Homework 3

Hanao Li

April 7, 2019

```
library(Hmisc)
library(data.table)
library(DT)
```

```
dat <- fread(input = "../Data/Homework 3 Data.csv", verbose = FALSE)
```

```
percentage.table <- function(x, digits = 1){
  tab <- table(x)
  percentage.tab <- 100*tab/(sum(tab))
  rounded.tab <- round(x = percentage.tab, digits = digits)
  return(rounded.tab)
}

mean.diff <- function(x, y){
  return(mean(x, na.rm=TRUE) - mean(y, na.rm=TRUE))
}

is.nan.function <- function(x){
  do.call(cbind, lapply(x, is.nan))
}
```

```
#Constant
id.name <- "id"
age.name <- "Age"
gender.name <- "Gender"
income.name <- "Income"
region.name <- "Region"
persona.name <- "Persona"
product.name <- "Product"
awareness.name <- "Awareness"
consideration.name <- "Consideration"
consumption.name <- "Consumption"
satisfaction.name <- "Satisfaction"
advocacy.name <- "Advocacy"
bp.pattern <- "BP_"
age.group.name <- "Age_Group"
income.group.name <- "Income_Group"
cuts.age <-c(18, 35, 50, 65, 120)
cuts.income <-1000*c(0, 50, 75, 100, 150, 250)

dat[, eval(age.group.name) := cut2(x = get(age.name), cuts = cuts.age)]
dat[, eval(income.group.name) := cut2(x = get(income.name), cuts = cuts.income)]
dat[, eval(satisfaction.name) := get(satisfaction.name) / 10]
unique.age.groups <-dat[, sort(unique(get(age.group.name)))]
unique.genders <-dat[, sort(unique(get(gender.name)))]
unique.income.groups <-dat[, sort(unique(get(income.group.name)))]
unique.regions <-dat[, sort(unique(get(region.name)))]
unique.personas <-dat[, sort(unique(get(persona.name)))]
unique.products <-dat[, unique(get(product.name))]
respondent.variables <-c(age.group.name, gender.name, income.group.name, region.name, persona.na
me)
states.of.engagement <-c(awareness.name, consideration.name, consumption.name, satisfaction.nam
e, advocacy.name)
bp.traits <-names(dat)[grep(pattern = bp.pattern, x = names(dat))]
```

Question 1

a)

```
#Age Group Percentage
percentage.table(x = dat[get(product.name) ==get(product.name)[1], get(age.group.name)])
```

```
## x
## [ 18, 35) [ 35, 50) [ 50, 65) [ 65,120]
##      31.1      26.9      27.1      14.9
```

```
#Gender Percentage
percentage.table(x = dat[get(product.name) ==get(product.name)[1], get(gender.name)])
```

```
## x
## Female   Male
##   53.4   46.6
```

```
#Income Group Percentage
percentage.table(x = dat[get(product.name) ==get(product.name)[1], get(income.group.name)])
```

```
## x
## [     0, 50000) [ 50000, 75000) [ 75000,100000) [100000,150000)
##            32.7            15.5            21.2            22.0
## [150000,250000]
##             8.6
```

```
#Region Percentage
percentage.table(x = dat[get(product.name) ==get(product.name)[1], get(region.name)])
```

```
## x
##   Midwest Northeast     South      West
##      26.2      25.6      24.1      24.1
```

```
#Persona Percentage
percentage.table(x = dat[get(product.name) ==get(product.name)[1], get(persona.name)])
```

```
## x
##      Ambivalent Adventurer     Consistent Compromiser
##                       24.5                       17.2
##    Materialistic Meditator        Outdoorsy Ombudsman
##                       11.5                       14.1
##   Precociously Preoccupied Technological Triumphalist
##                        9.3                       23.4
```

b)

Visual Display in Reporting Engine

Question 2

#a)

```
#Top 5 in the Northeast
awareness_rates <- dat[get(region.name) == "Northeast", .(Mean = round(100 * mean(get(awareness.
name), na.rm = TRUE), 1)), by = product.name]
setorderv(x = awareness_rates, cols = "Mean", order = -1)
awareness_rates[1:5]
```

```
##          Product Mean
## 1:  Maybe Mobile 84.7
## 2: Mobile Mayhem 79.0
## 3:     Next Text 73.8
## 4:       App Map 72.9
## 5:      Buzzdial 71.0
```

b)

```
#Top 5 Female Advocacy at least $100000
advocacy_rates <- dat[get(income.name) >= 100000 & get(gender.name) == "Female", .(Mean = round(
100 * mean(get(advocacy.name), na.rm = TRUE), 1)), by = product.name]
setorderv(x = advocacy_rates, cols = "Mean", order = -1)
advocacy_rates[1:5]
```

```
##            Product  Mean
## 1: Communic Nation 100.0
## 2:    Smartophonic  87.8
## 3:       Next Text  85.2
## 4:   Mobile Mayhem  85.2
## 5:      Triumphone  77.8
```

c)

