

# IntentionBehaviorGap

Yejin Hwang

3/7/2019

```
library(tidyverse)

## -- Attaching packages ----- tidyverse
## v ggplot2 3.1.0      v purrr  0.3.0
## v tibble  2.0.1      v dplyr  0.8.0.1
## v tidyr   0.8.2      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

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

## Recreating Table 2

```
# 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
```

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

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
## =====
```

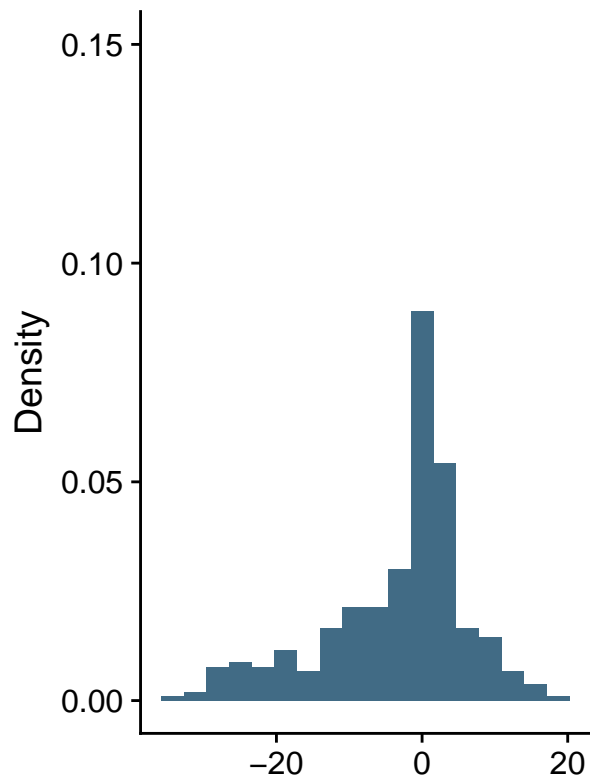
summary stats: tables 1 and 2 Table 3-Effects on Job Search Intensity. panel-data regression and se. outcome var winsorized=remove extreme values Table 4-Effects on Employment Outcomes. regression and se. Table 5-Effects on Frequency of Search-Channel Use

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

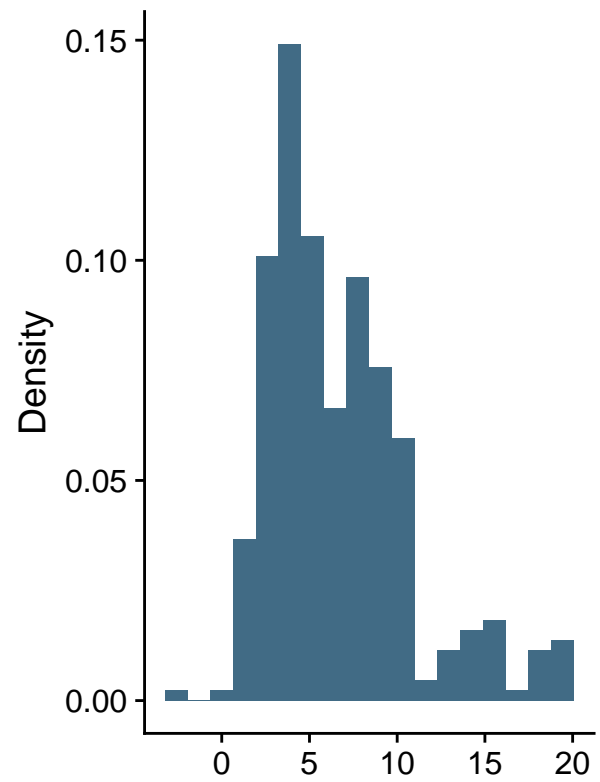
```
hours_diff_density <- ggplot(data1, aes(hours_diff, stat(density))) + geom_histogram(fill = "#416b85", b
apps_diff_density <- ggplot(data1, aes(apps_diff, stat(density))) + geom_histogram(fill = "#416b85", bi
plot_grid(hours_diff_density, apps_diff_density)
```

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

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



Hours difference: Goal - Baseline



Apps difference: Goal - Baseline