

IntentionBehaviorGap

Yejin Hwang

3/7/2019

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.1.0      v purrr  0.3.2
## v tibble  2.1.1      v dplyr  0.8.0.1
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(haven) #haven is part of tidyverse, but for some reason, this is needed.
library(ggplot2)
library(cowplot)

##
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggplot2':
##
##      ggsave

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

data1 <- read_dta("./data/final_data1.dta")
data2 <- read_dta("./data/final_data2.dta")
data3 <- read_dta("./data/final_data3.dta")
```

Recreating Table 2

```
# Read in data file, assign to MyData
MyData <- data1

# Select variables in Table 2 from MyData, put in dataframe d
vars <- c(34, 35, 36, 51, 52)
d <- data1[,vars]

# Compute means without missing (NA) values
summary(d)

##      goal_opport      goal_submit      goal_hours      activity_days
```

```
## Min. : 2.00 Min. : 2.000 Min. : 1.000 Min. :0.000
## 1st Qu.: 5.00 1st Qu.: 5.000 1st Qu.: 4.000 1st Qu.:2.000
## Median :10.00 Median : 7.000 Median : 7.000 Median :4.000
## Mean :10.35 Mean : 7.821 Mean : 8.458 Mean :3.751
## 3rd Qu.:12.00 3rd Qu.:10.000 3rd Qu.:10.000 3rd Qu.:5.000
## Max. :30.00 Max. :20.000 Max. :25.000 Max. :7.000
## NA's :761 NA's :760 NA's :755 NA's :743
## APcomplete
## Min. :0.0000
## 1st Qu.:1.0000
## Median :1.0000
## Mean :0.8881
## 3rd Qu.:1.0000
## Max. :1.0000
## NA's :698
```

```
sapply(d, mean, na.rm=TRUE)
```

```
## goal_opport goal_submit goal_hours activity_days APcomplete
## 10.3480826 7.8205882 8.4579710 3.7507003 0.8880597
```

```
sapply(d, sd, na.rm=TRUE)
```

```
## goal_opport goal_submit goal_hours activity_days APcomplete
## 6.4973655 4.3321042 5.7830548 2.3766256 0.3156859
```

Note: windsorized: how they deal with wacky means / outliers i.e. using x9 to replace x10 use windsor.mean function may be prewindsorized data

Tables:

```
controlGroup <- data1 %>% filter(treatment_group=="control") %>% select(age_yr, female_d, educ_yr, bs_b
stargazer::stargazer(controlGroup, type = "text")
```

```
##
## =====
## Statistic N Mean St. Dev. Min Pctl(25) Pctl(75) Max
## =====
```

summary stats: tables 1 and 2 Table 3-Effects on Job Search Intensity. panel-data regression and se. outcome var winsorized=remove extreme values Table 4-Effects on Employment Outcomes. regression and se. Table 5-Effects on Frequency of Search-Channel Use

Graphs: Figure 1. Intention-Behavior Gap: Difference at Baseline - density plot Hours and apps difference: Goal - Baseline

```
hours_diff_density <- ggplot(data1, aes(hours_diff, stat(density))) + geom_histogram(fill = "#416b85", b
apps_diff_density <- ggplot(data1, aes(apps_diff, stat(density))) + geom_histogram(fill = "#416b85", bin
plot_grid(hours_diff_density, apps_diff_density)
```

```
## Warning: Removed 769 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 763 rows containing non-finite values (stat_bin).
```

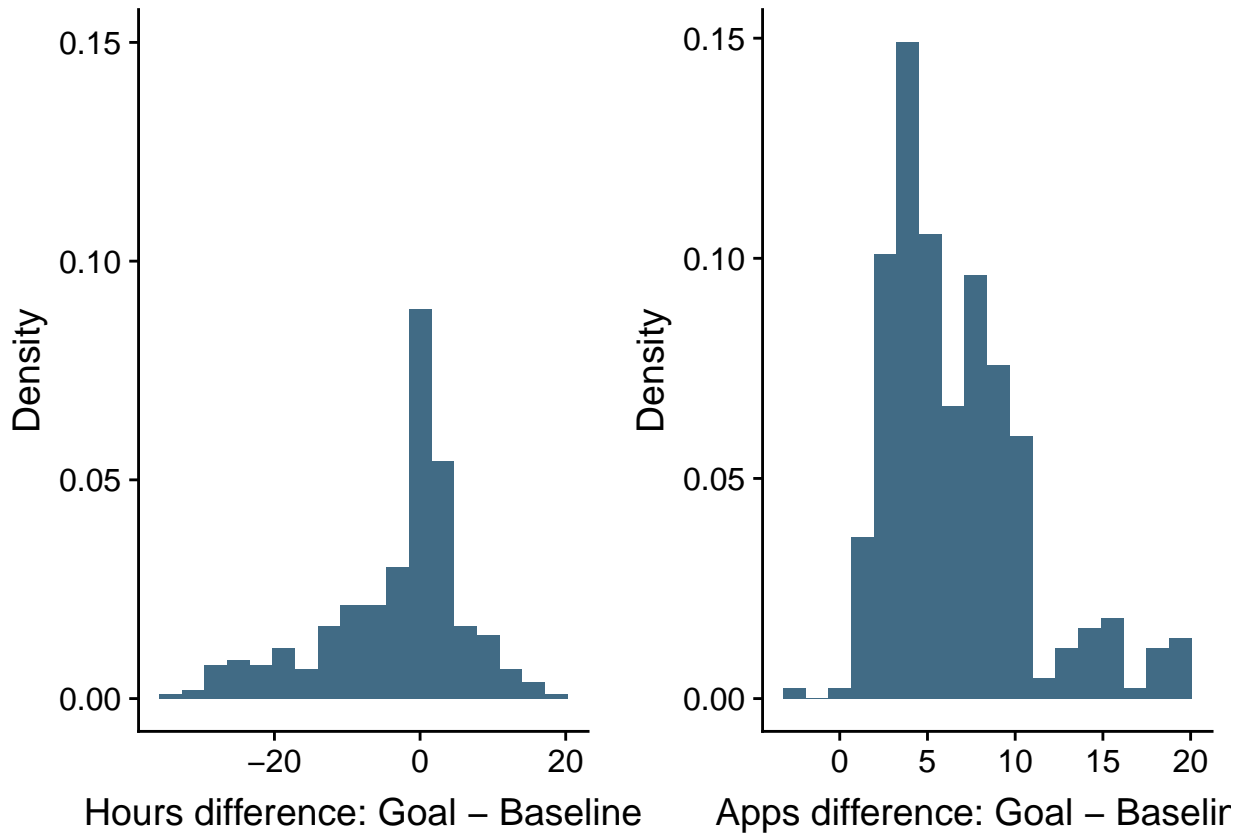


Table 5

