

# Package ‘BaylorEdPsych’

July 2, 2014

**Type** Package

**Title** R Package for Baylor University Educational Psychology Quantitative Courses

**Version** 0.5

**Date** 2012-05-08

**Author** A. Alexander Beaujean

**Maintainer** A. Alexander Beaujean <Alex\_Beaujean@Baylor.edu>

**Description** Functions and data used for Baylor University Educational Psychology Quantitative Courses

**Suggests** mvnmle, moments

**License** GPL (>= 2)

**Repository** CRAN

**Date/Publication** 2012-07-28 05:26:31

**NeedsCompilation** no

## R topics documented:

BaylorEdPsych-package . . . . .	2
BaylorEdPsych . . . . .	2
BilogData . . . . .	3
CohensD.rawData . . . . .	5
CohensD.unpairedT . . . . .	6
EndersTable1_1 . . . . .	7
EtaSq . . . . .	7
HedgesG.rawData . . . . .	8
LittleMCAR . . . . .	9
MLBOffense2011 . . . . .	10
MLBPitching2011 . . . . .	12
PseudoR2 . . . . .	14

<b>Index</b>	<b>16</b>
--------------	-----------

BaylorEdPsych-package *R functions/data for Baylor Ed Psych Department's Quantitative Courses*

---

**Description**

R functions and data for Baylor University Educational Psychology Department's Quantitative Courses

**Details**

Package: BaylorEdPsych  
Type: Package  
Version: 1.0  
Date: 2012-05-08  
License: GPL (>= 2)

**Author(s)**

A. Alexander Beaujean

Maintainer: A. Alexander Beaujean <Alex\_Beaujean@Baylor.edu>

---

BaylorEdPsych *Baylor Educational Psychology Quantitative Courses R Functions and Data*

---

**Description**

Baylor Educational Psychology Quantitative Courses R Functions and Data

**Value**

Data Sets:

[MLBOffense2011](#) 2011 Major League Baseball Offensive Data

[MLBPitching2011](#)

2011 Major League Baseball Pitching Data

[EndersTable1\\_1](#) Data from table 1.1 (p. 3) in Enders, C. K. (2011). *Applied missing data analysis*. New York: Guilford

Effect Sizes:

<a href="#">PseudoR2</a>	Various Pseudo- $R^2$ values for a regression with a dichotomous outcome
<a href="#">CohensD.rawData</a>	Cohen's d effect size from raw data
<a href="#">CohensD.unpairedT</a>	Cohen's d effect size from unpaired (independent samples) t test
<a href="#">HedgesG.rawData</a>	Hedge's g effect size from raw data
Missing Data:	
<a href="#">LittleMCAR</a>	Little's Missing Completely At Random (MCAR) chi-square test
Data Preperation:	
<a href="#">BilogData</a>	Coverts a R data frame to a fixed width data frame for use in BILOG-MG and exports it to a .dat file

**Author(s)**

A. Alexander Beaujean

**References**

Baylor Psychometric Laboratory: [http://blogs.baylor.edu/psychometric\\_lab/](http://blogs.baylor.edu/psychometric_lab/)

Baylor University Educational Psychology: <http://www.baylor.edu/soe/edp/>

---

BilogData

*Create Fixed Width Data File for use with BILOG-MG*

---

**Description**

Coverts a R data frame to a fixed width data frame for use in BILOG-MG and exports it to a .dat file. Similar to what the the est function in the irtoys package does, but allows for the use of your own ID variable instead of creating one.

**Usage**

```
BilogData(resp, data.name="mydata", location=NULL,
ret.val=FALSE, idvar=NULL)
```

**Arguments**

resp	A data frame consisting of item responses and possilby an identification variable
data.name	Name of the exported .dat file.
location	Where the created .dat file will be stored. Defaults to current working directory.
ret.val	If want the .dat to be printed on screen. Logical. Defaults to not priting on screen
idvar	If an ID variable is included in the resp data frame, give the name (e.g., idvar="ID"). Otherwise, a ID variable will be generated. ID variable can be up to 6 numeric characters in length.

**Details**

None.

**Value**

None. If `ret.val=TRUE` the exported `.dat` file will be printed on screen.

**Note**

None

**Author(s)**

A. Alexander Beaujean

**References**

Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG 3: Item analysis and test scoring with binary logistic models*. [Computer software]. Chicago, IL: Scientific Software International.

du Toit, M. (2003). *IRT from SSI*. Chicago: Scientific Software International.

