CMA3

2025-05-14

```
library(dplyr)
```

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

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

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

```
library(ggplot2)
library(broom)



# Compute the 80th percentile of the self-efficacy measure 'job_seek'
percentile_80 <- quantile(jobs2$job_seek, 0.80, na.rm = TRUE)

# Create a new variable for the mediator (job search self-efficacy) set at its 80th perc
entile
jobs2$job_seek_80 <- ifelse(jobs2$job_seek >= percentile_80, jobs2$job_seek, percentile_
80)

# Linear regression with outcome 'work1', treatment 'treat', mediator 'job_seek_80', and
covariates
linear_model <- lm(work1 ~ treat * job_seek_80 + econ_hard + sex + age + nonwhite + educ
+ income, data = jobs2)


summary(linear_model)
```

```
##
## Call:
## lm(formula = work1 ~ treat * job_seek_80 + econ_hard + sex +
##     age + nonwhite + educ + income, data = jobs2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6547 -0.3461 -0.2280  0.5598  0.9316
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -1.648141   1.134445  -1.453  0.14663
## treat              0.072485   1.347544   0.054  0.95711
## job_seek_80        0.495284   0.242098   2.046  0.04107 *
## econ_hard         -0.009338   0.016385  -0.570  0.56887
## sex               -0.083617   0.030861  -2.709  0.00687 **
## age               -0.007382   0.001511  -4.885 1.22e-06 ***
## nonwhite          -0.107474   0.041744  -2.575  0.01020 *
## educ              -0.004543   0.014447  -0.314  0.75326
## income            -0.004444   0.011931  -0.372  0.70961
## treat:job_seek_80 -0.005358   0.285517  -0.019  0.98503
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4576 on 892 degrees of freedom
## Multiple R-squared:  0.05932,    Adjusted R-squared:  0.04983
## F-statistic:  6.25 on 9 and 892 DF,  p-value: 1.353e-08
```

```
# Logistic regression with outcome 'work1' (logit model)
logit_model <- glm(work1 ~ treat * job_seek_80 + econ_hard + sex + age + nonwhite + educ
+ income,
                   data = jobs2, family = binomial(link = "logit"))


summary(logit_model)
```

```
##
## Call:
## glm(formula = work1 ~ treat * job_seek_80 + econ_hard + sex +
##     age + nonwhite + educ + income, family = binomial(link = "logit"),
##     data = jobs2)
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -10.345517   5.255091  -1.969  0.04899 *
## treat               1.390205   6.215221   0.224  0.82301
## job_seek_80         2.396185   1.121880   2.136  0.03269 *
## econ_hard          -0.039268   0.079423  -0.494  0.62101
## sex                -0.399169   0.147670  -2.703  0.00687 **
## age                -0.036573   0.007689  -4.757 1.97e-06 ***
## nonwhite           -0.529870   0.210419  -2.518  0.01180 *
## educ               -0.018174   0.069482  -0.262  0.79365
## income             -0.015089   0.057665  -0.262  0.79358
## treat:job_seek_80  -0.245351   1.314595  -0.187  0.85195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1140.3  on 901  degrees of freedom
## Residual deviance: 1085.4  on 892  degrees of freedom
## AIC: 1105.4
##
## Number of Fisher Scoring iterations: 4
```

```
# Extract point estimates for the controlled direct effect (CDE) from both models

# For the linear model, the controlled direct effect is the coefficient of the interacti
on term 'treat:job_seek_80'
linear_cde <- tidy(linear_model) %>%
  filter(term == "treat:job_seek_80") %>%
  select(estimate)

# For the logit model, the controlled direct effect is also the coefficient of the inter
action term 'treat:job_seek_80'
logit_cde <- tidy(logit_model) %>%
  filter(term == "treat:job_seek_80") %>%
  select(estimate)


cat("Controlled Direct Effect (Linear Model): ", linear_cde$estimate, "\n")
```

```
## Controlled Direct Effect (Linear Model):  -0.005357832
```

```
cat("Controlled Direct Effect (Logit Model): ", logit_cde$estimate, "\n")
```

```
## Controlled Direct Effect (Logit Model):  -0.2453505
```

```
library(haven)
library(dplyr)
library(boot)
library(broom)

# Set working directory
setwd("/Users/josephepstein/Desktop/Causal Mediation Analysis")


jobs2 <- read_dta("jobs2.dta")

