# Package 'dobson'

October 13, 2022

Type Package

Title Data from the GLM Book by Dobson and Barnett				
Version 0.4				
<b>Description</b> Example datasets from the book ``An Introduction to Generalised Linear Models" (Year: 2018, ISBN:9781138741515) by Dobson and Barnett.				
License GPL-2				
Encoding UTF-8				
LazyData true				
<b>Depends</b> R (>= $2.10$ )				
RoxygenNote 6.0.1				
NeedsCompilation no				
Author Adrian Barnett [aut, cre]				
Maintainer Adrian Barnett <a.barnett@qut.edu.au></a.barnett@qut.edu.au>				
Repository CRAN				
<b>Date/Publication</b> 2018-11-20 05:30:22 UTC				
R topics documented:				
achievement       2         aids       3         anthers       4         balanced       4         beetle       5         birthweight       6         carbohydrate       6         Cars       7				

2 achievement

Index		<b>29</b>
	waist	28
	vaccine	
	unbalanced	27
	ulcer	26
	tumor	25
	survival	25
	sugar	24
	stroke.wide	23
	senility	22
	remission	22
	poisson	21
	PLOS	21
	plasma	20
	plants	20
	plant.dried	19
	pasture	19
	moths	18
	mortality	17
	machine	10
	leukemia	16 16
	insurance	15
	housing	14
	hiroshima	14
	hepatitis	13
	graduates	
	failure	12
	ear	11
	dogs	10

achievement

Achievement data from table 6.15

# Description

Achievement scores after three training methods

# Usage

data(achievement)

aids 3

#### **Format**

```
A tibble with 21 observations and the following 3 variables. method training method (A, B or C) y achievement scores
```

x aptitude scores measured before training commenced

#### References

```
Winer, B. J. (1971). Statistical Principles in Experimental Design (2nd ed.).
```

# **Examples**

```
data(achievement)
summary(achievement)
```

aids

AIDS data from table 4.5

# Description

Numbers of cases of AIDS in Australia by date of diagnosis for successive 3-month periods from 1984 to 1988

## Usage

```
data(aids)
```

## **Format**

A tibble with 20 observations and the following 3 variables.

```
year year
quarter quarter of year
cases number of cases
```

#### **Source**

National Centre for HIV Epidemiology and Clinical Research 1994

```
data(aids)
summary(aids)
```

4 balanced

anthers

Embryogenic anthers data from table 7.2

# Description

Numbers of embryogenic anthers of the plant species Datura innoxia Mill obtained when anthers were prepared under several different conditions

# Usage

```
data(anthers)
```

#### **Format**

A tibble with 6 observations and the following 4 variables.

y numbers of embryogenic anthers

n number of anthers

storage storage condition, control or treatment

centrifuge centrifuging force (g)

#### References

Sangwan-Norrell, B. S. (1977). Androgenic stimulating factor in the anther and isolated pollen grain culture of Datura innoxia mill. *Journal of Experimental Biology* 28, 843–852.

# **Examples**

```
data(anthers)
summary(anthers)
```

balanced

Balanced data from table 6.12

## **Description**

Fictitious balanced data for a two-factor ANOVA with equal numbers of observations in each subgroup

## Usage

data(balanced)

beetle 5

## **Format**

```
A tibble with 12 observations and the following 3 variables.
```

```
factorA factor A factorB factor B data dependent data
```

# **Examples**

```
data(balanced)
summary(balanced)
```

beetle

Beetle data from table 7.2

# Description

Numbers of beetles dead after five hours exposure to gaseous carbon disulphide at various concentrations

# Usage

```
data(beetle)
```

## **Format**

A tibble with 6 observations and the following 3 variables.

```
x dose (log base 10 CS2mgl^-1)n number of beetles
```

y numbers killed

# References

Bliss, C. I. (1935). The calculation of the dose-mortality curve. *Annals of Applied Biology* 22, 134–167.

```
data(beetle)
summary(beetle)
```

6 carbohydrate

birthweight

Birthweight data from table 2.3

#### Description

Birthweight and gestational age for twelve boys and girls

## Usage

```
data(birthweight)
```

#### **Format**

A tibble with 12 observations and the following 4 variables.

```
boys gestational age boys gestational age (weeks)
boys weight boys birthweight (grams)
girls gestational age girls gestational age (weeks)
girls weight girls birthweight (grams)
```

# Examples

```
data(birthweight)
summary(birthweight)
```

carbohydrate

Carbohydrate data from table 6.3

# Description

Percentages of total calories obtained from complex carbohydrates, for twenty male insulin-dependent diabetics who had been on a high-carbohydrate diet for six months.

