample median of the observations in X.
min	The sample minimum of the observations in X.
max	The sample maximum of the observations in X.
s	The sample standard deviation of the observations in X.
var	The sample variance of the observations in X.
test	The test statistic for the null hypothesis μ
p.value	The p.value for the test statistic for μ
CIL	The lower bound of the constructed confidence interval for μ
CIU	The upper bound of the constructed confidence interval for μ
shapiro.statistic	The value of the Shapiro-Wilks Statistic for Normality.

shapiro.p.value	The P-value of the Shapiro-Wilks Statistic for Normality.
alpha	The desired Type I Error Rate for Hypothesis Tests and Confidence Intervals
mu0	The null hypothesis for the true population mean
.	
shapiro	Logical: TRUE returns the Shapiro-Wilks Test for normality
digits	Number of Digits to round calculations

Author(s)

Michael Rotondi, <mrotondi@uwo.ca>

References

Casella G and Berger RL. Statistical Inference (2nd Ed.) Duxbury: New York, 2002.

Examples

```
x <- rexp(100);  
univar(x);
```

Index

*Topic **design**

diffDetect, 3
epi2x2, 4
epiKappa, 6
mcNemar, 9
n4means, 10
n4props, 12
sensSpec, 13

*Topic **htest**

corXY, 2
epiKappa, 6
epiTTest, 7

*Topic **multivariate**

corXY, 2
epi2x2, 4
epiKappa, 6
epiTTest, 7
mcNemar, 9
sensSpec, 13

*Topic **univar**

univar, 15

Contingency Table (epi2x2), 4

corXY, 2

diffDetect, 3

epi2x2, 4, 10

epiKappa, 6

epiTTest, 7

kappa, 14

mcNemar, 6, 9

Mean Difference (diffDetect), 3

n4means, 10, 13

n4props, 11, 12

print.corXY (corXY), 2

print.diffDetect (diffDetect), 3

print.epi2x2 (epi2x2), 4

print.epiKappa (epiKappa), 6

print.epiTTest (epiTTest), 7

print.mcNemar (mcNemar), 9

print.n4means (n4means), 10

print.n4props (n4props), 12

print.sensSpec (sensSpec), 13

print.univar (univar), 15

sensSpec, 7, 13

summary.corXY (corXY), 2

summary.diffDetect (diffDetect), 3

summary.epi2x2 (epi2x2), 4

summary.epiKappa (epiKappa), 6

summary.epiTTest (epiTTest), 7

summary.mcNemar (mcNemar), 9

summary.n4means (n4means), 10

summary.n4props (n4props), 12

summary.sensSpec (sensSpec), 13

summary.univar (univar), 15

univar, 15