

# Package: CR2 (via r-universe)

November 2, 2024

**Title** Compute Cluster Robust Standard Errors with Degrees of Freedom Adjustments

**Version** 0.2.1

**Date** 2023-01-03

**Description** Estimate different types of cluster robust standard errors (CR0, CR1, CR2) with degrees of freedom adjustments. Standard errors are computed based on 'Liang and Zeger' (1986) <[doi:10.1093/biomet/73.1.13](https://doi.org/10.1093/biomet/73.1.13)> and Bell and 'McCaffrey' <<https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2002002/article/9058-eng.pdf?st=NxMjN1YZ>>. Functions used in Huang and Li <[doi:10.3758/s13428-021-01627-0](https://doi.org/10.3758/s13428-021-01627-0)>, Huang, 'Wiedermann', and 'Zhang' <[doi:10.1080/00273171.2022.2077290](https://doi.org/10.1080/00273171.2022.2077290)>, and Huang, 'Zhang', and Li (forthcoming: Journal of Research on Educational Effectiveness).

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.2.3

**URL** <https://github.com/flh3/CR2>

**BugReports** <https://github.com/flh3/CR2/issues>

**Depends** R (>= 2.10)

**Imports** stats, lme4, nlme, Matrix, methods, generics, magrittr, broom, dplyr, performance, tibble

**Repository** <https://flh3.r-universe.dev>

**RemoteUrl** <https://github.com/flh3/cr2>

**RemoteRef** HEAD

**RemoteSha** cf7e9f681803e963e46ce0d54aa6eaa247606e65

## Contents

clustSE	2
crct	3
getV	4
glance.CR2	5
gpadat	5
MatSqrtInverse	6
ncvMLM	7
robust_mixed	7
satdf	9
sch25	9
sharedat	10
tidy.CR2	11

<b>Index</b>	<b>12</b>
--------------	-----------

---

clustSE	<i>Cluster robust standard errors with degrees of freedom adjustments (for lm and glm objects)</i>
---------	--

---

### Description

Function to compute the CR0, CR1, CR2 cluster robust standard errors (SE) with Bell and McCaffrey (2002) degrees of freedom (df) adjustments. Useful when dealing with datasets with a few clusters. Shows output using different CR types and degrees of freedom choices (for comparative purposes only). For linear and logistic regression models (as well as other GLMs). Computes the BRL-S2 variant.

### Usage

```
clustSE(mod, clust = NULL, digits = 3, ztest = FALSE)
```

### Arguments

mod	The lm model object.
clust	The cluster variable (with quotes).
digits	Number of decimal places to display.
ztest	If a normal approximation should be used as the naive degrees of freedom. If FALSE, the between-within degrees of freedom will be used.

### Value

A data frame with the CR adjustments with p-values.

estimate	The regression coefficient.
se.unadj	The model-based (regular, unadjusted) SE.

CR0	Cluster robust SE based on Liang & Zeger (1986).
CR1	Cluster robust SE (using an adjustment based on number of clusters).
CR2	Cluster robust SE based on Bell and McCaffrey (2002).
tCR2	t statistic based on CR2.
dfn	Degrees of freedom(naive): can be infinite (z) or between-within (default). User specified.
dfBM	Degrees of freedom based on Bell and McCaffrey (2002).
pv.unadj	p value based on model-based standard errors.
CR0pv	p value based on CR0 SE with dfBM.
CR0pv.n	p value based on CR0 SE with naive df.
CR1pv	p value based on CR1 SE with dfBM.
CR1pv.n	p value based on CR1 SE with naive df.
CR2pv	p value based on CR2 SE with dfBM.
CR2pv.n	p value based on CR2 SE with naive df.

## References

- Bell, R., & McCaffrey, D. (2002). *Bias reduction in standard errors for linear regression with multi-stage samples*. *Survey Methodology*, 28, 169-182. ([link](#))
- Liang, K.Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13–22. [doi:10.1093/biomet/73.1.13](https://doi.org/10.1093/biomet/73.1.13)

## Examples

```
clustSE(lm(mpg ~ am + wt, data = mtcars), 'cyl')
data(sch25)
clustSE(lm(math ~ ses + minority + mses + mhmwk, data = sch25), 'schid')
```

---

crct	<i>Simulated data from 18 schools (from a cluster randomized controlled trial)</i>
------	--

---

## Description

Synthetic dataset used in the manuscript in the Journal of Research on Educational Effectiveness.

## Usage

```
data(crct)
```

**Format**

A data frame with 4233 rows and 12 variables:

**usid** Unique school identifier (the grouping variable).

**stype** School type (elementary, middle, or high school).

**trt** Treatment indicator. 1 = intervention; 0 = control.

**odr\_post** Office disciplinary referral outcome.

**odr\_pre** Office disciplinary referral (baseline).

**size** School enrollment size (to the nearest hundred).

**female** Student is female: 1 = yes.

**stype\_ms** Dummy code for school type; middle school.

**stype\_lem** Dummy code for school type; elementary school.

**stype\_hs** Dummy code for school type; high school.

**race\_Black** Dummy code for student race/ethnicity; Black student.

**race\_Hispanic** Dummy code for student race/ethnicity; Hispanic student.

---

getV

*Get V matrix for merMod objects*

---

**Description**

Function to extract V matrix.

