

# PSET 2 Dev and Data

Matthew Zhao

11/5/2022

```
library(tidyverse)
library(haven)
library(broom)
library(ggplot2)
library(fixest)
library(stargazer)
options(scipen=999)
```

## Q1

Access to savings accounts could promote longer term savings for more capital intensive projects/investments in their own businesses by creating a place where they can safely save their profits for a longer period.

## Q2

```
savings_dta <- read_dta("pset2_data/dataset_savings.dta")

savings_df <- savings_dta %>%
  mutate(active = ifelse(first6_num_trans_savings > 1, TRUE, FALSE),
         treatment_bg_boda = treatment*bg_boda,
         active_bg_boda = active*bg_boda,
         bg_boda_wave2 = (bg_boda & wave2),
         bg_malevendor_wave2 = (bg_malevendor & wave2),
         bg_malevendor_wave3 = (bg_malevendor & wave3),
         treament_bg_malevendor = treatment*bg_malevendor,
         active_bg_malevendor = active*bg_malevendor,
         literate_swahili = ifelse((bg_kis_read == 1 & bg_kis_write ==1),
                                   1,
                                   NA),
         literate_swahili = ifelse((bg_kis_read ==0 | bg_kis_write == 0),
                                   0,
                                   literate_swahili))
```

```
bal_controls = c(
  "wave1",
  "wave2",
  "wave3",
```

```

'bg_boda',
'bg_age' ,
'bg_married',
'bg_num_children',
'bg_educ',
'literate_swahili',
'bg_rosca',
'bg_rosca_contrib_lyr',
'bg_animalsvalue',
'bg_totalinc_lastweek',
'bg_loan_bank',
'bg_loan_friend',
'bg_healthstatus',
'per_hard_save',
'per_invest_choice2',
'per_somewhat_patient',
'per_time_consistent',
'per_pat_now_impact_later',
'per_maximpat',
'per_fwd_digit_score2',
'per_ravens_matrix',
'bg_durvalue_hh',
'per_hyperbolic'
)

```

```

longer_stats <- savings_df %>%
  drop_na(treatment, filled_log) %>%
  mutate_if(is.numeric, ~replace_na(., mean(., na.rm = TRUE))) %>%
  select(bg_gender, treatment, all_of(bal_controls)) %>%
  pivot_longer(cols=bal_controls,
               names_to='name', values_to='value')

```

## Note: Using an external vector in selections is ambiguous.  
 ## i Use 'all\_of(bal\_controls)' instead of 'bal\_controls' to silence this message.  
 ## i See <<https://tidyselect.r-lib.org/reference/faq-external-vector.html>>.  
 ## This message is displayed once per session.

```

grp_means <- longer_stats %>%
  group_by(bg_gender, treatment, name) %>%
  summarise(means = mean(value),
            .groups='keep') %>%
  mutate(bg_gender = ifelse(bg_gender==0, 'Female', 'Male'),
         treatment = ifelse(treatment==0, 'Control', 'Treatment')) %>%
  pivot_wider(names_from=c(treatment, bg_gender), values_from=means,
             names_sep=', ')
pvals <- longer_stats %>%
  group_by(bg_gender, name) %>%
  summarise(pval = t.test(value ~ treatment)$p.value,
            .groups='keep') %>%
  mutate(bg_gender = ifelse(bg_gender==0,
                           'pval_diff_female',
                           'pval_diff_male')) %>%
  pivot_wider(names_from=bg_gender, values_from=pval)

```

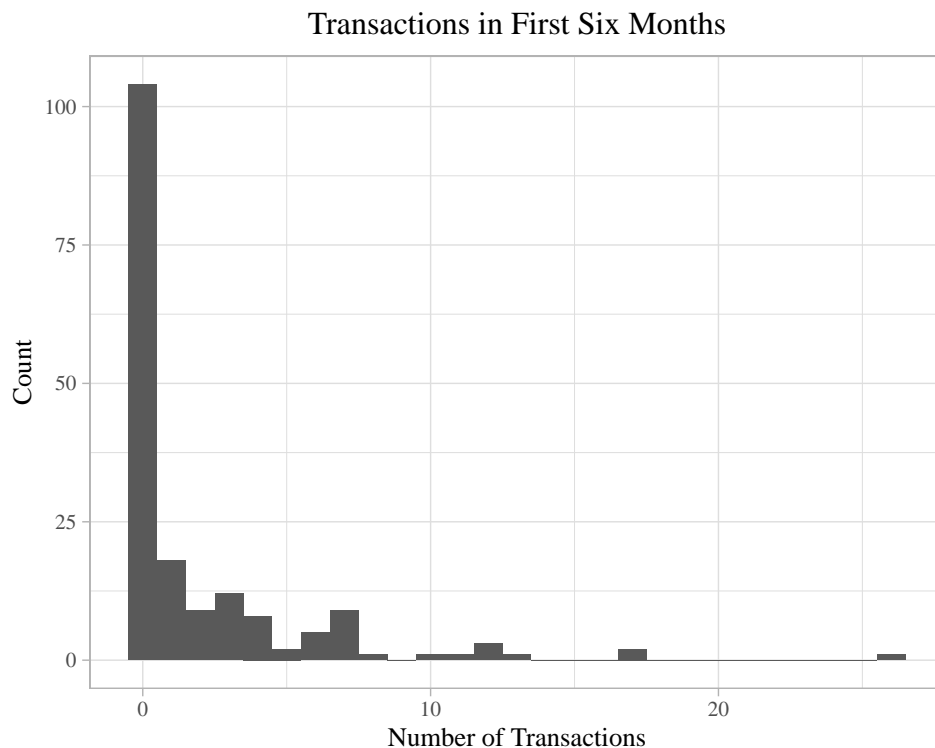
```
smy <- left_join(grp_means,pvals,by='name')
smy <- smy[match(bal_controls,smy$name),]
write_csv(smy,file='./pset2_outputs/table1.csv')
```

