Package: ICCbin (via r-universe)

May 12, 2026

Version 1.2.1

Date 2026-05-12

Type Package

Title Facilitates Clustered Binary Data Generation, and Estimation of Intracluster Correlation Coefficient (ICC) for Binary Data

Imports stats

Suggests lme4

Description Assists in generating binary clustered data, estimates of Intracluster Correlation coefficient (ICC) for binary response in 16 different methods, and 5 different types of confidence intervals.

License GPL (>= 2)

URL https://cran.r-project.org/package=ICCbin

BugReports https://github.com/makhtarh/ICCbin/issues

RoxygenNote 8.0.0

Encoding UTF-8

Repository https://makhtarh.r-universe.dev

Date/Publication 2026-05-12 11:16:45 UTC

RemoteUrl https://github.com/makhtarh/iccbin

RemoteRef HEAD

RemoteSha e3d538b731c3ef0a991be88a40a6a1ab941be1e8

Contents

iccbin *Estimates Intracluster Correlation coefficients (ICC) and its confidence intervals (CI)*

Description

Estimates Intracluster Correlation coefficients (ICC) in 16 different methods and its confidence intervals (CI) in 5 different methods given the data on cluster labels and outcomes

Usage

```
iccbin(
  cid,
  y,
  data = NULL,
 method = c("aov", "aovs", "keq", "kpr", "keqs", "kprs", "stab", "ub", "fc", "mak",
    "peq", "pgp", "ppr", "rm", "lin", "sim"),
  ci.type = c("aov", "wal", "fc", "peq", "rm"),
  alpha = 0.05,
  kappa = 0.45,
  nAGQ = 1,
  M = 1000
)
```

Arguments

cid	Column name indicating cluster id in the dataframe data
y	Column name indicating binary response in the dataframe data
data	A dataframe containing cid and y
method	The method to be used to compute ICC. A single or multiple methods can be used at a time. By default, all 16 methods will be used. See Details for more.
ci.type	Type of confidence interval to be computed. By default all 5 types will be reported. See Details for more
alpha	The significance level to be used while computing confidence interval. Default value is 0.05
kappa	Value of Kappa to be used in computing Stabilized ICC when the method stab is chosen. Default value is 0.45
nAGQ	An integer scaler, as in glmer function of package lme4, denoting the number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood. Used when the method lin is chosen. Default value is 1
M	Number of Monte Carlo replicates used in ICC computation method sim. Default is 1000

Details

If in the dataframe, the cluster id (`cid`) is not a factor, it will be changed to a factor and a warning message will be given

If estimate of ICC in any method is outside the interval [0, 1], the estimate and corresponding confidence interval (if appropriate) will not be provided and warning messages will be produced

If the lower limit of any confidence interval is below 0 and upper limit is above 1, they will be replaced by 0 and 1 respectively and a warning message will be produced

Method `aov` computes the analysis of variance estimate of ICC. This estimator was originally proposed for continuous variables, but various authors (e.g. Elston, 1977) have suggested its use for binary variables

Method `aovs` gives estimate of ICC using a modification of analysis of variance technique (see Fleiss, 1981)

Method `keq` computes moment estimate of ICC suggested by Kleinman (1973), uses equal weight $w_i = 1/k$, for each of k clusters

Method `kpr` computes moment estimate of ICC suggested by Kleinman (1973), uses weights proportional to cluster size $w_i = n_i/N$

Method `keqs` gives a modified moment estimate of ICC with equal weights (`keq`) (see Kleinman, 1973)

Method `kprs` gives a modified moment estimate of ICC with weights proportional to cluster size (`kpr`) (see Kleinman, 1973)

Method `stab` provides a stabilized estimate of ICC proposed by Tamura and Young (1987)

Method `ub` computes moment estimate of ICC from an unbiased estimating equation (see Yamamoto and Yanagimoto, 1992)

Method `fc` gives Fleiss-Cuzick estimate of ICC (see Fleiss and Cuzick, 1979)

Method `mak` computes Mak's estimate of ICC (see Mak, 1988)

Method `peq` computes weighted correlation estimate of ICC proposed by Karlin, Cameron, and Williams (1981) using equal weight to every pair of observations

Method `pgp` computes weighted correlation estimate of ICC proposed by Karlin, Cameron, and Williams (1981) using equal weight to each cluster irrespective of size

