Power Analysis

12/11/2023

1 Exploring the Influence of Personality Profiles on User Engagement in Dating Apps

1.1 Project

This study investigates the user personalities displayed on dating app profiles and their subsequent engagement levels. By comparing the effects of creating a neutral personality profile with crafting a highly humorous and intriguing personality description, we aim to examine the impact of varying personality portrayals in dating app profiles (independent of accompanying photos) on user interactions and engagement.

1.2 Samples

Our study utilized samples drawn from five distinct account profiles, each created by individual team members. Each team member generated two account profiles featuring the same photo: one with a factual description (control) and another with a humorous description (treatment). The team members, on a daily basis, swiped right on 10 candidates using the app and recorded the number of matches as our treatment effect. Consequently, our daily samples consisted of 10 profiles in the control group and 10 profiles in the treatment group, summing up to 10 * 5 profiles for each category.

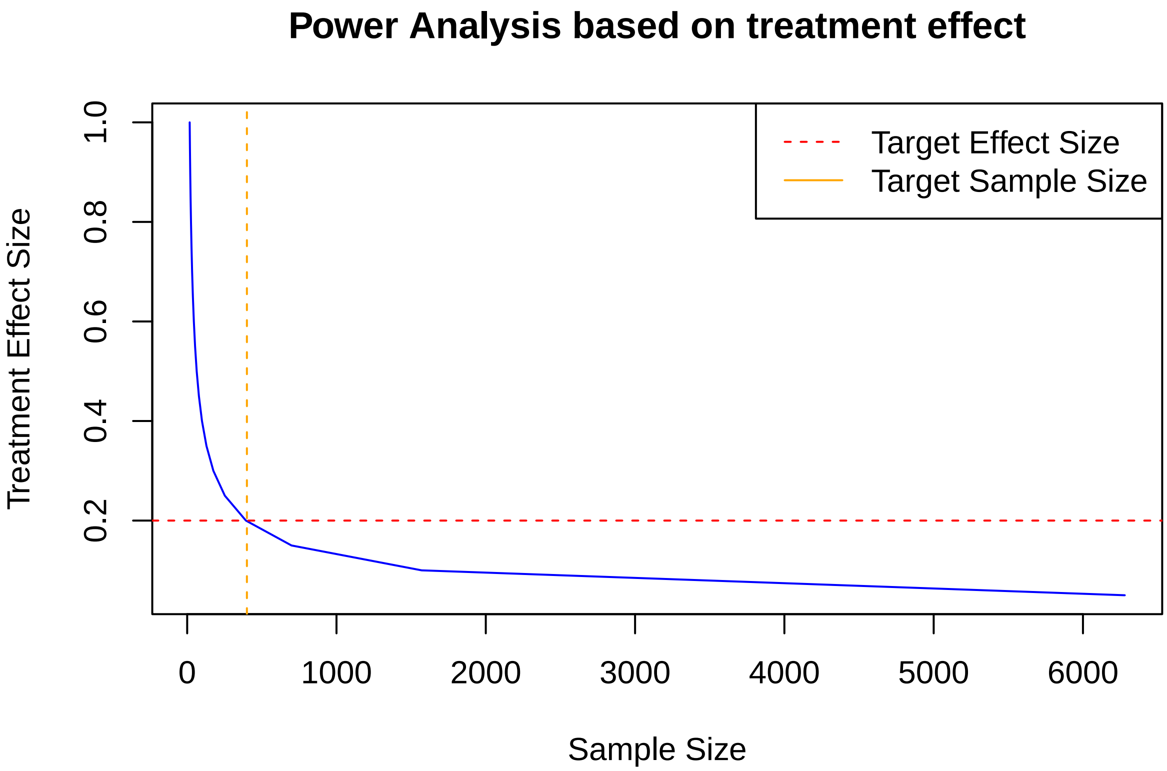
1.3 To simulate a power analysis based on treatment effect

Drawing from previous research published in (https://doi.org/10.1016/j.chb.2020.106667), it has been established that employing humor and offering complimentary messages in the context of online dating serves as a substantial predictor for men. This body of research suggests that incorporating humor into dating profiles of men is likely to result in a greater number of matches compared to profiles devoid of humor. In the upcoming experiment, we anticipate observing a positive impact from the treatment.

