W241 Final Project Google Forms Analysis

Anni Yao

2025-04-15

0.1 Libraries and Database

Data was collected from Friends and Coworkers (In-Network) on Google Forms

```
library(tidyverse)
library(scales)
library(effsize)
library(AER)
```

```
# File from Google Sheets
file_path <- "C:/Users/anniy/Downloads/coffee_drinking_dataset_for_analysis.csv"
df <- read.csv(file_path, stringsAsFactors = FALSE)

df <- df %>% filter(rowSums(is.na(.)) < ncol(.))
df <- df[!is.na(df[, 7]), ]

glimpse(df)
```

```
## Rows: 44
## Columns: 13
## $ Timestamp
## $ Name
## $ Email
## $ Do.you.drink.at.least.one.cup.of.coffee.4.times.a.week.
## $ What.time.did.you.wake.up.this.morning.
## $ What.time.did.you.drink.your.coffee.....A..On.day.1.you.need.to.drink.coffee.20.minutes.after.wakin
## $ How.many.hours.has.it.been.since.you.woke.up...For.testing.accuracy..please.complete.this.survey.6
## $ How.awake.do.you.feel.right.now...scale.1.5.
## $ How.difficult.was.it.to.focus.on.tasks.today....scale.1.5.
## $ How.physically.tired.do.you.feel.right.now.....scale.1.5.
## $ How.much.coffee.did.you.drink.today.
## $ Did.you.notice.any.differences.in.your.alertness.today.compared.to.yesterday...Open.ended.
## $ Any.additional.comments.about.your.experience.with.drinking.coffee.at.this.time...Open.ended.
```

URL to the Coffee Study https://tinyurl.com/coffeestudy2025

0.2 Data Cleaning:

```r
# Clean values from our excel sheet
colnames(df) <- make.names(colnames(df))

df <- df %>%
  rename(
    wake_time = What.time.did.you.wake.up.this.morning.,
    coffee_time = What.time.did.you.drink.your.coffee.....A..On.day.1.you.need.to.drink.coffee.20.minute
    hours_since_wake = How.many.hours.has.it.been.since.you.woke.up...For.testing.accuracy..please.compl
    alertness = How.awake.do.you.feel.right.now...scale.1.5.,
    focus_difficulty = How.difficult.was.it.to.focus.on.tasks.today....scale.1.5.,
    physical_tiredness = How.physically.tired.do.you.feel.right.now.....scale.1.5.
  )

extract_number <- function(x) {
  as.numeric(gsub("[^1-5]", "", x))
}

df$alertness <- extract_number(df$alertness)
df$focus_difficulty <- extract_number(df$focus_difficulty)
df$physical_tiredness <- extract_number(df$physical_tiredness)

# Create Alert Index Score (range 3-15): higher = more alert
df$alert_index <- df$alertness + df$focus_difficulty + df$physical_tiredness


df$coffee_amount_raw <- df$How.much.coffee.did.you.drink.today.

df$coffee_cups <- df$coffee_amount_raw %>%
  tolower() %>%
  str_extract("\\d+(\\.\\d+)?") %>%
  as.numeric()

# Convert time data and compute delay
df$hours_since_wake <- as.numeric(df$hours_since_wake)
df$wake_time <- as.POSIXct(df$wake_time, format = "%I:%M:%S %p")
df$coffee_time <- as.POSIXct(df$coffee_time, format = "%I:%M:%S %p")
df$coffee_delay_mins <- as.numeric(difftime(df$coffee_time, df$wake_time, units = "mins"))
```

We created an **Alert Index Score** by summing three 1–5 scale variables: 1) How awake participants felt, 2) How focused they were, and 3) How physically tired they felt (reversed: 5 = not tired at all).

The combined Alert Index ranges from 3 to 15 and provides a broader picture of perceived mental alertness.

0.3 Identify Compliant Participants

```r
# Define compliance flags
df$compliant_survey_time <- ifelse(df$hours_since_wake >= 5 & df$hours_since_wake <= 7, 1, 0)
df$compliant_coffee_time <- ifelse(
  df$coffee_delay_mins <= 30 | (df$coffee_delay_mins >= 90 & df$coffee_delay_mins <= 150),
  1, 0
)
```

```r
# Define treatment vs control group
df$treatment_group <- ifelse(df$coffee_delay_mins >= 90, 1, 0)
df$treatment_label <- ifelse(df$treatment_group == 1, "Treatment", "Control")

# Creating a new df to filter for compliant participants only
df_compliant <- df %>%
  filter(compliant_survey_time == 1 & compliant_coffee_time == 1)
```

Note: We defined compliance with the survey timing as completing the survey **between 5 and 7 hours** after waking, rather than exactly at the 6-hour mark. This allowed us to include participants who followed the instructions closely, even if not perfectly. We believe this +- 1 hour window may help us maintain data quality and ensure a bigger sample size for our analysis.