```
#what is the frequency
#data2
#stargazer: exclude variables
#b1_1_t: Empl. agency
#b1_2_t: dropped cv
#b1_3_t: placed ad
#b1_4_t: answered ad
#b1_5_t: searched online
#b1_6_t: Fam./Friends
#b4_t: employer responses
#filter by: ws_plus_d, ws_d
m1 <- lm(b1_1_t ~ ws_d + ws_plus_d + educ_yr + age_yr + female_d + bs_b15a + lang_xhosa_d + lang_venda_d)
m2 <- lm(b1_2_t ~ ws_d + ws_plus_d + educ_yr + age_yr + female_d + bs_b15a + lang_xhosa_d + lang_venda_d)
m3 <- lm(b1_3_t ~ ws_d + ws_plus_d + educ_yr + age_yr + female_d + bs_b15a + lang_xhosa_d + lang_venda_d)
m4 <- lm(b1_4_t ~ ws_d + ws_plus_d + educ_yr + age_yr + female_d + bs_b15a + lang_xhosa_d + lang_venda_d)
m5 <- lm(b1_5_t ~ ws_d + ws_plus_d + educ_yr + age_yr + female_d + bs_b15a + lang_xhosa_d + lang_venda_d)
m6 <- lm(b1_6_t ~ ws_d + ws_plus_d + educ_yr + age_yr + female_d + bs_b15a + lang_xhosa_d + lang_venda_d)
stargazer(m1, m2, m3, m4, m5, m6, column.labels = c("Empl. agency", "Dropped CV", "Placed ad", "Answered ad", "Searched online", "Fam./Friends", "Employer responses"))

##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
## % Date and time: Thu, Apr 25, 2019 - 11:16:58
## \begin{table}[!htbp] \centering
##   \caption{Table 5-Effects on Frequency of Search-Channel Use}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccccc}
##   \hline
##   \end{tabular}
## \end{table}
```

```

## \[-1.8ex]\hline
## \hline \[-1.8ex]
## & \multicolumn{6}{c}{\textit{Dependent variable:}} \\
## \cline{2-7}
## \[-1.8ex] & b1\_1\_t & b1\_2\_t & b1\_3\_t & b1\_4\_t & b1\_5\_t & b1\_6\_t \\
## & Empl. agency & Dropped CV & Placed ad & Answered ad & Searched online & Fam./friends \\
## \[-1.8ex] & (1) & (2) & (3) & (4) & (5) & (6) \\
## \hline \[-1.8ex]
## ws\_d &  $-\$0.016$  &  $-\$0.140$  &  $0.060$  &  $-\$0.068$  &  $0.058$  &  $-\$0.048$  \\
## &  $(0.130)$  &  $(0.117)$  &  $(0.131)$  &  $(0.111)$  &  $(0.103)$  &  $(0.087)$  \\
## & & & & & & \\
## ws\_plus\_d &  $0.362^{***}$  &  $0.253^{**}$  &  $0.153$  &  $0.301^{***}$  &  $0.410^{***}$  &  $-\$0.021$  \\
## &  $(0.111)$  &  $(0.100)$  &  $(0.112)$  &  $(0.095)$  &  $(0.088)$  &  $(0.075)$  \\
## & & & & & & \\
## \hline \[-1.8ex]
## Observations & 1,937 & 1,936 & 1,934 & 1,927 & 1,931 & 1,926 \\
## R2 & 0.093 & 0.069 & 0.028 & 0.088 & 0.364 & 0.045 \\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{6}{r}{ $^{*}p < 0.1$ ;  $^{**}p < 0.05$ ;  $^{***}p < 0.01$ } \\
## \end{tabular}
## \end{table}

```

#stargazer type = latex