

HW_4_HXD220000_MXB220061_KXV220007_MXS2200

2023-04-10

Authors : Mankirat Singh Bhamra MXB220061 - HariKrishna Dev HXD220000 - Krishnan Venkatesan KXV220007 - Medha Priyanga Saravanan MXS220057

Part I: Logistic Regression 1. Estimate a logistic regression model using “buyer” as the dependent variable and the following as predictor variables: * last * total_ * gender * child * youth * cook * do_it * reference * art * geog

Hint: To do this in R, first transform the buyer and gender variables into a 0/1 dummy variable using data.table syntax Then run the logistic regression command. Finally, ask R to create a new variable that contains the predicted probability of purchase for each consumer.

```
require(data.table)
```

```
## Loading required package: data.table
```

```
bbb <- fread("~/Library/Mobile Documents/com-apple~CloudDocs/School Work/Sem 2/BUAN 6337/BUAN_6337_Predictive_Analysis/HW/HW-4/BBB.csv",
             na.strings = c("NA", ""),
             sep = "auto",
             stringsAsFactors = FALSE,
             data.table = TRUE,)

summary(bbb)
```

```
##      acctnum      gender      state      zip
## Min.   :10001  Length:50000  Length:50000  Min.    : 801
## 1st Qu.:22501  Class :character  Class :character  1st Qu.: 7643
## Median :35000  Mode  :character  Mode  :character  Median :11208
## Mean   :35000                                Mean   :11531
## 3rd Qu.:47500                                3rd Qu.:16242
## Max.   :60000                                Max.   :21244
##      zip3      first      last      book_
## Min.    : 8.0  Min.    : 1.00  Min.    : 1.00  Min.    : 15.00
## 1st Qu.: 76.0  1st Qu.:11.00  1st Qu.: 7.00  1st Qu.: 15.00
## Median :112.0  Median :21.00  Median :11.00  Median : 28.00
## Mean   :114.9  Mean   :25.67  Mean   :12.36  Mean   : 46.83
## 3rd Qu.:162.0  3rd Qu.:35.00  3rd Qu.:15.00  3rd Qu.: 71.00
## Max.   :212.0  Max.   :99.00  Max.   :35.00  Max.   :158.00
##      nonbook_      total_      purch      child
## Min.    : 0.0  Min.    : 15.0  Min.    : 1.00  Min.    :0.0000
## 1st Qu.: 86.0  1st Qu.:128.0  1st Qu.: 1.00  1st Qu.:0.0000
## Median :162.0  Median :209.0  Median : 2.00  Median :1.0000
## Mean   :161.5  Mean   :208.3  Mean   : 3.89  Mean   :0.8545
## 3rd Qu.:237.0  3rd Qu.:284.0  3rd Qu.: 6.00  3rd Qu.:1.0000
## Max.   :354.0  Max.   :479.0  Max.   :12.00  Max.   :8.0000
##      youth      cook      do_it      reference
## Min.    :0.000  Min.    :0.0000  Min.    :0.0000  Min.    :0.0000
## 1st Qu.:0.000  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000
## Median :0.000  Median :1.0000  Median :0.0000  Median :0.0000
## Mean   :0.391  Mean   :0.9366  Mean   :0.4631  Mean   :0.3122
## 3rd Qu.:1.000  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:0.0000
## Max.    :6.000  Max.    :8.0000  Max.    :7.0000  Max.    :5.0000
##      art      geog      buyer
## Min.    :0.0000  Min.    :0.000  Length:50000
## 1st Qu.:0.0000  1st Qu.:0.000  Class :character
## Median :0.0000  Median :0.000  Mode  :character
## Mean   :0.3859  Mean   :0.547
## 3rd Qu.:1.0000  3rd Qu.:1.000
## Max.    :6.0000  Max.    :7.000
```

Changing buyer and gender variables into a flag

```
bbb[,gender := ifelse(bbb$gender == "M",1,0)]
bbb[,buyer := ifelse(bbb$buyer == "yes",1,0)]
```

Logistic regression

```
fmla <- buyer ~ last + total_ + gender + child + youth + cook + do_it + reference + art + geog
logit <- glm(fmla, data = bbb,family= binomial(link="logit"))
```

Predicted probabilities

```
bbb[, prob := predict(logit,type = "response")]
head(bbb)
```

```
##      acctnum gender state  zip zip3 first last book_ nonbook_ total_ purch child
## 1:   10001      1   NY 10605  106   49  29  109   248   357   10    3
## 2:   10002      1   NY 10960  109   39  27   35   103   138    3    0
## 3:   10003      0   PA 19146  191   19  15   25   147   172    2    0
## 4:   10004      0   NJ  7016   70    7   7   15   257   272    1    0
## 5:   10005      0   NY 10804  108   15  15   15   134   149    1    0
## 6:   10006      0   NY 11366  113    7   7   15    98   113    1    0
##      youth cook do_it reference art geog buyer      prob
## 1:      2    2    0          1  0    2    0 0.02002900
## 2:      1    0    1          0  0    1    0 0.01660684
## 3:      0    2    0          0  0    0    0 0.01582522
## 4:      0    0    0          1  0    0    0 0.07687632
## 5:      0    1    0          0  0    0    0 0.02012333
## 6:      1    0    0          0  0    0    1 0.04694578
```

2. Summarize and interpret the results (so that a marketing manager can understand them). Which variables are statistically significant? Which seem to be economically important? Interpret the odds-ratios for each of the predictors.

```
summary(logit)
```

```
##
## Call:
## glm(formula = fmla, family = binomial(link = "logit"), data = bbb)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4031  -0.4129  -0.2807  -0.1839   3.2650
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.3608301  0.0492961 -47.891 < 2e-16 ***
## last        -0.0947124  0.0027924 -33.918 < 2e-16 ***
## total_       0.0011160  0.0001982   5.630 1.80e-08 ***
## gender       0.7607204  0.0357608  21.272 < 2e-16 ***
## child       -0.1862162  0.0172824 -10.775 < 2e-16 ***
## youth       -0.1129745  0.0261087  -4.327 1.51e-05 ***
## cook        -0.2703210  0.0171283 -15.782 < 2e-16 ***
## do_it       -0.5391648  0.0269657 -19.994 < 2e-16 ***
## reference    0.2346876  0.0265583   8.837 < 2e-16 ***
## art          1.1555840  0.0221439  52.185 < 2e-16 ***
## geog         0.5742763  0.0186311  30.824 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 30355  on 49999  degrees of freedom
## Residual deviance: 24122  on 49989  degrees of freedom
## AIC: 24144
##
## Number of Fisher Scoring iterations: 6
```

When we look at the z values of the different terms, we can determine the impact of the each variable in the logistic model. The art category has the highest propensity of buyers as the |z value_art| is the highest. This is succeeded by the last time the person bought a book i.e. a buyer who hasn't bought a book in a while is more likely to buy a book now. Total value has the least significance among all the factors, which implies buyers don't care much about the total cost of their purchase.

Part II: Decile Analysis of Logistic Regression Results 1. Assign each customer to a decile based on his or her predicted probability of purchase with 'bucket 1' being the highest average purchase probability.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##   between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

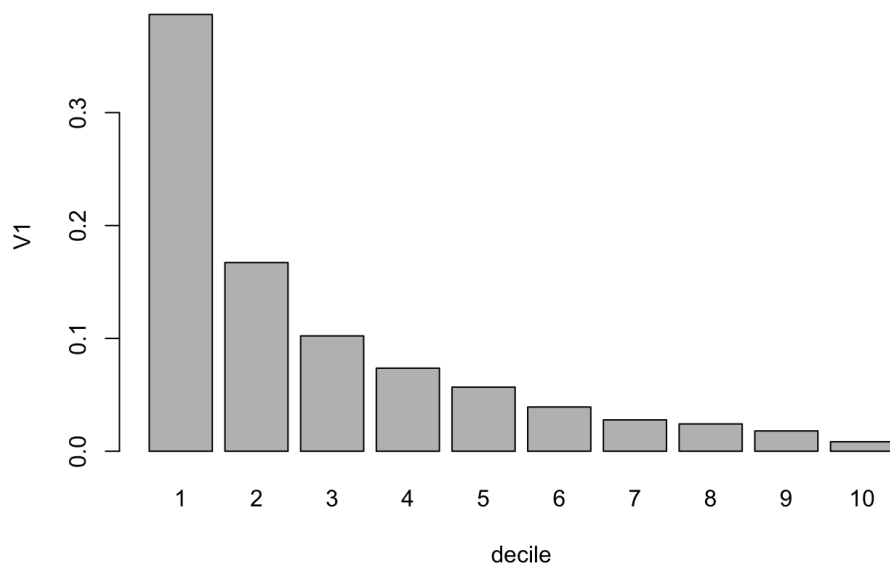
```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
bbb[,decile := ntile(1 - bbb$prob,10)]
bbb[,bucket := paste("Bucket ",as.character(decile))]
bbb[,decile := factor(decile,levels = c(1:10))]
```