0.4 Compliance Summary

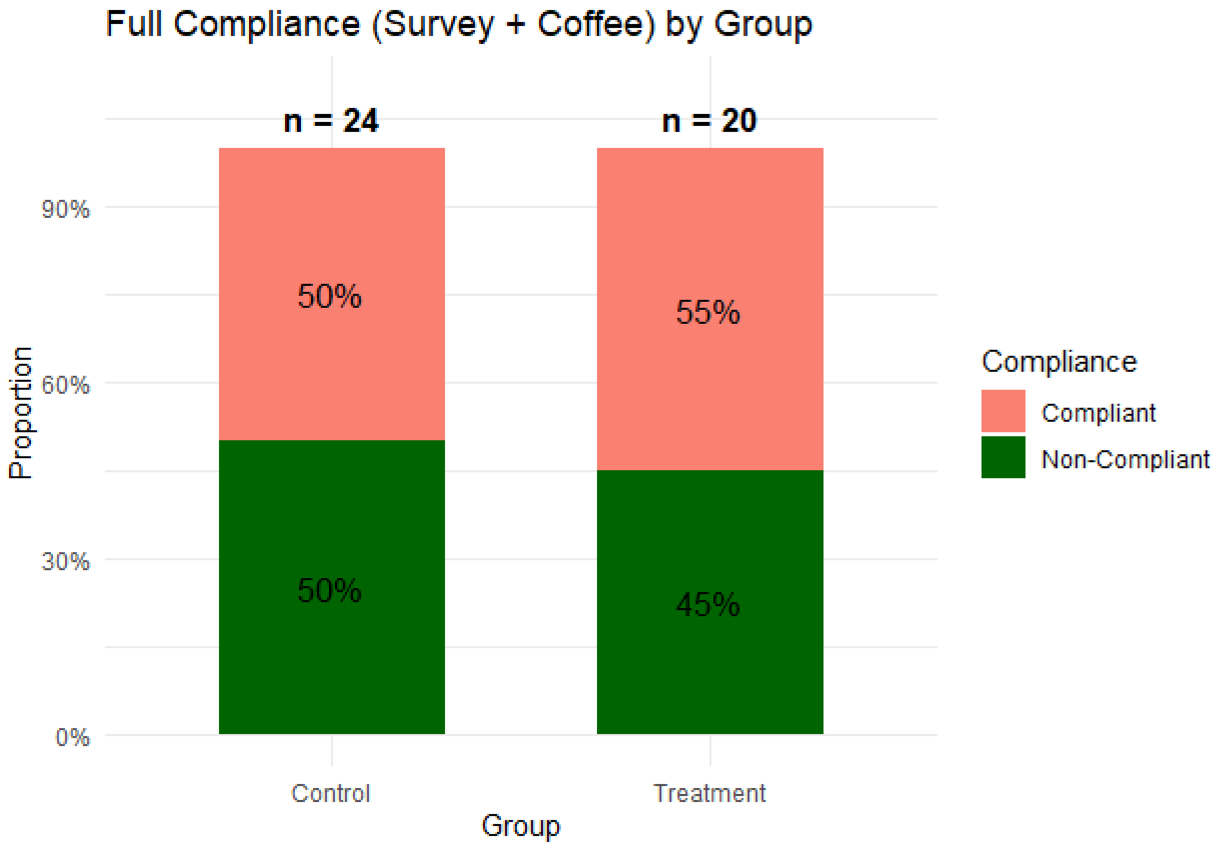
```r
df %>%
  summarise(
    total_responses = n(),
    compliant_survey = mean(compliant_survey_time, na.rm = TRUE),
    compliant_coffee = mean(compliant_coffee_time, na.rm = TRUE)
  )
```

```
##   total_responses compliant_survey compliant_coffee
## 1              44        0.5909091        0.8863636
```

```r
total_counts <- df %>%
  filter(!is.na(treatment_group)) %>%
  group_by(treatment_label) %>%
  summarise(total = n())

# Count compliant participants per group
compliant_counts <- df %>%
  filter(compliant_survey_time == 1 & compliant_coffee_time == 1) %>%
  group_by(treatment_label) %>%
  summarise(compliant = n())

# Merge the two summaries
group_summary <- left_join(total_counts, compliant_counts, by = "treatment_label") %>%
  mutate(compliant = replace_na(compliant, 0),
         compliance_rate = paste0(compliant, "/", total, " (", round(100 * compliant / total), "%)"))

print(group_summary)
```

```
## # A tibble: 2 x 4
##   treatment_label total compliant compliance_rate
##   <chr>           <int>     <int> <chr>
## 1 Control            24        12 12/24 (50%)
## 2 Treatment          20        11 11/20 (55%)
```

Our compliance rates were **50% in the control group** and **55% in the treatment group**. Only about half of participants followed both the coffee timing and survey timing instructions.