### Q3

This is an issue since significant p values mean that our treatment group may systematically differ from our control group, meaning that these significantly different variables could be potential confounders that prevent us from saying if the effect of treatment is causal or not.

### Q4

```
q4 <- savings_df %>%
  filter(treatment == 1) %>%
  drop_na(treatment,first6_num_trans_savings) %>%
  select(first6_num_trans_savings)
ggplot(q4, aes(x=first6_num_trans_savings)) +
  geom_histogram(binwidth = 1) + theme_light() +
  ggtitle("Transactions in First Six Months") +
  xlab("Number of Transactions") + ylab("Count") +
  theme(plot.title = element_text(family = 'serif', hjust = 0.5, size = 12),
        axis.title = element_text(family = 'serif', size = 10),
        axis.text = element_text(family = 'serif',size=8))
```



Q4a)

The authors include controls because individuals have different baseline characteristics as we saw in Q2/Q3 which could affect our estimate of the treatment effect if we don't account for them.

Q4b)

The summation term is summing the strata variables and interactions between year, gender indicator, and bodas indicator across years, capturing the total cumulative effect of different strata on the response.

Q4c)

```
outcome_variables = c(
  "active",
  "bank\\_savings",
  "animal\\_savings",
  "rosca\\_contrib"
)
q4c <- savings_df %>%
  drop_na(treatment,bg_gender,bg_boda,bg_malevendedor,
    wave2,wave3)
vars <- 'treatment + bg_gender + bg_boda + bg_malevendedor + bg_educ + wave2 + bg_gender*wave2 + bg_gender*wave3'
m1 <- lm(paste('active ~', vars),data=q4c)
m2 <- lm(paste('bank_savings ~', vars),data=q4c)
m3 <- lm(paste('animal_savings ~', vars),data=q4c)
m4 <- lm(paste('rosca_contrib ~', vars),data=q4c)
models <- list(m1,m2,m3,m4)
table2 <- stargazer(models,type='latex',digits=2,dep.var.labels=outcome_variables,
  title='Table 2: Impacts on Savings',keep = c('treatment'),header=F,suppress.errors = T)
```

Table 1: Table 2: Impacts on Savings

	<i>Dependent variable:</i>			
	active	bank_savings	animal_savings	rosca_contrib
	(1)	(2)	(3)	(4)
treatment	0.34*** (0.03)	10.79*** (3.59)	16.20* (9.47)	10.83 (7.46)
Observations	356	248	248	248
R <sup>2</sup>	0.24	0.07	0.06	0.14
Adjusted R <sup>2</sup>	0.22	0.04	0.03	0.12
Residual Std. Error	0.32 (df = 347)	27.67 (df = 239)	73.07 (df = 239)	57.51 (df = 239)
F Statistic	13.86*** (df = 8; 347)	2.30** (df = 8; 239)	1.90* (df = 8; 239)	5.05*** (df = 8; 239)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Q5)

Taking the coefficients at face value, being offered access to a bank account was associated with being 41% more likely to use a bank account on average, an average increase of 9.36 Ksh in average daily amount deposited into a bank account, an average increase of 16.79 in average daily purchases of animals, and an average increase of 7.81 KSH in average daily amount deposited into ROSCA.

The effect of offering access to bank accounts on savings is significant when measured by all except ROSCA contributions, with active bank account usage and bank savings significant at the 1% level and animal savings at the 10% level.

## Q6)

$\beta_2$  measures the average effect associated with being offered access to a bank account on our response of choice.  $\beta_2 + \gamma_2$  measures the average effect associated with being offered access to a bank account on our response of choice controlling for if the respondent is a male market vendor i.e. treatment effect for male market vendors/interaction of gender.  $\beta_2 + \delta_2$  measures the average effect associated with being offered access to a bank account on our response of choice controlling for if the respondent is a *boda* i.e. treatment effect for bodas.