W241 Final Project Google Forms Analysis

Anni Yao

2025-04-15

0.1 W241 Final Project Google Forms Analysis

```
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.waki
## $ 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.minut
    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$coffee_amount_raw <- df$How.much.coffee.did.you.drink.today.

# Clean and convert it to numeric
df$coffee_cups <- df$coffee_amount_raw %>%
  tolower() %>%
  str_extract("\\d+(\\.\\d+)?") %>%   # extract number like "1" or "1.5"
  as.numeric()

df$alertness <- extract_number(df$alertness)
df$focus_difficulty <- extract_number(df$focus_difficulty)
df$physical_tiredness <- extract_number(df$physical_tiredness)
df$hours_since_wake <- as.numeric(df$hours_since_wake)

# Convert time to datetime and calculate coffee delay
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"))
```

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
)

# 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

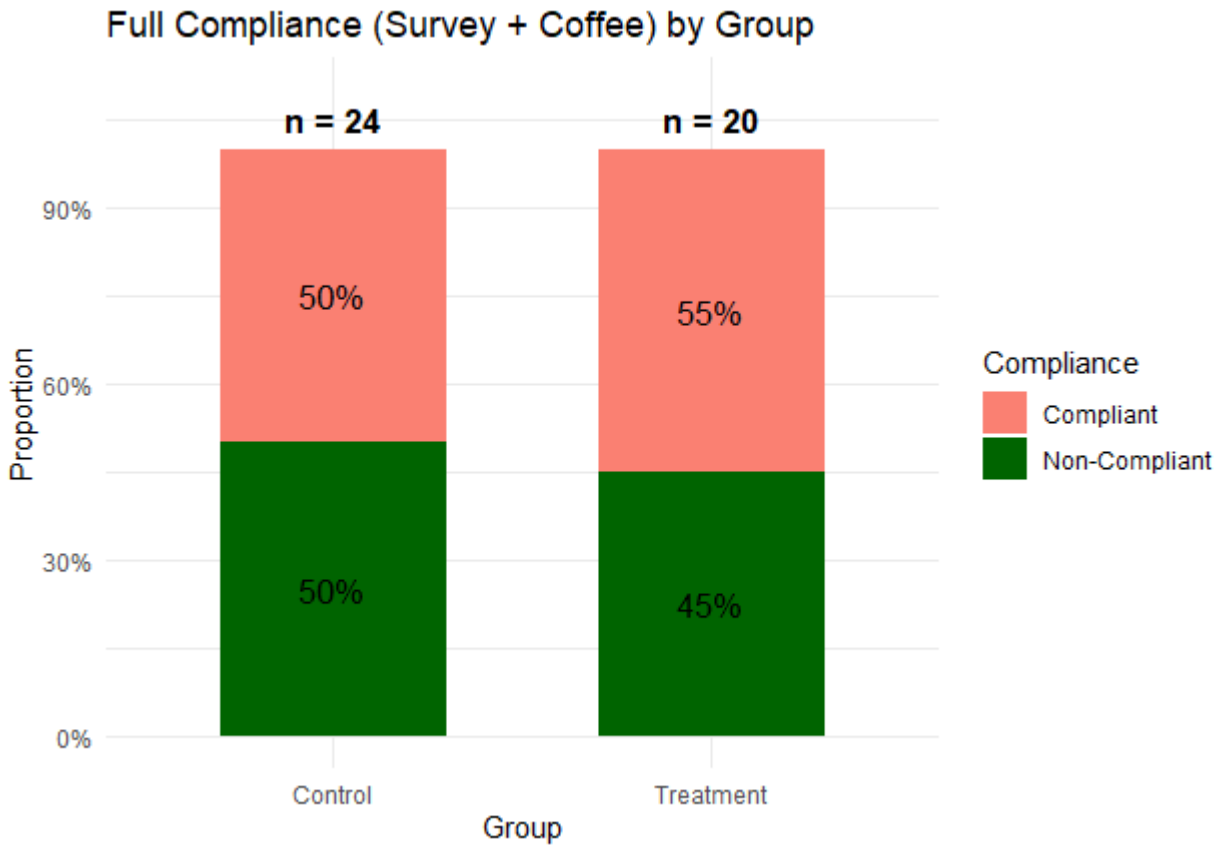
```
df %>%
  summarise(
    total_responses = n(),
    compliant_survey = mean(compliant_survey_time, na.rm = TRUE),
    compliant_coffee = mean(compliant_coffee_time, na.rm = TRUE)
  )
```