2. Create a bar chart plotting the average response rate by decile (as just defined above).

```
response_rates=bbb[,mean(buyer),by=decile]

barplot(V1~decile,data=response_rates)
```



3. Generate a table showing the number of customers, the number of buyers of “The Art History of Florence,” and the response rate to the offer by decile for the random sample (i.e. the 50,000 customers) in the dataset.

```
results <- bbb[, .(num_cust = .N, num_buy = sum(buyer), resp_rate = sum(buyer) / .N), by = decile]

head(results)
```

```
##   decile num_cust num_buy resp_rate
## 1:      8     5000     121   0.0242
## 2:      9     5000      90   0.0180
## 3:      4     5000     368   0.0736
## 4:      6     5000     196   0.0392
## 5:     10     5000      42   0.0084
## 6:      3     5000     511   0.1022
```

4. For the 50,000 customers in the dataset, generate a table showing the average values of the following variables by probability of purchase decile: Total \$ spent Months since last purchase, and Number of books purchased for each of the seven categories (i.e., children, youth, cookbooks, do-it-yourself, reference, art and geography).

```
avg_bbb <- bbb[, .(avg_total = mean(total_), avg_last = mean(last),
  avg_child = mean(child), avg_youth = mean(youth),
  avg_cook = mean(cook), avg_do_it = mean(do_it),
  avg_reference = mean(reference), avg_art = mean(art),
  avg_geog = mean(geog)),
  by = decile]

avg_bbb
```

```
##      decile avg_total avg_last avg_child avg_youth avg_cook avg_do_it
## 1:      8  191.5660  14.4184   0.8050   0.3604   0.9090   0.4480
## 2:      9  193.6108  17.8576   0.9606   0.4052   1.1182   0.6506
## 3:      4  207.6430   8.7828   0.7526   0.3626   0.7966   0.3404
## 4:      6  199.1302  10.9376   0.7480   0.3648   0.8648   0.3942
## 5:     10  204.3416  25.8684   1.0674   0.4630   1.3094   0.7722
## 6:      3  214.2284   8.6188   0.7910   0.3654   0.7960   0.3698
## 7:      1  257.3526   7.1944   1.0648   0.5138   1.0668   0.4714
## 8:      7  191.3296  12.3724   0.7608   0.3480   0.8362   0.4208
## 9:      5  199.1118   9.5732   0.7580   0.3338   0.8208   0.3698
## 10:     2  224.8692   7.9580   0.8364   0.3928   0.8482   0.3934
##      avg_reference avg_art avg_geog
## 1:      0.2050   0.1140   0.2544
## 2:      0.2524   0.1276   0.3160
## 3:      0.3082   0.3024   0.5404
## 4:      0.2588   0.1634   0.3862
## 5:      0.2476   0.0692   0.2916
## 6:      0.3832   0.4802   0.7010
## 7:      0.5628   1.5006   1.3308
## 8:      0.2274   0.1320   0.2946
## 9:      0.2724   0.2168   0.4638
## 10:     0.4046   0.7530   0.8908
```

