Hiking Survey Analysis

```
# Loading Packages
library(foreign)
library(data.table)
library(knitr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(naniar)
```

```
## Warning: package 'naniar' was built under R version 3.6.3
```

```
library(sandwich)
library(lmtest)
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 3.6.2
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(multiwayvcov)
```

```
## Warning: package 'multiwayvcov' was built under R version 3.6.2
```

```
library(pwr)
```

```
## Warning: package 'pwr' was built under R version 3.6.3
```

```
# Cleaning and Transforming the Survey Data

Hiking_Data <- read.csv(file = 'Survey_Data.csv')
# Hiking_Data
Hiking_Data <-
  Hiking_Data %>%
    rename(
        Start_Date = StartDate,
        Consent = Q456,
        Duration_Seconds = Duration..in.seconds.,
        Birth_Year = Q2,
        Gender = Q3,
        Gender_Self_Describe = Q3_4_TEXT,
        Height = Q223,
        Weight = Q224,
        Marital_Status = Q5,
        Have_Children = Q23,
        Education_Level = Q7,
        Physical_Activity_Level = Q8,
        Hiking_Backing_Experience_Level = Q16,
        Previous_Desert_Hiking_Experience = Q17,
        Hiking_Exclusion_Condition = Q211,
        Liters_of_Water_Control = Q217_1,
        Liters_of_Water_Treatment = Q659_1
        ) %>%
    filter(Consent == 'I consent, begin the study'
        & (!is.na(Liters_of_Water_Control) == TRUE | !is.na(Liters_of_Water_Treatment) == TRUE)
        & Hiking_Exclusion_Condition == 'No'
        ) %>%
    mutate_all(list(~na_if(.,""))) %>%
    mutate(
        Treatment_Flag = as.factor(ifelse(!is.na(Liters_of_Water_Treatment) == TRUE, 1, 0))
        ,Male_Flag = ifelse(Gender == 'Man', 1, 0)
        ,Duration_Seconds = as.numeric(as.character(Duration_Seconds))
        ,Weight = as.numeric(as.character(Weight))
        ,Height = as.numeric(as.character(Height))
        ,Liters_of_Water_Control = as.numeric(as.character(Liters_of_Water_Control))
        ,Liters_of_Water_Treatment = as.numeric(as.character(Liters_of_Water_Treatment))
        ,Birth_Year = as.numeric(as.character(Birth_Year))
        ,Liters_of_Water_Coalesced =  coalesce(Liters_of_Water_Control, Liters_of_Water_Treatment)
        ,Gender = coalesce(Gender, Gender_Self_Describe)
        ,Age = (2020 - Birth_Year)
        ,BMI = (Weight*703)/(Height**2)
         ) %>%
    replace_with_na_at(.vars = "BMI", condition = ~.x < 5) %>%
    select(
        Start_Date
        # ,Consent
        ,Duration_Seconds
        ,Birth_Year
        ,Gender
        ,Age
```

```
        ,Height
        ,Weight
        ,BMI
        ,Marital_Status
        ,Have_Children
        ,Education_Level
        ,Physical_Activity_Level
        ,Hiking_Backing_Experience_Level
        ,Previous_Desert_Hiking_Experience
        ,Hiking_Exclusion_Condition
        ,Liters_of_Water_Control
        ,Liters_of_Water_Treatment
        ,Liters_of_Water_Coalesced
        ,Treatment_Flag
    )

head(Hiking_Data)
```

```
##        Start_Date Duration_Seconds Birth_Year Gender Age Height Weight      BMI
## 1 4/12/2020 14:01              124       1991    Man  29     70    225 32.28061
## 2 4/12/2020 16:47              146       1991    Man  29     71    165 23.01032
## 3 4/12/2020 16:50              282       1980    Man  40     69    180 26.57845
## 4 4/12/2020 16:56              210       1990    Man  30     71    160 22.31303
## 5 4/12/2020 17:02              161       1988  Woman  32     62    115 21.03148
## 6 4/12/2020 17:04              227       1990    Man  30     66    145 23.40106
##   Marital_Status Have_Children                   Education_Level
## 1        Married            No                  College graduate
## 2         Single            No Graduate level (some or graduate)
## 3        Married           Yes Graduate level (some or graduate)
## 4         Single            No Graduate level (some or graduate)
## 5        Married           Yes Graduate level (some or graduate)
## 6        Married            No Graduate level (some or graduate)
##                 Physical_Activity_Level
## 1          Moderate (3-4 Times per week)
## 2           Minimal (1-2 Times per week)
## 3 Extensive (More than 4 Times per week)
## 4 Extensive (More than 4 Times per week)
## 5 Extensive (More than 4 Times per week)
## 6           Minimal (1-2 Times per week)
##                                                  Hiking_Backing_Experience_Level
## 1      Moderate (I hike or backpack a few