

# Experimental Power Analysis

## Does placing signs encouraging stairs over elevators increase stair use?

### ##Experiment Overview

In this experiment, we aim to examine whether the placement of a sign stating 'Use the stairs, improve your health' influences the choice between using stairs versus elevators. This document provides an overview of our power analysis to determine the necessary sample sizes to test our hypothesis.

## Testing Procedure

**Metrics:** Our metric of interest is the percentage of people choosing to use the elevator versus the stairs in both the control and treatment groups.

**Treatment:** The treatment group is exposed to a sign with the message "Use the stairs, improve your health," while the control group is not exposed to any sign.

**Statistical Power Approach:** A series of t-tests are conducted to compare elevator use between the control and treatment groups (Null-hypothesis: Treatment effect is zero). To ensure sufficient power, we simulate different sample sizes and calculate the proportion of experiments in which we would correctly reject the null hypothesis. The aim is to determine the minimum sample size required to detect significant differences at varying levels of treatment effects.

**Assumptions:** 1. Baseline Behavior: In the control group, 80% of participants use the elevator, with a standard deviation of 15%. 2. Treatment Effect: We assume the sign may reduce elevator use by 2% to 20%, depending on its effectiveness.

```
library(data.table)
```

```
## Warning: Paket 'data.table' wurde unter R Version 4.1.2 erstellt
```

```
library(here)
```

```
## Warning: Paket 'here' wurde unter R Version 4.1.3 erstellt
```

```
## here() starts at C:/Users/lenar/OneDrive/Dokumente/Statistik Master/Berkeley/Experiments/takethestairsyoulazies
```

```
library(ggplot2)
```

```
## Warning: Paket 'ggplot2' wurde unter R Version 4.1.2 erstellt
```

```
library(tidyr)
```

```
## Warning: Paket 'tidyr' wurde unter R Version 4.1.2 erstellt
```

```

set.seed(10)

#Create dummy data under the following assumptions
# control group has 80% of people taking elevator with a SD of 15%
# treatment effect is a 2-20% reduction in people taking the elevator'

d <- data.table(
  id = 1:1100
)
d[, D := sample(c('treatment', 'control'), size=.N, replace=T)]
d[D == "control", Y_control := rnorm(n=.N, mean=80, sd=15)]

#set up four possible treatment effect sizes [2, 8, 14, 20]
d[D == "treatment", Y2_treatment := rnorm(n=.N, mean=78, sd=15)]
d[D == "treatment", Y8_treatment := rnorm(n=.N, mean=72, sd=15)]
d[D == "treatment", Y14_treatment := rnorm(n=.N, mean=66, sd=15)]
d[D == "treatment", Y20_treatment := rnorm(n=.N, mean=60, sd=15)]

d[, Y2 := ifelse(D == "control", Y_control, Y2_treatment)]
d[, Y8 := ifelse(D == "control", Y_control, Y8_treatment)]
d[, Y14 := ifelse(D == "control", Y_control, Y14_treatment)]
d[, Y20 := ifelse(D == "control", Y_control, Y20_treatment)]

```

## Simulation of 10 Subjects

In case we could only observe 10 people – 5 in treatment, when we've been setting up the signs, and another 5 in control. this would be the power of our experiment across the different treatment effect sizes we think could be possible.

```

#initialize random sample of 10 people
d10 <- d[, .SD[sample(.N, 5)], keyby = D]
#set up place to store p values
p_values <- c()
#list out hypothetical treatment sizes to run t test on
treatments_to_test <- c("Y2", "Y8", "Y14", "Y20")

for (treatment_effect in treatments_to_test) {
  t_test_ten_people <- t.test(d10[D == 'control', ..treatment_effect],
                             d10[D == 'treatment', ..treatment_effect])
  p_values <- c(p_values, t_test_ten_people$p.value)
}

p_values

```

```
## [1] 0.1562463 0.9052187 0.6146585 0.1386144
```

[The p-value from our 10-person experiment with an effect size of 2% is: 0.1562463. The p-value from our 10-person experiment with an effect size of 8% is: 0.9052187. The p-value from our 10-person experiment with an effect size of 14% is: 0.6146585. The p-value from our 10-person experiment with an effect size of 20% is: 0.1386144.]

