# Package: ICCbin (via r-universe)

September 7, 2024

**Version** 1.2

**Date** 2019-03-01

**Type** Package

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

**Author** Akhtar Hossain [aut, cre], Hrishikesh Chakraborty [aut]

**Maintainer** Akhtar Hossain <mhossain@email.sc.edu>

**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/akhtarh/ICCbin/issues

**LazyData** TRUE

**RoxygenNote** 6.1.1

**Encoding** UTF-8

**Repository** https://akhtarh.r-universe.dev

**RemoteUrl** https://github.com/akhtarh/iccbin

**RemoteRef** HEAD

**RemoteSha** 6dafc1f6ae75914e62fde9d07661c4ca605de44e

## Contents

`iccbin` *Estimates Intracluster Correlation coefficients (ICC) and it's confidence intervals (CI)*

## Description

Estimates Intracluster Correlation coefficients (ICC) in 16 different methods and it's 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 it's 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 stabilizd 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 Simith'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 <mhossain@email.sc.edu>

Hirshikesh Chakraborty <rishi.c@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

## Examples

```
bccdata <- rcbin(prop = .4, prvar = .2, noc = 30, csize = 20, csvar = .2, rho = .2)
iccbin(cid = cid, y = y, data = bccdata)
iccbin(cid = cid, y = y, data = bccdata, method = c("aov", "fc"), ci.type = "fc")
```

---

# 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 <mhossain@email.sc.edu>

Hrishikesh Chakraborty <rishi.c@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

rcbin1 iccbin

## Examples

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

---

# rcbin1 *Generates correlated binary cluster data*

---

## Description

Generates correrlated binary cluster data given value of Intracluster Correlation, proportion of event and it's variance, number of clusters, cluster size and it's variance, and minimum cluster size

## Usage

```
rcbin1(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 in interest, default value is 0.5. See Detail |
| prvar | A numeric value between 0 and 1 denoting varince in assumed proportion of event (prvar), default value is 0. See Detail |
| noc | A positive numeric value telling the number of clusters to be generated |
| csize | A numeric value ($\geq$ 2) denoting cluster size desired |
| csvar | A positive numeric value denoting Variance of cluster size, default value is 0, see Detail |

| | |
|---|---|
| mincsize | A numeric value ($\geq 2$) denoting the minimum cluster size desired, default value is 2, see Detail |
| rho | A numeric value between 0 and 1 denoting desired level of Intracluster Correlation |

## Details

If supplied value of prvar is 0, the event proportion for all clusters is considered constant as supplied by prop. If supplied prvar is > 0, cluster specific event proportions are generated from Beta distribution with shape1 and shape2 parameters $a$ and $b$ respectively, see rbeta The shape parameters are obtained using supplied values of prop and prvar by solving the equations prop $= a/(a+b)$ and prvar $= ab/[(a+b)^2(1+a+b)]$

If supplied value of csvar is 0, cluster of equal size (csize) will be generated. For csvar > 0, will be generated from Normal or Negative Binomial dsitributions depending on relationship between csize and csvar. If csvar < csize, the varying cluster sizes will be generated from a Normal distribution with mean = csize and variacne = csvar (see rnorm). If csvar $\geq$ csize i.e. in the case of overdispersion, cluster sizes will be generated from Negative Bionomial distribution using mu = csize and size $=$ csize$/[$csize$($cscv$^2 - 1)]$ (see rnbinom), where cscv is the coefficient of variation of cluster sizes defined as $\sqrt{\text{csvar}}/\text{csize}$. If the size of any cluster is generated as less than 2, it will be replaced by the supplied value of minimum cluster size (mincsize) which has a default value of 2

## Value

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

## Author(s)

Akhtar Hossain <mhossain@email.sc.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 iccbin

## Examples

```
rcbin1(prop = .6, prvar = .1, noc = 100, csize = 10, csvar = 12, rho = 0.2, mincsize = 2)
```

# Index