# Usage

```
data(carbohydrate)
```

#### **Format**

```
A tibble with 20 observations and the following 4 variables.
```

```
carbohydrate percent of total calories obtained from complex carbohydrates age age in years
weight body weight relative to "ideal" weight for height
protein percentage of calories as protein
```

Cars 7

## **Source**

K. Webb

# **Examples**

```
data(carbohydrate)
summary(carbohydrate)
```

Cars

Cars data from table 8.1

# Description

Preferences for air conditioning and power steering in cars by gender and age.

# Usage

```
data(Cars)
```

#### **Format**

A tibble with 18 observations and the following 4 variables.

```
sex sex
age age group
response ordinal response
frequency frequency
```

#### References

McFadden, M., J. Powers, W. Brown, and M. Walker (2000). Vehicle and driver attributes affecting distance from the steering wheel in motor vehicles. *Human Factors* 42, 676–682.

```
data(Cars)
summary(Cars)
```

8 chronic

cholesterol

Cholesterol data from table 6.24

#### **Description**

Cholesterol, age and BMI for thirty women.

## Usage

```
data(cholesterol)
```

#### **Format**

A tibble with 30 observations and the following 3 variables.

```
chol serum cholesterol (millimoles per liter)
age age (years)
bmi body mass index (kg/m2)
```

## **Examples**

```
data(cholesterol)
summary(cholesterol)
```

chronic

Chronic health data from table 2.7

#### **Description**

Numbers of chronic medical conditions reported by samples of women living in large country towns (town group) or in more rural areas (country group) in New South Wales, Australia

# Usage

```
data(chronic)
```

#### **Format**

```
A data frame with 49 observations and the following 2 variables.
```

```
place place (town or country) number number of conditions
```

```
data(chronic)
summary(chronic)
```

cyclones 9

cyclones

Cyclone data from table 1.2

# Description

The number of tropical cyclones during a season from November to April in Northeastern Australia

# Usage

```
data(cyclones)
```

#### **Format**

A tibble with 13 observations and the following 3 variables.

```
years season years
season season number
number of cyclones
```

#### References

Dobson AJ and Stewart J (1974). Frequencies of tropical cyclones in the northeastern Australian area. *Australian Meteorological Magazine* 22, 27–36.

# **Examples**

```
data(cyclones)
summary(cyclones)
```

dobson

dobson: Example datasets from the book "An Introduction to Generalised Linear Models" (4th edition)

# Description

datasets from our book

10 dogs

doctors

Doctors data from table 9.1

# Description

Data from the famous doctors study of smoking conducted by Sir Richard Doll and colleagues

## Usage

```
data(doctors)
```

#### **Format**

A tibble with 10 observations and the following 4 variables.

```
age age group
smoking smoker or non-smoker
deaths number of deaths
person-years person years of of observation at the time of the analysis
```

#### References

Breslow, N. E. and N. E. Day (1987). *Statistical Methods in Cancer Research, Volume 2: The Design and Analysis of Cohort Studies.* Lyon: International Agency for Research on Cancer.

## **Examples**

```
data(doctors)
summary(doctors)
```

dogs

Dogs data from table 11.9

# Description

Measurements of left ventricular volume and parallel conductance volume on five dogs under eight different load conditions

#### Usage

```
data(dogs)
```

ear 11

#### **Format**

A tibble with 40 observations and the following 4 variables.

```
dog dog number
condition load condition
y left ventricular volume
x parallel conductance volume
```

#### References

Boltwood, C. M., R. Appleyard, and S. A. Glantz (1989). Left ventricular volume measurement by conductance catheter in intact dogs: the parallel conductance volume increases with end-systolic volume. *Circulation* 80, 1360–1377.