## With only ten subjects, what is our power?

We repeat this process – sampling 10 people from our dummy data and conducting the t-tests – one-thousand times to plot the distribution of p-values observed at each sample size.

```
#create a table to plot p-values per treatment effect
t_test_p_values <- data.table(
  id = 1:1000
)

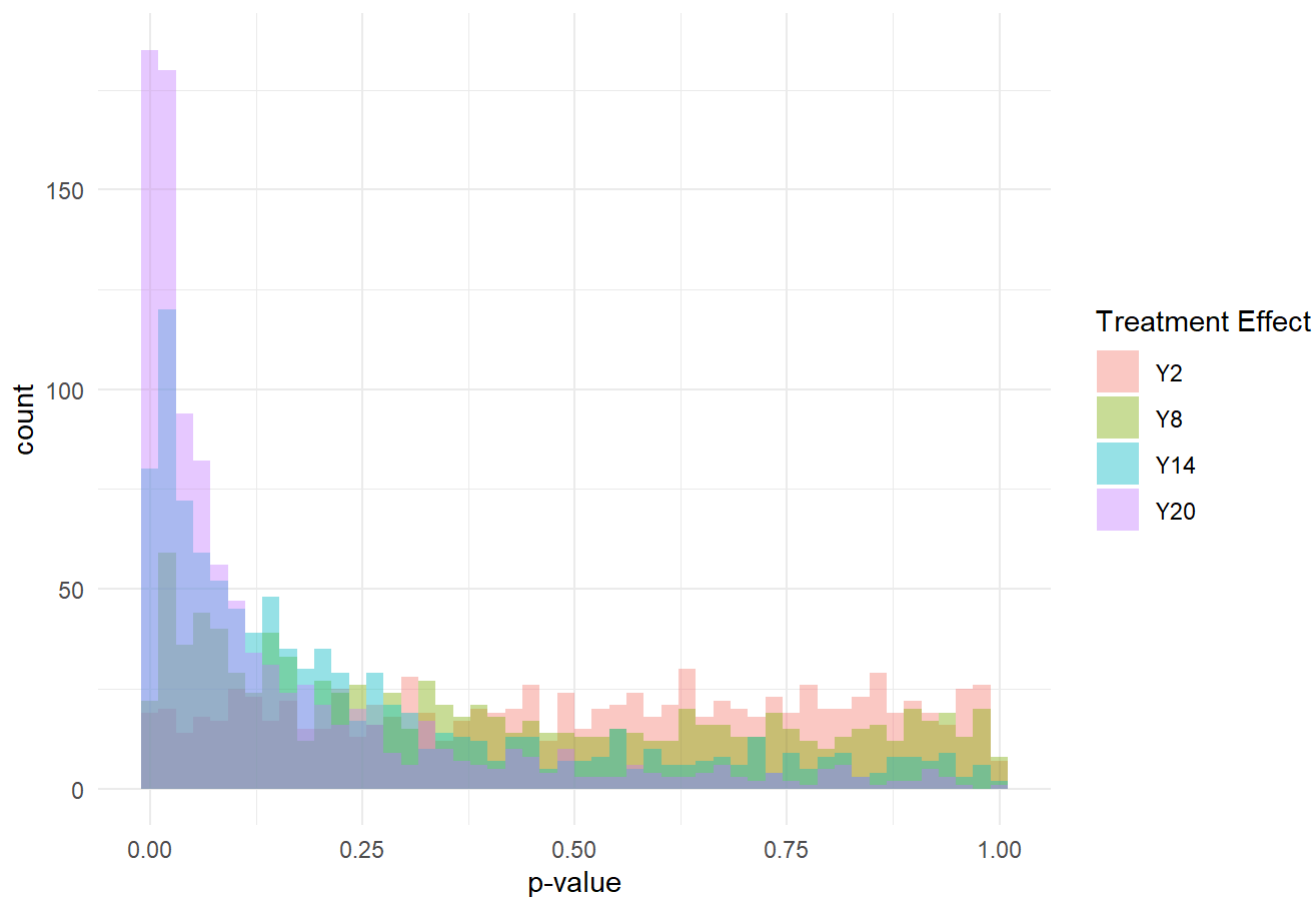
for (treatment_effect in treatments_to_test) {
  t_test_p_values_i <- c()
  for (i in 1:1000) {
    d10_i <- d[, .SD[sample(.N, 5)], keyby = D]

    t_test_ten_people_i <- t.test(d10_i[D == 'control', ..treatment_effect],
                                d10_i[D == 'treatment', ..treatment_effect])
    t_test_p_values_i <- c(t_test_p_values_i, t_test_ten_people_i$p.value)
  }
  t_test_p_values[, (treatment_effect) := t_test_p_values_i]
}
```