5. Summarize and interpret the decile analysis results. Are the patterns in the decile analysis consistent with your conclusions from the logistic regression? (Hint: graph some of the results in the previous question.)

```
# summarize average values by decile
require(psych)
```

```
## Loading required package: psych
```

```
results2 <- describeBy(bbb, bbb$decile)
results2
```

```
##
## Descriptive statistics by group
## group: 1
##      vars      n      mean      sd      median      trimmed      mad      min
## acctnum      1 5000 34921.04 14425.91 35123.00 34884.75 18556.22 10009.00
## gender        2 5000   0.58    0.49    1.00    0.60    0.00    0.00
## state*        3 5000   8.63    2.55   10.00    9.00    1.48    1.00
## zip           4 5000 11583.05 5660.24 11204.50 11627.76 5840.70 802.00
## zip3          5 5000 115.44   56.64   112.00   115.88   57.82    8.00
## first         6 5000  32.12   20.48   31.00   31.07   23.72    1.00
## last          7 5000   7.19    5.41    7.00    6.64    5.93    1.00
## book_         8 5000   78.45   42.76   82.00   78.50   60.79   15.00
## nonbook_      9 5000  178.90   86.78  184.00  180.55  108.23    1.00
## total_       10 5000  257.35  100.51  259.00  258.22  108.23   16.00
## purch        11 5000   6.51    3.78    7.00    6.51    5.93    1.00
## child        12 5000   1.06    1.20    1.00    0.88    1.48    0.00
## youth        13 5000   0.51    0.77    0.00    0.36    0.00    0.00
## cook         14 5000   1.07    1.19    1.00    0.89    1.48    0.00
## do_it        15 5000   0.47    0.74    0.00    0.31    0.00    0.00
## reference    16 5000   0.56    0.80    0.00    0.41    0.00    0.00
## art          17 5000   1.50    0.93    1.00    1.47    1.48    0.00
## geog         18 5000   1.33    1.25    1.00    1.19    1.48    0.00
## buyer       19 5000   0.39    0.49    0.00    0.36    0.00    0.00
## prob        20 5000   0.39    0.17    0.33    0.36    0.13    0.22
## decile*     21 5000   1.00    0.00    1.00    1.00    0.00    1.00
## bucket*    22 5000   1.00    0.00    1.00    1.00    0.00    1.00
##
##      max      range      skew      kurtosis      se
## acctnum 60000.00 49991.00 0.01    -1.22 204.01
## gender   1.00     1.00 -0.33    -1.89 0.01
## state*   15.00    14.00 -1.17    0.47 0.04
## zip      21239.00 20437.00 0.14    -0.91 80.05
## zip3     212.00   204.00 0.14    -0.91 0.80
## first    99.00    98.00 0.35    -0.69 0.29
## last     35.00    34.00 1.08     1.90 0.08
## book_    158.00   143.00 -0.04    -1.38 0.60
## nonbook_ 341.00   340.00 -0.14    -1.12 1.23
## total_   479.00   463.00 -0.07    -0.66 1.42
## purch     12.00    11.00 -0.05    -1.41 0.05
## child     7.00     7.00 1.10     0.75 0.02
## youth     5.00     5.00 1.57     2.59 0.01
## cook      7.00     7.00 1.09     0.74 0.02
## do_it     4.00     4.00 1.58     2.16 0.01
## reference 5.00     5.00 1.51     2.26 0.01
## art       6.00     6.00 0.59     0.41 0.01
## geog      7.00     7.00 0.95     0.73 0.02
## buyer     1.00     1.00 0.46    -1.79 0.01
## prob      0.98     0.77 1.34     1.18 0.00
## decile*   1.00     0.00 NaN      NaN 0.00
## bucket*   1.00     0.00 NaN      NaN 0.00
## -----
## group: 2
##      vars      n      mean      sd      median      trimmed      mad      min
## acctnum      1 5000 35123.62 14499.50 35129.50 35160.98 18711.15 10018.00
## gender        2 5000   0.51    0.50    1.00    0.51    0.00    0.00
## state*        3 5000   8.64    2.55   10.00    8.99    1.48    1.00
## zip           4 5000 11546.16 5688.68 11206.00 11558.02 5960.05 801.00
## zip3          5 5000 115.07   56.92   112.00   115.19   59.30    8.00
## first         6 5000  24.03   18.98   17.00   21.75   14.83    1.00
## last          7 5000   7.96    5.49    7.00    7.53    5.93    1.00
## book_         8 5000   54.89   39.90   29.00   50.16   20.76   15.00
## nonbook_      9 5000  169.98   87.83  173.00  171.36  112.68    0.00
## total_       10 5000  224.87  100.10  228.00  224.17  111.19   15.00
## purch        11 5000   4.52    3.61    2.00    4.09    1.48    1.00
## child        12 5000   0.84    1.13    0.00    0.62    0.00    0.00
## youth        13 5000   0.39    0.68    0.00    0.25    0.00    0.00
## cook         14 5000   0.85    1.15    0.00    0.62    0.00    0.00
## do_it        15 5000   0.39    0.70    0.00    0.23    0.00    0.00
## reference    16 5000   0.40    0.66    0.00    0.27    0.00    0.00
## art          17 5000   0.75    0.68    1.00    0.68    0.00    0.00
## geog         18 5000   0.89    0.95    1.00    0.76    1.48    0.00
## buyer       19 5000   0.17    0.37    0.00    0.08    0.00    0.00
## prob        20 5000   0.16    0.03    0.16    0.16    0.03    0.13
```

```
## decile*      21 5000      2.00      0.00      2.00      2.00      0.00      2.00
## bucket*      22 5000      1.00      0.00      1.00      1.00      0.00      1.00
##              max      range      skew      kurtosis      se
## acctnum      59999.00 49981.00 -0.02      -1.23 205.05
## gender        1.00      1.00 -0.04      -2.00  0.01
## state*        15.00      14.00 -1.10      0.56  0.04
## zip           21244.00 20443.00 0.18      -0.92 80.45
## zip3           212.00      204.00 0.18      -0.92  0.80
## first          99.00      98.00 0.96      0.17  0.27
## last           35.00      34.00 0.99      1.92  0.08
## book_          152.00     137.00 0.78      -0.78 0.56
## nonbook_       345.00     345.00 -0.09     -1.15 1.24
## total_         476.00     461.00 0.03      -0.67 1.42
## purch          12.00      11.00 0.77      -0.84 0.05
## child           7.00      7.00 1.47      1.92  0.02
## youth           5.00      5.00 1.91      3.93  0.01
## cook            6.00      6.00 1.45      1.59  0.02
## do_it           4.00      4.00 1.89      3.39  0.01
## reference       4.00      4.00 1.67      2.69  0.01
## art             4.00      4.00 0.61      0.32  0.01
## geog            6.00      6.00 1.12      1.33  0.01
## buyer           1.00      1.00 1.78      1.18  0.01
## prob            0.22      0.09 0.34      -1.07 0.00
## decile*         2.00      0.00  NaN      NaN  0.00
## bucket*         1.00      0.00  NaN      NaN  0.00
```

```
## -----
## group: 3
##              vars      n      mean      sd      median      trimmed      mad      min
## acctnum        1 5000 34938.53 14579.93 34847.0 34901.08 18740.81 10008.00
## gender          2 5000      0.45      0.50      0.0      0.44      0.00      0.00
## state*          3 5000      7.68      2.52      9.0      8.03      1.48      1.00
## zip             4 5000 11547.55 5625.85 11218.0 11569.55 5706.53 801.00
## zip3            5 5000     115.09     56.29     112.0     115.31     56.34      8.00
## first           6 5000     22.00     17.70     15.0     19.48     11.86      1.00
## last            7 5000      8.62      5.50      9.0      8.31      5.93      1.00
## book_           8 5000     47.32     36.62     29.0     41.44     20.76     15.00
## nonbook_        9 5000    166.90     87.57    170.0    167.91    111.19      0.00
## total_          10 5000    214.23     98.69    218.0    212.87    110.45     15.00
## purch           11 5000      3.89      3.35      2.0      3.35      1.48      1.00
## child           12 5000      0.79      1.10      0.0      0.57      0.00      0.00
## youth           13 5000      0.37      0.66      0.0      0.23      0.00      0.00
## cook            14 5000      0.80      1.11      0.0      0.58      0.00      0.00
## do_it           15 5000      0.37      0.67      0.0      0.22      0.00      0.00
## reference       16 5000      0.38      0.64      0.0      0.27      0.00      0.00
## art             17 5000      0.48      0.62      0.0      0.40      0.00      0.00
## geog            18 5000      0.70      0.81      1.0      0.58      1.48      0.00
## buyer           19 5000      0.10      0.30      0.0      0.00      0.00      0.00
## prob            20 5000      0.10      0.01      0.1      0.10      0.02      0.09
## decile*         21 5000      3.00      0.00      3.0      3.00      0.00      3.00
## bucket*         22 5000      1.00      0.00      1.0      1.00      0.00      1.00
##              max      range      skew      kurtosis      se
## acctnum      59997.00 49989.00 0.03      -1.21 206.19
## gender        1.00      1.00 0.20      -1.96  0.01
## state*        14.00      13.00 -1.16      0.65  0.04
## zip           21239.00 20438.00 0.16      -0.88 79.56
## zip3           212.00      204.00 0.16      -0.88  0.80
## first          99.00      98.00 1.24      1.01  0.25
## last           35.00      34.00 0.88      1.88  0.08
## book_          149.00     134.00 1.12      -0.02 0.52
## nonbook_       344.00     344.00 -0.08     -1.16 1.24
## total_         478.00     463.00 0.08      -0.66 1.40
## purch          12.00      11.00 1.12      -0.09 0.05
## child           7.00      7.00 1.61      2.64  0.02
## youth           5.00      5.00 2.11      5.26  0.01
## cook            7.00      7.00 1.61      2.59  0.02
## do_it           4.00      4.00 1.96      3.78  0.01
## reference       5.00      5.00 1.79      3.75  0.01
## art             4.00      4.00 1.06      0.78  0.01
## geog            5.00      5.00 1.23      1.82  0.01
## buyer           1.00      1.00 2.63      4.90  0.00
## prob            0.13      0.04 0.18      -1.18 0.00
## decile*         3.00      0.00  NaN      NaN  0.00
```

```

## bucket*      1.00      0.00      NaN      NaN      0.00
## -----
## group: 4
##      vars      n      mean      sd      median      trimmed      mad      min
## acctnum      1 5000 35085.50 14340.80 35022.50 35098.59 18354.59 10004.00
## gender        2 5000      0.37      0.48      0.00      0.34      0.00      0.00
## state*        3 5000      7.67      2.53      9.00      8.03      1.48      1.00
## zip           4 5000 11526.54 5591.09 11210.00 11537.33 5633.88 801.00
## zip3          5 5000  114.88  55.94  112.00  114.99  56.34  8.00
## first         6 5000  19.98  17.68  15.00  17.24  11.86  1.00
## last          7 5000   8.78   6.02   9.00   8.32   5.93  1.00
## book_         8 5000  41.63  34.97  28.00  34.72  19.27 15.00
## nonbook_      9 5000 166.02  86.89 168.00 167.20 109.71  0.00
## total_       10 5000 207.64  97.10 209.00 206.78 111.19 15.00
## purch        11 5000   3.40   3.20   2.00   2.77   1.48  1.00
## child        12 5000   0.75   1.03   0.00   0.56   0.00  0.00
## youth        13 5000   0.36   0.66   0.00   0.23   0.00  0.00
## cook         14 5000   0.80   1.09   0.00   0.58   0.00  0.00
## do_it        15 5000   0.34   0.66   0.00   0.19   0.00  0.00
## reference    16 5000   0.31   0.59   0.00   0.19   0.00  0.00
## art          17 5000   0.30   0.53   0.00   0.21   0.00  0.00
## geog         18 5000   0.54   0.76   0.00   0.39   0.00  0.00
## buyer       19 5000   0.07   0.26   0.00   0.00   0.00  0.00
## prob        20 5000   0.07   0.01   0.07   0.07   0.01  0.06
## decile*      21 5000   4.00   0.00   4.00   4.00   0.00  4.00
## bucket*      22 5000   1.00   0.00   1.00   1.00   0.00  1.00
##
##      max      range      skew      kurtosis      se
## acctnum 59992.00 49988.00  0.00    -1.19 202.81
## gender   1.00     1.00  0.55    -1.70  0.01
## state*   14.00    13.00 -1.18     0.69  0.04
## zip      21239.00 20438.00  0.19    -0.85 79.07
## zip3     212.00   204.00  0.19    -0.85  0.79
## first     97.00    96.00  1.38     1.52  0.25
## last      35.00    34.00  0.91     1.51  0.09
## book_    151.00   136.00  1.42     0.82  0.49
## nonbook_ 351.00   351.00 -0.08    -1.10  1.23
## total_   479.00   464.00  0.09    -0.63  1.37
## purch     12.00    11.00  1.40     0.70  0.05
## child     7.00     7.00  1.71     3.41  0.01
## youth     5.00     5.00  2.13     5.29  0.01
## cook       8.00     8.00  1.74     3.42  0.02
## do_it     5.00     5.00  2.23     5.57  0.01
## reference  4.00     4.00  2.12     5.04  0.01
## art        3.00     3.00  1.70     2.73  0.01
## geog       6.00     6.00  1.55     2.88  0.01
## buyer     1.00     1.00  3.26     8.66  0.00
## prob       0.09     0.02  0.17    -1.13  0.00
## decile*    4.00     0.00  NaN      NaN    0.00
## bucket*    1.00     0.00  NaN      NaN    0.00
## -----
## group: 5
##      vars      n      mean      sd      median      trimmed      mad      min
## acctnum      1 5000 34859.96 14292.97 34909.50 34850.35 18290.09 10011.00
## gender        2 5000      0.30      0.46      0.00      0.25      0.00      0.00
## state*        3 5000      7.66      2.50      9.00      8.01      1.48      1.00
## zip           4 5000 11606.16 5642.52 11214.00 11635.20 5645.74 820.00
## zip3          5 5000  115.68  56.46  112.00  115.97  56.34  8.00
## first         6 5000  20.04  17.36  15.00  17.08  11.86  1.00
## last          7 5000   9.57   6.04   9.00   9.00   5.93  1.00
## book_         8 5000  39.50  34.16  26.00  32.30  16.31 15.00
## nonbook_      9 5000 159.61  87.75 162.00 159.83 111.19  0.00
## total_       10 5000 199.11  98.36 201.00 197.32 114.16 15.00
## purch        11 5000   3.24   3.13   2.00   2.58   1.48  1.00
## child        12 5000   0.76   1.03   0.00   0.56   0.00  0.00
## youth        13 5000   0.33   0.62   0.00   0.21   0.00  0.00
## cook         14 5000   0.82   1.09   0.00   0.61   0.00  0.00
## do_it        15 5000   0.37   0.67   0.00   0.23   0.00  0.00
## reference    16 5000   0.27   0.55   0.00   0.16   0.00  0.00
## art          17 5000   0.22   0.47   0.00   0.12   0.00  0.00
## geog         18 5000   0.46   0.72   0.00   0.32   0.00  0.00
## buyer       19 5000   0.06   0.23   0.00   0.00   0.00  0.00
## prob        20 5000   0.06   0.00   0.06   0.06   0.01  0.05

