

# Package: Gendis2unmix (via r-universe)

August 26, 2024

**Type** Package

**Title** Calculates a generalized discriminant function to unmix two classes, typically sexes of birds

**Version** 0.1.1

**Author** Cajo J.F. ter Braak

**Maintainer** Cajo J.F. ter Braak <cajo.terbraak@wur.nl>

**Description** The goal of Gendis2unmix is to sex birds from a population on the basis of several measurements. The key feature is that the birds from different populations may differ in size but that within populations females are smaller than males (or reversely). The predict function for a set of unsexed birds from a new population therefore estimates a new cutoff value which thus depends on the sizes of the birds in the new population. In the training phase, a generalized discriminant function (GDF) is calculated from a birds of known sex of different populations, in which the algorithm uses a common within-covariance matrix across populations and sexes. In the prediction phase Gendis2unmix then applies the GDF to measurements of individuals of unknow sex or class. The cutoff value is determined by unmixing the distribution in terms of two normal distributions with unequal means and variances using an EM algorithm. The parametric approach taken in Gendis2unmix make it suitable for small number of samples in both the training and prediction phase (say 20-100 per sex/population).

**Depends** R (>= 3.5.0)

**License** GPL-3 | file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.2.3

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

**RemoteUrl** <https://github.com/CajoterBraak/Gendis2unmix>

**RemoteRef** HEAD

**RemoteSha** 4770925420eca0c1a1672cb81165f6f6569950e1

## Contents

fulmarin . . . . .	2
gendis . . . . .	3
JanMayenBirds . . . . .	5
predict.gendis . . . . .	5
print.gendis . . . . .	7
summary.gendis . . . . .	8
unmix . . . . .	9
<b>Index</b>	<b>11</b>

---

fulmarin	<i>Fulmarin petrels data</i>
----------	------------------------------

---

## Description

The dataframe `fulmarin` contains measurements on Fulmarine petrels with sex known from dissection or, for Snow Petrels, observation. The variables are as follows:

- population study site ID (integer)
  - 1 Northern Fulmar (*Fulmarus glacialis*), the Netherlands
  - 2 Northern Fulmar (*Fulmarus glacialis*), Jan Mayen
  - 3 Southern Fulmar (*Fulmarus glacialoides*), Ardery Island Antarctica
  - 4 Cape Petrel (*Daption capense*), Ardery Island Antarctica
  - 5 Antarctic Petrel (*Thalassoica antarctica*), Ardery Island, Antarctica
  - 6 Snow Petrel (*Pagodroma nivea*), Casey Station, Antarctica
- sex 0 is female; 1 is male
- HB Head Length (mm)
- BD2 Bill Depth at gonys (mm)
- TL Tarsus Length (mm)
- CL Culmen Length (mm).

## Author(s)

Jan Andries van Franeker (<[jan.vanfraneke@wur.nl](mailto:jan.vanfraneke@wur.nl)>)

## References

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. *The Auk* 110: pp 492-502, <https://doi.org/10.2307/4088413>  
<https://edepot.wur.nl/249350>

---

gendis	<i>Calculates a generalized discriminant function</i>
--------	---

---

### Description

gendis calculates a generalized discriminant function to distinguish two classes, typically sexes (male and female birds) based on measurements of a number of indicators for individuals from each of the two sexes from a series of different populations in which individuals may have a different mean size but a common-within covariance matrix.

### Usage

```
gendis(
  population = "population",
  sex = "sex",
  measurements = "other_variables",
  verbose = FALSE,
  data
)
```

### Arguments

population	a name of the variable for the populations in the data (default "population")
sex	a name of the variable indicating the two classes to distinguish in the data (default "sex") (0 vs 1 or "female" vs "male")
measurements	character ("other_variables", default) or character vector with names of measurement variables. gendis maintains the order of the names.
verbose	logical (default = FALSE)
data	data frame with variables

### Value

An object of class gendis which is a named list, among which,

population	name of variable indicating populations
sex	name of variable indicating the two sexes or classes
classnames	names for the classes of sex (level or value)
measurements	names of the variables in the GDF
GDF	the Generalized Discriminant Function, matrix with two columns differing in scaling of the GDF
mean.male	overall mean of males (the second level of factor(sex))
mean.female	overall mean of females (the first level of factor(sex))
within.sd	overall within standard deviation
cov_overall	overall within-group covariance matrix

means.male	mean of males per population
means.female	mean of females per population
within.sds	within standard deviation per population
ind_mv	number of males and females per population
cov_list	within-group covariance matrix per population
Nind	number of individuals
Np	number of populations

## References

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. *The Auk* 110: pp 492-502, <https://doi.org/10.2307/4088413>  
<https://edepot.wur.nl/249350>

## See Also

[predict.gendis](#), [summary.gendis](#), [print.gendis](#).