Method `ppr` computes weighted correlation estimate of ICC proposed by Karlin, Cameron, and Williams (1981) by weighting each pair according to the total number of pairs in which the individuals appear

Method `rm` estimates ICC using resampling method proposed by Chakraborty and Sen (2016)

Method `lin` estimates ICC using model linearization proposed by Goldstein et al. (2002)

Method `sim` estimates ICC using Monte Carlo simulation proposed by Goldstein et al. (2002)

CI type `aov` computes confidence interval for ICC using Smith's large sample approximation (see Smith, 1957)

CI type `wal` computes confidence interval for ICC using modified Wald test (see Zou and Donner, 2004).

CI type `fc` gives Fleiss-Cuzick confidence interval for ICC (see Fleiss and Cuzick, 1979; and Zou and Donner, 2004)

CI type peq estimates confidence interval for ICC based on direct calculation of correlation between observations within clusters (see Zou and Donner, 2004; and Wu, Crespi, and Wong, 2012)

CI type rm gives confidence interval for ICC using resampling method by Chakraborty and Sen (2016)

Value

estimates	A dataframe containing the name of methods used and corresponding estimates of Intracluster Correlation coefficients
ci	A dataframe containing names of confidence interval types and corresponding estimated confidence intervals

Author(s)

Akhtar Hossain <ahossain@live.com>

Hrishikesh Chakraborty <hrishikesh.chakraborty@duke.edu>

References

Chakraborty, H. and Sen, P.K., 2016. Resampling method to estimate intra-cluster correlation for clustered binary data. Communications in Statistics-Theory and Methods, 45(8), pp.2368-2377.

Elston, R.C., Hill, W.G. and Smith, C., 1977. Query: Estimating" Heritability" of a dichotomous trait. Biometrics, 33(1), pp.231-236.

Fleiss, J.L., Levin, B. and Paik, M.C., 2013. Statistical methods for rates and proportions. John Wiley & Sons.

Fleiss, J.L. and Cuzick, J., 1979. The reliability of dichotomous judgments: Unequal numbers of judges per subject. Applied Psychological Measurement, 3(4), pp.537-542.

Goldstein, H., Browne, W., Rasbash, J., 2002. Partitioning variation in multilevel models, Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1 (4), pp.223-231.

Karlin, S., Cameron, E.C. and Williams, P.T., 1981. Sibling and parent–offspring correlation estimation with variable family size. Proceedings of the National Academy of Sciences, 78(5), pp.2664-2668.

Kleinman, J.C., 1973. Proportions with extraneous variance: single and independent samples. Journal of the American Statistical Association, 68(341), pp.46-54.

Mak, T.K., 1988. Analysing intraclass correlation for dichotomous variables. Applied Statistics, pp.344-352.

Smith, C.A.B., 1957. On the estimation of intraclass correlation. Annals of human genetics, 21(4), pp.363-373.

Tamura, R.N. and Young, S.S., 1987. A stabilized moment estimator for the beta-binomial distribution. Biometrics, pp.813-824.

Wu, S., Crespi, C.M. and Wong, W.K., 2012. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. Contemporary clinical trials, 33(5), pp.869-880.

Yamamoto, E. and Yanagimoto, T., 1992. Moment estimators for the beta-binomial distribution. Journal of applied statistics, 19(2), pp.273-283.

Zou, G., Donner, A., 2004 Confidence interval estimation of the intraclass correlation coefficient for binary outcome data, Biometrics, 60(3), pp.807-811.

See Also

rcbin rcbin_flex

Examples

```
bccdata <- rcbin(prop = .4, prvar = .2, noc = 30, csize = 20, csvar = .2, rho = .2)

# Methods that do not require 'lme4'
iccbin(cid = cid, y = y, data = bccdata,
  method = c("aov", "aovs", "keq", "kpr", "keqs", "kprs", "stab", "ub",
             "fc", "mak", "peq", "pgp", "ppr", "rm"))

iccbin(cid = cid, y = y, data = bccdata, method = c("aov", "fc"), ci.type = "fc")


# All methods including 'lin' and 'sim' (requires 'lme4')
if (requireNamespace("lme4", quietly = TRUE)) {
  iccbin(cid = cid, y = y, data = bccdata)
}
```