```

```
## decile*      21 5000      5.00      0.00      5.00      5.00      0.00      5.00
## bucket*      22 5000      1.00      0.00      1.00      1.00      0.00      1.00
##              max      range      skew      kurtosis      se
## acctnum      59989.00 49978.00 0.00      -1.17 202.13
## gender        1.00      1.00      0.86      -1.26 0.01
## state*        14.00      13.00 -1.17      0.58 0.04
## zip           21239.00 20419.00 0.16      -0.88 79.80
## zip3           212.00      204.00 0.16      -0.88 0.80
## first          99.00      98.00 1.53      2.00 0.25
## last           35.00      34.00 1.06      1.81 0.09
## book_         156.00      141.00 1.55      1.21 0.48
## nonbook_       350.00      350.00 -0.02      -1.16 1.24
## total_        479.00      464.00 0.14      -0.65 1.39
## purch         12.00      11.00 1.53      1.06 0.04
## child          8.00      8.00 1.72      3.41 0.01
## youth          4.00      4.00 2.08      4.83 0.01
## cook           7.00      7.00 1.65      2.93 0.02
## do_it          5.00      5.00 2.13      5.34 0.01
## reference      4.00      4.00 2.21      5.43 0.01
## art            4.00      4.00 2.22      5.13 0.01
## geog           6.00      6.00 1.72      3.48 0.01
## buyer          1.00      1.00 3.83      12.66 0.00
## prob           0.06      0.02 0.11      -1.20 0.00
## decile*        5.00      0.00      NaN      NaN 0.00
## bucket*        1.00      0.00      NaN      NaN 0.00
```

```
## -----
## group: 6
##              vars      n      mean      sd      median      trimmed      mad      min
## acctnum        1 5000 34950.39 14471.13 34759.00 34928.20 18519.90 10006.00
## gender          2 5000      0.27      0.44      0.00      0.21      0.00      0.00
## state*          3 5000      7.71      2.49      9.00      8.06      1.48      1.00
## zip             4 5000 11552.83 5583.68 11203.00 11570.50 5759.16 804.00
## zip3            5 5000  115.14  55.87  112.00  115.32  56.34  8.00
## first           6 5000  21.10  16.37  15.00  18.20  8.90  3.00
## last            7 5000  10.94   5.73  11.00  10.40  5.93  1.00
## book_           8 5000  38.70  33.43  26.00  31.49  16.31  15.00
## nonbook_        9 5000  160.43  88.40  160.00  160.83  111.19  0.00
## total_         10 5000  199.13  98.91  200.00  197.40  114.16  15.00
## purch          11 5000   3.18   3.09   2.00   2.52   1.48   1.00
## child          12 5000   0.75   1.05   0.00   0.54   0.00   0.00
## youth          13 5000   0.36   0.66   0.00   0.23   0.00   0.00
## cook           14 5000   0.86   1.11   1.00   0.65   1.48   0.00
## do_it          15 5000   0.39   0.69   0.00   0.26   0.00   0.00
## reference      16 5000   0.26   0.54   0.00   0.14   0.00   0.00
## art            17 5000   0.16   0.42   0.00   0.06   0.00   0.00
## geog           18 5000   0.39   0.65   0.00   0.26   0.00   0.00
## buyer          19 5000   0.04   0.19   0.00   0.00   0.00   0.00
## prob           20 5000   0.04   0.00   0.04   0.04   0.00   0.04
## decile*        21 5000   6.00   0.00   6.00   6.00   0.00   6.00
## bucket*        22 5000   1.00   0.00   1.00   1.00   0.00   1.00
##              max      range      skew      kurtosis      se
## acctnum      59995.00 49989.00 0.01      -1.20 204.65
## gender        1.00      1.00 1.03      -0.95 0.01
## state*        14.00      13.00 -1.19      0.74 0.04
## zip           21239.00 20435.00 0.18      -0.86 78.97
## zip3           212.00      204.00 0.18      -0.86 0.79
## first          91.00      88.00 1.55      1.93 0.23
## last           35.00      34.00 1.08      1.96 0.08
## book_         146.00      131.00 1.61      1.41 0.47
## nonbook_       344.00      344.00 -0.02      -1.15 1.25
## total_        479.00      464.00 0.14      -0.67 1.40
## purch         12.00      11.00 1.58      1.26 0.04
## child          8.00      8.00 1.88      4.39 0.01
## youth          5.00      5.00 2.12      5.40 0.01
## cook           7.00      7.00 1.64      2.98 0.02
## do_it          5.00      5.00 2.05      5.14 0.01
## reference      5.00      5.00 2.39      6.88 0.01
## art            3.00      3.00 2.69      7.76 0.01
## geog           5.00      5.00 1.91      4.42 0.01
## buyer          1.00      1.00 4.75      20.54 0.00
## prob           0.05      0.01 0.08      -1.19 0.00
## decile*        6.00      0.00      NaN      NaN 0.00
```



```

## bucket*      1.00      0.00      NaN      NaN      0.00
## -----
## group: 7
##      vars      n      mean      sd      median      trimmed      mad      min
## acctnum      1 5000 34872.29 14539.19 34851.00 34861.62 18732.65 10010.00
## gender        2 5000      0.22      0.42      0.00      0.15      0.00      0.00
## state*        3 5000      7.64      2.52      9.00      7.99      1.48      1.00
## zip           4 5000 11417.21 5562.75 11020.00 11401.38 5336.62 820.00
## zip3          5 5000  113.78  55.66  110.00  113.62  53.37   8.00
## first         6 5000  21.71  15.64  15.00  18.74   8.90   3.00
## last          7 5000  12.37   5.70  11.00  11.66   2.97   1.00
## book_         8 5000  36.72  32.37  25.00  29.45  14.83  15.00
## nonbook_      9 5000 154.61  88.28 154.00 154.05 112.68   0.00
## total_       10 5000 191.33  99.33 192.00 188.53 117.13  15.00
## purch        11 5000   3.02   3.00   2.00   2.35   1.48   1.00
## child        12 5000   0.76   1.02   0.00   0.57   0.00   0.00
## youth        13 5000   0.35   0.63   0.00   0.23   0.00   0.00
## cook         14 5000   0.84   1.09   1.00   0.63   1.48   0.00
## do_it        15 5000   0.42   0.71   0.00   0.27   0.00   0.00
## reference    16 5000   0.23   0.50   0.00   0.12   0.00   0.00
## art          17 5000   0.13   0.37   0.00   0.03   0.00   0.00
## geog         18 5000   0.29   0.60   0.00   0.16   0.00   0.00
## buyer       19 5000   0.03   0.16   0.00   0.00   0.00   0.00
## prob        20 5000   0.03   0.00   0.03   0.03   0.00   0.03
## decile*      21 5000   7.00   0.00   7.00   7.00   0.00   7.00
## bucket*      22 5000   1.00   0.00   1.00   1.00   0.00   1.00
##
##      max      range      skew      kurtosis      se
## acctnum 59996.00 49986.00  0.00    -1.22 205.62
## gender   1.00     1.00  1.34    -0.20  0.01
## state*   14.00    13.00 -1.16     0.68  0.04
## zip      21239.00 20419.00  0.22    -0.82 78.67
## zip3     212.00   204.00  0.22    -0.82  0.79
## first    99.00    96.00  1.71     2.68  0.22
## last     35.00    34.00  1.40     2.75  0.08
## book_    145.00   130.00  1.69     1.70  0.46
## nonbook_ 350.00   350.00  0.04    -1.16  1.25
## total_   478.00   463.00  0.21    -0.69  1.40
## purch    12.00    11.00  1.66     1.54  0.04
## child     7.00     7.00  1.79     3.89  0.01
## youth     6.00     6.00  2.20     6.91  0.01
## cook      7.00     7.00  1.74     3.60  0.02
## do_it     5.00     5.00  1.91     4.03  0.01
## reference 5.00     5.00  2.47     7.32  0.01
## art       3.00     3.00  2.85     8.24  0.01
## geog      4.00     4.00  2.15     4.56  0.01
## buyer     1.00     1.00  5.74    30.99  0.00
## prob      0.04     0.01  0.09    -1.18  0.00
## decile*   7.00     0.00  NaN      NaN    0.00
## bucket*   1.00     0.00  NaN      NaN    0.00
## -----
## group: 8
##      vars      n      mean      sd      median      trimmed      mad      min
## acctnum      1 5000 34805.22 14380.86 34790.00 34778.73 18456.15 10001.00
## gender        2 5000   0.19   0.39   0.00   0.11   0.00   0.00
## state*        3 5000   8.65   2.55  10.00   9.01   1.48   1.00
## zip           4 5000 11534.26 5616.37 11201.00 11543.73 5733.21 802.00
## zip3          5 5000  114.96  56.20  112.00  115.05  56.34   8.00
## first         6 5000  24.16  15.59  17.00  21.16   5.93   7.00
## last          7 5000  14.42   5.77  13.00  13.69   2.97   1.00
## book_         8 5000  37.25  32.64  25.00  30.10  14.83  15.00
## nonbook_      9 5000 154.31  86.53 152.00 153.56 109.71   0.00
## total_       10 5000 191.57  97.49 189.00 188.98 112.68  15.00
## purch        11 5000   3.10   3.04   2.00   2.44   1.48   1.00
## child        12 5000   0.81   1.06   1.00   0.60   1.48   0.00
## youth        13 5000   0.36   0.65   0.00   0.23   0.00   0.00
## cook         14 5000   0.91   1.11   1.00   0.70   1.48   0.00
## do_it        15 5000   0.45   0.73   0.00   0.30   0.00   0.00
## reference    16 5000   0.21   0.50   0.00   0.08   0.00   0.00
## art          17 5000   0.11   0.34   0.00   0.01   0.00   0.00
## geog         18 5000   0.25   0.58   0.00   0.11   0.00   0.00
## buyer       19 5000   0.02   0.15   0.00   0.00   0.00   0.00
## prob        20 5000   0.02   0.00   0.02   0.02   0.00   0.02

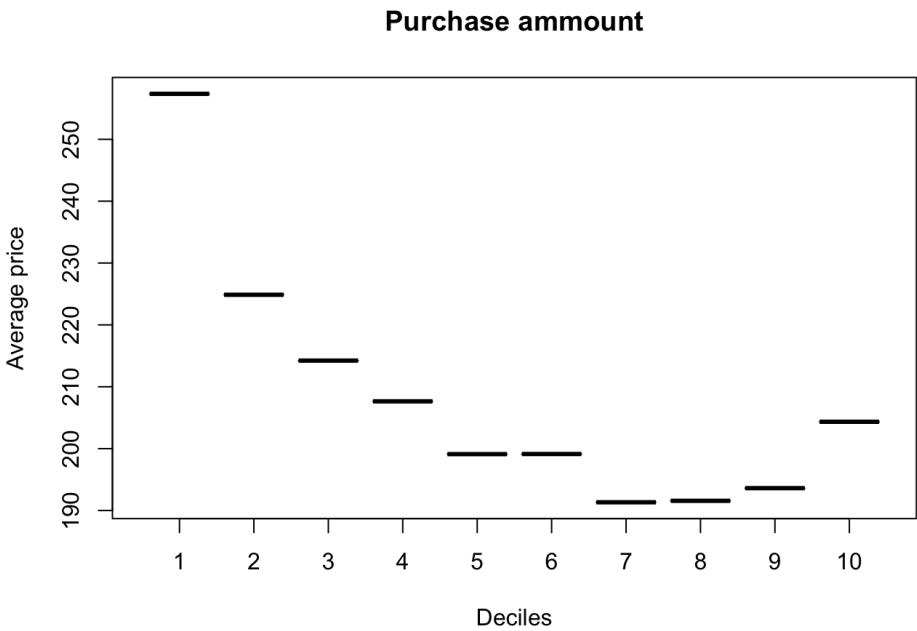
```

```
## decile*      21 5000      8.00      0.00      8.00      8.00      0.00      8.00
## bucket*      22 5000      1.00      0.00      1.00      1.00      0.00      1.00
##              max      range      skew      kurtosis      se
## acctnum      59993.00 49992.00 0.01      -1.19 203.38
## gender        1.00      1.00      1.61      0.58      0.01
## state*        15.00      14.00 -1.15      0.64      0.04
## zip           21239.00 20437.00 0.18      -0.86 79.43
## zip3           212.00      204.00 0.18      -0.86 0.79
## first          99.00      92.00      1.72      2.59 0.22
## last           35.00      34.00      1.25      2.11 0.08
## book_          144.00      129.00 1.63      1.49 0.46
## nonbook_       354.00      354.00 0.07      -1.11 1.22
## total_         479.00      464.00 0.22      -0.64 1.38
## purch          12.00      11.00 1.58      1.27 0.04
## child           8.00      8.00      1.78      4.01 0.01
## youth           5.00      5.00      2.08      5.32 0.01
## cook            7.00      7.00      1.55      2.62 0.02
## do_it           6.00      6.00      1.98      4.99 0.01
## reference       4.00      4.00      2.88      9.87 0.01
## art             3.00      3.00      3.13      10.49 0.00
## geog            4.00      4.00      2.49      6.40 0.01
## buyer           1.00      1.00      6.19      36.33 0.00
## prob            0.03      0.01      0.04      -1.18 0.00
## decile*         8.00      0.00      NaN      NaN 0.00
## bucket*         1.00      0.00      NaN      NaN 0.00
```