times a year, for at least a mile at a time)
## 2 Minimal (I've gone hiking or backpacking a few times, for at least a mile at a time)
## 3      Moderate (I hike or backpack a few times a year, for at least a mile at a time)
## 4              Extensive (I hike or backpack regularly, for at least a mile at a time)
## 5      Moderate (I hike or backpack a few times a year, for at least a mile at a time)
## 6      Moderate (I hike or backpack a few times a year, for at least a mile at a time)
##   Previous_Desert_Hiking_Experience Hiking_Exclusion_Condition
## 1                                No                        No
## 2                                No                        No
## 3                               Yes                        No
## 4                               Yes                        No
## 5                                No                        No
## 6                               Yes                        No
##   Liters_of_Water_Control Liters_of_Water_Treatment Liters_of_Water_Coalesced
```

```
## 1                    3.0                       NA                       3.0
## 2                     NA                        3                       3.0
## 3                     NA                        4                       4.0
## 4                     NA                        4                       4.0
## 5                    4.5                       NA                       4.5
## 6                    4.0                       NA                       4.0
##   Treatment_Flag
## 1              0
## 2              1
## 3              1
## 4              1
## 5              0
## 6              0
```

```
# Ensuring that we have adequate responses for both the treatment and control groups

Hiking_Data %>%
  group_by(Treatment_Flag)%>%
  tally()
```

```
## # A tibble: 2 x 2
##   Treatment_Flag     n
##   <fct>          <int>
## 1 0                 14
## 2 1                 17
```

```
# Simple T-Test to determine if there is a significant difference between the means of both groups

t_test <- t.test(Hiking_Data$Liters_of_Water_Control, Hiking_Data$Liters_of_Water_Treatment)
diff_in_mean <- t_test$estimate[2] - t_test$estimate[1]
p_value <- t_test$p.value
t_test
```

```
##
##  Welch Two Sample t-test
##
## data:  Hiking_Data$Liters_of_Water_Control and Hiking_Data$Liters_of_Water_Treatment
## t = 0.50561, df = 26.314, p-value = 0.6173
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9472276  1.5657150
## sample estimates:
## mean of x mean of y
##  3.985714  3.676471
```

Based on the p-value of 0.6173364 from the t-test, we can see that the -0.3092437 difference between the means is not statistically significant.

```
# Simple Linear Regression with Liters of Water Regressed on Treatment Group Dummy Variable
mod <- lm(Hiking_Data$Liters_of_Water_Coalesced ~ Hiking_Data$Treatment_Flag)
summary(mod)
```

```
##
## Call:
## lm(formula = Hiking_Data$Liters_of_Water_Coalesced ~ Hiking_Data$Treatment_Flag)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -2.1765 -1.3311  0.3143  0.9189  4.3235
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    3.9857     0.4475   8.906 8.53e-10 ***
## Hiking_Data$Treatment_Flag1   -0.3092     0.6043  -0.512    0.613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.674 on 29 degrees of freedom
## Multiple R-squared:  0.008949,   Adjusted R-squared:  -0.02523
## F-statistic: 0.2619 on 1 and 29 DF,  p-value: 0.6127
```

```r
# plot(mod)

n <- length(mod$effects)
r2 <- summary(mod)$r.squared
f2 <- r2/(1-r2)
power_test_post_hoc <-pwr.f2.test(u = 1, v = n - 1 - 1, f2 = f2, sig.level = .05)
power_test_a_priori <-pwr.f2.test(u = 1, f2 = f2, sig.level = .05, power = .8)
power_from_experiment <- power_test_post_hoc$power
required_n_for_a_priori <- ceiling(power_test_a_priori$v + power_test_a_priori$u + 1)

# Reference
# https://cran.r-project.org/web/packages/pwr/vignettes/pwr-vignette.html
```

As expected, we see the same p-value and difference in means between the two groups as we saw in the t-test. We can also see that we have a very low R-Squared value of 0.0089491. Using that R-Squared value to compute Cohens f2, we obtain an f2 value of 0.0090299. When we also factor in our low sample size of 31 observations and seek to reject the null at a significance level of .05, we obtain a power level of 0.0805043. It is also worth noting that given this effect size, we would need 872 observations to reject the null at a .05 significance level and with 80% power.