# Compute the 80th percentile of the self-efficacy measure 'job_seek'
percentile_80 <- quantile(jobs2$job_seek, 0.80, na.rm = TRUE)

# Create a new variable for the mediator (job search self-efficacy) set at its 80th perc
entile
jobs2$job_seek_80 <- ifelse(jobs2$job_seek >= percentile_80, jobs2$job_seek, percentile_
80)

# Define the function to fit the logit model and extract CDE
bootstrap_function <- function(data, indices) {
  # Resample the data
  resampled_data <- data[indices, ]

  # Fit logit model
  logit_model <- glm(work1 ~ treat * job_seek_80 + econ_hard + sex + age + nonwhite + ed
uc + income,
                     data = resampled_data, family = binomial(link = "logit"))

  # Extract the coefficient for the interaction term 'treat:job_seek_80' / CDE
  coef_interaction <- coef(logit_model)["treat:job_seek_80"]

  return(coef_interaction)
}

# Perform percentile bootstrap with 1000 replications
set.seed(123)  # For reproducibility
bootstrap_results <- boot(data = jobs2, statistic = bootstrap_function, R = 1000)

# Compute the 90% CI for the CDE
bootstrap_ci <- quantile(bootstrap_results$t, c(0.05, 0.95))

# CI
cat("90% Confidence Interval for the Controlled Direct Effect (CDE): ", bootstrap_ci,
"\n")
```

```
## 90% Confidence Interval for the Controlled Direct Effect (CDE):  -2.408939 2.159951
```

```
# Conduct hypothesis test (alpha = 0.1) for H0: CDE = 0 by computing p-value
bootstrap_test_stat <- bootstrap_results$t
p_value <- mean(bootstrap_test_stat >= 0)  # One-sided test for positive CDE
p_value_two_sided <- 2 * min(p_value, 1 - p_value)  # Two-sided test


cat("p-value for the hypothesis test H0: CDE = 0 (alpha = 0.1): ", p_value_two_sided,
"\n")
```

```
## p-value for the hypothesis test H0: CDE = 0 (alpha = 0.1):  0.922
```

```
bootstrap_mean <- mean(bootstrap_results$t)
cat("Mean CDE from bootstrap samples: ", bootstrap_mean, "\n")
```

```
## Mean CDE from bootstrap samples:  -0.1458079
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
setwd("/Users/josephepstein/Downloads")


library(haven)


plowUse_1_ <- read_dta("plowUse (1).dta")


View(plowUse_1_)
```

```
library(lmtest)  # For robust standard errors
```

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

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

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

```
str(plowUse_1_)
```

```
## tibble [234 × 57] (S3: tbl_df/tbl/data.frame)
##  $ isocode                 : chr [1:234] "ABW" "AFG" "AGO" "AIA" ...
##   ..- attr(*, "format.stata")= chr "%9s"
##  $ flfp2000                : num [1:234] NA 31.3 75.2 NA 50.2 NA 54.8 34.2 45.6 57.2
...
##   ..- attr(*, "label")= chr "Female labor force participation in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ female_ownership        : num [1:234] NA 2.8 56.6 NA 10.8 NA NA NA 38 31.8 ...
##   ..- attr(*, "label")= chr "Percent of firms with female ownership (in latest survey
year)"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ women_politics          : num [1:234] NA NA 16 NA 5 7 NA 0 28 3 ...
##   ..- attr(*, "label")= chr "Women in Politics in 2000, WDI"
##   ..- attr(*, "format.stata")= chr "%8.0g"
##  $ plow                    : num [1:234] 0 1 0 0 1 ...
##   ..- attr(*, "label")= chr "Animal plow cultivation variable (v39): Using Ethnologue
- pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ agricultural_suitability : num [1:234] 0.83 0.229 0.921 0.83 0.556 ...
##   ..- attr(*, "label")= chr "overall (millets, sorghum, wheat, barley, rye): share de
fined as suitable"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ tropical_climate        : num [1:234] 1 0.785 1 1 0.359 ...
##   ..- attr(*, "label")= chr "Frac land: tropics and subtropics: using Ethnologue - po
p weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ large_animals           : num [1:234] 1 1 0.999 1 1 ...
##   ..- attr(*, "label")= chr "presence of large animals"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ political_hierarchies   : num [1:234] 4 3.63 2.72 4 3.03 ...
##   ..- attr(*, "label")= chr "Jurisdictional hierarchy beyond local community (v33): U
sing Ethnologue - pop we"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ economic_complexity     : num [1:234] 7 6.2 6.7 7 5.04 ...