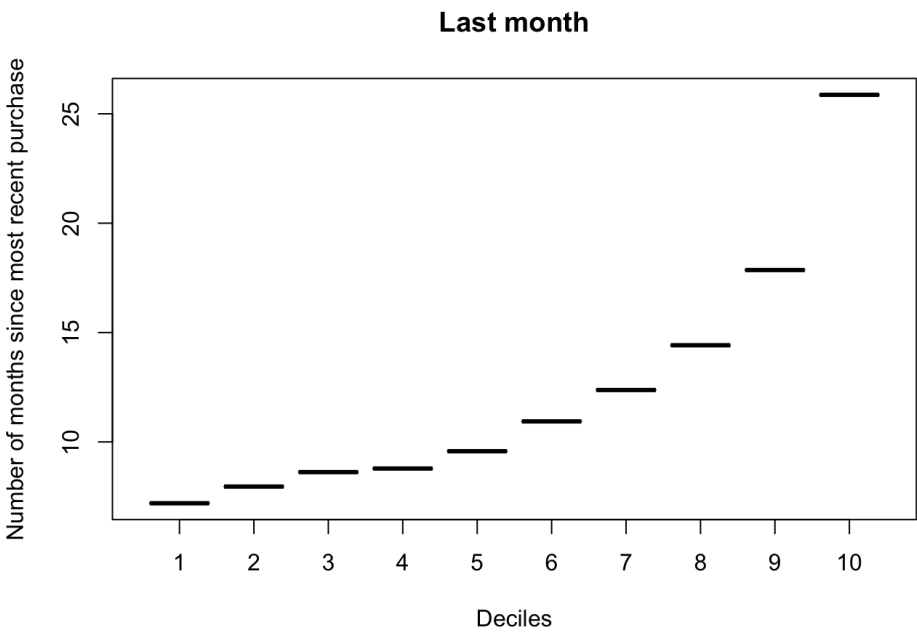
```
## -----
## group: 9
##              vars      n      mean      sd      median      trimmed      mad      min
## acctnum        1 5000 34939.76 14404.39 35047.00 34940.36 18357.55 10002.00
## gender          2 5000      0.23      0.42      0.00      0.16      0.00      0.00
## state*          3 5000      8.61      2.50      10.00      8.95      1.48      1.00
## zip             4 5000 11505.21 5634.26 11208.50 11508.40 5621.28 821.00
## zip3            5 5000 114.67      56.39      112.00      114.70      56.34      8.00
## first           6 5000 30.97      16.26      25.00      28.56      11.86      9.00
## last            7 5000 17.86      7.07      15.00      17.41      5.93      1.00
## book_           8 5000 44.96      35.97      26.00      39.00      16.31      15.00
## nonbook_        9 5000 148.65      88.83      141.00      146.36      111.19      0.00
## total_          10 5000 193.61      102.69      188.00      189.75      118.61      15.00
## purch           11 5000      3.83      3.38      2.00      3.29      1.48      1.00
## child           12 5000      0.96      1.21      1.00      0.73      1.48      0.00
## youth           13 5000      0.41      0.71      0.00      0.25      0.00      0.00
## cook            14 5000      1.12      1.28      1.00      0.91      1.48      0.00
## do_it           15 5000      0.65      0.87      0.00      0.49      0.00      0.00
## reference       16 5000      0.25      0.55      0.00      0.13      0.00      0.00
## art             17 5000      0.13      0.36      0.00      0.02      0.00      0.00
## geog            18 5000      0.32      0.60      0.00      0.19      0.00      0.00
## buyer           19 5000      0.02      0.13      0.00      0.00      0.00      0.00
## prob            20 5000      0.02      0.00      0.02      0.02      0.00      0.01
## decile*         21 5000      9.00      0.00      9.00      9.00      0.00      9.00
## bucket*         22 5000      1.00      0.00      1.00      1.00      0.00      1.00
##              max      range      skew      kurtosis      se
## acctnum      59990.00 49988.00 -0.01      -1.19 203.71
## gender        1.00      1.00      1.29      -0.35 0.01
## state*        15.00      14.00 -1.15      0.53 0.04
## zip           21239.00 20418.00 0.18      -0.84 79.68
## zip3           212.00      204.00 0.18      -0.84 0.80
## first          99.00      90.00      1.23      1.04 0.23
## last           35.00      34.00 0.54      -0.22 0.10
## book_          145.00      130.00 1.16      -0.03 0.51
## nonbook_       350.00      350.00 0.20      -1.10 1.26
## total_         477.00      462.00 0.28      -0.67 1.45
## purch          12.00      11.00 1.11      -0.18 0.05
## child           8.00      8.00      1.53      2.35 0.02
## youth           5.00      5.00      1.98      4.32 0.01
## cook            8.00      8.00      1.37      1.81 0.02
## do_it           6.00      6.00      1.55      2.71 0.01
## reference       5.00      5.00      2.50      7.26 0.01
## art             2.00      2.00      2.79      7.35 0.01
## geog            4.00      4.00      2.05      4.53 0.01
## buyer           1.00      1.00      7.25      50.55 0.00
## prob            0.02      0.01 -0.10      -1.19 0.00
## decile*         9.00      0.00      NaN      NaN 0.00
```

```
## bucket*      1.00      0.00      NaN      NaN      0.00
## -----
## group: 10
##      vars      n      mean      sd      median      trimmed      mad      min      max
## acctnum      1 5000 35508.69 14401.99 35593.00 35602.44 18410.93 10007 59988.00
## gender        2 5000      0.22      0.41      0.00      0.15      0.00      0      1.00
## state*        3 5000      7.64      2.55      9.00      7.98      1.48      1     14.00
## zip           4 5000 11492.65 5671.22 11209.00 11512.61 5835.51 822 21237.00
## zip3          5 5000  114.54   56.75  112.00  114.74   57.82      8    212.00
## first         6 5000  40.56   15.52   35.00   38.47   11.86     15    99.00
## last          7 5000  25.87    7.36   27.00   26.70    5.93      1    35.00
## book_         8 5000  48.89   38.49   27.00   43.47   17.79     15   145.00
## nonbook_      9 5000 155.46   87.83  151.00  154.02  109.71      0   351.00
## total_       10 5000  204.34  101.99  201.00  201.61  114.16     15   479.00
## purch        11 5000   4.22    3.63    2.00    3.73    1.48      1    12.00
## child        12 5000   1.07    1.27    1.00    0.85    1.48      0     7.00
## youth        13 5000   0.46    0.75    0.00    0.30    0.00      0     5.00
## cook         14 5000   1.31    1.46    1.00    1.07    1.48      0     8.00
## do_it        15 5000   0.77    1.02    0.00    0.59    0.00      0     7.00
## reference    16 5000   0.25    0.53    0.00    0.13    0.00      0     4.00
## art          17 5000   0.07    0.26    0.00    0.00    0.00      0     2.00
## geog         18 5000   0.29    0.57    0.00    0.18    0.00      0     4.00
## buyer       19 5000   0.01    0.09    0.00    0.00    0.00      0     1.00
## prob        20 5000   0.01    0.00    0.01    0.01    0.00      0     0.01
## decile*      21 5000  10.00    0.00   10.00   10.00    0.00     10    10.00
## bucket*     22 5000   1.00    0.00   1.00    1.00    0.00      1     1.00
##
##      range      skew      kurtosis      se
## acctnum 49981.00 -0.04      -1.19 203.67
## gender   1.00    1.36      -0.14  0.01
## state*   13.00  -1.11      0.50   0.04
## zip      20415.00 0.15      -0.88 80.20
## zip3     204.00  0.15      -0.88 0.80
## first    84.00   1.22      1.07 0.22
## last     34.00  -0.92      0.09 0.10
## book_    130.00  0.90      -0.64 0.54
## nonbook_ 351.00  0.12      -1.07 1.24
## total_   464.00  0.20      -0.64 1.44
## purch    11.00  0.85      -0.77 0.05
## child     7.00  1.38      1.69 0.02
## youth     5.00  1.81      3.64 0.01
## cook      8.00  1.24      1.05 0.02
## do_it     7.00  1.50      2.38 0.01
## reference 4.00  2.29      5.47 0.01
## art       2.00  3.75     13.73 0.00
## geog      4.00  2.11      4.79 0.01
## buyer     1.00 10.77     114.01 0.00
## prob      0.01 -0.08     -1.02 0.00
## decile*   0.00  NaN      NaN  0.00
## bucket*   0.00  NaN      NaN  0.00
```

```
plot(avg_bbb$decile,avg_bbb$avg_total,main = "Purchase ammount",xlab = "Deciles",ylab = "Average price")
```