## **Examples**

```
data(dogs)
summary(dogs)
```

ear

Ears data from table 11.10

#### **Description**

Numbers of ears clear of acute otitis media at 14 days by antibiotic treatment and age of the child. The children had acute otitis media in both ears.

#### Usage

```
data(ear)
```

#### **Format**

A tibble with 18 observations and the following 4 variables.

```
age child's age
treatment two treatments coded CEF and AMO
number clear number of clear ears
frequency faculty
```

#### Source

Rosner, B. (1989). Multivariate methods for clustered binary data with more than one level of nesting. *Journal of the American Statistical Association* 84, 373–380.

```
data(ear)
summary(ear)
```

12 graduates

failure

Failure time data from table 4.1

#### **Description**

Lifetimes of Kevlar epoxy strand pressure vessels at 70

#### Usage

```
data(failure)
```

#### **Format**

A tibble with 49 observations and the following variable.

lifetimes time to failure in hours

#### References

Andrews, D. F. and A. M. Herzberg (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer Verlag.

## **Examples**

```
data(failure)
summary(failure)
```

graduates

Graduate survival data from tables 7.16 and 7.17

#### **Description**

Survival 50 years after graduation of men and women who graduated each year from 1938 to 1947 from various Faculties of the University of Adelaide.

## Usage

```
data(graduates)
```

#### **Format**

A tibble with 60 observations and the following 5 variables.

```
year year of graduation
survive number of graduates who survived
total total number of graduates
faculty faculty
sex sex
```

hepatitis 13

## Source

J.A. Keats

# **Examples**

```
data(graduates)
summary(graduates)
```

hepatitis

Hepatitis data from table 10.5

## **Description**

Survival times in months of patients with chronic active hepatitis in a randomized controlled trial of prednisolone versus no treatment

## Usage

```
data(hepatitis)
```

#### **Format**

A tibble with 44 observations and the following 3 variables.

```
survival time survival time in months
censor censored, lost to follow up or died
group prednisolone or no treatment
```

# References

Altman DG, Bland JM (1998). Statistical notes: times to event (survival) data. *British Medical Journal* 317, 468–469.

```
data(hepatitis)
summary(hepatitis)
```

14 housing

hiroshima

Hiroshima data from table 7.14

## **Description**

The number of deaths from leukemia and other cancers among survivors of the Hiroshima atom bomb. The data are for deaths during the period 1950–1959 among survivors who were aged 25 to 64 years in 1950.

## Usage

```
data(hiroshima)
```

#### **Format**

A tibble with 6 observations and the following 4 variables.

```
radiation radiation dose (rads)
leukemia leukemia deaths
other cancer deaths from other cancers
total cancers total cancer deaths
```

#### References

Cox, D. R. and E. J. Snell (1981). *Applied Statistics: Principles and Examples*. London: Chapman & Hall.

Otake, M. (1979). Comparison of time risks based on a multinomial logistic response model in longitudinal studies. Technical Report No. 5, RERF, Hiroshima, Japan.

#### **Examples**

```
data(hiroshima)
summary(hiroshima)
```

housing

Housing data from table 8.5

#### **Description**

Data from an investigation into satisfaction with housing conditions in Copenhagen

#### Usage

```
data(housing)
```

insurance 15

#### **Format**

A tibble with 18 observations and the following 4 variables.

```
type housing type; tower block, apartment or house satisfaction satisfaction; low, medium or high contact contact with other residents; low or high frequency frequency
```

#### References

Madsen, M. (1971). Statistical analysis of multiple contingency tables. two examples. *Scandinavian Journal of Statistics* 3, 97–106.

## **Examples**

```
data(housing)
summary(housing)
```

insurance

Insurance data from table 9.13

# Description

Insurance claim data by car category, age group and district.

