BA830 Team Project-Team 13

Antonio Moral Cevallos, Bosoo Kim, Jiajian(Sylar) Guo, Manushi Patel, Ying(Amber) Wu, Yixuan Wang

4/29/2021

Introduction

(Amber)

```
library(data.table)
library(tidyverse)
library(lfe)
library(fixest)
library(lubridate)
library(stargazer)
library(modelsummary)
library(mltools)
library(knitr)
library(broom)
library(purrr)
# load dataset
treatment <- fread('Cognitive_Test_1.csv')
control<- fread('Cognitive_Test_2.csv')
```

```
# Data cleaning
treatment <- treatment[, treatment := 1]
control <- control[, treatment := 0]
total <- rbind(treatment, control)
total$StartDate <- as.Date(total$StartDate, format= "%Y-%m-%d")
total <- total[!(total$Finished=='False')]
total <- total[c(3:28,31:53), c("StartDate", "IPAddress", "Duration (in seconds)", "L
ocationLatitude", "LocationLongitude", "Q2", "Q3", "Q4", "Q5", "SC0", "treatment")]
IPadd <- split(total,total$IPAddress)
total2 <- data.frame()
for(x in IPadd){total2 <- rbind(total2,x[1,])}
total <- total2
```

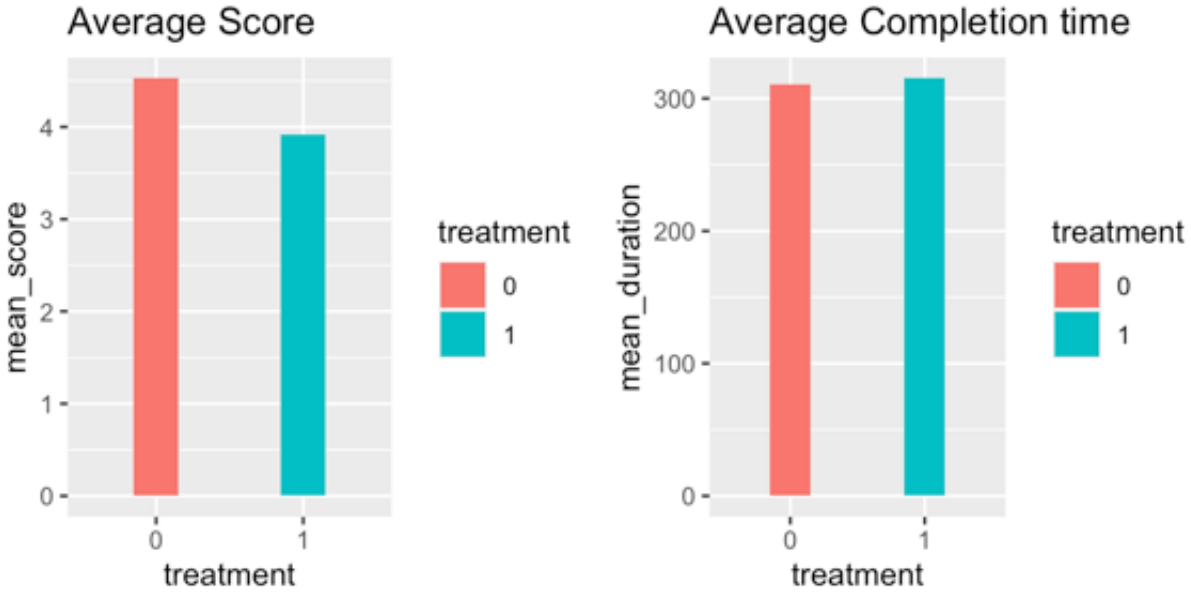
```
# change column names and data type
colnames(total)[c(1,3,6,7,8,9,10)] = c("date","duration","age","gender","GPA","work_e
xperience","score")
total$duration <- as.numeric(total$duration)
total$score <- as.numeric(total$score)
total$age <- as.factor(total$age)
total$gender <- as.factor(total$gender)
total$GPA <- as.factor(total$GPA)
total$work_experience <- as.factor(total$work_experience)
```

Method

(Manushi, Bosoo)

Data Analysis

(Antonio, Yixuan, Bosoo)



- t-test

```
t1 <- t.test(total[treatment == 1, score], total[treatment == 0, score])
t2 <- t.test(total[treatment == 1, duration], total[treatment == 0, duration])
tab <- map_df(list(t1, t2), tidy)
tab
```

```
## # A tibble: 2 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low conf.high
##      <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>     <dbl>
## 1   -0.603      3.92      4.53   -1.42     0.162      43.0    -1.46     0.252
## 2    4.64     315.      310.      0.0956   0.924      42.5   -93.3    103.
## # … with 2 more variables: method <chr>, alternative <chr>
```

Since p-value is larger than 0.05, the true difference in mean scores between the control group and treatment group is not statistically significant from 0. In other words, the mean scores between two groups are not significantly different. Similarly, the mean completion time between two groups are also not significantly different.

