W241_Project_PGSS_Campaign

Subha Vadakkumkoor, Walter Burge

August 3, 2018

```
knitr::opts_chunk$set(echo = TRUE)
```

```
library(data.table)
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
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
```

```
library(readr)
```

Loading data

Data imported from Salsa, excluding PII fields (name and email address) are read into R dataframes. All dataframes have the same structure and format. With each step of the treatment (Orignal email, Reminder1 and Reminder2), there are 2 files: list of people that were assigned the treatment (recieved the email) and list of people that responded to the treatment (donated money).

```
#Load data

#Original email
orig_email_rec<-read.csv('./data/BlastReport_Class Experiment Final Email_Recipients.csv')
setnames(orig_email_rec, old=c("Opened"), new=c("opened_orig_email"))

orig_email_resp<-read.csv('./data/BlastReport_Class Experiment Final Email_Conversions.csv')
sapply(orig_email_resp, class)
```

```
##    Supporter.ID     External.ID         Country           State
##        "factor"       "integer"        "factor"        "factor"
##            City  Reference.Name      Split.Name       Time.Sent
##        "factor"        "factor"        "factor"        "factor"
```

```
## Conversion.Date   Activity.Type   Activity.Name     Activity.ID
##        "factor"        "factor"        "factor"        "factor"
##          Amount   Donation.Type
##       "numeric"        "factor"
```

```
#Create an indicator and rename columns to reflect original email response (useful for merge later)
orig_email_resp$donated_after_orig_email=1
names(orig_email_resp)
```

```
##  [1] "Supporter.ID"             "External.ID"
##  [3] "Country"                  "State"
##  [5] "City"                     "Reference.Name"
##  [7] "Split.Name"               "Time.Sent"
##  [9] "Conversion.Date"          "Activity.Type"
## [11] "Activity.Name"            "Activity.ID"
## [13] "Amount"                   "Donation.Type"
## [15] "donated_after_orig_email"
```

```
setnames(orig_email_resp, old=c("Conversion.Date","Amount"), new=c("Orig_email_conversion_date", "orig_

#Reminder1
reminder1_rec<-read.csv('./data/BlastReport_Class experiment Reminder1_Recipients.csv')
setnames(reminder1_rec, old=c("Opened"), new=c("opened_reminder1"))
reminder1_rec_subset=reminder1_rec[,c("opened_reminder1","Supporter.ID")]
```

```
reminder1_resp<-read.csv('./data/BlastReport_Class experiment Reminder1_Conversions.csv')
reminder1_resp$donated_after_reminder1=1
setnames(reminder1_resp, old=c("Conversion.Date","Amount"), new=c("reminder1_conversion_date", "reminde

#Reminder2
reminder2_rec<-read.csv('./data/BlastReport_Class Experiment Reminder 2_Recipients.csv')
setnames(reminder2_rec, old=c("Opened"), new=c("opened_reminder2"))
reminder2_rec_subset=reminder2_rec[,c("opened_reminder2","Supporter.ID")]


reminder2_resp<-read.csv('./data/BlastReport_Class Experiment Reminder 2_Conversions.csv')
reminder2_resp$donated_after_reminder2=1
setnames(reminder2_resp, old=c("Conversion.Date","Amount"), new=c("reminder2_conversion_date", "reminde

#Load donor profile file
alumni_profile<-read.csv('./data/Alumni_profile_all.csv',colClasses = c("character","character","numeri
sapply(alumni_profile,class)
```

```
##         Constituent.Number           Constituent.UUID
##                "character"                "character"
##                  PGSS.Year               Faculty.Year
##                  "numeric"                "character"
##                TA.RD.Years              Received.Date
##                "character"                "character"
##                 Gift.Count         Total.Gift.Amounts
##                "character"                "character"
##        Largest.Gift.Amount             Last.Gift.Date
##                "character"                "character"
##      Last.Gift.Amount.Ever Months_since_last_donation
##                "character"                  "numeric"
##           last_gift_amount          total_gift_amount
```

```
##                  "numeric"                  "numeric"
##                 gift_count           LYBUNT_indicator
##                  "numeric"                  "numeric"
##           SYBUNT_Indicator              Never_donator
##                  "numeric"                  "numeric"
```

```
names(alumni_profile)
```

```
##  [1] "Constituent.Number"         "Constituent.UUID"
##  [3] "PGSS.Year"                  "Faculty.Year"
##  [5] "TA.RD.Years"                "Received.Date"
##  [7] "Gift.Count"                 "Total.Gift.Amounts"
##  [9] "Largest.Gift.Amount"        "Last.Gift.Date"
## [11] "Last.Gift.Amount.Ever"      "Months_since_last_donation"
## [13] "last_gift_amount"           "total_gift_amount"
## [15] "gift_count"                 "LYBUNT_indicator"
## [17] "SYBUNT_Indicator"           "Never_donator"
```

```
#Examine the layout of a representative file
cat("Fields in recipients file\n")
```

```
## Fields in recipients file
```

```
names(orig_email_rec)
```

```
##  [1] "Supporter.ID"            "External.ID"
##  [3] "Country"                 "State"
##  [5] "City"                    "Reference.Name"
##  [7] "Split.Name"              "Time.Sent"
##  [9] "Status"                  "opened_orig_email"
## [11] "Clicked"                 "Converted"
## [13] "Unsubscribed"            "First.Open.Date"
## [15] "Number.of.Links.Clicked" "Bounce.Category"
## [17] "Bounce.Code"
```

```
cat("\nFields in responder files\n")
```

```
##
## Fields in responder files
```

```
names(orig_email_resp)
```

```
##  [1] "Supporter.ID"               "External.ID"
##  [3] "Country"                    "State"
##  [5] "City"                       "Reference.Name"
##  [7] "Split.Name"                 "Time.Sent"
##  [9] "Orig_email_conversion_date" "Activity.Type"
## [11] "Activity.Name"              "Activity.ID"
## [13] "orig_email_amount"          "Donation.Type"
## [15] "donated_after_orig_email"
```

```
#Get dimensions of each file
cat("\nDimensions of each file\n")
```

```
##
## Dimensions of each file
```

```
dfList <- list(orig_email_rec,orig_email_resp,reminder1_rec,reminder1_resp,reminder2_rec,reminder2_resp
lapply(dfList,dim)
```

```
## [[1]]
## [1] 2110   17
##
## [[2]]
## [1] 25 15
##
## [[3]]
## [1] 2107   17
##
## [[4]]
## [1] 37 15
##
## [[5]]
## [1] 2111   17
##
## [[6]]
## [1] 34 15
```