## Examples

```
data("fulmarin")
names(fulmarin)
result <- gendis(population = "population", sex = "sex",
                 measurements = "other_variables", verbose = FALSE , data=fulmarin )
result$GDF
summary(result)
print(result)

# populations may have names:
fulmarin$pop <- factor(c("a1","a2","a3","a4","a5","a6")[fulmarin$population])
levels(fulmarin$pop)
names(fulmarin)
result2 <- gendis(population = "pop", sex = "sex",
                 measurements = c("HB","BD2","TL","CL"), verbose = FALSE , data=fulmarin )
# all equal should not give numeric differences.
#all.equal(result, result2)

result2$GDF - result$GDF
```

---

 JanMayenBirds

*Fulmarin petrels with unknown sex from Jan Mayen*


---

### Description

The data frame `JanMayenBirds` contains measurements on the Northern Fulmar petrels birds from the population at Jan Mayen. From the first 32 birds the sex is known by dissection, from the remaining 162 birds the sex is unknown.

- JAFCODE bird code (character)
- LOCATION location (character)
- DATE measurement date (character)
- DISSEX 0 is female; 1 is male
- HB Head Length (mm)
- BD2 Bill Depth at gonys (mm)
- TL Tarsus Length (mm)
- CL Culmen Length (mm).

### Author(s)

Jan Andries van Franeker (<[jan.vanfraneker@wur.nl](mailto:jan.vanfraneker@wur.nl)>)

### References

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. *The Auk* 110: pp 492-502, <https://doi.org/10.2307/4088413>  
<https://edepot.wur.nl/249350>

---

 predict.gendis

*Predict function using a generalized discriminant function*


---

### Description

`predict.gendis` applies a generalized discriminant function created with `gendis` to predict the sex (class) of each individual with measurements in `newdata`. From the `gendis` object, the coefficients that define the generalized discriminant function (GDF) are applied to the `newdata` to obtain the discriminant scores.

### Usage

```
## S3 method for class 'gendis'
predict(object, newdata, type = object$sex, verbose = FALSE, ...)
```

**Arguments**

object	an object of class <code>gendis</code> , typically created with <code>gendis</code>
newdata	a data frame with measurements on (new) individuals with variables used to create object. The data should be from a single population. If your data are from multiple populations, use <code>predict</code> for each subset (i.e. for each population).
type	what to predict: the sex or class of each individual (default), the generalized discriminant scores with cutpoint ("GDF" or "GDFscore") or the full output of the unmixing algorithm <code>unmix</code> ("cutpoint")
verbose	logical (default = FALSE). If TRUE a plot of the density of the GDF is produced.
...	other optional arguments

**Details**

The discriminant score are a linear combination of the variables in `newdata` that are shared with the variables used to create the object. The linear combination is defined by the GDF coefficients. The discriminant scores are subjected to an unmixing algorithm. This algorithm (`unmix`) generates a cutpoint below which individuals are predicted to be female (level 1 of `factor(sex)`) and above which they are predicted to be male (level 2 of `factor(sex)`). The cutpoint is at the point of intersection of two normal densities with unequal means and variances fitted to the discriminant scores (see `unmix` for details).

**Value**

See argument `type`.

**References**

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. *The Auk* 110: pp 492-502, <https://doi.org/10.2307/4088413>  
<https://edepot.wur.nl/249350>

**Examples**

```
data("fulmarin")
str(fulmarin)
result <- gendis(population = "population", sex = "sex",
                 measurements = "other_variables", verbose = FALSE , data=fulmarin )
data("JanMayenBirds")
sex.predict <- predict(result, newdata = JanMayenBirds, verbose = TRUE)
# one false prediction: (number 32)
data.frame(sex = JanMayenBirds$DISSEX, sex.predict)[seq(from=2, to = 37, by =5),]

predict(result, JanMayenBirds )
# same as default above
predict(result, JanMayenBirds, type = result$sex, verbose = FALSE)
# GDF score with cutpoint
predict(result, JanMayenBirds, type = "GDF", verbose = FALSE)
# unmix results only
predict(result, JanMayenBirds, type = "cutpoint", verbose = TRUE)
```

---

print.gendis	<i>Printing results a generalized discriminant analysis</i>
--------------	---

---

## Description

print.gendis prints the results of [gendis](#) in more detail than `summary.gendis`.

## Usage

```
## S3 method for class 'gendis'  
print(x, ...)
```

## Arguments

x                    an object of class `gendis`, created by [gendis](#).  
...                   other optional arguments

## Value

list of within-sex correlations matrices per population (invisible)

## References

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. *The Auk* 110: pp 492-502ter Braak (2019)

## See Also

[gendis](#), [summary.gendis](#), [predict.gendis](#).