## Usage

```
data(insurance)
```

#### **Format**

A tibble with 32 observations and the following 5 variables.

```
car car insurance category
age age group
district district where policy holder lived; 1=major city, 0=elsewhere
y number of claims
n number of insurance policies
```

#### References

Baxter, L. A., S. M. Coutts, and G. A. F. Ross (1980). Applications of linear models in motor insurance. Zurich, pp. 11–29. *Proceedings of the 21st International Congress of Actuaries*.

```
data(insurance)
summary(insurance)
```

16 machine

leukemia

Leukemia data from table 4.6

## **Description**

Survival times and white blood cell count for seventeen patients suffering from leukemia

## Usage

```
data(leukemia)
```

#### **Format**

A tibble with 17 observations and the following 2 variables.

```
time time to death in weeks
wbc log base 10 initial white blood cell count
```

#### References

Cox, D. R. and E. J. Snell (1981). *Applied Statistics: Principles and Examples*. London: Chapman & Hall.

# **Examples**

```
data(leukemia)
summary(leukemia)
```

machine

Machine data from table 6.26

# Description

Weights of machine components made by workers on different days

# Usage

```
data(machine)
```

#### **Format**

A tibble with 44 observations and the following 3 variables.

```
day day number 1 or 2
worker worker nunber 1 to 4
weight weight in grams
```

melanoma 17

#### **Examples**

```
data(machine)
summary(machine)
```

melanoma

Melanoma data from table 9.4

## **Description**

A cross-sectional study of patients with a form of skin cancer called malignant melanoma

## Usage

```
data(melanoma)
```

#### **Format**

A tibble with 12 observations and the following 3 variables.

```
type tumor type
site site of cancer
frequency frequency
```

## References

Roberts, G., A. L. Martyn, A. J. Dobson, and W. H. McCarthy (1981). Tumour thickness and histological type in malignant melanoma in New South Wales, Australia, 1970–76. *Pathology* 13, 763–770.

# **Examples**

```
data(melanoma)
summary(melanoma)
```

mortality

Mortality data from table 3.2

## **Description**

Numbers of deaths from coronary heart disease and population sizes by 5-year age groups for men in the Hunter region of New South Wales, Australia in 1991.

#### Usage

```
data(mortality)
```

18 moths

## **Format**

```
A tibble with 8 observations and the following 3 variables.

age group age group (years)

deaths number of deaths
```

# **Examples**

```
data(mortality)
summary(mortality)
```

population population size

moths

Moths data from table 1.4

## **Description**

Numbers of females and males in the progeny of 16 female light brown apple moths in Muswell-brook, New South Wales, Australia

## Usage

```
data(moths)
```

#### **Format**

A tibble with 16 observations and the following 3 variables.

```
group progeny group
females number of females
males number of males
```

# References

Lewis T (1987). Uneven sex ratios in the light brown apple moth: a problem in outlier allocation. In D. J. Hand and B. S. Everitt (Eds.), *The Statistical Consultant in Action*. Cambridge: Cambridge University Press.

```
data(moths)
summary(moths)
```

pasture 19

pasture

Pasture data from table 6.23

# Description

Response of a grass and legume pasture system to various quantities of phosphorus fertilizer

## Usage

```
data(pasture)
```

#### **Format**

A tibble with 27 observations and the following 2 variables.

K phosphorus levels (kilograms per hectare) yield total yield of grass and legume together (kilograms per hectare)

#### **Source**

D. F. Sinclair

## **Examples**

```
data(pasture)
summary(pasture)
```

plant.dried

Plant data from table 6.9

# Description

Dried weights of plants from three different growing conditions in long format

#### Usage

```
data(plant.dried)
```

#### **Format**

```
A tibble with 30 observations and the following 2 variables. group one of three treatment groups weight dried weight of plants
```

```
data(plant.dried)
summary(plant.dried)
```

20 plasma

plants

Plant weight data from table 2.7

# Description

Dried weight of plants grown under two conditions.

## Usage

```
data(plants)
```

#### **Format**

A tibble with 20 observations and the following 2 variables.

treatment weights of treatment plants in grams control weights of control plants in grams

# **Examples**

```
data(plants)
summary(plants)
```

plasma

Plasma phosphate data from table 6.25

## **Description**

Plasma phosphate levels in obese and control participants one hour after a standard glucose tolerance test.

## Usage

```
data(plasma)
```

#### **Format**

A tibble with 31 observations and the following 2 variables.

