Hypothesis Testing and Mediation Analysis

Your Name

2024-12-12

Required Libraries and Data Import

```
required_packages <- c(
    "rio", "ggplot2", "vtable", "plot3D",
    "corrplot", "regclass", "olsrr", "mediation", "multilevel", "stargazer"
)

for (i in 1:length(required_packages)) {
    if (!required_packages[i] %in% installed.packages()) {
        install.packages(required_packages[i])
    }
}

library(ggplot2)
library(rio)
```

```
## Warning: package 'rio' was built under R version 4.3.3
```

```
# Import data
df <- rio::import("BE603_gr_10.csv")
```

Hypothesis 1: Community Effect (Mediation Analysis)

Models and Bootstrap

```
# Model definitions
model_T <- lm(collected_funds ~ updates_count, data = df)
model_M <- lm(comments_count ~ updates_count, data = df)
model_Y <- lm(collected_funds ~ comments_count + updates_count, data = df)

# Print summaries
print("------------------MODEL T------------------")
```

```
## [1] "------------------MODEL T------------------"
```

```
print(summary(model_T))
```

```
##
## Call:
## lm(formula = collected_funds ~ updates_count, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -13759  -1727  -1020    513  38742
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2477.08      50.84   48.72   <2e-16 ***
## updates_count    84.65       5.32   15.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3200 on 4998 degrees of freedom
## Multiple R-squared:  0.04821,    Adjusted R-squared:  0.04802
## F-statistic: 253.2 on 1 and 4998 DF,  p-value: < 2.2e-16
```

```
print("------------------MODEL M------------------")
```

```
## [1] "------------------MODEL M------------------"
```

```
print(summary(model_M))
```

```
##
## Call:
## lm(formula = comments_count ~ updates_count, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -110.90  -10.39   -5.39    4.12  429.71
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.38881    0.32864   40.74   <2e-16 ***
## updates_count  0.76694    0.03439   22.30   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.68 on 4998 degrees of freedom
## Multiple R-squared:  0.09049,    Adjusted R-squared:  0.09031
## F-statistic: 497.3 on 1 and 4998 DF,  p-value: < 2.2e-16
```

```
print("------------------MODEL Y------------------")
```

```
## [1] "------------------MODEL Y------------------"
```

```
print(summary(model_Y))
```

```
##
## Call:
## lm(formula = collected_funds ~ comments_count + updates_count,
##     data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -23874  -1187   -633    434  33202
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1295.505     48.198  26.879  < 2e-16 ***
## comments_count   88.251      1.797  49.099  < 2e-16 ***
## updates_count    16.970      4.582   3.703 0.000215 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2628 on 4997 degrees of freedom
## Multiple R-squared:  0.358,  Adjusted R-squared:  0.3577
## F-statistic:  1393 on 2 and 4997 DF,  p-value: < 2.2e-16
```

```
# Mediation analysis with bootstrap
library(mediation)
```

```
## Loading required package: MASS
```

```
## Loading required package: Matrix
```

```
## Loading required package: mvtnorm
```

```
## Warning: package 'mvtnorm' was built under R version 4.3.3
```

```
## Loading required package: sandwich
```

```
## Warning: package 'sandwich' was built under R version 4.3.3
```

```
## mediation: Causal Mediation Analysis
## Version: 4.5.0
```

```
model_bootstrap <- mediation::mediate(
    model.m = model_M,
    model.y = model_Y,
    treat = "updates_count",
    mediator = "comments_count",
    boot = TRUE,
    sims = 500
)
```

```
## Running nonparametric bootstrap
```

```
print("------------------BOOTSTRAP------------------")
```

```
## [1] "------------------BOOTSTRAP------------------"
```

```
print(summary(model_bootstrap))
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                Estimate 95% CI Lower 95% CI Upper p-value
## ACME             67.683       53.961        85.08  <2e-16 ***
## ADE              16.970        2.709        33.06   0.028 *
## Total Effect     84.654       66.508       109.69  <2e-16 ***
## Prop. Mediated    0.800        0.668         0.96  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 5000
##
##
## Simulations: 500
```

```r
# Save mediation plot
png("H1_mediation_analysis_plot.png")
plot(model_bootstrap)
dev.off()
```

```
## pdf
##   2
```