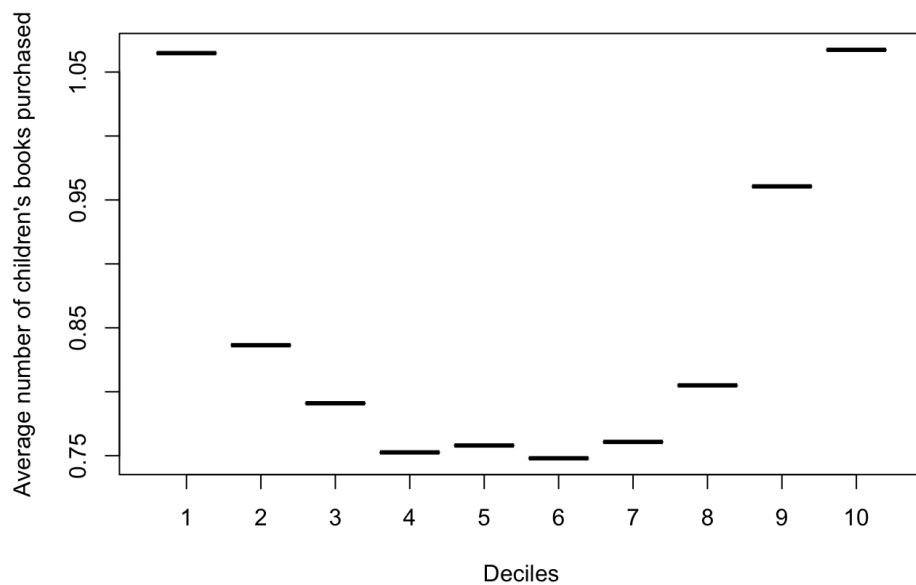


```
plot(avg_bbb$decile,avg_bbb$avg_last,main = "Last month",xlab = "Deciles",ylab = "Number of months since most recent purchase")
```