## Visual analysis

```
# Reshape the data from wide to long format
t_test_p_values_long <- melt(t_test_p_values, id.vars = "id",
                             variable.name = "treatment_effect",
                             value.name = "p_value")

# Plot the 4 series (columns) as histograms
ggplot(t_test_p_values_long, aes(x = p_value, fill = treatment_effect)) +
  geom_histogram(bins = 50, alpha = 0.4, position = "identity") +
  labs(
    title = 'p-value distribution for n = 10',
    x = 'p-value',
    fill = 'Treatment Effect'
  ) +
  theme_minimal()
```

p-value distribution for  $n = 10$ 

So we see, a sample of only 5 observations in each group would be not sufficient.

## Moar power!

Simulate the percentage of times we would reject the null across different effect sizes and different sample sizes.

```

#create a data table to record t-test rejects across varying sample sizes
experiment_simulations <- data.table(
  sample_size = integer(),
  t_test_rejects = numeric(),
  treatment_effect = character()
)

#make a vector to iterate through different numbers of people in the sample
people_in_sample <- c(seq(10,1000, by = 100)) #sequence 10-100 increment by 100

#three nested for loops this is embarrassing.

#loop over the different possible treatment effects
for (treatment_effect in treatments_to_test) {

  #loop over the possible number of people we could have in our sample
  for (i in 1:(length(people_in_sample))) {
    t_test_p_values_i <- rep(NA, 1000)

    #simulate 100 randomized experiments with each setting for n (sample size)
    for (j in 1:1000) {
      d_i_j <- d[, .SD[sample(.N, people_in_sample[i]/2)], keyby = D]
      t_test_i_people_j <- t.test(d_i_j[D == 'control', ..treatment_effect],
                                d_i_j[D == 'treatment', ..treatment_effect])
      t_test_p_values_i[j] <- t_test_i_people_j$p.value
    }
    t_test_rejects_i <- sum(t_test_p_values_i < 0.05) / length(t_test_p_values_i)
    new_row = data.table(sample_size = people_in_sample[i],
                        t_test_rejects = t_test_rejects_i,
                        treatment_effect = treatment_effect)
    experiment_simulations <- rbind(experiment_simulations, new_row)
  }
}

```

```

ggplot(experiment_simulations) +
  aes(x = sample_size, y = t_test_rejects, color = treatment_effect) +
  geom_line(linewidth = 1, alpha = 0.5) + # Creates the line
  geom_point(size = 1) + # Adds points at each sample size
  labs(
    title = "Proportion of T-Test Rejections by Sample Size",
    x = "Sample Size",
    y = "Proportion of Rejections (p < 0.05)"
  ) +
  theme_minimal()

```

```
## Warning: Ignoring unknown parameters: linewidth
```