##   ..- attr(*, "label")= chr "Settlement patterns (v30)"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ ln_income               : num [1:234] NA NA 6.49 NA 7.09 ...
##   ..- attr(*, "label")= chr "ln (income)"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ ln_income_squared       : num [1:234] NA NA 42.1 NA 50.3 ...
##   ..- attr(*, "label")= chr "ln (income) ^2"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ year_obs                : num [1:234] 1950 1631 1926 1950 1889 ...
##   ..- attr(*, "label")= chr "Year of observation (v102): Using Ethnologue - pop weigh
ted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ plow_aborigenal         : num [1:234] 0 1 0 0 1 ...
##   ..- attr(*, "label")= chr "Group 3 for animal plow cultivation variable (v39): Usin
g Ethnologue - pop weigh"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ hunting                 : num [1:234] 0.025 0.0422 0.1331 0.025 0.025 ...
##   ..- attr(*, "label")= chr "Dependence on hunting for subsistence (v2): Using Ethnol
```

```
ogue - pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ husbandry               : num [1:234] 0.1 0.365 0.178 0.1 0.397 ...
##   ..- attr(*, "label")= chr "Dependence on animal husbandry for subsistence (v4): Usi
ng Ethnologue - pop weig"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ terrain_slope           : num [1:234] 4.82 24.97 18.94 4.82 31.41 ...
##   ..- attr(*, "label")= chr "Avg terrain slope - pop weighted"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ soil_depth              : num [1:234] 0.893 0.262 0.985 0.893 0.781 ...
##   ..- attr(*, "label")= chr "Frac land: no depth constraints (plate 21: 40 cutoff) -
pop weighted"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ avg_temperature         : num [1:234] 26.35 8.63 20.45 26.35 11.06 ...
##   ..- attr(*, "label")= chr "Avg temperature over the year (C) - pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ avg_precipitation       : num [1:234] 96.4 23.2 95.5 96.4 79.3 ...
##   ..- attr(*, "label")= chr "Avg monthly precipiation (mm) - pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ country                 : chr [1:234] "Aruba" "Afghanistan" "Angola" "" ...
##   ..- attr(*, "label")= chr "country name"
##   ..- attr(*, "format.stata")= chr "%46s"
##  $ communist_dummy         : num [1:234] 0 1 1 NA 1 NA 0 0 0 1 ...
##   ..- attr(*, "label")= chr "Communism indicator variable"
##   ..- attr(*, "format.stata")= chr "%8.0g"
##  $ continent               : chr [1:234] "North America" "Asia" "Africa" "North Amer
ica" ...
##   ..- attr(*, "label")= chr "Continent"
##   ..- attr(*, "format.stata")= chr "%13s"
##  $ rugged                  : num [1:234] 0.4624 2.5181 0.8582 0.0126 3.4271 ...
##   ..- attr(*, "label")= chr "Ruggedness (Terrain Ruggedness Index, 100 m.)"
##   ..- attr(*, "format.stata")= chr "%12.0g"
##  $ european_descent        : num [1:234] NA 0 2 NA 100 ...
##   ..- attr(*, "label")= chr "% European descent"
##   ..- attr(*, "format.stata")= chr "%9.3f"
##  $ cath00                  : num [1:234] NA 0 0.621 NA 0.168 ...
##   ..- attr(*, "label")= chr "CATH00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ prot00                  : num [1:234] NA 0 0.15 NA 0.006 ...
##   ..- attr(*, "label")= chr "PROT00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ othchrist00             : num [1:234] NA 0 0.169 NA 0.02 ...
##   ..- attr(*, "label")= chr "OTHCHRIST00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ muslim00                : num [1:234] NA 0.981 0 NA 0.388 ...
##   ..- attr(*, "label")= chr "MUSLIM00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ hindu00                 : num [1:234] NA 0.004 0 NA 0 ...
##   ..- attr(*, "label")= chr "HINDU00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ trade_GDP               : num [1:234] 54.2 NA 120.1 NA 36.6 ...
##   ..- attr(*, "format.stata")= chr "%8.0g"
```

```
##  $ years_civil_conflict     : num [1:234] NA 21 31 NA 0 0 NA 0 16 3 ...
##   ..- attr(*, "label")= chr "Years of civil conflict, 1800-2007 - from COW"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ years_interstate_conflict: num [1:234] NA 1 2 NA 0 0 NA 1 9 2 ...