Randomization check

	Treatment	Control	P-value
Under 20	0.0384615	0.1052632	0.515617
Between 20-30	0.9230769	0.8947368	1.000000
Over 30	0.0384615	0.0000000	0.000000

	Treatment	Control	P-value
Male	0.6153846	0.3157895	0.0022497
Female	0.3461538	0.6315789	0.0037487
NonBinary	0.0384615	0.0000000	0.0000000

	Treatment	Control	P-value
Under 3.0	0.0384615	0.0000000	0.0000000
Between 3.0-3.5	0.1923077	0.5263158	0.0006283
Over 3.5	0.7692308	0.4736842	0.0027657

	Treatment	Control	P-value
No Experience	0.3461538	0.3684211	1
Under 2 years	0.4230769	0.4210526	1
Between 2-5	0.1923077	0.2105263	1
Between 6-9+	0.0384615	0.0000000	0

Regression of Score on treatment

```
score_reg <- lm(score~treatment, data=total)
score_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Sta
t", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	4.526	0.340	13.303	0.000
treatment	-0.603	0.448	-1.348	0.185

- Controlling for Age

```
score_age_reg <- feols(score~treatment + age, data=total, se='white')
score_age_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T
-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	2.583	0.713	3.620	0.001
treatment	-0.748	0.391	-1.914	0.063
age>30	2.165	0.749	2.891	0.006
age20-30	2.172	0.723	3.004	0.005

- Controlling for Gender

```
score_gender_reg <- feols(score~treatment + gender, data=total, se='white')
score_gender_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE",
"T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	4.481	0.292	15.335	0.000
treatment	-0.678	0.519	-1.305	0.199
genderMale	0.057	0.533	0.108	0.915
genderNon-binary / third gender	2.197	0.508	4.325	0.000
genderPrefer not to say	0.519	0.292	1.777	0.083

- Controlling for GPA

```
score_gpa_reg <- feols(score~treatment + GPA, data=total, se='white')
score_gpa_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T
-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	3.606	0.424	8.510	0.000
treatment	-0.606	0.424	-1.430	0.160
GPA>3.50	0.984	0.382	2.578	0.014
GPA3.00 - 3.50	0.863	0.412	2.092	0.043

- Controlling for Work Experience

```
score_exp_reg <- feols(score~treatment + work_experience, data=total, se='white')
score_exp_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T
-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	4.560	0.429	10.633	0.000
treatment	-0.603	0.434	-1.391	0.172
work_experience0 years	-0.470	0.529	-0.889	0.379
work_experience2 - 5 years	0.664	0.544	1.221	0.229
work_experience6 - 9 years	0.043	0.409	0.106	0.916

- Controlling for All covariates

```
score_all_reg <- feols(score~treatment + age + gender + GPA + work_experience, data=t
otal, se='white')
score_all_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T
-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	0.839	1.078	0.778	0.442
treatment	-0.646	0.485	-1.333	0.191

age>30	2.019	1.077	1.874	0.070
age20-30	1.954	0.906	2.156	0.038
genderMale	0.026	0.581	0.044	0.965
genderNon-binary / third gender	2.064	0.758	2.725	0.010
genderPrefer not to say	0.706	0.335	2.108	0.042
GPA>3.50	1.789	0.673	2.658	0.012
GPA3.00 - 3.50	1.988	0.825	2.410	0.021
work_experience0 years	-0.487	0.569	-0.855	0.398
work_experience2 - 5 years	0.853	0.632	1.349	0.186

Regression of completion time on treatment

```
duration_reg <- feols(duration~treatment, data=total, se='white')
duration_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	310.474	33.163	9.362	0.000
treatment	4.642	48.536	0.096	0.924

- Controlling for Age

```
duration_age_reg <- feols(duration~treatment + age, data=total, se='white')
duration_age_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE", "T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	244.131	97.794	2.496	0.017
treatment	-7.393	49.258	-0.150	0.881
age>30	258.262	100.750	2.563	0.014
age20-30	74.148	99.127	0.748	0.459

- Controlling for Gender

```
duration_gender_reg <- feols(duration~treatment + gender, data=total, se='white')
duration_gender_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "S
E", "T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	319.023	39.133	8.152	0.000
treatment	-3.832	48.228	-0.079	0.937
genderMale	-2.736	49.398	-0.055	0.956
genderNon-binary / third gender	41.809	35.056	1.193	0.240
genderPrefer not to say	-146.023	39.133	-3.731	0.001

- Controlling for GPA

```
duration_gpa_reg <- feols(duration~treatment + GPA, data=total, se='white')
duration_gpa_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE",
"T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	312.840	45.249	6.914	0.000
treatment	7.160	45.249	0.158	0.875
GPA>3.50	-6.743	41.953	-0.161	0.873
GPA3.00 - 3.50	1.573	42.562	0.037	0.971

- Controlling for Work Experience

```
duration_exp_reg <- feols(duration~treatment + work_experience, data=total, se='white
')
duration_exp_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE",
"T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	334.571	51.636	6.479	0.000
treatment	-3.350	50.426	-0.066	0.947
work_experience0 years	-40.874	58.760	-0.696	0.491

work_experience2 - 5 years	-42.932	64.313	-0.668	0.508
work_experience6 - 9 years	163.779	56.781	2.884	0.006

- Controlling for All covariates

```
duration_all_reg <- feols(duration~treatment + age + gender + GPA + work_experience,
data=total, se='white')
duration_all_reg %>% tidy() %>% kable(col.names = c("Predictor", "Coefficient", "SE",
"T-Stat", "P-Value"), digits = c(0, 3, 3, 3, 3), align = 'c')
```

Predictor	Coefficient	SE	T-Stat	P-Value
(Intercept)	301.039	131.732	2.285	0.029
treatment	-21.693	44.671	-0.486	0.630
age>30	265.148	116.742	2.271	0.030
age20-30	93.544	122.479	0.764	0.450
genderMale	9.294	48.805	0.190	0.850
genderNon-binary / third gender	33.604	70.516	0.477	0.637
genderPrefer not to say	-145.559	61.437	-2.369	0.024
GPA>3.50	-49.494	58.948	-0.840	0.407
GPA3.00 - 3.50	-44.008	94.563	-0.465	0.645
work_experience0 years	-32.016	66.221	-0.483	0.632
work_experience2 - 5 years	-52.890	86.304	-0.613	0.544

Limitations

(Sylar)

Conclusion

(Antonio)