

Ethan Ducharme

Report 2.3

MTH 332 Gary Davis

Analysis of students

preliminary year at college

To Predict a class grade

Abstract

Using glm as a function in r the author of this paper predicts a binary outcome of passing a class based on a generated summary statistic of numeric representations of how they did qualitatively on their freshman orientation

Intro

The author loads the required libraries including dplyr and ggplot2. The author reads a modified table into RStudio as a programming Environment the table has been modified to fit the standard of camel case as this adds to the authors future ability to use function calls to access the data in the table. He then proceeds to link the desired columns (11,23,22,25,24,26) together to form a Data Frame in R using the dplyr package. He then sums across the rows of the table to form a summary statistic called "Predictor" he does so using The rowSums() function call. He then uses is.na and index sub 76 to remove NA type data. He then blends the data using the cbind function call on x and y. Then using a function call to the glm() function he trains a logistic model to predict future course grades using the predict function to generate a logit model of

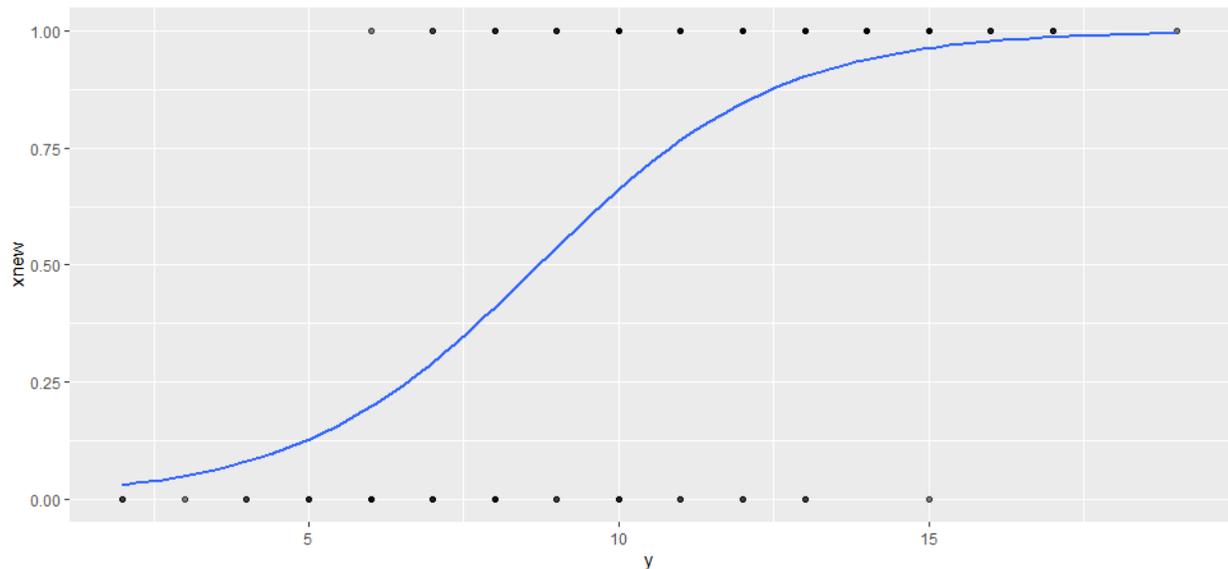
the probabilities of a certain new student passing a class based on their predictor score. He follows up on the data by using ggplot2 to plot a summary of the data along with the generated curve.

Analysis

The data shows that at around a predictor score of ~ 8.5 . The probability at a score of 8.5 or greater is more likely to pass. At a predictor score of 8.5 or less, it is more likely you fail. The likelihood for all values is represented by the following graph.

Conclusion

The data is fit to the logistic model fits as the likelihood of completing the course is well predicted by the predictor variable as a summary statistic.



R code:

```
install.packages("dplyr")
```

```
install.packages("ggplot2")
```

```
library("dplyr")
```

```
library("ggplot2")
```

```
college_prelim_year <- read.csv("C:/Users/disco/OneDrive - University of Massachusetts  
Dartmouth/Docs/Senior Year PT1/Semester 2 Spring/MTH 332 Statistics/Report  
5/college_prelim_year.csv")
```

```
new <- college_prelim_year[,c(11,23,22,25,24,26)]
```

```
Predictor <- rowSums(new) #form the correct summary stat
```

```
xold = college_prelim_year$CompletedCourse[!is.na(college_prelim_year$CompletedCourse)]
```

```
xnew <- xold[ -c(76) ]
```

```
y = Predictor[!is.na(Predictor)]
```

```
blended = cbind(xnew,y)
```

```
model <- glm(xnew ~ y, family=binomial)
```

```
newdata <- predict(model, type="response")
```

```
plot(xnew ~ y, col="steelblue")
```

```
ggplot(, aes(x=y, y=xnew)) +
```

```
  geom_point(alpha=.5) +
```

```
stat_smooth(method="glm", se=FALSE, method.args = list(family=binomial))
```