How can we improve this?

Our group discussed a few ways after the experiment to improve future compliance rates.

For any future experiments with human participants, we can:

1. Send Automated text/email reminders right before their target coffee time or survey time
2. Include a Mobile survey link with push notifications (although I'm not sure if google forms would have that feature)
3. Add Incentives for full compliance, like a gift card raffle or bonus points. Anni has offered a $25 starbucks giftcard to only one random participant.
4. We can use a third party App-based logging tool with built-in timers to guide them through Day A and Day B. We would need to search other survey tools to identify the right one

0.5 Comparative Statistics

```
t.test(alert_index ~ treatment_group, data = df_compliant)
```

```
##
##  Welch Two Sample t-test
##
## data:  alert_index by treatment_group
```

```
## t = 0.39335, df = 19.479, p-value = 0.6983
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -1.796766  2.630100
## sample estimates:
## mean in group 0 mean in group 1
##        12.41667        12.00000
```

We ran a Welch's t-test and found no significant difference in alertness between the **control group (mean = 12.42)** and the **treatment group (mean = 12.00)**, with a p-value of 0.70. The 95% confidence interval ranged from -1.80 to +2.63, meaning the true effect could go either way.

Overall: Not a clear impact based on current sample size

```r
library(effsize)
cohen.d(alert_index ~ treatment_group, data = df_compliant)
```

```
##
## Cohen's d
##
## d estimate: 0.1656184 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.7039455  1.0351823
```

Cohen's d was **0.17**, which is considered a **negligible effect size**. The 95% confidence interval ranged from -0.70 to +1.04 shows us that the difference in alertness between groups was very small overall with uncertainty.

0.6 Regression Analysis

```r
model <- lm(alert_index ~ treatment_group + hours_since_wake + coffee_cups, data = df_compliant)
summary(model)
```

```
##
## Call:
## lm(formula = alert_index ~ treatment_group + hours_since_wake +
##     coffee_cups, data = df_compliant)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.9094 -0.9738  0.3995  2.0906  3.0906
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.65956    8.57537   1.360    0.192
## treatment_group  -0.38164    1.22167  -0.312    0.759
## hours_since_wake  0.06441    1.35422   0.048    0.963
## coffee_cups       0.24507    0.90737   0.270    0.790
##
## Residual standard error: 2.59 on 17 degrees of freedom
```

```
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.01568,    Adjusted R-squared:  -0.158
## F-statistic: 0.09026 on 3 and 17 DF,  p-value: 0.9644
```

The linear regression model shows that participants who delayed their coffee had **-0.38 lower alertness scores** on average compared to the control group. Although, this effect was still only **marginally significant (p = 0.759)**, with a **standard error of 1.22**. Because the standard error is much larger than the estimate itself, it's likely that the true effect could be anywhere from much more negative to even positive.

- Adding how much cups of coffee participants drank that day to the model didn't improve predictive power significantly. This suggests that timing of caffeine may be more important than the quantity consumed
- While we observed a possible trend that delaying caffeine may improve alertness, our sample size of compliant participants using the google sheets was small. This limited our ability to detect a statistically significant effect with our models.
- Based on our earlier power calculation, we would definitely **need more participants** (especially compliant ones) to reliably detect an effect of this size. With more data, the standard errors would likely shrink, making it easier to confirm whether this trend holds up.

```r
# ITT: Use everyone, grouped by assigned treatment
model_itt <- lm(alert_index ~ treatment_group, data = df)
summary(model_itt)
```

```
##
## Call:
## lm(formula = alert_index ~ treatment_group, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7000 -0.8333  0.1667  2.1667  3.3000
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      11.8333     0.4772  24.798   <2e-16 ***
## treatment_group  -0.1333     0.7078  -0.188    0.851
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.338 on 42 degrees of freedom
## Multiple R-squared:  0.0008442,  Adjusted R-squared:  -0.02295
## F-statistic: 0.03549 on 1 and 42 DF,  p-value: 0.8515
```

```r
df$received_treatment <- ifelse(df$coffee_delay_mins >= 90, 1, 0)
cace_model <- ivreg(alert_index ~ received_treatment | treatment_group, data = df)
summary(cace_model)
```

