

# Package ‘EigenCorr’

August 11, 2011

**Version** 0.2

**Date** 2011-08-10

**License** GPL version 2 or newer

**Description** Compute p-values of EigenCorr1, EigenCorr2 and Tracy-Widom to select principal components for adjusting population stratification.

**Title** EigenCorr

**Author** Seunggeun, Lee <sglee@hsph.harvard.edu>

**Maintainer** Seunggeun, Lee <sglee@hsph.harvard.edu>

**Depends** R (>= 2.0.0)

**Suggests**

**URL** <http://www.r-project.org>

## R topics documented:

EigenCorr . . . . .	1
EigenCorr_Data . . . . .	3
Read_PCA . . . . .	3
<b>Index</b>	<b>5</b>

EigenCorr

*Compute scores and p-values of EigenCorrs***Description**

EigenCorr computes p-values of EigenCorr1, EigenCorr2 and Tracy-Widom, which can be used for selecting principal components to adjust population stratification.

**Usage**

```
EigenCorr( Y, PC, EVAL, N.Pval=100, N.Simu=10000)
```

**Arguments**

Y	an output phenotype vector.
PC	a matrix of principal components. The kth column should be the kth principal components. The order of rows and columns should be matched with Y and EVAL vectors. It need not to have all principal components.
EVAL	an vector of all eigenvalues. It should have all eigenvalues!!
N.Pval	an integer, the number of principal components whose p-values will be computed. If the number of principal components is smaller than N.Pval, p-values of all principal components (except last 2) will be computed (default=100).
N.Simu	an integer, the number of simulations to generate NULL distribution (default=10000).

**Value**

EigenCorr returns an dataframe object with 5 columns.

Gamma	Correlations between Y and PCs
EigenCorr_Score	EigenCorr scores. EigenCorr score is defined as multiplication of square correlation and eigenvalue.
EigenCorr1	P-values of EigenCorr1. They are computed based on the simulated null distribution.
EigenCorr2	P-values of EigenCorr2.
TW	P-values of eigenvalues based on the Tracy-Widom distribution.

**Author(s)**

Seunggeun Lee

**References**

Lee, S., Wright, F.A. and Zou, F. (2010) Control of population stratification by correlation-selected principal components. *Biometrics*, in press.

**Examples**

```
library(EigenCorr)
data(EigenCorr_Data) # EigenCorr example dataset

Y<-EigenCorr_Data$Y
PC<-EigenCorr_Data$PC
EVAL<-EigenCorr_Data$EVAL

EigenCorr( Y,PC,EVAL,N.Pval=50)
```

---

EigenCorr\_Data      *Example data for EigenCorr*

---

**Description**

This is an example dataset for EigenCorr. It contains outcome phenotypes, principal components and eigenvalues.

**Usage**

```
EigenCorr_Data
```

**Format**

EigenCorr\_Data contains the following objects:

**Y** a vector of the outcome phenotypes

**PC** a matrix of the principal components

**EVAL** a vector of the eigenvalues

---

Read\_PCA      *Read EigenStrat and ShrukenPCA output files*

---

**Description**

Read\_PCA reads EigenStrat and ShrukenPCA main output file. This file is formatted as: 1) First line indicates number of principal components (k), 2) Next k lines have top k eigenvalues, and 3) N additional lines for k principal components where N is the number of samples.

**Usage**

```
Read_PCA(filename)
```

**Arguments**

`filename` an EigenStrat or ShrukenPCA output file.

**Value**

Read\_PCA returns a list object.

`K` an integer, the number of principal components.

`EVAL` a eigen value vector.

`PC` a matrix of `K` principal components. The `k`th column is the `k`th principal component.

**Author(s)**

Seunggeun Lee

# Index

EigenCorr, [1](#)  
EigenCorr\_Data, [3](#)  
Read\_PCA, [3](#)