The provided code and plot are derived from simulating various treatment effect sizes, ranging from 5% to 100% with a 5% incremental increase, while aiming to achieve a consistent power level of 0.8. This illustrates how the necessary sample size changes with varying treatment effect sizes.

```r
# Set parameters
alpha <- 0.05           # Significance level
power <- 0.8            # Desired power
treatment_effect_sizes <- seq(0.05, 1, by = 0.05)  # Range of treatment effect sizes


# Initialize variables
target_sample_size <- NA
target_effect_size <- NA

# Function to calculate power for a given effect size and sample size
calculate_power <- function(effect_size, sample_size) {
  pwr.t.test(d = effect_size, n = sample_size, sig.level = alpha, type = "two.sample")$power
}
```

```r
# Iterate through treatment effect sizes and find the combination that achieves the desired power
for (effect_size in treatment_effect_sizes) {
  # Use a range of sample sizes for exploration
  for (sample_size in seq(10, 500, by = 10)) {
    current_power <- calculate_power(effect_size, sample_size)

    if (current_power >= power) {
      target_sample_size <- sample_size
      target_effect_size <- effect_size
      break
    }
  }

  if (!is.na(target_sample_size)) {
    break
  }
}

# Function to calculate sample size for a given treatment effect size
calculate_sample_size <- function(effect_size, alpha, power) {
  pwr.t.test(d = effect_size, sig.level = alpha, power = power, type = "two.sample")$n
}

# Calculate sample sizes for each treatment effect size
sample_sizes <- sapply(treatment_effect_sizes, calculate_sample_size, alpha = alpha, power = power)

# Plot the results
plot(sample_sizes, treatment_effect_sizes, type = "l", col = "blue",
     xlab = "Sample Size", ylab = "Treatment Effect Size",
     main = "Power Analysis based on treatment effect")

# Add a horizontal line for the desired power
abline(h = target_effect_size, col = "red", lty = 2)
abline(v = target_sample_size, col = "orange", lty = 2)

# Add legend
legend("topright", legend = c("Target Effect Size", "Target Sample Size"),
       col = c("red", "orange"), lty = c(2, 1))
```



1.3.1 Summary on power analysis based on treatment effect

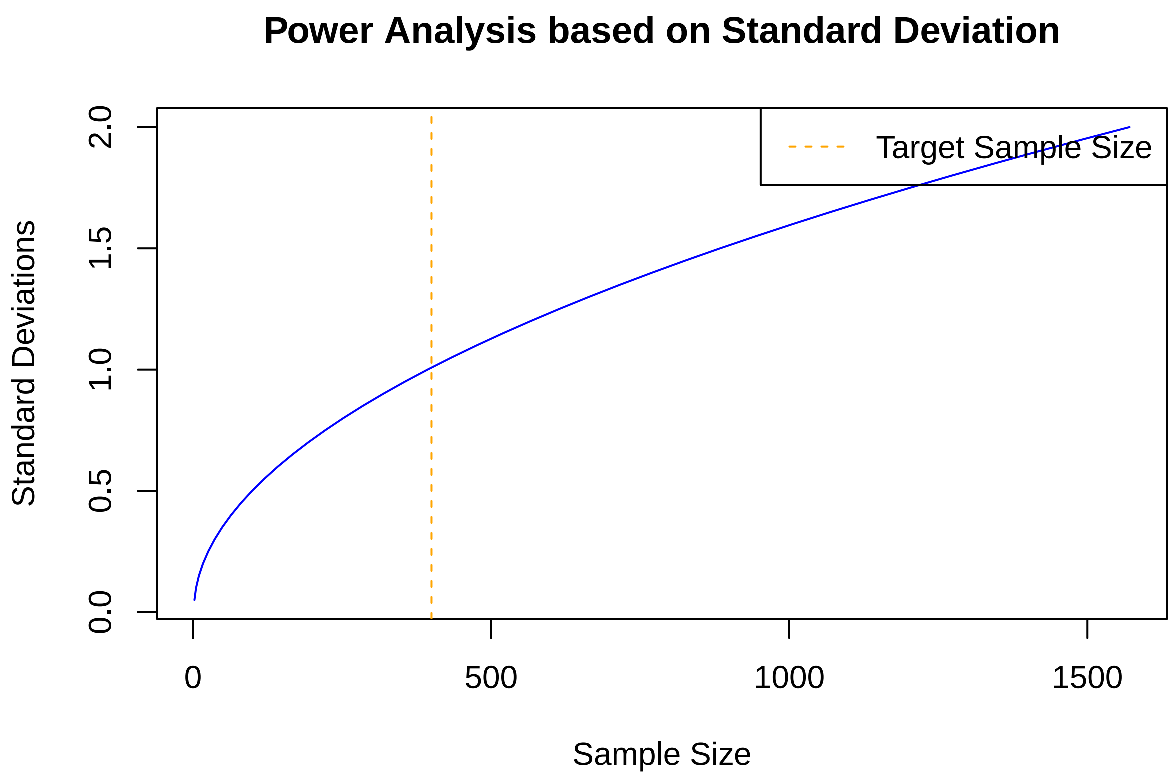
As demonstrated in the graph above, a smaller effect size typically requires a larger sample size to achieve the same level of statistical power. Conversely, a larger effect size can be detected with a smaller sample size. According to the analysis, in order to ensure a high probability (80%) of detecting a treatment effect of 20% (In this case, a treatment effect of 20% suggests that the study is designed to detect a 20% difference between groups, which could be a treatment group and a control group), a sample size of 400 is recommended for the statistical test being conducted.