##   ..- attr(*, "label")= chr "Years of interstate conflict, 1800-2007 - from COW"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ agr_va_gdp2000           : num [1:234] NA NA 5.66 NA 29.13 ...
##   ..- attr(*, "label")= chr "Value Added Agriculture/GDP in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ man_va_gdp2000           : num [1:234] NA NA 2.89 NA 11.44 ...
##   ..- attr(*, "label")= chr "Value Added Manufacturing/GDP in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ serv_va_gdp2000          : num [1:234] NA NA 22.2 NA 51.9 ...
##   ..- attr(*, "label")= chr "Value Added in Service/GDP in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ polity2_2000             : num [1:234] NA -7 -3 NA 5 NA NA -8 8 5 ...
##   ..- attr(*, "label")= chr "Polity 2 measure taken from the Polity IV dataset"
##   ..- attr(*, "format.stata")= chr "%8.0g"
##  $ plow_negative_crops      : num [1:234] 0.632 0.22 0.751 0.632 0.333 ...
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ plow_positive_crops      : num [1:234] 0 0.816 0.286 0 0.974 ...
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ region                   : chr [1:234] "North America" "Asia" "Sub-Saharan Africa"
"North America" ...
##   ..- attr(*, "label")= chr "continent with Sub-Saharan African Dummy"
##   ..- attr(*, "format.stata")= chr "%18s"
##  $ intensity_agriculture    : num [1:234] 0 0.7005 0.0135 0 1 ...
##   ..- attr(*, "label")= chr "intensity of agriculture"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ abs_inherit              : num [1:234] 0 0.000142 0.069202 0 0 ...
##   ..- attr(*, "label")= chr "absence of inheritance"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ matrilocal               : num [1:234] 0 0 0 0 0 ...
##   ..- attr(*, "label")= chr "matrilocal societies"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ patrilocal               : num [1:234] 1 1 0.648 1 0.987 ...
##   ..- attr(*, "label")= chr "patrilocal societies"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ nuclear_fam              : num [1:234] 0 0.000618 0.115161 0 0.015537 ...
##   ..- attr(*, "label")= chr "nuclear family"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ extended_fam             : num [1:234] 1 0.9994 0.0622 1 0.9845 ...
##   ..- attr(*, "label")= chr "extended family"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ oil_pc                   : num [1:234] NA 0 0.0719 NA 0 ...
##   ..- attr(*, "label")= chr "oil production/GDP"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ terrslope2               : num [1:234] 23.2 623.6 358.6 23.2 986.3 ...
##   ..- attr(*, "label")= chr "terrain slope squared"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ soil2                    : num [1:234] 0.7969 0.0684 0.9704 0.7969 0.6098 ...
##   ..- attr(*, "label")= chr "Soil depth squared"
```

```
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ avg_temp2               : num [1:234] 694.4 74.5 418 694.4 122.2 ...
##   ..- attr(*, "label")= chr "average temperature squared"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ avg_precip2             : num [1:234] 9292 538 9127 9292 6294 ...
##   ..- attr(*, "label")= chr "average precipitation squared"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ slope_soil              : num [1:234] 4.3 6.53 18.65 4.3 24.53 ...
##   ..- attr(*, "label")= chr "interaction slope*soil depth"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ slope_temp              : num [1:234] 127 216 387 127 347 ...
##   ..- attr(*, "label")= chr "interaction slope*temperature"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ slope_precip            : num [1:234] 465 579 1809 465 2492 ...
##   ..- attr(*, "label")= chr "interaction slope*precipitation"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ soil_temp               : num [1:234] 23.52 2.26 20.14 23.52 8.63 ...
##   ..- attr(*, "label")= chr "interaction soil*temperature"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ soil_precip             : num [1:234] 86.05 6.07 94.11 86.05 61.95 ...
##   ..- attr(*, "label")= chr "interaction soil*precipitation"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ temp_precip             : num [1:234] 2540 200 1953 2540 877 ...
##   ..- attr(*, "label")= chr "interaction temperature*precipitation"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  - attr(*, "notes")= chr [1:18] "this file has been compiled by Antonio Spilimbergo w
ith the help of Rafael Romeu and Adrian De La Garza. It is "| __truncated__ "This data s
et contains the data necessary to replicate Table 1 in James D. Fearon and David D. Lait
in, Ethnicit"| __truncated__ "gdpppp95 data for Bosnia/Herzegovina; Bhutan; Brunei; */"
"this file contains a subset of the COW Inter-State War Data, 1816-1997 The files were o
btained from the site ht"| __truncated__ ...
