

final project

```
# wide to long data
# ID might need to be factor
watero_df <- water_df %>%
  select(id, age, height, caloric_intake, pre_sleep, pre_exercise, pre_weight, pre_composition, pre_water)
  mutate(postperiod = 0)

watero_df <- rename(watero_df, c(pre_weight="weight", pre_composition = "composition", pre_water = "water"))

water1_df <- water_df %>%
  select(id, age, height, caloric_intake, pre_sleep, pre_exercise, post_weight, post_composition, post_water)
  mutate(postperiod = 1)

water1_df <- rename(water1_df, c(post_weight="weight", post_composition = "composition", post_water = "water"))

waterdata_df <- rbind(watero_df, water1_df)

write.csv(waterdata_df, file = "waterdata.csv")
```

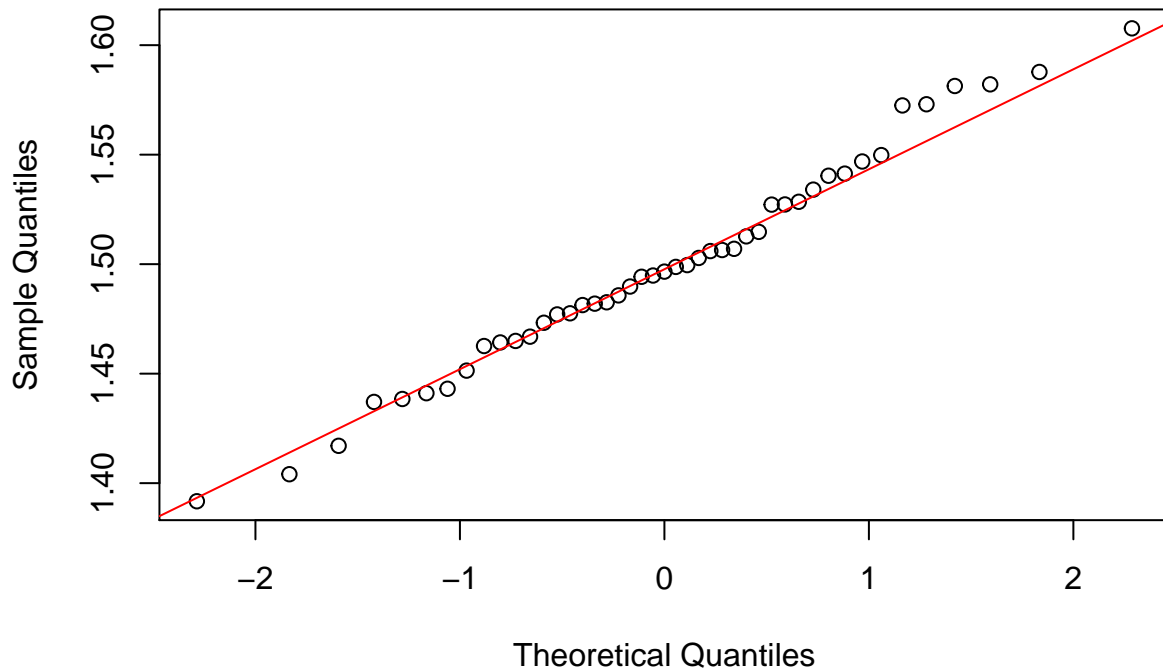
1. Non-compliance, whether some girls from the treatment group were not treated, whether some girls from the control group received treatment

```
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
```

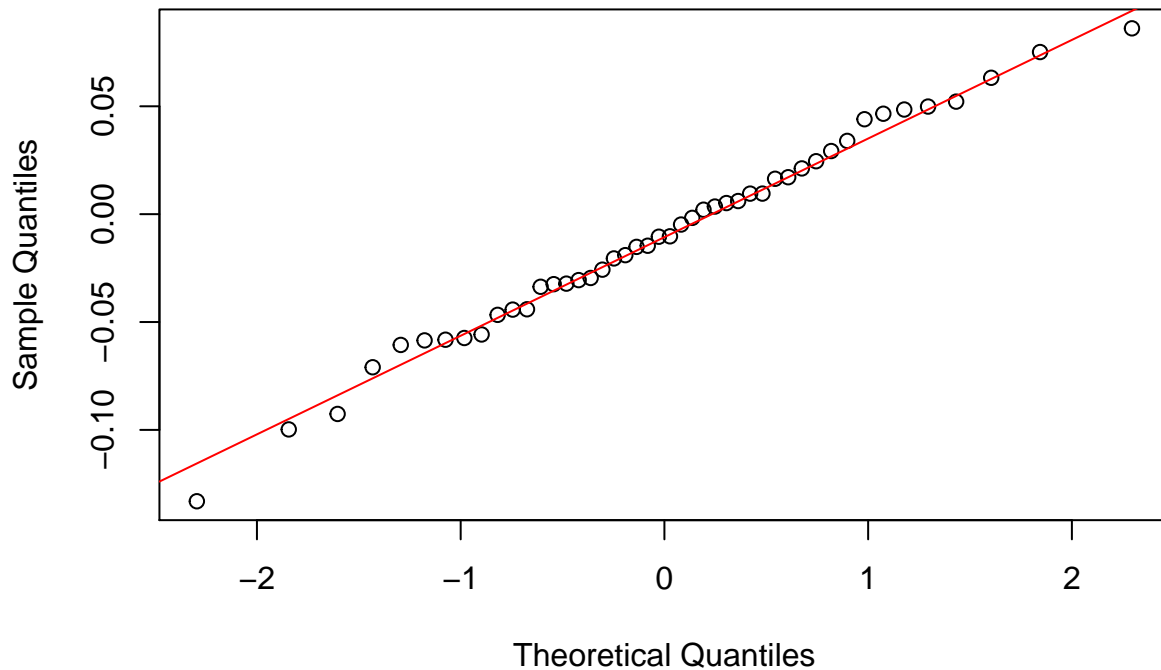
shapiro test suggests it's normally distributed; ttest can be interpreted as there's no non-compliance

```
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.

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

```
water_df <- water_df %>%
  filter(drop_out == 0) %>%
  mutate(treatment = relevel(factor(Di), ref = "0"))
```

```
# Estimating differences between groups:
m_list <- list()
m_list[[1]] <- felm(age ~ treatment | 0 | 0 | id, data = water_df)
m_list[[2]] <- felm(height ~ treatment | 0 | 0 | id, data = water_df)
m_list[[3]] <- felm(caloric_intake ~ treatment | 0 | 0 | id, data = water_df)
```

```
# Creating Table:
stargazer(m_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",
no.space = TRUE)
```

```
##
## Balance by Pre-Treatment Covariates
## =====
##                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
```

```
m_list2 <- list()
m_list2[[1]] <- felm(pre_sleep ~ treatment | 0 | 0 | id, data = water_df)
m_list2[[2]] <- felm(pre_exercise ~ treatment | 0 | 0 | id, data = water_df)
m_list2[[3]] <- felm(pre_water ~ treatment | 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    8.33        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
```

```
m_list[[7]] <- felm(pre_weight ~ treatment | 0 | 0 | id, data = water_df)
m_list[[8]] <- felm(pre_composition ~ treatment | 0 | 0 | id, data = water_df)
```

```
““
```

```
<<<<<< HEAD
```

```
===== >>>>>> origin/master
```