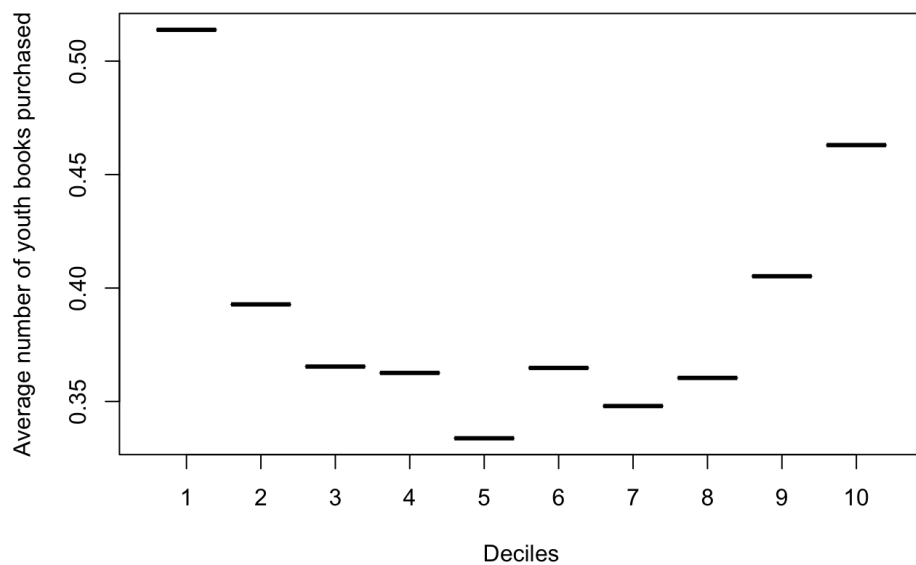
```
plot(avg_bbb$decile,avg_bbb$avg_child,main = "Children Book",xlab = "Deciles",ylab = "Average number of children's books purchased")
```

Children Book



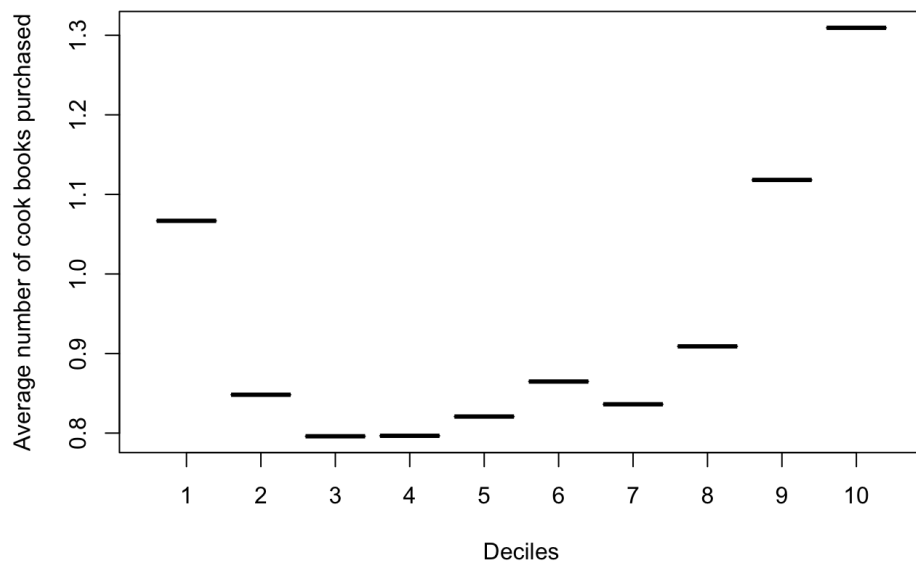
```
plot(avg_bbb$decile,avg_bbb$avg_youth,main = "Youth books",xlab = "Deciles",ylab = "Average number of youth books purchased")
```

Youth books



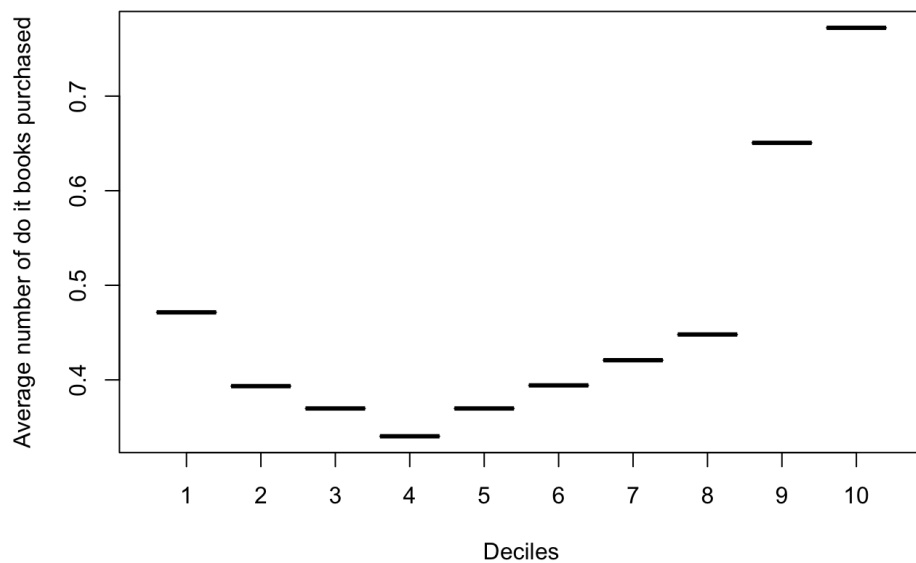
```
plot(avg_bbb$decile,avg_bbb$avg_cook,main = "Cook books",xlab = "Deciles",ylab = "Average number of cook books purchased")
```

Cook books



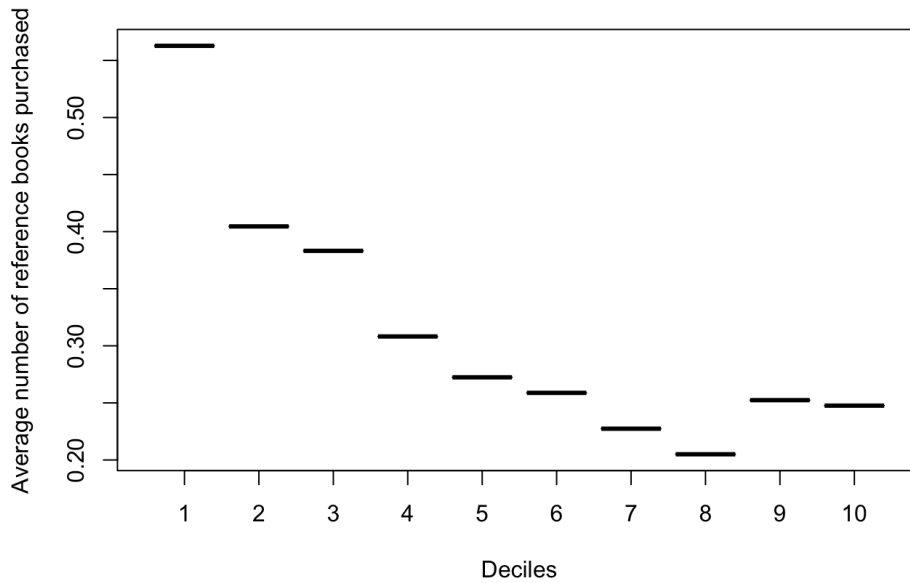
```
plot(avg_bbb$decile,avg_bbb$avg_do_it,main = "Do it books",xlab = "Deciles",ylab = "Average number of do it books purchased")
```