Visual Display in Reporting Engine

Question 3

```
#Inverted Average Score Function for Negative Traits
IAS <- function(x){
  10 - mean(x, na.rm = TRUE)
}
```

a)

```
#Separate Into Positive and Negative Traits
bp_positive <- bp.traits[1:8]
bp_negative <- bp.traits[9:12]
#Top 5 Brands overall
dat_positive <- dat[, lapply(X = .SD, FUN = "mean", na.rm = TRUE), by = product.name, .SDcols =
 bp_positive]
dat_negative <- dat[, lapply(X = .SD, FUN = "IAS"), by = product.name, .SDcols = bp_negative]
dat_all <- merge(dat_positive, dat_negative)
overall_perception <- dat_all[, .(Average = round(rowMeans(.SD), 1)), by = product.name, .SDcols
= bp.traits]
setorderv(x = overall_perception, cols = "Average", order = -1)
overall_perception[1:5]
```

```
##         Product Average
## 1:      Screenz     7.2
## 2:  Cellularity     6.4
## 3:    Next Text     6.3
## 4:   No Buttons     6.2
## 5: Smartophonic     6.1
```

b)

Visual Display in Reporting Engine

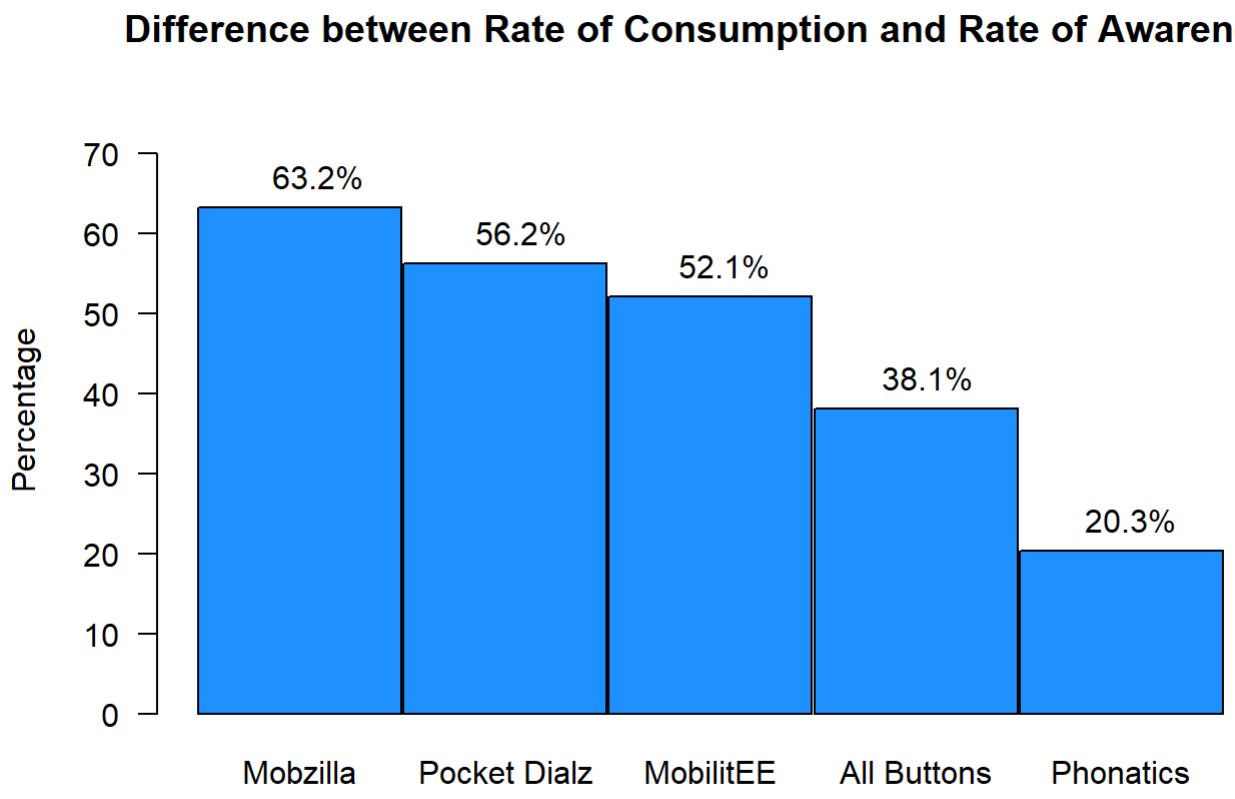
Question 4

a)

```
diff_con_awa <- dat[, .(Difference = round(mean.diff(get(consumption.name), get(awareness.name))
* 100, 1)), by = product.name]
setorderv(diff_con_awa, cols = "Difference", order = -1)
diff_con_awa[1:5]
```

```
##         Product Difference
## 1:     Mobzilla       63.2
## 2: Pocket Dialz       56.2
## 3:    MobilitEE       52.1
## 4:  All Buttons       38.1
## 5:    Phonatics       20.3
```