```

```
# Run the first regression to estimate residuals for the mediator (ln_income)
mediator_model <- lm(ln_income ~ plow + agricultural_suitability + tropical_climate +
                     large_animals + rugged + polity2_2000, data = plowUse_1_,
                     na.action = na.exclude)

# residuals of mediator
plowUse_1_$resid_mediator <- residuals(mediator_model)

# Run the second regression to estimate the residuals for women in politics-outcome
outcome_model <- lm(women_politics ~ plow + agricultural_suitability + tropical_climate
+
                    large_animals + rugged + polity2_2000, data = plowUse_1_,
                    na.action = na.exclude)

# residuals of outcome
plowUse_1_$resid_outcome <- residuals(outcome_model)

# indirect effect estimate: Regression of residualized outcome on residualized mediator
indirect_model <- lm(resid_outcome ~ resid_mediator + plow + resid_mediator:plow, data =
plowUse_1_)

# direct effect estimate: Regression of residualized outcome on exposure (plow)
direct_model <- lm(resid_outcome ~ plow + resid_mediator + resid_mediator:plow, data = p
lowUse_1_)


summary(indirect_model)
```

```
##
## Call:
## lm(formula = resid_outcome ~ resid_mediator + plow + resid_mediator:plow,
##     data = plowUse_1_)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.1382  -5.7614  -0.7043   3.1711  23.6597
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.02672    1.03750  -0.026  0.97950
## resid_mediator       2.95157    1.10325   2.675  0.00847 **
## plow                -0.61864    1.42963  -0.433  0.66596
## resid_mediator:plow -1.25791    1.34486  -0.935  0.35141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.562 on 125 degrees of freedom
##   (105 observations deleted due to missingness)
## Multiple R-squared:  0.1064, Adjusted R-squared:  0.08495
## F-statistic: 4.961 on 3 and 125 DF,  p-value: 0.002749
```

```
summary(direct_model)
```

```
##
## Call:
## lm(formula = resid_outcome ~ plow + resid_mediator + resid_mediator:plow,
##     data = plowUse_1_)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.1382  -5.7614  -0.7043   3.1711  23.6597
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.02672    1.03750  -0.026  0.97950
## plow                -0.61864    1.42963  -0.433  0.66596
## resid_mediator       2.95157    1.10325   2.675  0.00847 **
## plow:resid_mediator -1.25791    1.34486  -0.935  0.35141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.562 on 125 degrees of freedom
##   (105 observations deleted due to missingness)
## Multiple R-squared:  0.1064, Adjusted R-squared:  0.08495
## F-statistic: 4.961 on 3 and 125 DF,  p-value: 0.002749
```

```
library(haven)
library(lmtest)
library(boot)


setwd("/Users/josephepstein/Downloads")  # Adjust the path to your dataset
plowUse_1_ <- read_dta("plowUse (1).dta")


str(plowUse_1_)
```

```
## tibble [234 × 57] (S3: tbl_df/tbl/data.frame)
##  $ isocode                 : chr [1:234] "ABW" "AFG" "AGO" "AIA" ...
##   ..- attr(*, "format.stata")= chr "%9s"
##  $ flfp2000                : num [1:234] NA 31.3 75.2 NA 50.2 NA 54.8 34.2 45.6 57.2
...
##   ..- attr(*, "label")= chr "Female labor force participation in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ female_ownership        : num [1:234] NA 2.8 56.6 NA 10.8 NA NA NA 38 31.8 ...
##   ..- attr(*, "label")= chr "Percent of firms with female ownership (in latest survey
year)"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ women_politics          : num [1:234] NA NA 16 NA 5 7 NA 0 28 3 ...
##   ..- attr(*, "label")= chr "Women in Politics in 2000, WDI"
##   ..- attr(*, "format.stata")= chr "%8.0g"
##  $ plow                    : num [1:234] 0 1 0 0 1 ...
##   ..- attr(*, "label")= chr "Animal plow cultivation variable (v39): Using Ethnologue
- pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ agricultural_suitability : num [1:234] 0.83 0.229 0.921 0.83 0.556 ...
##   ..- attr(*, "label")= chr "overall (millets, sorghum, wheat, barley, rye): share de
fined as suitable"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ tropical_climate        : num [1:234] 1 0.785 1 1 0.359 ...
##   ..- attr(*, "label")= chr "Frac land: tropics and subtropics: using Ethnologue - po
p weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ large_animals           : num [1:234] 1 1 0.999 1 1 ...