Partchev, I. (2012). *irtoys: Simple interface to the estimation and plotting of IRT models* (version 0.1.5). Flanders, Belgium: KU Leuven. Retrieved from <http://cran.r-project.org/web/packages/irtoys/irtoys.pdf>

**See Also**

[irtoys-package](#)

**Examples**

```
#Simulate dichotomous data with ID
datt<-cbind(rnorm(100), rnorm(100), rnorm(100))
d.datt<-dim(datt)
datt<-as.numeric(datt > 0)
dim(datt)<-d.datt
datt<-data.frame(ID=seq(1,100), datt)

#Export \emph{mydata.dat} file to working directory
BilogData(datt, idvar="ID")

#Print the exported data frame, \emph{Test5.dat}, on screen, too
BilogData(datt, data.name = "Test5", ret.val=TRUE, idvar="ID")
```

---

CohensD.rawData	<i>Calculate's Cohen's d</i>
-----------------	------------------------------

---

**Description**

Calculate's Cohen's d

**Usage**

```
CohensD.rawData(E.data, C.data)
```

**Arguments**

E.data	Data from group 1 (Experimental/Treatment Group)
C.data	Data from group 2 (Control Group)

**Value**

Cohen's d effect size

**Author(s)**

A. Alexander Beaujean

**References**

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

**Examples**

```
#Simulate Data
set.seed(4565)
Treat<-rnorm(100,.5,1)
set.seed(45651)
Control<-rnorm(100,0,1)

CohensD.rawData(Treat,Control)
```

---

CohensD.unpairedT      *Cohen's d from an unpaired (independed samples) t-test*

---

**Description**

Calculate's Cohen's d using results from an unpaired (independed samples) t-test

**Usage**

```
CohensD.unpairedT(t.val, n1, n2)
```

**Arguments**

t.val	t statistic from unpaired/independed samples t-test
n1	Sample size for group 1
n2	Sample size for group 2

**Value**

Cohen's d effect size

**Author(s)**

A. Alexander Beaujean

**References**

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

**See Also**

[t.test](#)

**Examples**

```
#Simulate data
set.seed(4565)
Treat<-rnorm(100,.5,1)
set.seed(45651)
Control<-rnorm(100,0,1)
t<-t.test(Treat,Control, var.equal = TRUE)$statistic
nt<-length(Treat)
nc<-length(Control)

CohensD.unpairedT(t, nt, nc)
```

---

EndersTable1_1	<i>Data from Table 1.1 (p. 3) of Enders (2011) Applied Missing Data Analysis</i>
----------------	--

---

**Description**

Data from Table 1.1 (p. 3) of Enders (2011) *Applied Missing Data Analysis*.

**Usage**

```
data(EndersTable1_1)
```

**Format**

A data frame with 20 observations on the following 3 variables.

IQ IQ

JP Job Performace

WB Psychological Well Being

**Source**

Enders, C. K. (2011). *Applied missing data analysis*. New York: Guilford.

**Examples**

```
data(EndersTable1_1)
```

---

EtaSq	<i>(Partial) Eta Squared</i>
-------	------------------------------

---

**Description**

Returns eta squared and partial eta squared values for aov objects

**Usage**

```
EtaSq(x)
```

**Arguments**

x Analysis of variance (aov) object

**Value**

Eta^2  $\eta^2$  value (SS Effect/SS Total)

Partial Eta^2 Partial  $\eta^2$  value (SS Effect/SS Residual)

**Note**

$\eta^2$  overestimates the variance explained in the population (i.e., it is a biased estimator), but as the sample size gets larger the amount of bias gets smaller. Grissom and Kim (2012, p. 182–183) suggest alternative estimators.

**Author(s)**

A. Alexander Beaujean

**References**

Grissom, R. J. & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications* (2nd ed.). New York: Routledge

**See Also**

[aov](#)

**Examples**

```
data(MLBOffense2011)
groupRBI<-aov(RBI~Tm, data=MLBOffense2011, subset=c(Tm=="ARI" | Tm=="TOT" | Tm=="SFG"))
EtaSq(groupRBI)
```

---

HedgesG.rawData      *Hedges' g from raw data*

---

**Description**

Calculates Hedges' g effect size from raw data

**Usage**

```
HedgesG.rawData(E.data, C.data)
```

**Arguments**

E.data	Data from group 1
C.data	Data from group 2

**Details**

Similar to Cohen's d, but a different scaling. The interpretation is the same (i.e., group differences in standard deviation units)

**Value**

Hedges' g



**Author(s)**

A. Alexander Beaujean

**References**

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

**Examples**

```
#Simulate Data
set.seed(4565)
Treat<-rnorm(100,.5,1)
set.seed(45651)
Control<-rnorm(100,0,1)

HedgesG.rawData(Treat,Control)
```

---

LittleMCAR

*Little's missing completely at random (MCAR) test*

---

**Description**

Uses Little's test to assess for missing completely at random for multivariate data with missing values

**Usage**

```
LittleMCAR(x)
```

**Arguments**

x                    A data frame or data matrix of no more than 50 variables

**Details**

Depending on the sample size and number of missing data patterns, this function could take a long time to run.