```
barplot(height = diff_con_awa[1:5, Difference], names.arg = diff_con_awa[1:5, Product], space=0.
01, las =1, main =sprintf('Difference between Rate of Consumption and Rate of Awareness'), ylab
 ="Percentage", ylim =c(0, 1.2*max(diff_con_awa[1:5, Difference], na.rm =TRUE)),col ="dodgerblu
e")
text(x =-0.4+1:diff_con_awa[1:5, .N] *(1+0), y = diff_con_awa[1:5, Difference], labels =sprintf(
"%.1f%%", diff_con_awa[1:5, Difference]), pos =3)
```

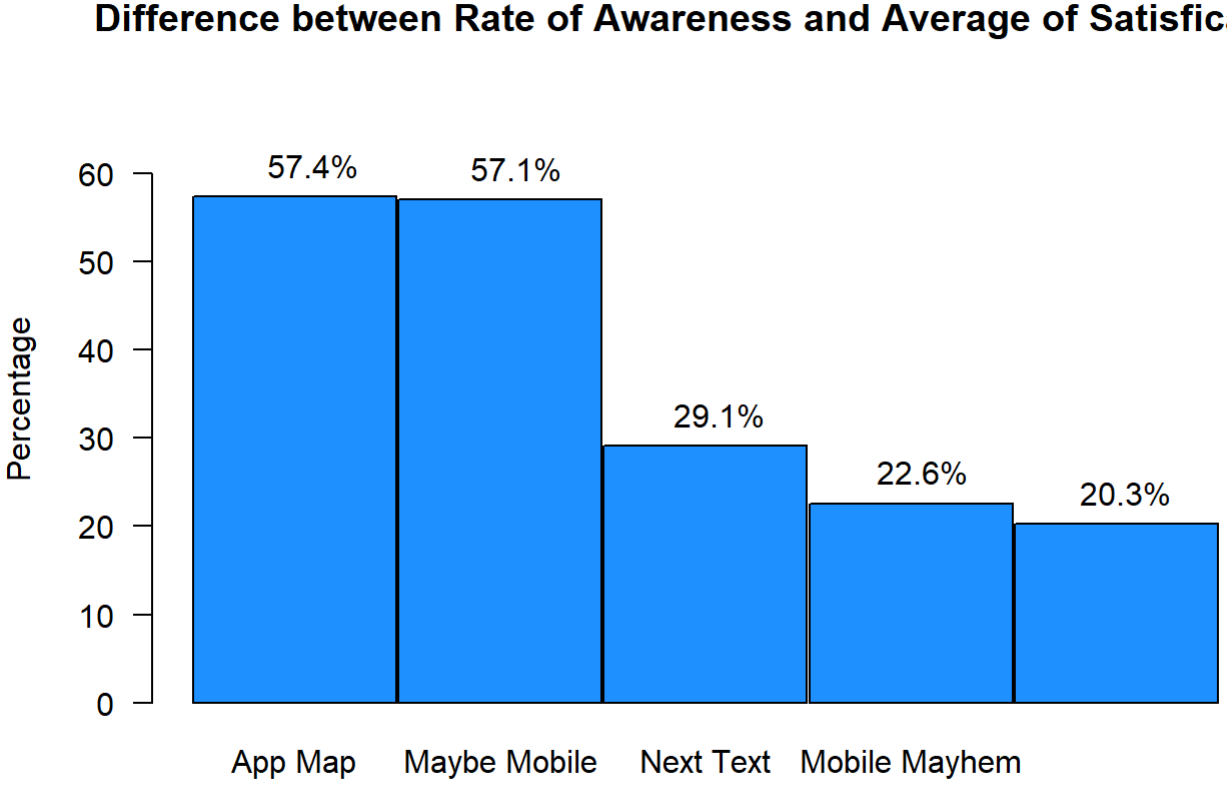


b)

```
diff_awa_sat <- dat[, .(Difference = round(mean.diff(get(awareness.name), get(satisfaction.nam
e)) * 100, 1)), by = product.name]
setorderv(diff_awa_sat, cols = "Difference", order = -1)
diff_awa_sat[1:5]
```

```
##          Product Difference
## 1:       App Map       57.4
## 2:  Maybe Mobile       57.1
## 3:     Next Text       29.1
## 4: Mobile Mayhem       22.6
## 5:   Cellularity       20.3
```

```
barplot(height = diff_awa_sat[1:5, Difference], names.arg = diff_awa_sat[1:5, Product], space=0.
01, las =1, main =sprintf('Difference between Rate of Awareness and Average of Satisfication'),
 ylab ="Percentage", ylim =c(0, 1.2*max(diff_awa_sat[1:5, Difference], na.rm =TRUE)),col ="dodge
rblue")
text(x =-0.4+1:diff_awa_sat[1:5, .N] *(1+0), y = diff_awa_sat[1:5, Difference], labels =sprintf(
"%.1f%%", diff_awa_sat[1:5, Difference]), pos =3)
```



c)