Visualization

```r
# Plotting comments vs updates
M_plot <- ggplot(df) +
    aes(y = comments_count, x = updates_count) +
    geom_hline(yintercept = 1, linetype = "dashed", alpha = 0.5) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", formula = "y ~ x", se = TRUE) +
    scale_x_log10() +
    scale_y_log10() +
    annotation_logticks(sides = "bl") +
    labs(
        y = "Comments Count",
        x = "Updates Count"
    ) +
    theme_classic()

ggsave("H1_mediating_plot.png", M_plot, width = 10, height = 6, dpi = 300)
```

```
## Warning in scale_x_log10(): log-10 transformation introduced infinite values.
```

```
## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
```

```
## Warning in scale_x_log10(): log-10 transformation introduced infinite values.
```

```
## Warning in scale_y_log10(): log-10 transformation introduced infinite values.
```

```
## Warning: Removed 2113 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```r
# Plotting collected funds vs updates
T_plot <- ggplot(df) +
    aes(y = collected_funds, x = updates_count) +
    geom_hline(yintercept = 1, linetype = "dashed", alpha = 0.5) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", formula = "y ~ x", se = TRUE) +
    labs(
        y = "Collected funds",
        x = "Updates Count"
    ) +
    theme_classic()
```

3D Visualization

```r
library(plot3D)
# 3D scatter plot with regression plane
x <- df$updates_count
y <- df$comments_count
```

```r
z <- df$collected_funds

fit <- lm(z ~ x + y)

# Create grid for regression plane
grid.lines <- 40
x.pred <- seq(min(x), max(x), length.out = grid.lines)
y.pred <- seq(min(y), max(y), length.out = grid.lines)
xy <- expand.grid(x = x.pred, y = y.pred)
z.pred <- matrix(predict(fit, newdata = xy), nrow = grid.lines, ncol = grid.lines)

png("h1_3d_plot.png", width = 800, height = 800)
scatter3D(x, y, z,
    pch = 19, cex = 1, colvar = NULL, col = "red",
    theta = 20, phi = 10, bty = "b",
    xlab = "Updates Count", ylab = "Comments Count",
    zlab = "Collected funds",
    surf = list(
        x = x.pred, y = y.pred, z = z.pred,
        facets = TRUE, fit = predict(fit),
        col = ramp.col(
            col = c("dodgerblue3", "seagreen2"),
            n = 300, alpha = 0.9
        ),
        border = "black"
    ),
    main = "Updates, Comments, and Collected funds"
)
dev.off()
```

```
## pdf
##   2
```

Hypothesis 2: Social and Negative Emotion Interaction

Model

```r
model_1 <- lm(collected_funds ~ social * negemo, data = df)
print(summary(model_1))
```

```
##
## Call:
## lm(formula = collected_funds ~ social * negemo, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3566  -1860  -1140    487  38531
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2903.47      75.59  38.410  < 2e-16 ***
## social         -193.04     148.26  -1.302  0.19296
## negemo         -107.46      66.14  -1.625  0.10429
## social:negemo   302.13     112.03   2.697  0.00702 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3278 on 4996 degrees of freedom
## Multiple R-squared:  0.001619,   Adjusted R-squared:  0.001019
## F-statistic:   2.7 on 3 and 4996 DF,  p-value: 0.04411
```

Visualization

```
df$social_factor <- as.factor(df$social)

h2_plot <- ggplot(df) +
    aes(y = collected_funds, x = negemo, color = social_factor) +
    geom_hline(yintercept = 1, linetype = "dashed", alpha = 0.5) +
    geom_point(alpha = 0.5) +
    geom_smooth(method = "lm", formula = "y ~ x", se = TRUE) +
    labs(
        y = "Fundraising Ratio (Collected/Goal)",
        x = "Negative emotion",
        color = "Social (0/1)"
    ) +
    scale_color_discrete(labels = c("No Social", "Social")) +
    theme_classic()

ggsave("h2_interaction_plot.png", h2_plot, width = 10, height = 6, dpi = 300)
```

Hypothesis 3: Early Success

Model and Visualization

```
model_1 <- lm(collected_funds ~ goal, data = df)
print(summary(model_1))
```

```
##
## Call:
## lm(formula = collected_funds ~ goal, data = df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -6310  -1515   -659    747  35676
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.446e+03  6.899e+01   20.96   <2e-16 ***
## goal        2.260e-01  8.645e-03   26.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3076 on 4998 degrees of freedom
## Multiple R-squared:  0.1203, Adjusted R-squared:  0.1201
## F-statistic: 683.6 on 1 and 4998 DF,  p-value: < 2.2e-16
```

```
h3_plot <- ggplot(df) +
  aes(y = collected_funds, x = goal) +
```

```
  geom_hline(yintercept = 1, linetype = "dashed", alpha = 0.5) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", formula = "y ~ x", se = TRUE) +
  labs(
    y = "Collected funds",
    x = "Goal",
    title = "Correlation between goal and collected funds"
  ) +
  theme_classic()

ggsave("H3_fundraisingratio_reachin30.png", h3_plot, width = 8, height = 6, dpi = 300)
```