##   ..- attr(*, "label")= chr "presence of large animals"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ political_hierarchies   : num [1:234] 4 3.63 2.72 4 3.03 ...
##   ..- attr(*, "label")= chr "Jurisdictional hierarchy beyond local community (v33): U
sing Ethnologue - pop we"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ economic_complexity     : num [1:234] 7 6.2 6.7 7 5.04 ...
##   ..- attr(*, "label")= chr "Settlement patterns (v30)"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ ln_income               : num [1:234] NA NA 6.49 NA 7.09 ...
##   ..- attr(*, "label")= chr "ln (income)"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ ln_income_squared       : num [1:234] NA NA 42.1 NA 50.3 ...
##   ..- attr(*, "label")= chr "ln (income) ^2"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ year_obs                : num [1:234] 1950 1631 1926 1950 1889 ...
##   ..- attr(*, "label")= chr "Year of observation (v102): Using Ethnologue - pop weigh
ted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ plow_aborigenal         : num [1:234] 0 1 0 0 1 ...
##   ..- attr(*, "label")= chr "Group 3 for animal plow cultivation variable (v39): Usin
g Ethnologue - pop weigh"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ hunting                 : num [1:234] 0.025 0.0422 0.1331 0.025 0.025 ...
##   ..- attr(*, "label")= chr "Dependence on hunting for subsistence (v2): Using Ethnol
```

```
ogue - pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ husbandry               : num [1:234] 0.1 0.365 0.178 0.1 0.397 ...
##   ..- attr(*, "label")= chr "Dependence on animal husbandry for subsistence (v4): Usi
ng Ethnologue - pop weig"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ terrain_slope           : num [1:234] 4.82 24.97 18.94 4.82 31.41 ...
##   ..- attr(*, "label")= chr "Avg terrain slope - pop weighted"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ soil_depth              : num [1:234] 0.893 0.262 0.985 0.893 0.781 ...
##   ..- attr(*, "label")= chr "Frac land: no depth constraints (plate 21: 40 cutoff) -
pop weighted"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ avg_temperature         : num [1:234] 26.35 8.63 20.45 26.35 11.06 ...
##   ..- attr(*, "label")= chr "Avg temperature over the year (C) - pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ avg_precipitation       : num [1:234] 96.4 23.2 95.5 96.4 79.3 ...
##   ..- attr(*, "label")= chr "Avg monthly precipiation (mm) - pop weighted"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ country                 : chr [1:234] "Aruba" "Afghanistan" "Angola" "" ...
##   ..- attr(*, "label")= chr "country name"
##   ..- attr(*, "format.stata")= chr "%46s"
##  $ communist_dummy         : num [1:234] 0 1 1 NA 1 NA 0 0 0 1 ...
##   ..- attr(*, "label")= chr "Communism indicator variable"
##   ..- attr(*, "format.stata")= chr "%8.0g"
##  $ continent               : chr [1:234] "North America" "Asia" "Africa" "North Amer
ica" ...
##   ..- attr(*, "label")= chr "Continent"
##   ..- attr(*, "format.stata")= chr "%13s"
##  $ rugged                  : num [1:234] 0.4624 2.5181 0.8582 0.0126 3.4271 ...
##   ..- attr(*, "label")= chr "Ruggedness (Terrain Ruggedness Index, 100 m.)"
##   ..- attr(*, "format.stata")= chr "%12.0g"
##  $ european_descent        : num [1:234] NA 0 2 NA 100 ...
##   ..- attr(*, "label")= chr "% European descent"
##   ..- attr(*, "format.stata")= chr "%9.3f"
##  $ cath00                  : num [1:234] NA 0 0.621 NA 0.168 ...
##   ..- attr(*, "label")= chr "CATH00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ prot00                  : num [1:234] NA 0 0.15 NA 0.006 ...
##   ..- attr(*, "label")= chr "PROT00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ othchrist00             : num [1:234] NA 0 0.169 NA 0.02 ...
##   ..- attr(*, "label")= chr "OTHCHRIST00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ muslim00                : num [1:234] NA 0.981 0 NA 0.388 ...
##   ..- attr(*, "label")= chr "MUSLIM00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ hindu00                 : num [1:234] NA 0.004 0 NA 0 ...
##   ..- attr(*, "label")= chr "HINDU00"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ trade_GDP               : num [1:234] 54.2 NA 120.1 NA 36.6 ...