Now we merge the orignal rec and resp datasets with the responders from reminder1 and reminder2. We assume that the reminders were sent to same people that original emails were sent to. Some of the fields like "opened", etc of reminders are not captured as they may not be required just yet and can be added later if needed.

```
#Merge original recipients email with Alumni profile
names(alumni_profile)
```

```
##  [1] "Constituent.Number"         "Constituent.UUID"
##  [3] "PGSS.Year"                  "Faculty.Year"
##  [5] "TA.RD.Years"                "Received.Date"
##  [7] "Gift.Count"                 "Total.Gift.Amounts"
##  [9] "Largest.Gift.Amount"        "Last.Gift.Date"
## [11] "Last.Gift.Amount.Ever"      "Months_since_last_donation"
## [13] "last_gift_amount"           "total_gift_amount"
## [15] "gift_count"                 "LYBUNT_indicator"
## [17] "SYBUNT_Indicator"           "Never_donator"
```

```
merged<-merge(orig_email_rec,alumni_profile,by.x="Supporter.ID",by.y="Constituent.UUID",all.x=TRUE)

#Merge with the original email response
merged<-merge(merged,orig_email_resp[,c("Supporter.ID","Orig_email_conversion_date","orig_email_amount"

cat("\nNum of rows",nrow(merged))
```

```
##
## Num of rows 2110
```

```
#Merge with the first reminder response
merged<-merge(merged,reminder1_rec_subset,by="Supporter.ID",all.x=TRUE)

merged<-merge(merged,reminder1_resp[,c("Supporter.ID","reminder1_conversion_date","reminder1_amount","d

cat("\nNum of rows",nrow(merged))
```

```
##
## Num of rows 2110
```

```r
#Merge with the second reminder response
merged<-merge(merged,reminder2_rec_subset,by="Supporter.ID",all.x=TRUE)

merged<-merge(merged,reminder2_resp[,c("Supporter.ID","reminder2_conversion_date","reminder2_amount","d

cat("\nNum of rows",nrow(merged))
```

```
##
## Num of rows 2110
```

```r
#Set NA's in indicators to 0
merged[(is.na(merged$donated_after_orig_email)),]$donated_after_orig_email=0
merged[(is.na(merged$donated_after_reminder1)),]$donated_after_reminder1=0
merged[(is.na(merged$donated_after_reminder2)),]$donated_after_reminder2=0
```