Group group; H-O=Hyperinsulinemic obsese, N-O=Non-hyperinsulinemic obese or C=Control phosphate plasma inorganic phosphate level (mg/dl)

```
data(plasma)
summary(plasma)
```

PLOS 21

**PLOS** 

PLOS Medicine data from figure 6.7

# Description

Data from 878 journal articles published in PLOS Medicine between 2011 and 2015

# Usage

```
data(PLOS)
```

#### **Format**

```
A data. frame with 878 observations and the following 2 variables. nchar title length authors number of authors, truncated to 30
```

# **Examples**

```
data(PLOS)
summary(PLOS)
```

poisson

Poisson data from table 4.3

## **Description**

Artificial data for a Poisson regression example

## Usage

```
data(poisson)
```

#### **Format**

A tibble with 9 observations and the following two variables.

```
x covariate
```

y dependent counts

```
data(poisson)
summary(poisson)
```

22 senility

remission

Remission data from table 10.1

## **Description**

Times to remission of leukemia patients

# Usage

```
data(remission)
```

#### **Format**

A tibble with 42 observations and the following 3 variables.

```
time time in weeks
group group; C=control, T=treatment
censored censored; 0=No, 1=Yes
```

#### References

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52, 203–223.

#### **Examples**

```
data(remission)
summary(remission)
```

senility

Senility data from table 7.8

# Description

Data from a sample of elderly people given a psychiatric examination to determine whether symptoms of senility were present together with their score on a subset of the Wechsler Adult Intelligent Scale (WAIS).

## Usage

```
data(senility)
```

# Format

A tibble with 54 observations and the following 2 variables.

- x WAIS score
- s symptoms of senility present; 1=yes, 0=no

stroke.wide 23

#### **Examples**

```
data(senility)
summary(senility)
```

stroke.wide

Stroke data from table 11.1

## **Description**

Longitudinal data from an experiment to promote the recovery of stroke patients in wide format. The response variable is the Bartel index with higher scores meaning better outcomes and a maximum score of 100.

## Usage

```
data(stroke.wide)
```

#### **Format**

A tibble with 24 observations and the following 10 variables.

```
Subject subject number
```

Group group; A=new occupational therapy intervention, B = existing stroke rehabilitation program in the same hospital as A, C = usual care in a different hospital

week1 Bartel index in week 1

week2 Bartel index in week 2

week3 Bartel index in week 3

week4 Bartel index in week 4

week5 Bartel index in week 5

week6 Bartel index in week 6

week7 Bartel index in week 7

week8 Bartel index in week 8

#### **Source**

C. Cropper, University of Queensland

24 sugar

#### **Examples**

```
data(stroke.wide)
summary(stroke.wide)

# To transform data from wide to long format use
## Not run:
library(reshape2)
stroke = melt(data=stroke.wide, id.vars=c('Subject','Group'),
   value.name='ability', variable.name='week')
stroke$time = as.numeric(gsub('week', '', stroke$week))

## End(Not run)
```

sugar

Sugar data from table 6.22

# Description

Average apparent per capita consumption of sugar (in kg per year) in Australia, as refined sugar and in manufactured foods

# Usage

```
data(sugar)
```

## **Format**

```
A tibble with 6 observations and the following 3 variables.

period period in years

refined refined sugar

manufactured Sugar in manufactured food
```

# Source

Australian Bureau of Statistics 1998

```
data(sugar)
summary(sugar)
```

survival 25

survival

Survival data from table 10.1

## **Description**

Survival times for leukemia patients

#### Usage

```
data(survival)
```

#### **Format**

A tibble with 33 observations and the following 3 variables.

```
survival time survival time in weeks WBC white blood cell count
AG test result; +=positive, -=negative
```

#### References

Feigl, P. and M. Zelen (1965). Estimation of exponential probabilities with concomitant information. *Biometrics* 21, 826–838.

## **Examples**

```
data(survival)
summary(survival)
```

tumor

Tumor data from table 8.6

## **Description**

Tumor responses of male and female patients receiving treatment for small-cell lung cancer

#### Usage

```
data(tumor)
```

#### **Format**

A tibble with 16 observations and the following 4 variables.

```
treatment treatment; sequential or alternating
sex sex
response four category ordinal response
frequency
```

26 ulcer

#### References

Holtbrugger, W. and M. Schumacher (1991). A comparison of regression models for the analysis of ordered categorical data. *Applied Statistics* 40, 249–259.