1.4 To simulate a power analysis based on data dispersion

The presented code and plot result from simulating different levels of dispersion (standard deviation) ranging from 0.05 to 2, with a 5% incremental increase. The goal is to maintain a consistent power level of 0.8. In this simulation, we have assumed an Average Treatment Effect (ATE) of 20%, as suggested in a prior simulation. This demonstrates the increasing required sample size as the standard deviation dispersion becomes higher.

```
# Install and load the necessary library

# Set parameters
alpha <- 0.05           # Significance level
power <- 0.8            # Desired power
mean1 <- 0.5            # Mean of group 1(control)
mean2 <- 0.7           # Mean of group 2(humor)
std_devs <- seq(0.05, 2, by = 0.05)  # Range of standard deviations to explore

# Initialize variables
target_sample_size <- NA
target_std_dev <- NA

# Function to calculate power for a given standard deviation and sample size
```

```r
calculate_power <- function(std_dev, sample_size) {
  d <- (mean2 - mean1) / std_dev  # Cohen's d effect size
  pwr.t.test(d = d, n = sample_size, sig.level = alpha, type = "two.sample")$power
}

# Iterate through standard deviations and find the combination that achieves the desired power
for (std_dev in std_devs) {
  # Use a range of sample sizes for exploration
  for (sample_size in seq(10, 500, by = 10)) {
    current_power <- calculate_power(std_dev, sample_size)

    if (current_power >= power) {
      target_sample_size <- sample_size
      target_std_dev <- std_dev
      break
    }
  }

  if (!is.na(target_sample_size)) {
    break
  }
}
# Function to calculate sample size for a given standard deviation
calculate_sample_size <- function(std_dev, alpha, power) {
  d <- (mean2 - mean1) / std_dev  # Cohen's d effect size
  pwr.t.test(d = d, sig.level = alpha, power = power, type = "two.sample")$n
}

# Calculate sample sizes for each standard deviation
sample_sizes <- sapply(std_devs, calculate_sample_size, alpha = alpha, power = power)

# Plot the results
plot(sample_sizes, std_devs, type = "l", col = "blue",
     xlab = "Sample Size", ylab = "Standard Deviations",
     main = "Power Analysis based on Standard Deviation")

# Add a horizontal line for the desired power
abline(v = 400, col = "orange", lty = 2)

# Add legend
legend("topright", legend = c("Target Sample Size"),
       col = c("orange"), lty = c(2, 1))
```