Let us create dependant and covariates

```r
#Check for MULTIPLE DONATIONS

#Define treatment indicator
merged$treat<-ifelse(merged$Split.Name %in% c("Split A"),1,0)
table(merged$Split.Name,merged$treat)
```

```
##
##              0    1
##   Split A    0 1055
##   Split B 1055    0
```

```r
#Total donation amount
merged$orig_email_amount_copy=merged$orig_email_amount
merged$reminder1_amount_copy=merged$reminder1_amount
merged$reminder2_amount_copy=merged$reminder2_amount

merged$orig_email_amount=ifelse(is.na(merged$orig_email_amount_copy),0,merged$orig_email_amount)
merged$reminder1_amount=ifelse(is.na(merged$reminder1_amount_copy),0,merged$reminder1_amount_copy)
merged$reminder2_amount=ifelse(is.na(merged$reminder2_amount_copy),0,merged$reminder2_amount_copy)

merged$total_donation_amount=merged$orig_email_amount+merged$reminder1_amount+merged$reminder2_amount
summary(merged$total_donation_amount)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    0.000    0.000    0.000    5.583    0.000 4000.000
```

```r
summary(merged$orig_email_amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   1.121   0.000 500.000
```

```r
summary(merged$reminder1_amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   1.062   0.000 500.000
```

```r
summary(merged$reminder2_amount)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0     0.0     0.0     3.4     0.0  4000.0
```

```
#Days till donation
merged$donation_date=coalesce(as.Date(merged$Orig_email_conversion_date),as.Date(merged$reminder1_conve
merged[(is.na(merged$donation_date)),]$donation_date=as.Date('2018-7-24')

merged$days_till_donation=merged$donation_date-as.Date(merged$Time.Sent)
#merged[(merged$donation_date>0),]
table(merged$days_till_donation)
```

```
##
##   -1    0    1    2    3    4    5    6    7    8    9
## 2019    6   11    1    1   26    8    2   32    3    1
```

```
#Donation response indicator
merged$donated_any_time=0
merged[merged$days_till_donation>0,]$donated_any_time=1

#Define non-compliance. What about people who opened but did not click or contribute?

table(merged$opened_orig_email,merged$donated_after_reminder1)
```

```
##
##            0    1
##   FALSE 1243    7
##   TRUE   832   28
```

```
table(merged$opened_orig_email,merged$donated_after_reminder2)
```

```
##
##            0    1
##   FALSE 1240   10
##   TRUE   838   22
```

```
merged$compliant=0
merged[merged$opened_orig_email,]$compliant=1
merged$compliant=ifelse(merged$opened_reminder1,1,0)
merged$compliant=ifelse(merged$opened_reminder2,1,0)
table(merged$opened_reminder1)
```

```
##
## FALSE  TRUE
##  1183   921
```

```
#Need to add more indicators : read both original and reminder and responded only after reminder, etc
```

Stats

```
cat("Response rate after original email")
```

```
## Response rate after original email
```

```
table(merged$donated_after_orig_email,merged$treat)
```

```
##
##        0    1
##   0 1048 1038
##   1    7   17
```

```
cat("Response rate after reminder1")
```

```
## Response rate after reminder1
```

```
table(merged$donated_after_reminder1,merged$treat)
```

```
##
##        0    1
##   0 1039 1036
##   1   16   19
```

```
cat("Response rate after reminder2")
```

```
## Response rate after reminder2
```

```
table(merged$donated_after_reminder2,merged$treat)
```

```
##
##        0    1
##   0 1039 1039
##   1   16   16
```

```
table(merged$opened_orig_email,merged$donated_after_reminder1)
```

```
##
##            0    1
##   FALSE 1243    7
##   TRUE   832   28
```

```
table(merged$opened_orig_email,merged$donated_after_reminder2)
```

```
##
##            0    1
##   FALSE 1240   10
##   TRUE   838   22
```