```r
# Multiple Regression with Liters of Water Regressed on Treatment Group Dummy Variable and all Covariat
mod2 <- lm(Hiking_Data$Liters_of_Water_Coalesced ~ Hiking_Data$Treatment_Flag
           + Hiking_Data$Gender
           + Hiking_Data$BMI
           + Hiking_Data$Age
           + Hiking_Data$Marital_Status
           + Hiking_Data$Have_Children
           + Hiking_Data$Education_Level
           + Hiking_Data$Physical_Activity_Level
           + Hiking_Data$Hiking_Backing_Experience_Level
           + Hiking_Data$Previous_Desert_Hiking_Experience
          )
summary(mod2)
```

```
##
## Call:
## lm(formula = Hiking_Data$Liters_of_Water_Coalesced ~ Hiking_Data$Treatment_Flag +
##     Hiking_Data$Gender + Hiking_Data$BMI + Hiking_Data$Age +
##     Hiking_Data$Marital_Status + Hiking_Data$Have_Children +
##     Hiking_Data$Education_Level + Hiking_Data$Physical_Activity_Level +
```

```
##     Hiking_Data$Hiking_Backing_Experience_Level + Hiking_Data$Previous_Desert_Hiking_Experience)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8271 -0.9820  0.0000  0.9619  2.8584
##
## Coefficients:
##
## (Intercept)
## Hiking_Data$Treatment_Flag1
## Hiking_Data$GenderWoman
## Hiking_Data$BMI
## Hiking_Data$Age
## Hiking_Data$Marital_StatusSingle
## Hiking_Data$Have_ChildrenYes
## Hiking_Data$Education_LevelGraduate level (some or graduate)
## Hiking_Data$Education_LevelSome college
## Hiking_Data$Physical_Activity_LevelMinimal (1-2 Times per week)
## Hiking_Data$Physical_Activity_LevelModerate (3-4 Times per week)
## Hiking_Data$Hiking_Backing_Experience_LevelMinimal (I've gone hiking or backpacking a few times, for
## Hiking_Data$Hiking_Backing_Experience_LevelModerate (I hike or backpack a few times a year, for at le
## Hiking_Data$Previous_Desert_Hiking_ExperienceYes
##
## (Intercept)
## Hiking_Data$Treatment_Flag1
## Hiking_Data$GenderWoman
## Hiking_Data$BMI
## Hiking_Data$Age
## Hiking_Data$Marital_StatusSingle
## Hiking_Data$Have_ChildrenYes
## Hiking_Data$Education_LevelGraduate level (some or graduate)
## Hiking_Data$Education_LevelSome college
## Hiking_Data$Physical_Activity_LevelMinimal (1-2 Times per week)
## Hiking_Data$Physical_Activity_LevelModerate (3-4 Times per week)
## Hiking_Data$Hiking_Backing_Experience_LevelMinimal (I've gone hiking or backpacking a few times, for
## Hiking_Data$Hiking_Backing_Experience_LevelModerate (I hike or backpack a few times a year, for at le
## Hiking_Data$Previous_Desert_Hiking_ExperienceYes
##
## (Intercept)
## Hiking_Data$Treatment_Flag1
## Hiking_Data$GenderWoman
## Hiking_Data$BMI
## Hiking_Data$Age
## Hiking_Data$Marital_StatusSingle
## Hiking_Data$Have_ChildrenYes
## Hiking_Data$Education_LevelGraduate level (some or graduate)
## Hiking_Data$Education_LevelSome college
## Hiking_Data$Physical_Activity_LevelMinimal (1-2 Times per week)
## Hiking_Data$Physical_Activity_LevelModerate (3-4 Times per week)
## Hiking_Data$Hiking_Backing_Experience_LevelMinimal (I've gone hiking or backpacking a few times, for
## Hiking_Data$Hiking_Backing_Experience_LevelModerate (I hike or backpack a few times a year, for at le
## Hiking_Data$Previous_Desert_Hiking_ExperienceYes
##
## (Intercept)
```

```
## Hiking_Data$Treatment_Flag1
## Hiking_Data$GenderWoman
## Hiking_Data$BMI
## Hiking_Data$Age
## Hiking_Data$Marital_StatusSingle
## Hiking_Data$Have_ChildrenYes
## Hiking_Data$Education_LevelGraduate level (some or graduate)
## Hiking_Data$Education_LevelSome college
## Hiking_Data$Physical_Activity_LevelMinimal (1-2 Times per week)
## Hiking_Data$Physical_Activity_LevelModerate (3-4 Times per week)
## Hiking_Data$Hiking_Backing_Experience_LevelMinimal (I've gone hiking or backpacking a few times, for
## Hiking_Data$Hiking_Backing_Experience_LevelModerate (I hike or backpack a few times a year, for at le
## Hiking_Data$Previous_Desert_Hiking_ExperienceYes
##
## (Intercept)
## Hiking_Data$Treatment_Flag1
## Hiking_Data$GenderWoman
## Hiking_Data$BMI
## Hiking_Data$Age
## Hiking_Data$Marital_StatusSingle
## Hiking_Data$Have_ChildrenYes
## Hiking_Data$Education_LevelGraduate level (some or graduate)
## Hiking_Data$Education_LevelSome college
## Hiking_Data$Physical_Activity_LevelMinimal (1-2 Times per week)
## Hiking_Data$Physical_Activity_LevelModerate (3-4 Times per week)
## Hiking_Data$Hiking_Backing_Experience_LevelMinimal (I've gone hiking or backpacking a few times, for
## Hiking_Data$Hiking_Backing_Experience_LevelModerate (I hike or backpack a few times a year, for at le
## Hiking_Data$Previous_Desert_Hiking_ExperienceYes
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.896 on 15 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.307,  Adjusted R-squared:  -0.2936
## F-statistic: 0.5112 on 13 and 15 DF,  p-value: 0.8843
```

```
# plot(mod2)
```

The model does not improve when adding in the additional covariates, and this is likely due to the fact that we are working with such few observations. Until additional survey volume can be collected and the analysis re-performed, the results of this regression should be interpreted with caution if not discarded entirely.