1.4.1 Summary on power analysis based on data dispersion

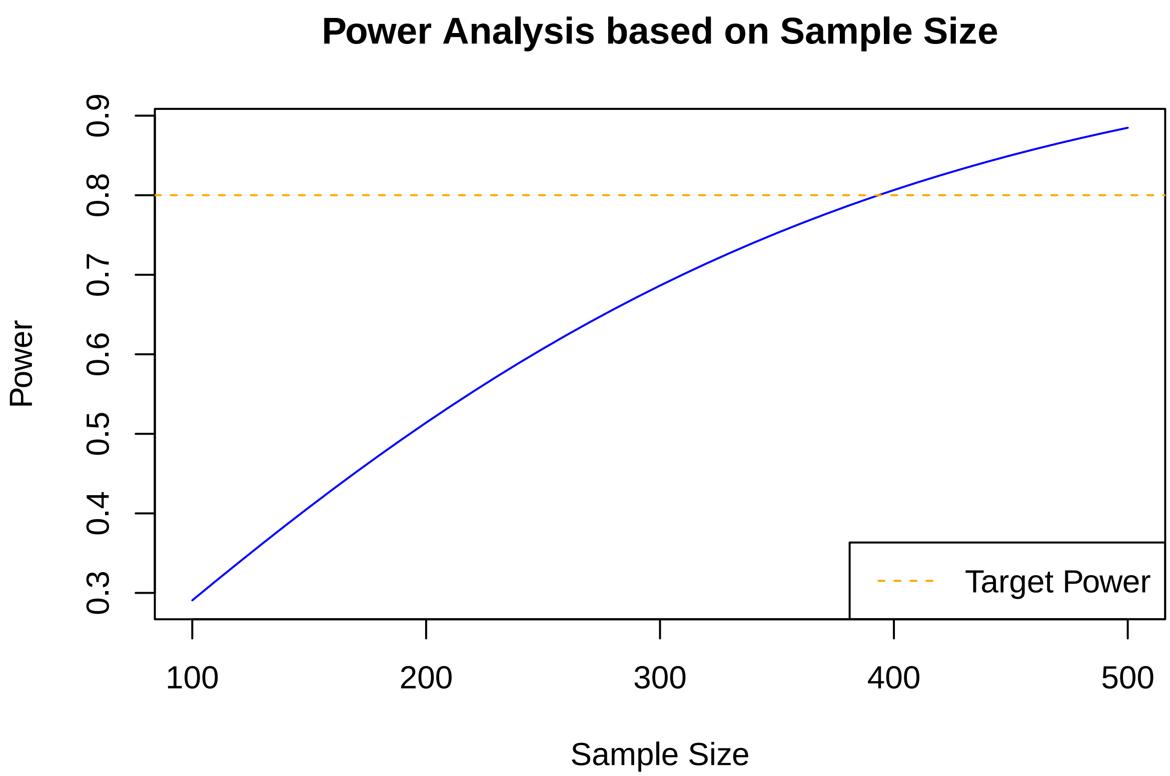
Standard deviation is a measure of the spread or dispersion of data. A larger standard deviation suggests more variability in the data. To increase the chances of detecting a true effect in the presence of higher variability, a larger sample size is needed. This is because a larger sample size provides more information and helps to better estimate the underlying population parameters. Therefore, as the standard deviation (dispersion) becomes higher, it becomes more challenging to detect smaller effects, and a larger sample size is required to achieve the desired level of statistical power.

1.5 To simulate a power analysis based on sample size

The presented code and plot result from simulating different sample sizes ranging from 10 to 500, with 10 unit incremental increase. The goal is to reach a power level of 0.8. In this simulation, we have assumed an Average Treatment Effect (ATE) of 20% and a Standard Deviation of 1.0, as suggested in a prior simulation. This demonstrates the increasing required sample size to reach a power level of 0.8 or higher.

```
# Install and load the necessary library

# Set parameters
alpha <- 0.05            # Significance level
mean1 <- 0.5             # Mean of group 1(control)
mean2 <- 0.7            # Mean of group 2(humor)
std_dev <- 1.0         #standard deviation
sample_size <- seq(100, 500, by = 10) # Range of sample sizes to explore

# Initialize variables
power <- NA

# Function to calculate power for a given standard deviation and sample size
calculate_power <- function(std_dev, sample_size, alpha) {
```

```r
  d <- (mean2 - mean1) / std_dev  # Cohen's d effect size
  pwr.t.test(d = d, n = sample_size, sig.level = alpha, type = "two.sample")$power
}

# For loop to calculate power for a given sample size

for(i in 1:length(sample_size)) {
  power[i] <- calculate_power(std_dev, sample_size[i],alpha)
}


# Plot the results
plot(sample_size, power, type = "l", col = "blue",
     xlab = "Sample Size", ylab = "Power",
     main = "Power Analysis based on Sample Size")

# Add a horizontal line for the desired power
abline(h = .80, col = "orange", lty = 2)

# Add legend
legend("bottomright", legend = c("Target Power"),
       col = c("orange"), lty = c(2, 1))
```



1.5.1 Summary on power analysis based on sample size

Sample size is a measure of the quantity of experiment data available. A larger sample size suggests higher power when standard deviation and average treatment effect are constant. To increase the chances of detecting a true effect when standard deviation is 1 and average treatment effect is 20%, a larger sample size is needed. This is because a larger sample size provides more information and helps to better estimate the underlying population parameters. Therefore, if the sample size is too small, it becomes more challenging to detect a true effect. A 400 sample size is required to achieve the desired level of statistical power (0.80) when standard deviation and average treatment effect are held constant.