##   ..- attr(*, "format.stata")= chr "%8.0g"
```

```
##  $ years_civil_conflict     : num [1:234] NA 21 31 NA 0 0 NA 0 16 3 ...
##   ..- attr(*, "label")= chr "Years of civil conflict, 1800-2007 - from COW"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ years_interstate_conflict: num [1:234] NA 1 2 NA 0 0 NA 1 9 2 ...
##   ..- attr(*, "label")= chr "Years of interstate conflict, 1800-2007 - from COW"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ agr_va_gdp2000           : num [1:234] NA NA 5.66 NA 29.13 ...
##   ..- attr(*, "label")= chr "Value Added Agriculture/GDP in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ man_va_gdp2000           : num [1:234] NA NA 2.89 NA 11.44 ...
##   ..- attr(*, "label")= chr "Value Added Manufacturing/GDP in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ serv_va_gdp2000          : num [1:234] NA NA 22.2 NA 51.9 ...
##   ..- attr(*, "label")= chr "Value Added in Service/GDP in 2000"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ polity2_2000             : num [1:234] NA -7 -3 NA 5 NA NA -8 8 5 ...
##   ..- attr(*, "label")= chr "Polity 2 measure taken from the Polity IV dataset"
##   ..- attr(*, "format.stata")= chr "%8.0g"
##  $ plow_negative_crops      : num [1:234] 0.632 0.22 0.751 0.632 0.333 ...
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ plow_positive_crops      : num [1:234] 0 0.816 0.286 0 0.974 ...
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ region                   : chr [1:234] "North America" "Asia" "Sub-Saharan Africa"
"North America" ...
##   ..- attr(*, "label")= chr "continent with Sub-Saharan African Dummy"
##   ..- attr(*, "format.stata")= chr "%18s"
##  $ intensity_agriculture    : num [1:234] 0 0.7005 0.0135 0 1 ...
##   ..- attr(*, "label")= chr "intensity of agriculture"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ abs_inherit              : num [1:234] 0 0.000142 0.069202 0 0 ...
##   ..- attr(*, "label")= chr "absence of inheritance"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ matrilocal               : num [1:234] 0 0 0 0 0 ...
##   ..- attr(*, "label")= chr "matrilocal societies"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ patrilocal               : num [1:234] 1 1 0.648 1 0.987 ...
##   ..- attr(*, "label")= chr "patrilocal societies"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ nuclear_fam              : num [1:234] 0 0.000618 0.115161 0 0.015537 ...
##   ..- attr(*, "label")= chr "nuclear family"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ extended_fam             : num [1:234] 1 0.9994 0.0622 1 0.9845 ...
##   ..- attr(*, "label")= chr "extended family"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ oil_pc                   : num [1:234] NA 0 0.0719 NA 0 ...
##   ..- attr(*, "label")= chr "oil production/GDP"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ terrslope2               : num [1:234] 23.2 623.6 358.6 23.2 986.3 ...
##   ..- attr(*, "label")= chr "terrain slope squared"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ soil2                    : num [1:234] 0.7969 0.0684 0.9704 0.7969 0.6098 ...
##   ..- attr(*, "label")= chr "Soil depth squared"
```

```
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ avg_temp2               : num [1:234] 694.4 74.5 418 694.4 122.2 ...
##   ..- attr(*, "label")= chr "average temperature squared"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ avg_precip2             : num [1:234] 9292 538 9127 9292 6294 ...
##   ..- attr(*, "label")= chr "average precipitation squared"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ slope_soil              : num [1:234] 4.3 6.53 18.65 4.3 24.53 ...
##   ..- attr(*, "label")= chr "interaction slope*soil depth"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ slope_temp              : num [1:234] 127 216 387 127 347 ...
##   ..- attr(*, "label")= chr "interaction slope*temperature"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ slope_precip            : num [1:234] 465 579 1809 465 2492 ...
##   ..- attr(*, "label")= chr "interaction slope*precipitation"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ soil_temp               : num [1:234] 23.52 2.26 20.14 23.52 8.63 ...
##   ..- attr(*, "label")= chr "interaction soil*temperature"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ soil_precip             : num [1:234] 86.05 6.07 94.11 86.05 61.95 ...
##   ..- attr(*, "label")= chr "interaction soil*precipitation"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  $ temp_precip             : num [1:234] 2540 200 1953 2540 877 ...