Initial regression trial

```
#names(merged)
merged$treat_reminder=merged$treat * merged$opened_reminder1 * merged$opened_reminder2
merged$opened_atleast_one_reminder=merged$opened_reminder1 * merged$opened_reminder2
#Need to add more covariates
merged$donated_any_time <- factor(merged$donated_any_time)
reg_response<-glm(donated_any_time~treat+opened_atleast_one_reminder+PGSS.Year+last_gift_amount+SYBUNT_
print(summary(reg_response))
```

```
##
## Call:
## glm(formula = donated_any_time ~ treat + opened_atleast_one_reminder +
##     PGSS.Year + last_gift_amount + SYBUNT_Indicator + LYBUNT_indicator +
##     gift_count, family = "binomial", data = merged)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0452  -0.2712  -0.1671  -0.0937   3.3050
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 -4.513e+01  2.423e+01  -1.863 0.062529 .
## treat                        2.856e-01  2.366e-01   1.207 0.227470
## opened_atleast_one_reminder  1.749e+00  2.516e-01   6.951 3.61e-12 ***
```

```
## PGSS.Year                    1.981e-02  1.212e-02   1.635 0.102016
## last_gift_amount            -1.756e-05  2.051e-04  -0.086 0.931772
## SYBUNT_Indicator             1.226e+00  4.309e-01   2.846 0.004432 **
## LYBUNT_indicator             2.049e+00  3.618e-01   5.663 1.49e-08 ***
## gift_count                   4.769e-02  1.339e-02   3.562 0.000368 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 705.20  on 2097  degrees of freedom
## Residual deviance: 561.46  on 2090  degrees of freedom
##   (12 observations deleted due to missingness)
## AIC: 577.46
##
## Number of Fisher Scoring iterations: 7
```

```
#Add robust errors

names(merged)
```

```
##  [1] "Supporter.ID"                "External.ID"
##  [3] "Country"                     "State"
##  [5] "City"                        "Reference.Name"
##  [7] "Split.Name"                  "Time.Sent"
##  [9] "Status"                      "opened_orig_email"
## [11] "Clicked"                     "Converted"
## [13] "Unsubscribed"                "First.Open.Date"
## [15] "Number.of.Links.Clicked"     "Bounce.Category"
## [17] "Bounce.Code"                 "Constituent.Number"
## [19] "PGSS.Year"                   "Faculty.Year"
## [21] "TA.RD.Years"                 "Received.Date"
## [23] "Gift.Count"                  "Total.Gift.Amounts"
## [25] "Largest.Gift.Amount"         "Last.Gift.Date"
## [27] "Last.Gift.Amount.Ever"       "Months_since_last_donation"
## [29] "last_gift_amount"            "total_gift_amount"
## [31] "gift_count"                  "LYBUNT_indicator"
## [33] "SYBUNT_Indicator"            "Never_donator"
## [35] "Orig_email_conversion_date"  "orig_email_amount"
## [37] "donated_after_orig_email"    "opened_reminder1"
## [39] "reminder1_conversion_date"   "reminder1_amount"
## [41] "donated_after_reminder1"     "opened_reminder2"
## [43] "reminder2_conversion_date"   "reminder2_amount"
## [45] "donated_after_reminder2"     "treat"
## [47] "orig_email_amount_copy"      "reminder1_amount_copy"
## [49] "reminder2_amount_copy"       "total_donation_amount"
## [51] "donation_date"               "days_till_donation"
## [53] "donated_any_time"            "compliant"
## [55] "treat_reminder"              "opened_atleast_one_reminder"
```

```
#Add other regressions here

#Determining effect of two different splits on donation after original e-mail
just_msg_effect_immediate<-lm(merged$donated_after_orig_email~merged$Split.Name)
print("Original treatment effect on immediate response:")
```

```
## [1] "Original treatment effect on immediate response:"
```

```
print(summary(just_msg_effect_immediate))
```

```
##
## Call:
## lm(formula = merged$donated_after_orig_email ~ merged$Split.Name)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.01611 -0.01611 -0.00664 -0.00664  0.99336
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.016114   0.003263   4.938  8.5e-07 ***
## merged$Split.NameSplit B -0.009479   0.004615  -2.054   0.0401 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.106 on 2108 degrees of freedom
## Multiple R-squared:  0.001997,   Adjusted R-squared:  0.001524
## F-statistic: 4.219 on 1 and 2108 DF,  p-value: 0.0401
```

```
#Need to add more covariates
merged2<-merged[merged$total_donation_amount>0, ]
summary(merged$total_donation_amount)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##    0.000    0.000    0.000    5.583    0.000 4000.000
```

```
reg_amount<-lm(total_donation_amount~treat+opened_atleast_one_reminder+PGSS.Year+last_gift_amount+SYBUN
print(summary(reg_amount))
```

```
##
## Call:
## lm(formula = total_donation_amount ~ treat + opened_atleast_one_reminder +
##     PGSS.Year + last_gift_amount + SYBUNT_Indicator + LYBUNT_indicator +
##     gift_count, data = merged2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -736.80  -35.87   10.39   43.71  432.57
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -3.889e+03  3.228e+03  -1.205   0.2318
## treat                         2.708e+01  3.448e+01   0.785   0.4345
## opened_atleast_one_reminder -5.539e+01  3.691e+01  -1.501   0.1373
## PGSS.Year                     1.987e+00  1.617e+00   1.229   0.2226
## last_gift_amount              9.144e-01  3.994e-02  22.895   <2e-16 ***
## SYBUNT_Indicator             -6.150e+01  6.504e+01  -0.946   0.3472
## LYBUNT_indicator             -9.557e+01  5.579e+01  -1.713   0.0905 .
## gift_count                   -2.658e+00  2.893e+00  -0.919   0.3610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 157.8 on 81 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.8764, Adjusted R-squared:  0.8657
## F-statistic: 82.06 on 7 and 81 DF,  p-value: < 2.2e-16
```

```
reg_delay<-lm(days_till_donation~treat+treat_reminder+opened_reminder1+opened_reminder2+PGSS.Year+Months
#print(summary(reg_delay))

sum(merged$total_donation_amount)
```

```
## [1] 11780.61
```

```
table(merged$opened_orig_email,merged$opened_reminder1)
```

```
##
##         FALSE TRUE
##   FALSE   944  303
##   TRUE    239  618
```