

# Package: abnormality (via r-universe)

September 6, 2024

**Type** Package

**Title** Measure a Subject's Abnormality with Respect to a Reference Population

**Version** 0.1.0

**Maintainer** Michael Marks <mmarks13@gmail.com>

**Description** Contains the functions to implement the methodology and considerations laid out by Marks et al. in the manuscript Measuring Abnormality in High Dimensional Spaces: Applications in Biomechanical Gait Analysis. As of 2/27/2018 this paper has been submitted and is under scientific review. Using high-dimensional datasets to measure a subject's overall level of abnormality as compared to a reference population is often needed in outcomes research. Utilizing applications in instrumented gait analysis, that article demonstrates how using data that is inherently non-independent to measure overall abnormality may bias results. A methodology is introduced to address this bias to accurately measure overall abnormality in high dimensional spaces. While this methodology is in line with previous literature, it differs in two major ways. Advantageously, it can be applied to datasets in which the number of observations is less than the number of features/variables, and it can be abstracted to practically any number of domains or dimensions. After applying the proposed methodology to the original data, the researcher is left with a set of uncorrelated variables (i.e. principal components) with which overall abnormality can be measured without bias. Different considerations are discussed in that article in deciding the appropriate number of principal components to keep and the aggregate distance measure to utilize.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 5.0.1

**Imports** MASS (>= 7.3.0), Matrix

**NeedsCompilation** no  
**Author** Michael Marks [aut, cre]  
**Date/Publication** 2018-03-13 17:42:56 UTC  
**Repository** https://mmarks13.r-universe.dev  
**RemoteUrl** https://github.com/mmmarks13/abnormality  
**RemoteRef** HEAD  
**RemoteSha** c30a9d7f5e660477b4715ad3e4728458179e4591

## Contents

generate_correlated_matrix . . . . .	2
overall_abnormality . . . . .	3
<b>Index</b>	<b>4</b>

---

generate\_correlated\_matrix  
*Generate a matrix of correlated variables*

---

## Description

Generate a matrix of correlated variables

## Usage

```
generate_correlated_matrix(n, p, corr, constant_cov_matrix = T, mean = 0)
```

## Arguments

n	number of observations
p	number of features/variables
corr	the correlation coefficient ( $-1 < r < 1$ )
constant_cov_matrix	should the value of corr be constant in the covariance matrix, or should corr be the average value in the covariance matrix.
mean	the mean value of the generated variables.

## Value

an  $n \times p$  matrix

## Examples

```
Subject <- generate_correlated_matrix(1, 100, corr = .75, constant_cov_matrix = TRUE)
Reference_Population <- generate_correlated_matrix(100, 100, corr = .75, constant_cov_matrix = TRUE)
```

---

overall\_abnormality    *Measure a Subject's Abnormality with Respect to a Reference Population*

---

### Description

Measure a Subject's Abnormality with Respect to a Reference Population

### Usage

```
overall_abnormality(Subj, Ref, stopping_rule = "Kaiser-Guttman",
  dist_measure = "MAD", TVE = 1, k = 2)
```

### Arguments

Subj	a vector of length n
Ref	an n x p matrix containing the reference population.
stopping_rule	the stopping rule to use when deciding the number of principal components to retain. Options include: c("Kaiser-Guttman", "brStick", "TVE").
dist_measure	the aggregate distance measure to use. Options include: c("MAD", "Euclidean", "Manhattan", "RMSE", "Lk-Norm")
TVE	a numeric value between 0 and 1. The minimum total variance explained for the retained principal components. This will only be used if "TVE" is chosen as the stopping_rule.
k	the value of k if Lk-Norm is chosen as a distance measure

### Value

An unbiased measure of overall abnormality of the subject as compared to the reference population based on the parameters supplied.

### Examples

```
p = 100
Subj <- rep(1, p)
Reference_Population <- generate_correlated_matrix(100, p, corr = 0.75, constant_cov_matrix = TRUE)
overall_abnormality(Subj, Reference_Population)
overall_abnormality(Subj, Reference_Population, dist_measure = "Euclidean")
overall_abnormality(Subj, Reference_Population, stopping_rule = "TVE", TVE = .90)
overall_abnormality(Subj, Reference_Population, dist_measure = "Lk-Norm", k = .5, stopping_rule = "brStick")
```

# Index

`generate_correlated_matrix`, [2](#)

`overall_abnormality`, [3](#)