W241_Project_PGSS_Campaign

Subha Vadakkumkoor, Walter Burge

August 3, 2018

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

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).

```
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)
```

```
#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("pened_orig_email"))

orig_email_resp<-read.csv('./data/BlastReport_Class Experiment Final Email_Conversions.csv')
#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_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_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
```

```
#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"                  "pened_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 with the original email response
merged<-merge(orig_email_rec,orig_email_resp[,c("Supporter.ID","Orig_email_conversion_date","orig_email

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

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

```
#Merge with the first reminder response
merged<-merge(merged,reminder1_resp[,c("Supporter.ID","reminder1_conversion_date","reminder1_amount","d

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

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

```
#Merge with the second reminder response
merged<-merge(merged,reminder2_resp[,c("Supporter.ID","reminder2_conversion_date","reminder2_amount","d

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

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

```
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


#Total donation amount
merged$total_donation_amount=merged$orig_email_amount+merged$reminder1_amount+merged$reminder2_amount
```

```
## Warning in Ops.factor(merged$orig_email_amount, merged$reminder1_amount):
## '+' not meaningful for factors
```

```
## Warning in Ops.factor(merged$orig_email_amount + merged$reminder1_amount, :
## '+' not meaningful for factors
```

```
#CHECK FOR MULTIPLE DONATIONS

#Create time of donation indicator - when did they donate

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

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

table(merged$donation_status)
```

```
## < table of extent 0 >
```

```
#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
```

Initial regression trial

```
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
```

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

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

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

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

```r
#ADD DAYS SINCE DONATION
```

Initial regression trial

```r
reg1<-lm(donated_after_orig_email~treat, data=merged)
print(summary(reg1))
```

```
##
## Call:
## lm(formula = donated_after_orig_email ~ treat, data = merged)
##
## 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) 0.006635   0.003263   2.033   0.0421 *
## treat       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
```

```r
#Add other regressions here
```