---

rcbin *Generates correlated binary cluster data*

Description

Generates correrlated binary cluster data given value of Intracluster Correlation, proportion of event, perceent of variation in event proportion, number of clusters, cluster size and percent of variation in cluster size

Usage

```
rcbin(prop = 0.5, prvar = 0, noc, csize, csvar = 0, rho)
```

Arguments

prop	A numeric value between 0 and 1 denoting assumed proportion of event in interest, default value is 0.5. See Detail
prvar	A numeric value between 0 and 1 denoting percent of variation in assumed proportion of event (prvar), default value is 0. See Detail
noc	A numeric value telling the number of clusters to be generated

csize	A numeric value denoting desired cluster size. See Deatil
csvar	A numeric value between 0 and 1 denoting percent of variation in cluster sizes (csize), default value is 0. See Detail
rho	A numeric value between 0 and 1 denoting desired level of Intracluster Correlation

Details

The minimum and maximum values of event proportion (prop) will be taken as 0 and 1 respectively in cases where it exceeds the valid limits (0, 1) due to larger value of percent variation (prvar) supplied

The minimum value of cluster size (csize) will be taken as 2 in cases where it goes below 2 due to larger value of percent variation (csvar) supplied

Value

A dataframe with two columns presenting cluster id (cid) and a binary response (y) variables

Author(s)

Akhtar Hossain <ahossain@live.com>

Hrishikesh Chakraborty <hrishikesh.chakraborty@duke.edu>

References

Lunn, A.D. and Davies, S.J., 1998. A note on generating correlated binary variables. Biometrika, 85(2), pp.487-490.

See Also

rcbin_flex for an enhanced version using Beta and Negative Binomial distributions with actual variance inputs. iccbin

Examples

```
rcbin(prop = .4, prvar = .2, noc = 30, csize = 20, csvar = .2, rho = .2)
```

---

rcbin_flex *Generates correlated binary cluster data with flexible distributional options*

Description

Generates correlated binary cluster data given value of Intracluster Correlation, proportion of event and its variance, number of clusters, cluster size and its variance, and minimum cluster size. Compared to rcbin, this function uses a Beta distribution for cluster-specific event proportions and supports overdispersed cluster sizes via the Negative Binomial distribution. Parameters prvar and csvar represent actual variances, not percentages.

Usage

```
rcbin_flex(prop = 0.5, prvar = 0, noc, csize, csvar = 0, mincsize = 2, rho)
```

Arguments

prop	A numeric value between 0 and 1 denoting assumed proportion of event of interest, default value is 0.5. See Details.
prvar	A non-negative numeric value (strictly less than prop*(1-prop)) denoting variance in assumed proportion of event (prop), default value is 0. See Details.
noc	A positive numeric value giving the number of clusters to be generated.
csize	A numeric value (≥ 2) denoting desired cluster size.
csvar	A non-negative numeric value denoting variance of cluster size, default value is 0. See Details.
mincsize	A numeric value (≥ 2) denoting the minimum cluster size, default value is 2. See Details.
rho	A numeric value between 0 and 1 denoting desired level of Intracluster Correlation.

Details

If prvar is 0, the event proportion is constant across all clusters as supplied by prop. If prvar > 0, cluster-specific event proportions are generated from a Beta distribution with shape parameters a and b obtained by solving prop $= a/(a+b)$ and prvar $= ab/[(a+b)^2(1+a+b)]$. See rbeta.

If csvar is 0, all clusters have equal size csize. For csvar > 0, cluster sizes are generated from a Normal distribution when csvar < csize (see rnorm), or from a Negative Binomial distribution when csvar $\geq$ csize to handle overdispersion (see rnbinom). Any cluster size below mincsize is replaced by mincsize.

Unlike rcbin, the parameters prvar and csvar represent actual variances, not percentages of prop or csize.

Value

A dataframe with two columns: cluster id (cid) and binary response (y).

Author(s)

Akhtar Hossain <ahossain@live.com>

References

Lunn, A.D. and Davies, S.J., 1998. A note on generating correlated binary variables. Biometrika, 85(2), pp.487-490.

See Also

rcbin iccbin

Examples

```
rcbin_flex(prop = .6, prvar = .05, noc = 100, csize = 10, csvar = 12,
           rho = 0.2, mincsize = 2)
```

Index