

# final project

1. Balance table: first balance table for two groups before treatment; then balance table for two groups considering sample attrition.

```
##
## Balance by Pre-Treatment Covariates
## =====
##                Dependent variable:
##                -----
##                age      height    caloric
## -----
## Control Mean    19.82      1.60     1,927.40
##                p = 0.00* p = 0.00* p = 0.00*
## treatment       0.60      -0.004    10.52
##                p = 0.08   p = 0.70   p = 0.78
## -----
## Observations    100        100      100
## =====
## Note:                                                    NA
##                *p<0.05; Clustered SEs

m_list2 <- list()
m_list2[[1]] <- felm(pre_sleep ~ Di | 0 | 0 | id, data = water_df)
m_list2[[2]] <- felm(pre_exercise ~ Di | 0 | 0 | id, data = water_df)
m_list2[[3]] <- felm(pre_water ~ Di | 0 | 0 | id, data = water_df)

stargazer(m_list2,
  type = "text",
  digits = 2,
  intercept.bottom = FALSE,
  dep.var.labels = c("sleep", "exercise", "water"),
  covariate.labels = c("Control Mean", "treatment"),
  model.numbers = FALSE,
  report = "vcp*",
  star.cutoffs = c(.05, NA, NA),
  notes = "$^{*}$p$<$0.05; Clustered SEs",
  keep.stat = "n",
  title = "Balance by Pre-Treatment Covariates",
  no.space = TRUE)

##
## Balance by Pre-Treatment Covariates
## =====
##                Dependent variable:
##                -----
##                sleep    exercise    water
## -----
## Control Mean    8.40        2.46     1.09
##                p = 0.00* p = 0.00* p = 0.00*
## treatment       -0.62       0.17     -0.06
##                p = 0.03* p = 0.67   p = 0.38
## -----
## Observations    100        100      100
```

```
## =====
## Note: NA
## *p<0.05; Clustered SEs
##
## Balance by Pre-Treatment Covariates
## =====
## Dependent variable:
## -----
## weight composition
## -----
## Control Mean 66.14 77.76
## p = 0.00* p = 0.00*
## treatment 0.05 1.38
## p = 0.91 p = 0.55
## -----
## Observations 100 100
## =====
## Note: NA
## *p<0.05; Clustered SEs
```

2. Dropping out Checking the coveriates after removing dropout Checking the dropout rates

```
#remove dropouts
waterdrop_df <- water_df %>% filter(drop_out==0)

#Checking coveriates

d_list <- list()
d_list[[1]] <- febm(age ~ Di | 0 | 0 | id, data = waterdrop_df)
d_list[[2]] <- febm(height ~ Di | 0 | 0 | id, data = waterdrop_df)
d_list[[3]] <- febm(caloric_intake ~ Di | 0 | 0 | id, data = waterdrop_df)

# Creating Table:
stargazer(d_list,
type = "text",
digits = 2,
intercept.bottom = FALSE,
dep.var.labels = c("age", "height", "caloric_intake"),
covariate.labels = c("Control Mean", "treatment"), model.numbers = FALSE,
report = "vcp*",
star.cutoffs = c(.05, NA, NA),
notes = "$^{*}$p$<$0.05; Clustered SEs",
keep.stat = "n",
title = "Balance by Pre-Treatment Covariates After Removing Dropout",
no.space = TRUE)
```

## Balance by Pre-Treatment Covariates After Removing Dropout

Dependent variable:		
age	height	caloric
Control Mean 19.80 1.60 1,930.52		

p = 0.00\* p = 0.00\* p = 0.00\*  
 treatment 0.53 -0.004 7.48  
 p = 0.14 p = 0.72 p = 0.85

---

## Observations 91 91 91

Note: NA \*p<0.05; Clustered SEs

```
##
## Balance by Pre-Treatment Covariates After Removing Dropout
## =====
##               Dependent variable:
##               -----
##               sleep    exercise    water
## -----
## Control Mean    8.33        2.41        1.05
##               p = 0.00*    p = 0.00*    p = 0.00*
## treatment       -0.55        0.14       -0.05
##               p = 0.07    p = 0.74    p = 0.50
## -----
## Observations    91          91          91
## =====
## Note:                                     NA
##               *p<0.05; Clustered SEs

##
## Balance by Pre-Treatment Covariates After Removing Dropout
## =====
##               Dependent variable:
##               -----
##               weight    composition
## -----
## Control Mean    66.18        77.69
##               p = 0.00*    p = 0.00*
## treatment       0.004        1.28
##               p = 1.00    p = 0.61
## -----
## Observations    91          91
## =====
## Note:                                     NA
##               *p<0.05; Clustered SEs
```

