Power Analysis

Generate Data

Upon analysis of multiple related studies on cigarette litter, the team chose three treatments that had an impact on the number of cigarette butts (CB) on the ground. The first case is whether the environment is dirty or clean. The team will stage CBs on the ground as treatment. The next case is whether there are accessible bins available to see if there is a reduction in litter. The final case is whether there are visible signs directing people to the available bins.

For the first treatment case, the team referenced mean and standard deviation on a study that was conducted that analyzed the effects of clean environments on a college campus [1]. In this study, the team staged the areas of interest with CBs for the treatment condition. In order to apply this to our own experiment, we had to increase the number of CB findings since their recordings happened every 2 hours. Realistically, our team will be collecting data daily so we scaled the findings by 3 to 5 times more CB to align with people's typical work / recreational schedule.

As for the second treatment, the team would like to use the bin as a treatment because Smith and Novotny reported that 2/3 of smokers are not inclined to throw CB in the garbage bin if they are more than 10 paces away [2]. We created our simulation by centering a random normal sample with the mean around 1 for the smokers that use the bin to dispose of CB and 1.5 for the smokers that won't be using the bin [because the bin is more than 10 paces further away [1]. As we do not know the standard deviation of the data, we are assuming 1 and adding a noise as the same method mentioned in the staged treatment.

Lastly, for the signs treatment, the team would like to differentiate the effectiveness of a sign as a factor in proper CB disposal. Smith and Novotny reported that 72% of smokers who notices the installment of new receptacles do use them; however, only 20% of smokers do notice the installment [2]. To this effect, we would like to display messages about the nearest garbage cans as well as the detrimental effects of littering CB. For this data, we generated in the same manner as the bin's random normal sample with mean of 1 for the sign subjects and 1.7 for the ones who do not see the signs.

When the team is performing the full experiment, we intend to capture additional factors to see if there is correlation. These variables include the location of the data collection, the average temperature that day, and other weather conditions. Since all of the cities that the team members live in are northern, we expect relatively similar weather conditions, but the data will still be collected as reference.

Generate ATE for the 3 scenarios

```
CB_staged_group_mean <- d[ , .(group_mean = mean(NumCBs_staged)), keyby = .(staged)]
CB_staged_ate        <- CB_staged_group_mean[ , diff(group_mean)]

CB_bin_group_mean <- d[ , .(group_mean = mean(NumCBs_bin)), keyby = .(bin)]
CB_bin_ate        <- CB_bin_group_mean[ , diff(group_mean)]

CB_signs_group_mean <- d[ , .(group_mean = mean(NumCBs_sign)), keyby = .(signs)]
CB_signs_ate        <- CB_signs_group_mean[ , diff(group_mean)]


c(CB_staged_ate, CB_bin_ate, CB_signs_ate)
```

```
## [1]  0.8 -0.3 -1.2
```

Power Calculation: Treatment Effect of staged CB

Sample 1 - Staged CBs Treamtent

```
#resampling function for the power analysis
cb_sample <- function(data = d, size_in_each = 5) {
  data[ , .(NumCBs_staged = sample(NumCBs_staged, size_in_each, replace = TRUE)), by = staged] %$%
    t.test(NumCBs_staged ~ staged) %$%
    p.value
}
```

```
percentages_to_sample <- seq(from = 10, to = 30, by = 1)
sample_per_condition_staged  <- (d[ , .N] * .5 * percentages_to_sample) %>%
  round()
achieved_power_staged <- NA

for(i in 1:length(sample_per_condition_staged)) {

  p_values <- replicate(
    n = 1000,
    expr = cb_sample(data = d, size_in_each = sample_per_condition_staged[i])
    )

  achieved_power_staged[i] <- mean(p_values < 0.05)
}
```

Sample 2 - Bins Treatment

```
#resampling function for the power analysis
cb_sample_bin <- function(data = d, size_in_each = 5) {
  data[ , .(NumCBs_bin = sample(NumCBs_bin, size_in_each, replace = TRUE)), by = bin] %$%
    t.test(NumCBs_bin ~ bin) %$%
    p.value
}
```

```
percentages_to_sample <- seq(from = 10, to = 30, by = 1)
sample_per_condition_bin  <- (d[ , .N] * .5 * percentages_to_sample) %>%
  round()
achieved_power_bin <- NA

for(i in 1:length(sample_per_condition_bin)) {
  p_values <- replicate(
    n = 1000,
    expr = cb_sample_bin(data = d, size_in_each = sample_per_condition_bin[i])
    )

  achieved_power_bin[i] <- mean(p_values < 0.05)

}
```

Sample 3 - Signs Treatment

[1]

```
#resampling function for the power analysis
cb_sample_sign <- function(data = d, size_in_each = 5) {
  data[ , .(NumCBs_sign = sample(NumCBs_sign, size_in_each, replace = TRUE)), by = signs] %$%
    t.test(NumCBs_sign ~ signs) %$%
    p.value
}
```