**Usage**

```
getV(x)
```

**Arguments**

x                      lme4 object

**Value**

V matrix (weight) for multilevel models

---

glance.CR2	<i>Glance at goodness-of-fit statistics</i>
------------	---

---

**Description**

Helper function used to obtain supporting fit statistics for multilevel models. The R2s are computed using the performance package.

**Usage**

```
## S3 method for class 'CR2'
glance(x, ...)
```

**Arguments**

x	A CR2 object.
...	Unused, included for generic consistency only.

**Value**

glance returns one row with the columns:

nobs	the number of observations
sigma	the square root of the estimated residual variance
logLik	the data's log-likelihood under the model
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
r2.marginal	marginal R2 based on fixed effects only using method of Nakagawa and Schielzeth (2013)
r2.conditional	conditional R2 based on fixed and random effects using method of Nakagawa and Schielzeth (2013)

---

gpadat	<i>Grade point average (GPA) data of students from 25 schools</i>
--------	---

---

**Description**

For investigating heteroskedasticity.

**Usage**

```
data(gpadat)
```

**Format**

A data frame with 8,956 rows and 18 variables:

**gpa** Grade point average. 1 = D ... 4 = A.

**female** Gender. Female = 1.

**race** Student race/ethnicity (factor).

**dis** Disability status (1 = yes/0 = no).

**frpl** Free/reduced price lunch status.

**race\_w** Dummy coded race (White).

**race\_a** Dummy coded race (Asian).

**race\_b** Dummy coded race (Black).

**race\_h** Dummy coded race (Hispanic).

**race\_o** Dummy coded race (Other).

**per\_asian** Group-aggregated Asian variable.

**per\_black** Group-aggregated Black variable.

**per\_hisp** Group-aggregated Hispanic variable.

**per\_other** Group-aggregated Other variable.

**per\_fem** Group-aggregated female variable.

**per\_dis** Group-aggregated disability variable.

**per\_frpl** Group-aggregated frpl variable.

**schoolid** School identifier (cluster variable).

---

MatSqrtInverse

*Compute the inverse square root of a matrix*

---

**Description**

From Imbens and Kolesar (2016).

**Usage**

MatSqrtInverse(A)

**Arguments**

A                      The matrix object.

**Value**

Returns a matrix.

---

ncvMLM	<i>Testing for nonconstant variance (ncv)</i>
--------	---

---

### Description

Function to detect heteroscedasticity in two-level random intercept models. Uses a generalization of the Breusch-Pagan-type (using squared residuals) and Levene-type test (using the absolute value of residuals). Note: this will not tell you if including random slopes are warranted (for that, use the `robust_mixed`) function and compare differences in model-based and robust standard errors.

### Usage

```
ncvMLM(mx, bp = TRUE)
```

### Arguments

<code>mx</code>	The lme or merMod model object.
<code>bp</code>	Computes a Breusch-Pagan-type test (TRUE). If FALSE computes a Levene-type test.

### Value

A p-value ( $p < .05$  suggests heteroskedasticity).

### References

Huang, F., Wiedermann, W., & Zhang, B. (2022). Accounting for Heteroskedasticity Resulting from Between-group Differences in Multilevel Models. *Multivariate Behavioral Research*.

### Examples

```
require(lme4)
data(sch25)
ncvMLM(lmer(math ~ byhomewk + male + ses + (1|schid), data = sch25)) #supported
ncvMLM(lmer(math ~ byhomewk + male + ses + minority + (1|schid), data = sch25)) #hetero
```

---

<code>robust_mixed</code>	<i>Cluster robust standard errors with degrees of freedom adjustments for lmerMod/lme objects</i>
---------------------------	---

---

### Description

Function to compute the CR2/CR0 cluster robust standard errors (SE) with Bell and McCaffrey (2002) degrees of freedom (dof) adjustments. Suitable even with a low number of clusters. The model based (mb) and cluster robust standard errors are shown for comparison purposes.

**Usage**

```
robust_mixed(m1, digits = 3, type = "CR2", satt = TRUE, Gname = NULL)
```

**Arguments**

m1	The lmerMod or lme model object.
digits	Number of decimal places to display.
type	Type of cluster robust standard error to use ("CR2" or "CR0").
satt	If Satterthwaite degrees of freedom are to be computed (if not, between-within df are used).
Gname	Group/cluster name if more than two levels of clustering (does not work with lme).

**Value**

A data frame (results) with the cluster robust adjustments with p-values.

Estimate	The regression coefficient.
mb.se	The model-based (regular, unadjusted) SE.
cr.se	The cluster robust standard error.
df	degrees of freedom: Satterthwaite or between-within.
p.val	p-value using CR0/CR2 standard error.
stars	stars showing statistical significance.