Visual Display in Reporting Engine

Question 5

a)

```
awa_agg <- merge(dat[get(product.name) == "Buzzdial"], dat[get(product.name) != "Buzzdial", .(Ag
gregated = mean(get(awareness.name), na.rm = TRUE)), by = id.name])
awa_agg$Aggregated[is.nan(awa_agg$Aggregated)] <- 0
awa_mod <- glm(Awareness ~ Age_Group + Income_Group + Gender + Region + Persona + Aggregated, fa
mily = binomial, data = awa_agg)
awa_value <- round(cbind(exp(cbind(coef(awa_mod), confint(awa_mod))), coef(summary(awa_mod))[, 4
]), 3)
```

```
## Waiting for profiling to be done...
```

```
colnames(awa_value) <- c("Odds Ratio", "2.5%", "97.5%", "P-value")
datatable(awa_value)
```

Show 10 ▼ **entries** **Search:**

	Odds Ratio	2.5%	97.5%	P-value

	Odds Ratio	2.5%	97.5%	P-value
(Intercept)	1.364	1.068	1.742	0.013
Age_Group[35, 50)	1.809	1.62	2.021	0
Age_Group[50, 65)	0.878	0.791	0.975	0.015
Age_Group[65,120]	1.275	1.122	1.45	0
Income_Group[50000, 75000)	0.919	0.81	1.043	0.191
Income_Group[75000,100000)	0.971	0.866	1.089	0.612
Income_Group[100000,150000)	0.927	0.828	1.038	0.188
Income_Group[150000,250000]	0.88	0.753	1.03	0.11
GenderMale	0.864	0.795	0.938	0
RegionNortheast	2.035	1.812	2.286	0

Showing 1 to 10 of 18 entries Previous 1 2 Next

From the model result, we could see that for a threshold value of 0.05, intercept, age group, gender, Region Northeast and Region West, Persona Consistent Compromiser and significant variables. Aggregated is not significant since its pvalue is 0.226. It has a odds ratio 95% confidence from 0.495 to 1.181 and 1 is included in the interval and we can say awareness of other products does not have an impact on the awareness of this product.

b)

```
sat_agg <- merge(dat[get(product.name) == "Buzzdial"], dat[get(product.name) != "Buzzdial", .(Ag
gregated = mean(get(satisfaction.name), na.rm = TRUE)), by = id.name])
sat_agg$Aggregated[is.nan(sat_agg$Aggregated)] <- 0
sat_mod <- lm(Satisfaction ~ Age_Group + Income_Group + Gender + Region + Persona + Aggregated,
 data = sat_agg)
sat_value <- round(cbind(coef(sat_mod), confint(sat_mod), summary(sat_mod)$coefficients[, 4]), 3
)
colnames(sat_value) <- c("Coefficients", "2.5%", "97.5%", "P-value")
datatable(sat_value)
```

Show 10 entries Search:

	Coefficients	2.5%	97.5%	P-value
(Intercept)	0.725	0.711	0.74	0
Age_Group[35, 50)	0.043	0.033	0.053	0

	Coefficients	2.5%	97.5%	P-value
Age_Group[50, 65)	0.055	0.044	0.066	0
Age_Group[65,120]	0.062	0.049	0.076	0
Income_Group[50000, 75000)	-0.004	-0.016	0.008	0.529
Income_Group[75000,100000)	-0.003	-0.014	0.008	0.57
Income_Group[100000,150000)	0.002	-0.009	0.013	0.692
Income_Group[150000,250000]	-0.002	-0.017	0.014	0.843
GenderMale	-0.002	-0.01	0.006	0.695
RegionNortheast	-0.061	-0.073	-0.05	0

Showing 1 to 10 of 18 entries Previous 1 2 Next

From the model result, we could see that for a threshold value of 0.05, intercept, age group, Region, Persona and Aggregated variables are all significant. It has a coeficients 95% confidence interval from 0.044 to 0.079 and 0 is not included and we can say satisfaction of other products does have a impositive effact on the satisfaction of this product.

c)

Visual Display in Reporting Engine