

-

Tutorial

R Tutorial (R-Tutorial.html)

ggplot2

ggplot2 Short Tutorial (ggplot2-Tutorial-With-R.html)

ggplot2 Tutorial 1 - Intro (Complete-Ggplot2-Tutorial-Part1-With-R-Code.html)

ggplot2 Tutorial 2 - Theme (Complete-Ggplot2-Tutorial-Part2-Customizing-Theme-With-R-Code.html)

ggplot2 Tutorial 3 - Masterlist (Top50-Ggplot2-Visualizations-MasterList-R-Code.html)

ggplot2 Quickref (ggplot2-cheatsheet.html)

Foundations

Linear Regression (Linear-Regression.html)

Statistical Tests (Statistical-Tests-in-R.html)

Missing Value Treatment (Missing-Value-Treatment-With-R.html)

Outlier Analysis (Outlier-Treatment-With-R.html)

Feature Selection (Variable-Selection-and-Importance-With-R.html)

Model Selection (Model-Selection-in-R.html)

Logistic Regression (Logistic-Regression-With-R.html)

Advanced Linear Regression (Environments.html)

Advanced Regression Models

Advanced Regression Models (adv-regression-models.html)

Time Series

Time Series Analysis (Time-Series-Analysis-With-R.html)

Time Series Forecasting (Time-Series-Forecasting-With-R.html)

More Time Series Forecasting ([Time-Series-Forecasting-With-R-part2.html](#))

High Performance Computing

Parallel computing ([Parallel-Computing-With-R.html](#))

Strategies to Speedup R code ([Strategies-To-Improve-And-Speedup-R-Code.html](#))

Useful Techniques

Association Mining ([Association-Mining-With-R.html](#))

Multi Dimensional Scaling ([Multi-Dimensional-Scaling-With-R.html](#))

Optimization ([Profiling.html](#))

InformationValue package ([Information-Value-With-R.html](#))

Stay up-to-date. Subscribe!

(https://docs.google.com/forms/d/1xkMYkLNFU9U39Dd8S_2JC0p8B5t6_Yq6zUQjanQQJpY/viewform)

Chat! (<https://docs.google.com/forms/d/13GrkCFcNa-TOIIIQghsz2SIEbc-YqY9eJX02B19I5Ow/viewform>)

Contents

Use Cases

Example: Gasoline Yield

Beta regression is commonly used when you want to model Y that are probabilities themselves.

This is evident when the value of Y is a *proportion* that ranges between 0 to 1. The data points of Y variable typically represent a proportion of events that form a subset of the total population (assuming that it follows a beta distribution).

Use Cases

1. From GasolineYield data: Proportion of crude oil converted to gasoline after distillation and fractionation
2. Proportion of individuals infected with 'xyz' when exposed to various levels of artificial preservative agent.

Example: Gasoline Yield

Lets predict the gasoline yield as a function of batch and temperature. The example below shows an example implementation of beta regression using the GasolineYield data from betareg package.

```

library (betareg)
data("GasolineYield", package = "betareg") # initialize data
inputData <- GasolineYield # plug-in your data here
trainingIndex <- c(1:(nrow(inputData)-1)) # create row indices of training data
trainingData <- inputData[trainingIndex, ] # training data
testData <- inputData[-trainingIndex, ] # test data
betaMod <- betareg(yield ~ batch + temp, data = trainingData) # train model. Tune var na
mes.
summary (betaMod) # model summary
predict (betaMod, testData) # predict on test data (0.19 vs actual 0.18)

#> Call:
#> betareg(formula = yield ~ batch + temp, data = GasolineYield)
#>
#> Standardized weighted residuals 2:
#>      Min      1Q  Median      3Q      Max
#> -2.8750 -0.8149  0.1601  0.8384  2.0483
#>
#> Coefficients (mean model with logit link):
#>
#>      Estimate Std. Error z value Pr(>|z|)
#> (Intercept) -6.1595710  0.1823247 -33.784 < 2e-16
#> batch1      1.7277289  0.1012294  17.067 < 2e-16
#> batch2      1.3225969  0.1179020  11.218 < 2e-16
#> batch3      1.5723099  0.1161045  13.542 < 2e-16
#> batch4      1.0597141  0.1023598  10.353 < 2e-16
#> batch5      1.1337518  0.1035232  10.952 < 2e-16
#> batch6      1.0401618  0.1060365   9.809 < 2e-16
#> batch7      0.5436922  0.1091275   4.982 6.29e-07
#> batch8      0.4959007  0.1089257   4.553 5.30e-06
#> batch9      0.3857930  0.1185933   3.253 0.00114 **
#> temp        0.0109669  0.0004126  26.577 < 2e-16
#>
#> Phi coefficients (precision model with identity link):
#>
#>      Estimate Std. Error z value Pr(>|z|)
#> (phi)    440.3      110.0   4.002 6.29e-05
#>
#> Signif. codes:  0 ' ' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
#>  
#> Type of estimator: ML (maximum likelihood)  
#> Log-likelihood: 84.8 on 12 Df  
#> Pseudo R-squared: 0.9617  
#> Number of iterations: 51 (BFGS) + 3 (Fisher scoring)
```

This page is based on the examples available in Beta regression vignette (<http://cran.r-project.org/web/packages/betareg/vignettes/betareg.pdf>).

© 2016-17 Selva Prabhakaran. Powered by jekyll (<http://jekyllrb.com/>), knitr (<http://yihui.name/knitr/>), and pandoc (<http://johnmacfarlane.net/pandoc/>). This work is licensed under the Creative Commons License. (<http://creativecommons.org/licenses/by-nc/3.0/>)