**Author(s)**

Francis Huang, <huangf@missouri.edu>

Bixi Zhang, <bixizhang@missouri.edu>

**References**

*Bell, R., & McCaffrey, D. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology, 28, 169-182. ([link](#))*

*Liang, K.Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73(1), 13-22. ([link](#))*

**Examples**

```
require(lme4)
data(sch25, package = 'CR2')
robust_mixed(lmer(math ~ male + minority + mses + mhmwk + (1|schid), data = sch25))
```



---

satdf	<i>Compute Satterthwaite degrees of freedom</i>
-------	---

---

**Description**

Function to compute empirical degrees of freedom based on Bell and McCaffrey (2002).

**Usage**

```
satdf(m1, type = "none", Vinv2, Vm2, br2, Gname = NULL)
```

**Arguments**

m1	The lmerMod or lme model object.
type	The type of cluster robust correction used (i.e., CR2 or none).
Vinv2	Inverse of the variance matrix.
Vm2	The variance matrix.
br2	The bread component.
Gname	The group (clustering variable) name'

**Value**

Returns a vector of degrees of freedom.

**Author(s)**

Francis Huang, <huangf@missouri.edu>  
Bixi Zhang, <bixizhang@missouri.edu>

---

sch25	<i>Data from 25 schools (based on the NELS dataset)</i>
-------	---

---

**Description**

For examining the association between amount homework done per week and math outcome.

**Usage**

```
data(sch25)
```

**Format**

A data frame with 546 rows and 8 variables:

**schid** The school identifier (the grouping variable)

**ses** Student-level socioeconomic status

**byhomewk** Total amount of time the student spent on homework per week. 1 = None, 2 = Less than one hour, 3 = 1 hour, 4 = 2 hours, 5 = 3 hours, 6 = 4-6 hours, 7 = 7 - 9 hours, 8 = 10 or more

**math** Mathematics score.

**male** Dummy coded gender, 1 = male, 0 = female

**minority** Dummy coded minority status, 1 = yes, 0 = no

**mSES** Aggregated socioeconomic status at the school level

**mhmwk** Aggregated time spent on homework at the school level

**Source**

<https://nces.ed.gov/pubs92/92030.pdf>

---

sharedat

*Data from Project SHARE*

---

**Description**

Project SHARE (Sexual Health and Relationships) was a cluster randomized trial (CRT) in Scotland carried out to measure the impact of a school-based sexual health program (Wight et al., 2002).

**Usage**

`data(sharedat)`

**Format**

A data frame with 5399 observations and 7 variables.

**school** The cluster variable

**sex** factor indicating F or M

**arm** treatment arm = 1 vs control = 0

**kscore** Pupil knowledge of sexual health

**idno** student id number

**sc** factor showing the highest social class of the father or mother based on occupation (coded 10: I (highest), 20: II, 31: III non-manual, 32: III manual, 40: IV, 50: V (lowest), 99: not coded).

**zscore** standardized knowledge score

**Source**

[doi:10.7910/DVN/YXMQZM](https://doi.org/10.7910/DVN/YXMQZM) Harvard dataverse

**References**

Moulton, L. (2015). *readme.txt contains an overall explanation of the data sets.* Harvard. [doi:10.7910/DVN/YXMQZM](https://doi.org/10.7910/DVN/YXMQZM)

Wight, D., Raab, G. M., Henderson, M., Abraham, C., Buston, K., Hart, G., & Scott, S. (2002). *Limits of teacher delivered sex education: Interim behavioural outcomes from randomised trial.* *BMJ*, 324, 1430. [doi:10.1136/bmj.324.7351.1430](https://doi.org/10.1136/bmj.324.7351.1430)

**Examples**

```
data(sharedat)
```

---

tidy.CR2

*Tidy a CR2 object*

---

**Description**

Tidy a CR2 object

**Usage**

```
## S3 method for class 'CR2'
tidy(x, conf.int = FALSE, conf.level = 0.95, ...)
```

**Arguments**

x	A CR2 object.
conf.int	Logical indicating whether or not to include a confidence interval in the tidied output. Defaults to FALSE.
conf.level	The confidence level to use for the confidence interval if conf.int = TRUE. Must be strictly greater than 0 and less than 1. Defaults to 0.95, which corresponds to a 95 percent confidence interval.
...	Unused, included for generic consistency only.

**Value**

A tidy `tibble::tibble()` summarizing component-level information about the model

# Index

## \* datasets

- crct, [3](#)
- gpadat, [5](#)
- sch25, [9](#)
- sharedat, [10](#)

clustSE, [2](#)

crct, [3](#)

getV, [4](#)

glance.CR2, [5](#)

gpadat, [5](#)

MatSqrtInverse, [6](#)

ncvMLM, [7](#)

robust\_mixed, [7](#)

satdf, [9](#)

sch25, [9](#)

sharedat, [10](#)

tibble::tibble(), [11](#)

tidy.CR2, [11](#)