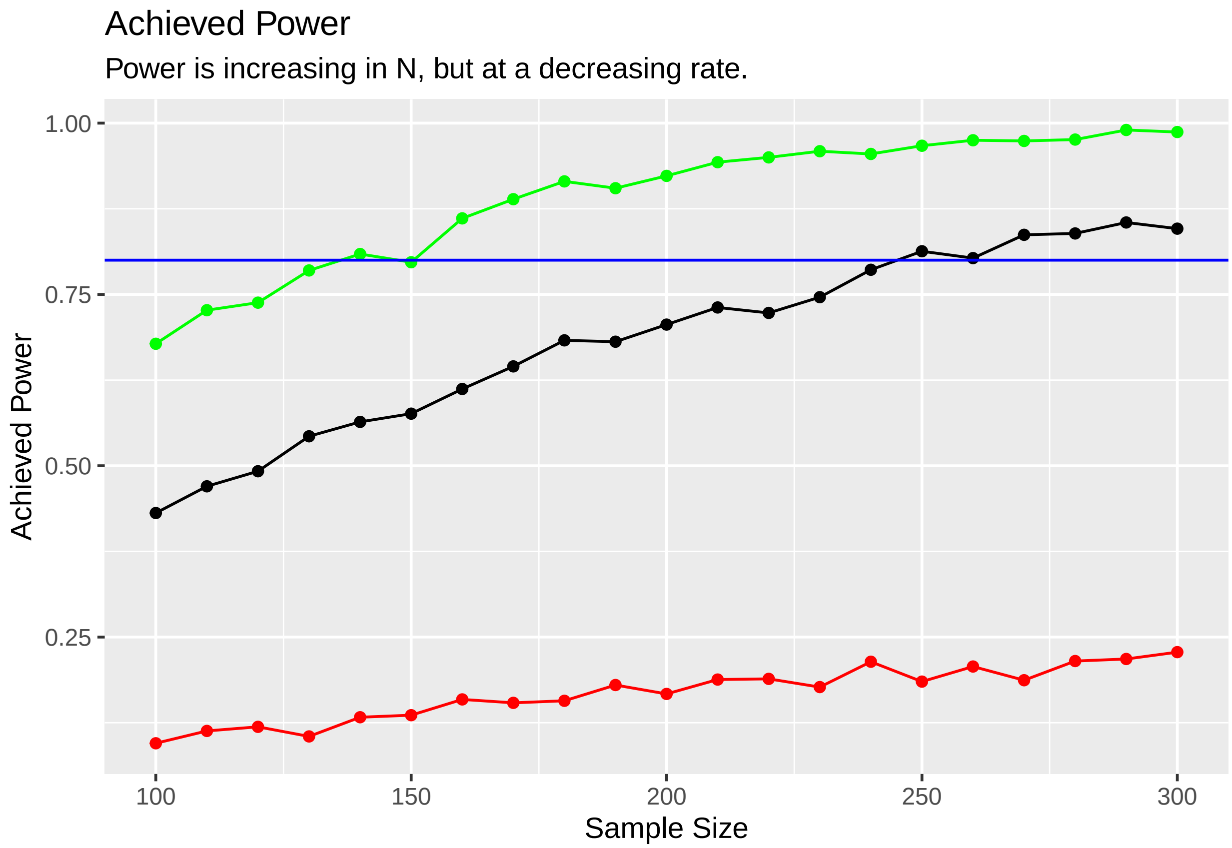
```
percentages_to_sample <- seq(from = 10, to = 30, by = 1)
sample_per_condition_sign  <- (d[ , .N] * .5 * percentages_to_sample) %>%
  round()
achieved_power_sign <- NA

for(i in 1:length(sample_per_condition_sign)) {
  p_values <- replicate(
    n = 1000,
    expr = cb_sample_sign(data = d, size_in_each = sample_per_condition_sign[i])
    )

  achieved_power_sign[i] <- mean(p_values < 0.05)
}
```

Power Calculation Visual Analysis

In Sample 3, the presence of signs in the smoking area and Sample 1 staging CBs in the smoking areas achieved the necessary power of 0.8, which are reasonable treatment effects we would consider in the further experiment design. For treatment using signs, ~150 sample size is required to achieve 0.8; treatment using staging CBs, ~250 samplze size is required to achieve desired power of 0.8. However, we may reconsider the bin or no bin approach since it achieved low power and gained very little power as we increased the sample size.

```
ggplot() +
  geom_line(aes(x = sample_per_condition_staged, y = achieved_power_staged, colour ="Staged"), color="bl
  geom_line(aes(x = sample_per_condition_bin, y = achieved_power_bin, colour ="Bin"), color = "red") +
  geom_line(aes(x = sample_per_condition_sign, y = achieved_power_sign, colour ="Sign"), color = "green
  geom_point(aes(x = sample_per_condition_staged, y = achieved_power_staged), color="black") +
  geom_point(aes(x = sample_per_condition_bin, y = achieved_power_bin), color = "red") +
  geom_point(aes(x = sample_per_condition_sign, y = achieved_power_sign), color = "green") +
  geom_hline(yintercept = 0.8, color = 'blue', size=0.5)+
  labs(
    title    = 'Achieved Power',
    subtitle = 'Power is increasing in N, but at a decreasing rate.',
    x        = 'Sample Size',
    y        = 'Achieved Power'
  )
```

[1] code was replicated for Sample 2 and 3 following the same technique as in Sample 1



References

[1] Julian Sagebiel et al 2020 Environ. Res. Commun. 2 091002

[2] Smith, E. A., & Novotny, T. E. (2011). Whose butt is it? Tobacco industry research about smokers and cigarette butt waste. Tobacco control, 20(Suppl 1), i2-i9.