Power calculations

Daphne

2025-09-16

Preliminary set up

```r
library(tidyverse)
library(pwr)
load("/Users/drutna/Library/CloudStorage/OneDrive-UniversitätZürichUZH/UZH Carbon Neutrality/4_Data/7_LI
pilot_bl_co2 <- tCO2_Baseline_Costs
pilot_el_co2 <- tCO2_Program_Costs
```

Power calculations for DID analysis

Calculating the s.d. and correlation from pilot data

We use pilot data to estimate s.d. at bl and el, the correlation between bl and el, and the s.d. of change scores.

Clean and merge the pilot data:

```r
# Clean bl data

# Convert waste to tons
pilot_bl_co2$Waste1 <- pilot_bl_co2$Waste1/1000

# If waste exceeds 10 tons, set to NA (data entry error)
pilot_bl_co2 <- pilot_bl_co2 %>%
    mutate(Waste1 = ifelse(Waste1 > 10, NA, Waste1))

# If fume cupboards exceeds 10 tons, divide by 1000 (data entry error)
pilot_bl_co2 <- pilot_bl_co2 %>%
    mutate(`Fume Cupboards1` =
    ifelse(`Fume Cupboards1` > 10, `Fume Cupboards1`/1000, `Fume Cupboards1`))


# Create id and total_co2
pilot_bl_co2 <- pilot_bl_co2 %>%
  mutate(id = row_number()) %>%
  filter(if_any(-id, ~ !is.na(.))) %>%
  mutate(
    total_co2 = rowSums(across(-id), na.rm = TRUE)
  )
# Clean el data
```

```r
# Convert waste to tons
pilot_el_co2$Waste2 <- pilot_el_co2$Waste2/1000

# If waste exceeds 10 tons, set to NA (seems like an error in data entry)
pilot_el_co2 <- pilot_el_co2 %>%
    mutate(Waste2 = ifelse(Waste2 > 10, NA, Waste2))

# If fume cupboards exceeds 10 tons, divide by 1000 (data entry error)
pilot_el_co2 <- pilot_el_co2 %>%
    mutate(`Fume Cupboards2` =
    ifelse(`Fume Cupboards2` > 10, `Fume Cupboards2`/1000, `Fume Cupboards2`))

# Create id and total_co2
pilot_el_co2 <- pilot_el_co2 %>%
  mutate(id = row_number()) %>%
  filter(if_any(-id, ~ !is.na(.))) %>%
  mutate(
    total_co2 = rowSums(across(-id), na.rm = TRUE)
  )

# Merge
merged_data <- full_join(
  pilot_bl_co2,
  pilot_el_co2,
  by = "id",
  suffix = c("_bl", "_el")
)
#Drop rows with NA in either total_co2_bl or total_co2_el
merged_data <- merged_data %>%
    filter(!is.na(total_co2_bl) & !is.na(total_co2_el))

head(merged_data)
```

```
## # A tibble: 6 x 21
##   Waste1 `Fume Cupboards1` `ULT Freezers1` Refrigerators1 Water1
##    <dbl>             <dbl>           <dbl>          <dbl>  <dbl>
## 1  5.96              1.21            0.42           0.53     NA
## 2 NA                 3.57            0.09          NA        NA
## 3  3.18              0.118           0.23           0.21     NA
## 4 NA                NA               0.22          NA        NA
## 5  0.945            NA               0.2            0.05     NA
## 6  1.05              0.62            0.07           0.07     NA
## # i 16 more variables: `Biosafety Cabinets1` <dbl>, `-20C Freezers1` <dbl>,
## #   IT1 <dbl>, `Open/other1` <dbl>, id <int>, total_co2_bl <dbl>, Waste2 <dbl>,
## #   `Fume Cupboards2` <dbl>, `ULT Freezers2` <dbl>, Refrigerators2 <dbl>,
## #   Water2 <dbl>, `Biosafety Cabinets2` <dbl>, `-20C Freezers2` <dbl>,
## #   IT2 <dbl>, `Open/other2` <dbl>, total_co2_el <dbl>
```

Calculate the s.d. at bl and el, the correlation, the sd of change scores, and the estimated impact:

```r
sd_bl <- sd(merged_data$total_co2_bl, na.rm = TRUE)
sd_el <- sd(merged_data$total_co2_el, na.rm = TRUE)
```

```r
rho <- cor(merged_data$total_co2_bl, merged_data$total_co2_el)
sigma_d <- sqrt(sd_bl^2 + sd_el^2 - 2 * rho * sd_bl * sd_el)
beta <- mean(merged_data$total_co2_bl, na.rm = TRUE) - mean(merged_data$total_co2_el, na.rm = TRUE)
print(paste("SD at bl:", round(sd_bl, 2)))
```

```
## [1] "SD at bl: 3.52"
```

```r
print(paste("SD at el:", round(sd_el, 2)))
```

```
## [1] "SD at el: 2.47"
```

```r
print(paste("Correlation between bl and el:", round(rho, 2)))
```

```
## [1] "Correlation between bl and el: 0.57"
```

```r
print(paste("SD of change scores:", round(sigma_d, 2)))
```

```
## [1] "SD of change scores: 2.92"
```

```r
print(paste("Estimated impact (mean change):", round(beta, 2)))
```

```
## [1] "Estimated impact (mean change): 1.5"
```

Calculating power, MDE and sample size

Calculate the power (using pilot data) and MDE:

```r
# Parameters
alpha <- 0.05 # significance level
n <- 100 # number of lab groups per arm
power <- 0.8 # desired power

# Calculate power
power_res <- pwr.t.test(
    n = n,
    d = beta / sigma_d,
    sig.level = alpha,
    type = "two.sample",
    alternative = "two.sided"
)
power <- as.numeric(power_res[4])

print(paste("Estimated power based on UZH pilot data:",
      round(power,3)))
```

```
## [1] "Estimated power based on UZH pilot data: 0.95"
```

```r
# Calculate MDE for 80% power
mde_res <- pwr.t.test(
    n = n,
    power = 0.8,
    sig.level = alpha,
    type = "two.sample",
    alternative = "two.sided"
)
mde <- as.numeric(mde_res[2])

print(paste("Minimum detectable effect size (Cohen's d):", round(mde, 3)))
```

```
## [1] "Minimum detectable effect size (Cohen's d): 0.398"
```

```r
# Calculate sample size for 80% power with estimated effect size
n_res <- pwr.t.test(
    d = beta / sigma_d,
    power = 0.8,
    sig.level = alpha,
    type = "two.sample",
    alternative = "two.sided"
)
n <- as.numeric(n_res[1])

print(paste("Number of lab groups per arm:", ceiling(n)))
```

```
## [1] "Number of lab groups per arm: 61"
```