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

```
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(alertness ~ treatment_group, data = df_compliant)
```

```
##
##  Welch Two Sample t-test
##
## data:  alertness by treatment_group
## t = -0.28814, df = 17.922, p-value = 0.7765
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.8796327  0.6675115
## sample estimates:
## mean in group 0 mean in group 1
##        4.166667        4.272727
```

We ran a Welch's t-test and found no significant difference in alertness between the control group **(mean = 4.17)** and the treatment group (mean = 4.27), with a p-value of 0.78. The 95% confidence interval ranged from -0.88 to +0.67, meaning the true effect could go either way.

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

```
##
## Cohen's d
##
## d estimate: -0.1220871 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.9909737  0.7467995
```

Cohen's d was **-0.12**, which is considered a **negligible effect size**. This shows us that the difference in alertness between groups was very small overall.

0.6 Regression Analysis

```
model <- lm(alertness ~ treatment_group + focus_difficulty + physical_tiredness + hours_since_wake + co
summary(model)
```

```
##
## Call:
## lm(formula = alertness ~ treatment_group + focus_difficulty +
##     physical_tiredness + hours_since_wake + coffee_cups, data = df_compliant)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54051 -0.27362  0.02837  0.22219  0.67300
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.3326     1.3857  -0.240  0.81355
## treatment_group      0.3554     0.2024   1.756  0.09951 .
## focus_difficulty     0.3756     0.1394   2.695  0.01661 *
## physical_tiredness   0.5710     0.1717   3.325  0.00462 **
## hours_since_wake     0.0649     0.2081   0.312  0.75939
## coffee_cups          0.1284     0.1392   0.922  0.37095
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3958 on 15 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.8299, Adjusted R-squared:  0.7731
## F-statistic: 14.63 on 5 and 15 DF,  p-value: 2.595e-05
```

The linear regression model shows that participants who delayed their coffee had **0.36 higher alertness scores** on average compared to the control group, although this effect was still only **marginally significant ($p = 0.10$)**, with a **standard error of 0.20**.

- focus_difficulty and physical_tiredness goes into the Alert index score and should have a correlation.
- 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.

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

```
##
## Call:
## lm(formula = alertness ~ treatment_group, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.91667 -0.22917  0.08333  1.00000  1.08333
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.91667    0.19373   20.22   <2e-16 ***
## treatment_group  0.08333    0.28735    0.29    0.773
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9491 on 42 degrees of freedom
## Multiple R-squared:  0.001998,   Adjusted R-squared:  -0.02176
## F-statistic: 0.0841 on 1 and 42 DF,  p-value: 0.7732
```

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

```
##
## Call:
## ivreg(formula = alertness ~ received_treatment | treatment_group,
##     data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.91667 -0.22917  0.08333  1.00000  1.08333
```

```
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.91667    0.19373   20.22   <2e-16 ***
## received_treatment   0.08333    0.28735    0.29    0.773
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9491 on 42 degrees of freedom
## Multiple R-Squared: 0.001998,    Adjusted R-squared: -0.02176
## Wald test: 0.0841 on 1 and 42 DF,  p-value: 0.7732
```

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 coffee led to a +0.08 point increase in alertness on average, but this difference was **not statistically significant** ($p = 0.77$).
- **CACE result:** Among those who actually followed the assigned coffee delay (in compliance), the estimated effect was also +0.08, and was also **not statistically significant** ($p = 0.77$).