Package 'NegativeBinomialRegression'

April 17, 2024

Title Assessing the Usefulness of Negative Binomial Regression

Version 0.0.1.2000

Description This package is for assessing the usefulness of Negative Binomial Regression among different datasets. It was created as part of the final project for Math 3190 Fundamentals of Data Science Spring 2024 at Southern Utah University.

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Depends tidyverse,
R (>= 2.10),
MASS

Suggests knitr,
rmarkdown,
ggplot2

Imports kableExtra,
shiny,
shinyjs

LazyData true

VignetteBuilder knitr

R topics documented:

`bikes_bridges` *Bicycle Counts for East River Bridges*

Description

Daily total of bike counts conducted monthly on the Brooklyn Bridge, Manhattan Bridge, Williamsburg Bridge, and Queensboro Bridge.

Usage

```
bikes_bridges
```

Format

A data frame containing 215 observations on 2 date variables, 2 temperature ranges, precipitation amounts, 4 different bridges and a total number of bikes.

date Factor with each date,

day factor with the day,

temp_high number with high temperature recorded each day,

temp_low number with low temperature recorded each day,

precipitation precipitation percentages for each day,

Brooklyn_bridge number of bikes that crossed the Brooklyn bridge,

Manhattan_bridge number of bikes that crossed the Manhattan bridge,

Williamsburg_bridge number of bikes that crossed the Williamsburg bridge,

Queensboro_bridge number of bikes that crossed the Queensboro bridge,

total total number of bikes that crossed every bridge for the day,

Details

The data are from the New York Department of Transportation for 2017

Source

NYCDOT Bicycle Counts 2017.

https://data.cityofnewyork.us/Transportation/Bicycle-Counts-for-East-River-Bridges-Historical-/gua4-p9wg/about_data

`bike_rentals` *Bike Rentals Data*

Description

The bike_rentals dataset contains information on bike-sharing rentals over the past 2 years, including environmental and seasonal settings.

Usage

```
bike_rentals
```

Format

A data frame with 17379 rows and 16 variables:

instant record index,

dteday date,

season season (1:spring, 2:summer, 3:fall, 4:winter),

yr year (0: 2011, 1:2012),

mnth month (1 to 12),

hr hour (0 to 23),

holiday whether the day is holiday or not,

weekday day of the week,

workingday if day is neither weekend nor holiday then 1, otherwise is 0,

weathersit weather situation:

- **1** Clear, Few clouds, Partly cloudy,
- **2** Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist,
- **3** Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds,
- **4** Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog,

temp Normalized temperature in Celsius (values are divided by 41),

atemp Normalized feeling temperature in Celsius (values are divided by 50),

hum Normalized humidity (values are divided by 100),

windspeed Normalized wind speed (values are divided by 67),

casual count of casual users,

registered count of registered users,

cnt count of total rental bikes including both casual and registered,

droughts *Periods Between Rain Events*

Description

Data collected at Winnipeg International Airport (Canada) on periods (in days) between rain events.

Usage

```
droughts
```

Format

A data frame with 2 columns:

rownames instance of each entry,

length the length of time from the completion of the last rain event to the beginning of the next rain event,

year the calendar year.

Examples

```
## Not run:
  boxplot(length ~ year, data = droughts)
  boxplot(log(length) ~ year, data = droughts)
  hist(droughts$length, main = "Winnipeg Droughts", xlab = "length (in days)")
 hist(log(droughts$length), main = "Winnipeg Droughts", xlab = "length (in days, log scale)")

## End(Not run)
```

restaurant_inspections *Data on Restaurant Inspections*

Description

The restaurant_inspections data contains data on restaurant health inspections performed in Anchorage, Alaska.

Usage

```
restaurant_inspections
```

Format

A data frame with 27178 rows and 5 variables:

rownames instance of each entry,

business_name Name of restaurant/chain,

inspection_score Health Inspection Score,

Year Year of inspection,

NumberofLocations Number of locations in restaurant chain,

Weekend Bool indicating if the inspection was performed on a weekend.

Details

This data set is used in the Regression chapter of The Effect.

Source

Camus, Louis-Ashley. 2020. Kaggle

References

Huntington-Klein. 2021. The Effect: An Introduction to Research Design and Causality

---

`runNegBin` *Negative Binomial App*

Description

This function allows the Negative Binomial shiny app to run. The app gives demonstrations of Negative Binomial distributions as well as some interactivity with different regression that the model can predict. It shows also provides differences between the Negative Binomial regression and other common regressive techniques.

Usage

```
runNegBin()
```

`ship_accidents` *Ship Accidents*

Description

Data on ship accidents.

Usage

```
ship_accidents
```

Format

A data frame containing 40 observations on 5 ship types in 4 vintages and 2 service periods.

type factor with levels "A" to "E" for the different ship types,

construction factor with levels "1960-64", "1965-69", "1970-74", "1975-79" for the periods of construction,

operation factor with levels "1960-74", "1975-79" for the periods of operation,

service aggregate months of service,

incidents number of damage incidents.

Details

The data are from McCullagh and Nelder (1989, p. 205, Table 6.2) and were also used by Greene (2003, Ch. 21), see below.

Source

Online complements to Greene (2003).

https://pages.stern.nyu.edu/~wgreene/Text/tables/tablelist5.htm

References

Greene, W.H. (2003). Econometric Analysis, 5th edition. Upper Saddle River, NJ: Prentice Hall.

McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models, 2nd edition. London: Chapman & Hall.

See Also

Greene2003, AER package

Index