## A tibble: 2 × 2

```
Di dropout_rate
1 0 0.08 2 1 0.10
```

3. Non-compliance, whether some girls from the treatment group were not treated, whether some girls from the control group received treatment

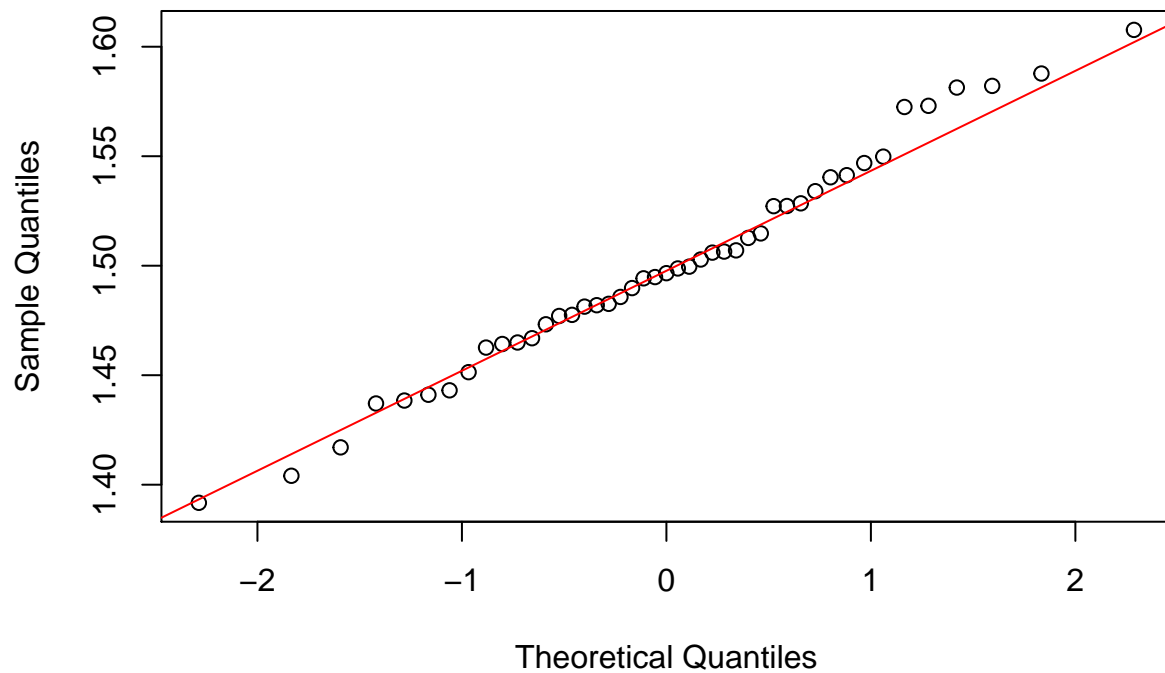
```
treatment <- water_df %>%
  mutate(extrawater = post_water - pre_water) %>%
```

```
filter(Di == 1)
shapiro.test(treatment$extrawater)
```

```
##
## Shapiro-Wilk normality test
##
## data: treatment$extrawater
## W = 0.98701, p-value = 0.8887
```

```
require(graphics)
qqnorm(treatment$extrawater);qqline(treatment$extrawater, col = 2)
```

## Normal Q-Q Plot



```
t.test(treatment$extrawater, mu=1.5)
```

```
##
## One Sample t-test
##
## data: treatment$extrawater
## t = -0.10417, df = 44, p-value = 0.9175
## alternative hypothesis: true mean is not equal to 1.5
## 95 percent confidence interval:
##  1.484351 1.514111
## sample estimates:
## mean of x
##  1.499231
```

*# shapiro test suggests it's normally distributed; ttest can be interpreted as there's no non-compliance*