## Examples

```
data("fulmarin")  
names(fulmarin)  
result <- gendis(population = "population", sex = "sex",  
                  measurements = "other_variables", verbose = FALSE , data=fulmarin )  
result$GDF  
summary(result)  
print(result)  
  
# populations may have names:  
fulmarin$pop <- factor(c("a1","a2","a3","a4","a5","a6")[fulmarin$population])  
levels(fulmarin$pop)  
names(fulmarin)  
result2 <- gendis(population = "pop", sex = "sex",
```

```
        measurements = c("HB","BD2","TL","CL"), verbose = FALSE , data=fulmarin )
# all equal should not give numeric differences.
#all.equal(result, result2)

result2$GDF - result$GDF
```

---

summary.gendis

*Summary of a generalized discriminant analysis*

---

## Description

summary.gendis summarizes the results of [gendis](#).

## Usage

```
## S3 method for class 'gendis'
summary(object, ...)
```

## Arguments

object            an object of class gendis, created by [gendis](#).  
...                other optional arguments.

## Value

GDF

## References

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. The Auk 110: pp 492-502  
ter Braak (2019)

## See Also

[gendis](#), [print.gendis](#), [predict.gendis](#).

## Examples

```
data("fulmarin")
names(fulmarin)
result <- gendis(population = "population", sex = "sex",
                 measurements = "other_variables", verbose = FALSE , data=fulmarin )
result$GDF
summary(result)
print(result)
```



```

# populations may have names:
fulmarin$pop <- factor(c("a1", "a2", "a3", "a4", "a5", "a6")[fulmarin$population])
levels(fulmarin$pop)
names(fulmarin)
result2 <- gendis(population = "pop", sex = "sex",
                  measurements = c("HB", "BD2", "TL", "CL"), verbose = FALSE, data=fulmarin)
# all equal should not give numeric differences.
#all.equal(result, result2)

result2$GDF - result$GDF

```

---

unmix	<i>Unmixing a distribution by decomposing it in two normal ones with unequal means and variances.</i>
-------	---

---

## Description

unmix generates a cutpoint below which individuals are predicted to be female (level 1 of factor(sex)) and above which they are predicted to be male (level 2 of factor(sex)). The cutpoint is at the point of intersection of two normal densities with unequal means and variances fitted to argument x. This function is used internally in the `predict.gendis` function.

## Usage

```
unmix(x, verbose = FALSE)
```

## Arguments

x	a numeric vector of discriminant scores with optional attribute "classnames", e.g. c("female", "male")
verbose	logical (default = FALSE)

## Details

unmix is an EM algorithm following example 4.3.2 of Titterton et al. (1985). Alternatively, library flexmix could have been used.

## Value

A list consisting of

- cutpoint point of equal density of the normal distributions
- p1 estimated probability of class 0 ("female"), informally: fraction of individuals in class 0
- p2 estimated probability of class 1 ("female"), informally: fraction of individuals in class 0
- m1 estimated mean of the normal distribution of class 0

- $m_2$  estimated mean of the normal distribution of class 1
- $v_1$  estimated variance of the normal distribution of class 0
- $v_2$  estimated variance of the normal distribution of class 1

## References

Titterton, D.M., Smith, A.F.M. & Makov, U.E. (1985). Statistical analysis of finite mixture distributions, Wiley, 1985. pages 86/87, example 4.3.2

van Franeker, J A. ter Braak, C J F. 1993. A generalized discriminant for sexing fulmarine petrels from external measurements. The Auk 110: pp 492-502, <https://doi.org/10.2307/4088413>  
<https://edepot.wur.nl/249350>

## Examples

```
data("fulmarin")
result <- gendis(population = "population", sex = "sex",
                 measurements = c("HB", "BD2", "TL", "CL"), verbose = FALSE , data=fulmarin )
data("JanMayenBirds")
#get the measurements in the generalized discriminant function (GDF) from the new data
newdata <- as.matrix(JanMayenBirds[, c("HB", "BD2", "TL", "CL")])
# combine the measurements using the coefficients of the GDF
GDFscores <- newdata%% result$GDF[,2]
attr(GDFscores,which = "classnames") <- result$classnames
# note the attribute classnames with the names to be used in the printout
# for first and second level of the factor sex
# Calculate the cutpoint using unmix instead of predict.gendis
unmix(GDFscores,verbose = TRUE)
```

# Index

fulmarin, [2](#)

gendis, [3](#), [5–8](#)

JanMayenBirds, [5](#)

predict.gendis, [4](#), [5](#), [7–9](#)

print.gendis, [4](#), [7](#), [8](#)

summary.gendis, [4](#), [7](#), [8](#)

unmix, [6](#), [9](#)