##   ..- attr(*, "label")= chr "interaction temperature*precipitation"
##   ..- attr(*, "format.stata")= chr "%9.0g"
##  - attr(*, "notes")= chr [1:18] "this file has been compiled by Antonio Spilimbergo w
ith the help of Rafael Romeu and Adrian De La Garza. It is "| __truncated__ "This data s
et contains the data necessary to replicate Table 1 in James D. Fearon and David D. Lait
in, Ethnicit"| __truncated__ "gdpppp95 data for Bosnia/Herzegovina; Bhutan; Brunei; */"
"this file contains a subset of the COW Inter-State War Data, 1816-1997 The files were o
btained from the site ht"| __truncated__ ...
```

```
#  first regression to estimate residuals for the mediator (ln_income)
mediator_model <- lm(ln_income ~ plow + agricultural_suitability + tropical_climate +
                     large_animals + rugged + polity2_2000, data = plowUse_1_,
                     na.action = na.exclude)

# Get residuals of mediator
plowUse_1_$resid_mediator <- residuals(mediator_model)

# second regression to estimate residuals for the outcome (women in politics)
outcome_model <- lm(women_politics ~ plow + agricultural_suitability + tropical_climate
+
                    large_animals + rugged + polity2_2000, data = plowUse_1_,
                    na.action = na.exclude)

# residuals of the outcome
plowUse_1_$resid_outcome <- residuals(outcome_model)

# Define a function to compute the indirect and direct effects for a bootstrap sample
compute_effects <- function(data, indices) {
  # Resample the data
  resampled_data <- data[indices, ]

  # Re-run the regression for the mediator and outcome
  mediator_model_boot <- lm(ln_income ~ plow + agricultural_suitability + tropical_clima
te +
                            large_animals + rugged + polity2_2000, data = resampled_dat
a,
                            na.action = na.exclude)
  resampled_data$resid_mediator <- residuals(mediator_model_boot)

  outcome_model_boot <- lm(women_politics ~ plow + agricultural_suitability + tropical_c
limate +
                           large_animals + rugged + polity2_2000, data = resampled_data,
                           na.action = na.exclude)
  resampled_data$resid_outcome <- residuals(outcome_model_boot)

  # Estimate the indirect effect: Regression of residualized outcome on residualized med
iator
  indirect_model_boot <- lm(resid_outcome ~ resid_mediator + plow + resid_mediator:plow,
data = resampled_data)
  indirect_effect <- coef(indirect_model_boot)["resid_mediator"]

  # Estimate the direct effect: Regression of residualized outcome on exposure (plow)
  direct_model_boot <- lm(resid_outcome ~ plow + resid_mediator + resid_mediator:plow, d
ata = resampled_data)
  direct_effect <- coef(direct_model_boot)["plow"]

  return(c(direct_effect, indirect_effect))
}

# Perform the bootstrap with 1000 replications
set.seed(123)  # For reproducibility
```

```
bootstrap_results <- boot(data = plowUse_1_, statistic = compute_effects, R = 1000)

# View the bootstrap results
bootstrap_results
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = plowUse_1_, statistic = compute_effects, R = 1000)
##
##
## Bootstrap Statistics :
##       original      bias    std. error
## t1* -0.6186369  0.02234738   0.4999843
## t2*  2.9515704 -0.06749184   0.9191308
```

```
# Calculate 95% confidence intervals for the direct and indirect effects
direct_ci <- quantile(bootstrap_results$t[, 1], probs = c(0.025, 0.975))
indirect_ci <- quantile(bootstrap_results$t[, 2], probs = c(0.025, 0.975))

# CI
cat("95% Confidence Interval for Direct Effect:", direct_ci, "\n")
```

```
## 95% Confidence Interval for Direct Effect: -1.614154 0.2533579
```

```
cat("95% Confidence Interval for Indirect Effect:", indirect_ci, "\n")
```

```
## 95% Confidence Interval for Indirect Effect: 1.277137 4.837867
```

```
# Calculate the p-value for the test of the null hypothesis that the effects are zero
# null hypothesis: Direct effect = 0
p_value_direct <- mean(bootstrap_results$t[, 1] >= 0 | bootstrap_results$t[, 1] <= 0)

# null hypothesis: Indirect effect = 0
p_value_indirect <- mean(bootstrap_results$t[, 2] >= 0 | bootstrap_results$t[, 2] <= 0)


cat("p-value for Direct Effect = 0:", p_value_direct, "\n")
```

```
## p-value for Direct Effect = 0: 1
```

```
cat("p-value for Indirect Effect = 0:", p_value_indirect, "\n")
```

```
## p-value for Indirect Effect = 0: 1
```