

Oefeningen Discriminantanalyse

Vraag 1: Zoek in de Canvas-cursus de dataset **bordeaux** en plaats die in een dataframe.

Ter info:

bordeaux	<i>Bordeaux Wines Dataset</i>
----------	-------------------------------

Description

Quality measures of wines from Bordeaux, France

Format

A data frame with 34 observations on the following 6 variables.

year	year of harvest
temperature	sum of daily average temperatures (in celsius degrees)
sun	duration of insolation (in hours)
heat	number of super-hot days
rain	rain level (in millimeters)
quality	wine quality: a factor with levels bad, good, and medium

References

Chapter 10: Analyse Discriminante, page 353.
Tenenhaus M. (2007) *Statistique*. Dunod, Paris.

- Pas een discriminatanalyse toe op deze data set waarbij *quality* als afhankelijke variabele optreedt en *temperature*, *sun*, *heat* en *rain* als de onafhankelijke variabelen.
- Hoeveel dimensies heeft de discriminantanalyse?
- Maak een plot van de resultaten van de discriminantanalyse. Wat wordt er hier gevisualiseerd?
- Pas het predict commando even toe op de oorspronkelijke gegevens. Hoeveel percent van de wijnen krijgen dezelfde label opgeplakt via de discriminantanalyse als geobserveerd?

Vraag 2: Zoek in de Canvas-cursus de dataset **birthwt** en plaats die in een dataframe.

Ter info:

birthwt	<i>Risk Factors Associated with Low Infant Birth Weight</i>
---------	---

Description
The birthwt data frame has 189 rows and 10 columns. The data were collected at Baystate Medical Center, Springfield, Mass during 1986.
Usage
birthwt
Format
This data frame contains the following columns:
low indicator of birth weight less than 2.5 kg.
age mother's age in years.
lwt mother's weight in pounds at last menstrual period.
race mother's race (1 = white, 2 = black, 3 = other).
smoke smoking status during pregnancy.
ptl number of previous premature labours.
ht history of hypertension.
ui presence of uterine irritability.
ftv number of physician visits during the first trimester.
bwt birth weight in grams.
Source
Hosmer, D.W. and Lemeshow, S. (1989) <i>Applied Logistic Regression</i> . New York: Wiley
References
Venables, W. N. and Ripley, B. D. (2002) <i>Modern Applied Statistics with S</i> . Fourth edition. Springer.

- Analyseer de type van gegevens in deze dataset.
- Welke kolommen kunnen er als afhankelijke variabele voor een discriminant analyse worden gebruikt?
- Welke kolommen kunnen er als onafhankelijke variabelen voor een discriminant analyse worden gebruikt?
- Pas een discriminatanalyse toe op deze data set (maakt gebruik van je keuzes hierboven).
- Hoeveel dimensies heeft de discriminantanalyse?
- Maak een plot van de resultaten van de discriminantanalyse.

Vraag 3: Zoek op Canvas de dataset Cars93 en plaats die in een dataframe.

Ter info:

Cars93

Data from 93 Cars on Sale in the USA in 1993

Description

The Cars93 data frame has 93 rows and 27 columns.

Usage

Cars93

Format

This data frame contains the following columns:

Manufacturer Manufacturer.

Model Model.

Type Type: a factor with levels "Small", "Sporty", "Compact", "Midsize", "Large" and "Van".

Min.Price Minimum Price (in \(\$1,000): price for a basic version.

Price Midrange Price (in \(\$1,000): average of Min.Price and Max.Price.

Max.Price Maximum Price (in \(\$1,000): price for "a premium version".

MPG.city City MPG (miles per US gallon by EPA rating).

MPG.highway Highway MPG.

AirBags Air Bags standard. Factor: none, driver only, or driver & passenger.

DriveTrain Drive train type: rear wheel, front wheel or 4WD; (factor).

Cylinders Number of cylinders (missing for Mazda RX-7, which has a rotary engine).

EngineSize Engine size (litres).

Horsepower Horsepower (maximum).

RPM RPM (revs per minute at maximum horsepower).

Rev.per.mile Engine revolutions per mile (in highest gear).

Man.trans.avail Is a manual transmission version available? (yes or no, Factor).

Fuel.tank.capacity Fuel tank capacity (US gallons).

Passengers Passenger capacity (persons)

Length Length (inches).
 Wheelbase Wheelbase (inches).
 Width Width (inches).
 Turn.circle U-turn space (feet).
 Rear.seat.room Rear seat room (inches) (missing for 2-seater vehicles).
 Luggage.room Luggage capacity (cubic feet) (missing for vans).
 Weight Weight (pounds).
 Origin Of non-USA or USA company origins? (factor).
 Make Combination of Manufacturer and Model (character).

Details

Cars were selected at random from among 1993 passenger car models that were listed in both the *Consumer Reports* issue and the *PACE Buying Guide*. Pickup trucks and Sport/Utility vehicles were eliminated due to incomplete information in the *Consumer Reports* source. Duplicate models (e.g., Dodge Shadow and Plymouth Sundance) were listed at most once.

Further description can be found in Lock (1993).

Source

Lock, R. H. (1993) 1993 New Car Data. *Journal of Statistics Education* 1(1). <http://www.amstat.org/publications/jse/v1n1/datasets.lock.html>.

References

Venables, W. N. and Ripley, B. D. (1999) *Modern Applied Statistics with S-PLUS*. Third Edition. Springer.

-
- Analyseer de type van gegevens in deze dataset.
 - Splits de data set in twee: de eerste 90 rijen als 'leer data set' waarom je een discriminant analyse gaat toepassen (zie e) en de laatste 3 rijen om na te gaan of de bekomen discriminant analyse deze auto's in de juiste klasse plaatst (zie g).
 - Welke kolommen kunnen er als afhankelijke variabele voor een discriminant analyse worden gebruikt?
 - Welke kolommen kunnen er als onafhankelijke variabelen voor een discriminant analyse worden gebruikt?
 - Pas een discriminatanalyse toe op deze data set. Neem als afhankelijke variabele de kolom Type en als onafhankelijke variabelen alle bruikbare kolomen (zie d).
 - Hoeveel dimensies heeft de discriminantanalyse?
 - Pas de bekomen discriminatanalyse met behulp van predict-commando toe op de laatste 3 rijen van de oorspronkelijke gegevens (zie b). Worden alle 3 de wagens in de juiste klasse geplaatst?