**Value**

chi.square	Chi-square value
df	Degrees of freedom used for chi-square
missing.patterns	Number of missing data patterns
amount.missing	Amount and percent of missing data
data	The data, organized by missing data patterns

**Author(s)**

A. Alexander Beaujean

**References**

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404), 1198–1202.

**See Also**

[md.pattern](#)

**Examples**

```
data(EndersTable1_1)
LittleMCAR(EndersTable1_1)
```

---

MLBOffense2011

*2011 Major League Baseball offensive data*

---

**Description**

Offensive data from Major League Baseball players in 2011 who were not traded during the season

**Usage**

```
data(MLBOffense2011)
```

**Format**

A data frame with 814 observations (only 688 had at least 1 At Bat [AB]) on the following 30 variables:

FName Player's first name

LName Player's last name

HR Home runs

Year Year of season

Age Player's age during year

Tm Player's team

Lg League in which player plays (NL = National League; AL = American League)

G Games (Regulation length games in which a player plays)

PA Plate appearances (PA = AB+BB+IBB+HBP+SH+SF)

AB Number of official at bats (AB = PA-BB-IBB-HBP-SH-SF)

R Runs scored

H Hits ( $H = B1+B2+B3+HR$ )  
B1 Singles  
B2 Doubles  
B3 Triples  
RBI Runs batted in  
BB Walks  
IBB Intentional walks  
SO Strike outs  
HBP Hit by pitch  
SH Sacrifice bunts  
SF Sacrifice flies  
GDP Number of times player grounded into a double play  
SB Stolen bases  
CS Number of times player was caught stealing a base  
BA Batting average ( $BA = H/AB$ )  
OBP On-base percentage ( $OBP = (H+BB+HBP)/(AB+BB+HBP+SF)$ )  
SLG Slugging percentage ( $SLG = (B1)+(2*B2)+3*B3)+(4*HR)/AB$ )  
OPS On-base percentage plus slugging percentage ( $OPS = OBP+SLG$ )  
TB Total bases ( $TB = (B1)+(2*B2)+(3*B3)+(4*HR)$ )

**Details**

Data from Retrosheet

**Author(s)**

Thanks to Aaron Baggett for compiling the data

**Source**

The information used here was obtained free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at [www.retrosheet.org](http://www.retrosheet.org).

**Examples**

```
data(MLBOffense2011)
```

---

 MLBPitching2011

 2011 Major League Baseball pitching data
 

---

**Description**

Offensive data from some Major League Baseball pitchers in 2011 who were not traded during the season

**Usage**

`data(MLBPitching2011)`

**Format**

A data frame with 642 observations (531 with a Win-Loss Percent [WLP]) on the following 44 variables.

Fname Player's first name

Lname Player's last name

SO Strike outs

Year Year of season

Age Player's age

Tm Player's team

Lg League in which player plays (NL = National League; AL = American League)

G Games (Regulation length games in which a player plays)

GS Games started

CG Complete games

SHO Shutouts (Games in which a team allows no runs)

GF Games finished (Games in which a relief pitcher is the last to appear)

W Wins

L Losses

WLP Win-Loss Percentage ( $WLP = W/(W+L)$ )

SV Saves

IP Innings pitched

H Hits allowed

R Runs allowed

ER Earned runs allowed

BB Walks allowed

ERA Earned run average ( $ERA = 9*ER/IP$ )

ERAP Adjusted earned run average ( $ERAP = 100*(lgERA/ERA)$ )

HR Homeruns allowed  
 BF Batters faced  
 AB At bats  
 B1 Singles Allowed  
 B2 Doubles allowed  
 B3 Triples allowed  
 IBB Intentional walks issued  
 HBP Hit by pitch  
 SH Sacrifice bunts allowed  
 SF Sacrifice flys allowed  
 GDP Ground ball double-plays induced  
 SB Stolen bases allowed  
 CS Caught stealing  
 PO Pickoffs  
 BK Balks  
 WP Wild pitches  
 BA Opponent's batting average ( $BA = H/AB$ )  
 OBP Opponent's on-base percentage ( $OBP = (H+BB+HBP)/(AB+BB+HBP+SF)$ )  
 SLG Opponent's slugging percentage ( $SLG = (B1)+(2*B2)+3*B3)+(4*HR)/AB$ )  
 OPS Opponent's on-base plus slugging percentage ( $OPS = OBP+SLG$ )  
 Pit Pitches thrown ( $Pit = str+balls\ thrown$ )  
 Str Strikes thrown

**Details**

Data from Retrosheet

**Author(s)**

Thanks to Aaron Baggett for compiling the data

**Source**

The information used here was obtained free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at [www.retrosheet.org](http://www.retrosheet.org).