#### **Examples**

```
data(tumor)
summary(tumor)
```

ulcer

Ulcer data from table 9.7

## **Description**

Data from a retrospective case-control study. A group of ulcer patients was compared with a group of control patients not known to have peptic ulcer, but who were similar to the ulcer patients with respect to age, sex and socioeconomic status.

## Usage

```
data(ulcer)
```

#### **Format**

A tibble with 8 observations and the following 4 variables.

```
ulcer type of ulcer
case-control case or control
aspirin aspirin user
frequency frequency
```

#### References

Duggan, J. M., A. J. Dobson, H. Johnson, and P. P. Fahey (1986). Peptic ulcer and non-steroidal anti-inflammatory agents. *Gut* 27, 929–933.

```
data(ulcer)
summary(ulcer)
```

unbalanced 27

unbalanced

Unbalanced data from table 6.27

## **Description**

Unbalanced data from a fictitious two-factor experiment

## Usage

```
data(unbalanced)
```

#### **Format**

A tibble with 10 observations and the following 3 variables.

factorA factor A factorB factorB data dependent data

# **Examples**

data(unbalanced)
summary(unbalanced)

vaccine

Vaccine data from table 9.6

# **Description**

Data from a vaccine trial.

## Usage

```
data(vaccine)
```

## **Format**

A tibble with 6 observations and the following 3 variables.

```
treatment treatment group response response to treatment frequency frequency
```

## Source

R.S. Gillett

28 waist

#### **Examples**

```
data(vaccine)
summary(vaccine)
```

waist

Waist loss data from table 2.8

# Description

The weights, in kilograms, of twenty men before and after participation in a "waist loss" program

## Usage

```
data(waist)
```

## **Format**

A tibble with 20 observations and the following 3 variables.

```
man man number
before weight before in kgs
after weight after in kgs
```

#### References

Egger, G., G. Fisher, S. Piers, K. Bedford, G. Morseau, S. Sabasio, B. Taipim, G. Bani, M. Assan, and P. Mills (1999). Abdominal obesity reduction in Indigenous men. *International Journal of Obesity* 23, 564–569.

```
data(waist)
summary(waist)
```

# **Index**

* datasets	unbalanced, 27
achievement, 2	vaccine, 27
aids, 3	waist, 28
anthers, 4	
balanced, 4	achievement, 2
beetle, 5	aids, 3
birthweight, 6	anthers, 4
carbohydrate, 6	
Cars, 7	balanced, 4
cholesterol, 8	beetle, 5
chronic, 8	birthweight, 6
cyclones, 9	
doctors, 10	carbohydrate, 6
dogs, 10	Cars, 7
ear, 11	cholesterol, 8
failure, 12	chronic, 8
graduates, 12	cyclones, 9
hepatitis, 13	dobson, 9
hiroshima, 14	dobson-package (dobson), 9
housing, 14	doctors, 10
insurance, 15	doctors, 10
leukemia, 16	u0g3, 10
machine, 16	ear, 11
melanoma, 17	
mortality, 17	failure, 12
moths, 18	
pasture, 19	graduates, 12
plant.dried, 19	
plants, 20	hepatitis, 13
plasma, 20	hiroshima, 14
PLOS, 21	housing, 14
poisson, 21	
remission, 22	insurance, 15
senility, 22	leukemia, 16
stroke.wide, 23	Teukellita, 10
sugar, 24	machine, 16
survival, 25	melanoma, 17
tumor, 25	mortality, 17
ulcer, 26	moths, 18
a_001, _0	1110 (110, 10

30 INDEX

```
pasture, 19
plant.dried, 19
plants, 20
plasma, 20
PLOS, 21
poisson, 21
remission, 22
senility, 22
stroke.wide, 23
sugar, 24
survival, 25
tumor, 25
ulcer, 26
unbalanced, \color{red} \textbf{27}
vaccine, 27
\quad \text{waist}, \textcolor{red}{28}
```