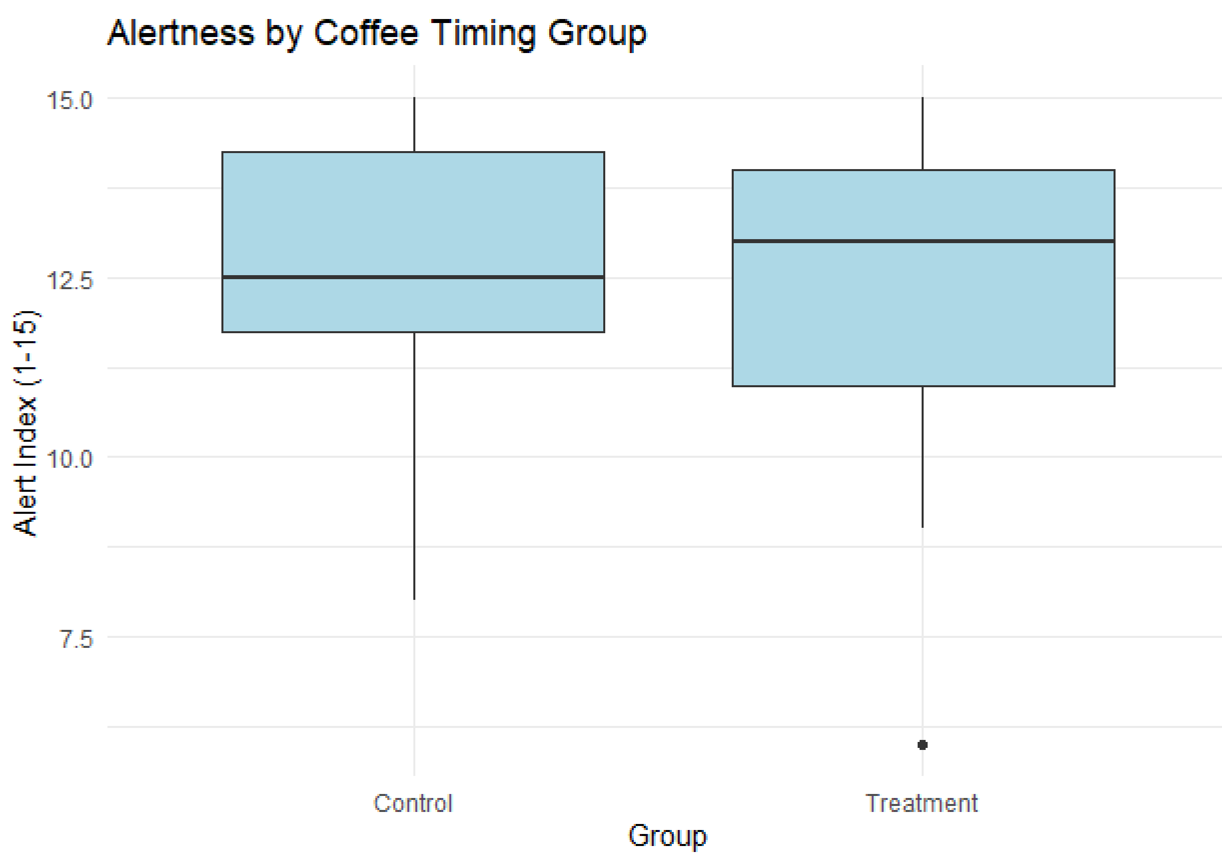
```
##
## Call:
## ivreg(formula = alert_index ~ received_treatment | treatment_group,
##     data = df)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7000 -0.8333  0.1667  2.1667  3.3000
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         11.8333     0.4772  24.798   <2e-16 ***
## received_treatment  -0.1333     0.7078  -0.188    0.851
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.338 on 42 degrees of freedom
## Multiple R-Squared: 0.0008442,   Adjusted R-squared: -0.02295
## Wald test: 0.03549 on 1 and 42 DF,  p-value: 0.8515
```

We tested whether assigning participants to delay their coffee (vs. drinking it right away) led to higher alertness, using both Intention-to-Treat (ITT) and Complier Average Causal Effect (CACE) approaches.

- **ITT result:** Assigning someone to delay their coffee led to a –**0.13** point decrease in their Alert Index Score on average, but this effect was **not statistically significant** (p = 0.85).
- **CACE result:** Among participants who actually followed the assigned delay, the estimated effect was also –**0.13**, and again **not statistically significant** (p = 0.85).

0.7 Visual Box Plot of Treatment and Control Groups



Overall: From our Google sheets results, while we didn't find strong evidence that delaying caffeine boosts alertness, this study helped us design and test a controlled experiment from end to end. With more participants, better tools (ex. Having mobile reminders & facilitators), incentives, and longer tracking, we believe this experiment could be refined and replicated for clearer insights.