**Examples**

`data(MLBPitching2011)`

PseudoR2

*Pseudo-R2 Statistics***Description**

Returns various pseudo- $R^2$  statistics (as well as Akaike's [corrected] information criterion) from a glm object. Should mimic those returned from a logistic/probit regression in *Stata* when using `fitstat`

**Usage**

```
PseudoR2(glmModel)
```

**Arguments**

`glmModel` Object from a glm model, preferably logistic regression, e.g., `family=binomial(link="logit")`

**Details**

None

**Value**

<code>McFadden</code>	McFadden Pseudo- $R^2$
<code>Adj.McFadden</code>	McFadden Adjusted Pseudo- $R^2$
<code>Cox.Snell</code>	Cox and Snell Pseudo- $R^2$ (also known as ML Pseudo- $R^2$ )
<code>Nagelkerke</code>	Nagelkerke Pseudo- $R^2$
<code>McKelvey.Zavoina</code>	McKelvey and Zavoina Pseudo- $R^2$
<code>Efron</code>	Efron Pseudo- $R^2$
<code>Count</code>	Count Pseudo- $R^2$ , number of correctly classified cases, using $\hat{\pi} > .50$ as the cutoff
<code>Adj.Count</code>	Adjusted Count Pseudo- $R^2$
<code>AIC</code>	Akaike's information criterion
<code>Corrected.AIC</code>	Corrected Akaike information criterion

**Note**

There are many documented problems with using pseudo-R2 values (e.g., Long, 1997). Use the values judiciously.

**Author(s)**

A. Alexander Beaujean

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer-Verlag.
- Efron, B. (1978). Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, 73(361), 113–121.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Hoboken, NJ: Wiley.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- McFadden, D. (1979). Quantitative methods for analysing travel behavior of individuals: Some recent developments. In D. A. Hensher & P. R. Stopher (Eds.), *Behavioural travel modelling* (pp. 279–318). London: Croom Helm.
- McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, 4(1), 103–120
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691–692.

## See Also

[glm](#)

## Examples

```
data(MLBOffense2011)
MLBOffense2011$NL<-ifelse(MLBOffense2011$LG=="NL", 1,0)
#predict MLB league membership from RBI and slugging
model1<-glm(NL~RBI + SLG, data=MLBOffense2011, family=binomial(link="logit"))
PseudoR2(model1)
```

# Index

- \*Topic **BILOR-MG**
  - BilogData, 3
- \*Topic **Baseball**
  - MLBOffense2011, 10
- \*Topic **Cohen's d**
  - CohensD.rawData, 5
  - CohensD.unpairedT, 6
- \*Topic **Effect Sizes**
  - BaylorEdPsych, 2
- \*Topic **Effect Size**
  - CohensD.rawData, 5
  - CohensD.unpairedT, 6
  - EtaSq, 7
  - HedgesG.rawData, 8
- \*Topic **Eta Squared**
  - EtaSq, 7
- \*Topic **Hedges' g**
  - HedgesG.rawData, 8
- \*Topic **Item Response Theory**
  - BilogData, 3
- \*Topic **Little's Chi-Square**
  - LittleMCAR, 9
- \*Topic **Logistic Regression**
  - PseudoR2, 14
- \*Topic **MCAR**
  - LittleMCAR, 9
- \*Topic **Missing Completely At Random**
  - LittleMCAR, 9
- \*Topic **Missing Data**
  - BaylorEdPsych, 2
  - EndersTable1\_1, 7
- \*Topic **Missing**
  - LittleMCAR, 9
- \*Topic **Partial Eta Squared**
  - EtaSq, 7
- \*Topic **Pseudo-R2**
  - BaylorEdPsych, 2
  - PseudoR2, 14
- \*Topic **R2**
  - PseudoR2, 14
- \*Topic **datasets**
  - EndersTable1\_1, 7
  - MLBOffense2011, 10
  - MLBPitching2011, 12
- ANOVA (EtaSq), 7
- aov, 8
- BaylorEdPsych, 2
- BaylorEdPsych-package, 2
- BilogData, 3, 3
- CohensD.rawData, 3, 5
- CohensD.unpairedT, 3, 6
- Effect Size (EtaSq), 7
- EndersTable1\_1, 2, 7
- EtaSq, 7
- glm, 15
- HedgesG.rawData, 3, 8
- irtoys-package, 4
- LittleMCAR, 3, 9
- md.pattern, 10
- MLBOffense2011, 2, 10
- MLBPitching2011, 2, 12
- PseudoR2, 3, 14
- t.test, 6