Do it books



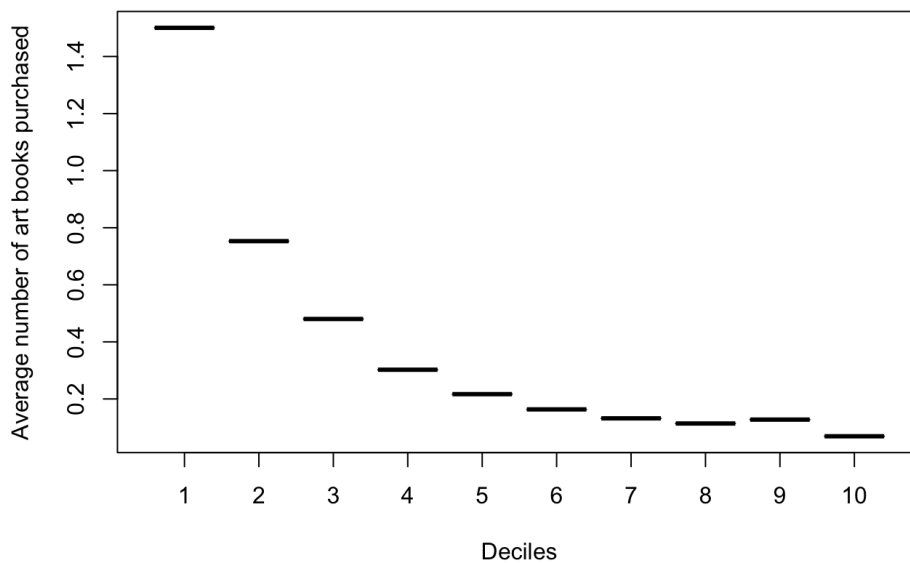
```
plot(avg_bbb$decile,avg_bbb$avg_reference,main = "Reference books purchased",xlab = "Deciles",ylab = "Average number of reference books purchased")
```

Reference books purchased



```
plot(avg_bbb$decile,avg_bbb$avg_art,main = "Art books",xlab = "Deciles",ylab = "Average number of art books purchased")
```

Art books



Total: We can see that Top deciles

purchase the most in terms of purchase amount and it decreases as the decile progress. Note that there is small rise in Decile 10 but that could be due to population bias.

Last month: We can see that recency of purchase heavily influences probability of purchase as it more likely to purchase the book if its been a long time since the customers last purchase.

Children category: We see that its evenly distributed around decile 5, where customers are more likely to buy the book when they are closer to decile 1 or 10.

Youth category: We see that its almost evenly distributed around decile 5 but it peaks at Decile 1.

Cook category: We see that its evenly distributed around decile 5, where customers are more likely to buy the book when they are closer to decile 10.

Do it category: We can see that do it books are most purchased by Decile 10 and the probability of do it books being bought dips at decile 4.

Reference category: We can see that do it books are most purchased by Decile 1 and the probability of do it books being bought dips at decile 8.

Art books: People who buy the book have a higher probability of buying other art books.

Part III: Profitability Analysis Use the following cost information to assess the profitability of using logistic regression to determine which of the remaining 500,000 customers should receive a specific offer: Cost to mail offer to customer: 0.5\$ Selling price (shipping included): 18\$ Wholesale price paid by BookBinders: 9\$ Shipping costs: 3\$ 1. What is the breakeven response rate?

The breakeven response rate can be calculated as follows: Revenue per customer = Selling price - Wholesale price - Shipping costs = \$18.00 - \$9.00 - \$3.00 = \$6.00 Breakeven cost per response = Cost to mail offer to customer / Revenue per customer = \$0.50 / \$6.00 = **0.0833** Therefore, the breakeven response rate is **8.33%** (i.e., we need at least 8.33% of the targeted customers to respond to the offer to break even).

2. For the customers in the dataset, create a new variable (call it "mailto_logit") with a value of 1 if the customer's predicted probability is greater than or equal to the breakeven response rate and 0 otherwise. Hint: You can use data.table's conditional assignment syntax

```
# calculate breakeven response rate
revenue_per_customer <- 18 - 9 - 3
breakeven_cost_per_response <- 0.5 / revenue_per_customer
breakeven_response_rate <- breakeven_cost_per_response * 100
breakeven_response_rate
```

```
## [1] 8.333333
```

```
# create new variable "mailto_logit"
bbb[, mailto_logit := ifelse(prob >= breakeven_cost_per_response, 1, 0)]
bbb
```

```
##      acctnum gender state  zip zip3 first last book_ nonbook_ total_ purch
##      1:  10001      1   NY 10605  106   49  29  109    248    357    10
##      2:  10002      1   NY 10960  109   39  27   35    103    138     3
##      3:  10003      0   PA 19146  191   19  15   25    147    172     2
##      4:  10004      0   NJ  7016   70    7   7   15    257    272     1
##      5:  10005      0   NY 10804  108   15  15   15    134    149     1
##      ---
## 49996:  59996      0   NY 11967  119    9   9   15     12     27     1
## 49997:  59997      0   NJ  8882   88   25   5   79    294    373     7
## 49998:  59998      1   NJ  7410   74    3   3   15    178    193     1
## 49999:  59999      1   NJ  7090   70   49  29   98    246    344     8
## 50000:  60000      1   NY 11355  113   29   1   60    125    185     5
##      child youth cook do_it reference art geog buyer      prob decile
##      1:    3    2    2    0          1  0    2    0 0.02002900      8
##      2:    0    1    0    1          0  0    1    0 0.01660684      9
##      3:    0    0    2    0          0  0    0    0 0.01582522      9
##      4:    0    0    0    0          1  0    0    0 0.07687632      4
##      5:    0    0    1    0          0  0    0    0 0.02012333      8
##      ---
## 49996:    1    0    0    0          0  0    0    0 0.03326801      7
## 49997:    3    0    1    1          0  1    1    0 0.11340915      3
## 49998:    0    0    0    0          1  0    0    0 0.19245092      2
## 49999:    2    0    1    0          2  1    2    0 0.13798918      2
## 50000:    1    0    1    0          0  1    2    0 0.58885780      1
##      bucket mailto_logit
##      1: Bucket  8          0
##      2: Bucket  9          0
##      3: Bucket  9          0
##      4: Bucket  4          0
##      5: Bucket  8          0
##      ---
## 49996: Bucket  7          0
## 49997: Bucket  3          1
## 49998: Bucket  2          1
## 49999: Bucket  2          1
## 50000: Bucket  1          1
```

3. Out of the 50,000 test sample, how many customers should have received the targeting promotion mail for "The Art History of Florence" based on the breakeven response rate (i.e., the number of mailto_logit ==1 in the data)? Also, among those who would have targeted, what would have been the response rate (i.e., mean of buyer among mailto_logit ==1)? How much higher is this response rate relative to the overall response rate in the data?

```
# calculate number of customers to receive offer
num_customers <- sum(bbb$mailto_logit)
num_customers
```



```
## [1] 15560
```

```
# calculate response rate among those who receive offer
response_rate <- mean(bbb$buyer[bbb$mailto_logit == 1])
response_rate
```

```
## [1] 0.2135604
```

```
# calculate expected profit from campaign
expected_profit <- num_customers * (18 - 9 - 3 - 0.5)
expected_profit
```

```
## [1] 85580
```

4. Consider that there are 500,000 remaining customers for the roll-out (excluding 50,000 test group in the current data. Assuming our test group is similar to the roll-out group (i.e., our test group is representative of the roll-out group), what is the expected number of buyers of 'The Art History of Florence' if we do targeted mailing based on the breakeven response rate? Hint: Count the number of buyers among the targeted (mailto_logit==1) in the test group and multiply it by 10 since the roll-out sample is 10 times larger. Alternatively, you can use the response rate.

```
# calculate expected number of buyers in roll-out group
expected_num_buyers <- num_customers * (mean(bbb$buyer[bbb$mailto_logit == 1]) / mean(bbb$mailto_logit))
expected_num_buyers_rollout <- expected_num_buyers * 10
expected_num_buyers_rollout
```

```
## [1] 106780.2
```