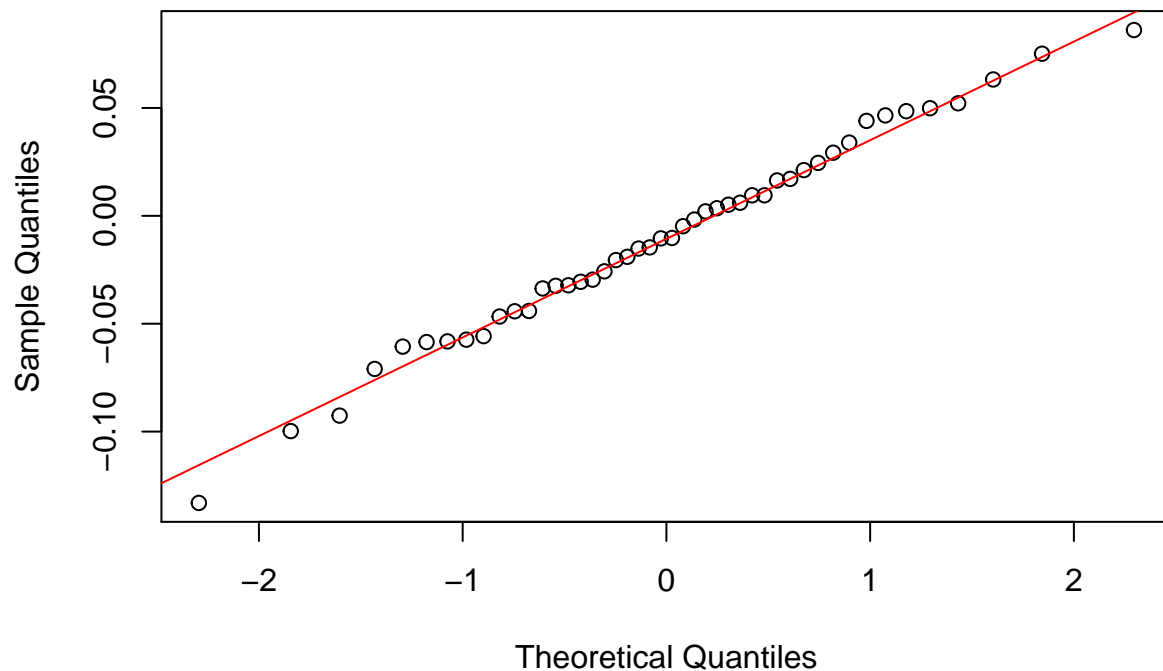
```
control <- water_df %>%
  mutate(extrawater = post_water - pre_water) %>%
  filter(Di == 0)
```

```
shapiro.test(control$extrawater)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  control$extrawater  
## W = 0.99092, p-value = 0.9741
```

```
require(graphics)  
qqnorm(control$extrawater);qqline(control$extrawater, col = 2)
```

### Normal Q-Q Plot



```
t.test(control$extrawater, mu=0)
```

```
##  
##  One Sample t-test  
##  
## data:  control$extrawater  
## t = -1.4386, df = 45, p-value = 0.1572  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
##  -0.023929538  0.003988794  
## sample estimates:  
##    mean of x  
## -0.009970372
```

*# shapiro test suggests it's normally distributed; ttest can be interpreted as there's no non-compliance*

The null-hypothesis of the shapiro test is that the population is normally distributed. Thus, if the p-value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population; in other words, the data are not normal.

Conclusion: only the treatment group received treatment and all of them received treatment; no one in the

control group received any treatment and the amount of water they drunk did not differ significantly from what they had drunk before the experiment.

#### 4. Estimate ATE

```
waterdata_df <- read.csv("waterdata.csv")
waterdata_df <- waterdata_df %>%
  filter(drop_out == 0) %>%
  group_by(Di, postperiod) %>% summarise(avg_weight = round(mean(weight, na.rm = TRUE), 2), avg_comp = 1)

#Estimate ATE
waterdata_df %>% group_by(Di) %>%
  summarise (ATE_weight = diff(avg_weight),
            ATE_comp = diff(avg_comp),
            ATE_BMI = diff(avg_BMI))
```

A tibble: 2 × 4

	Di	ATE_weight	ATE_comp	ATE_BMI
1	0	-0.78	-4.02	-0.31
2	1	-1.87	-9.80	-0.74