Mediation Analysis in R

Notes:

(1) This PDF is part of YouTube tutorials. This PDF is for individual, personal usage only.

Method 1: https://youtu.be/p2XbncjiA6k

Method 2: https://youtu.be/MIIF-ICF52Y

(2) The author accepts no responsibility for the topicality, correctness, completeness or quality of the information provided.

Method 1

Baron & Kenny (1986) 3 regressions

```
# read data
Mediation_data <-
  read.csv('https://raw.githubusercontent.com/TidyPython/Mediation_analysis/main/mediation_hypothetical_

# show the head of the data
head(Mediation_data)
```

```
##          X M  Y
## 1 65.00000 3  6
## 2 65.35354 2  6
## 3 65.70707 2  7
## 4 66.06061 2  5
## 5 66.41414 2  5
## 6 66.76768 4 14
```

```
# c path linear regression
c_path = lm(Y~X, data=Mediation_data)
summary(c_path)
```

```
##
## Call:
## lm(formula = Y ~ X, data = Mediation_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.0137 -2.5361  0.1512  1.9639  7.7747
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.92914    2.84540   0.678  0.49938
## X            0.09977    0.03423   2.915  0.00441 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.493 on 98 degrees of freedom
## Multiple R-squared:  0.07978,    Adjusted R-squared:  0.07039
## F-statistic: 8.496 on 1 and 98 DF,  p-value: 0.004411
```

```
# a path linear regression
a_path = lm(M~X, data=Mediation_data)
summary(a_path)
```

```
##
## Call:
## lm(formula = M ~ X, data = Mediation_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.00796 -0.83457  0.01555  0.75988  2.92151
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.64685    0.92465   0.700  0.48586
## X            0.03325    0.01112   2.989  0.00353 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.135 on 98 degrees of freedom
## Multiple R-squared:  0.08356,    Adjusted R-squared:  0.07421
## F-statistic: 8.936 on 1 and 98 DF,  p-value: 0.003535
```

```
# b path linear regression
b_path = lm(Y~M+X, data=Mediation_data)
summary(b_path)
```

```
##
## Call:
## lm(formula = Y ~ M + X, data = Mediation_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.05014 -0.88156  0.02536  0.92353  2.12187
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.032734   0.871195   0.038    0.970
## M           2.931769   0.094939  30.881   <2e-16 ***
## X           0.002286   0.010920   0.209    0.835
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.067 on 97 degrees of freedom
## Multiple R-squared:  0.915,  Adjusted R-squared:  0.9133
## F-statistic: 522.3 on 2 and 97 DF,  p-value: < 2.2e-16
```

```r
# calculate the indirect effect
indirect_effect<-a_path$coefficients[2]*b_path$coefficients[2]
print(indirect_effect)
```

```
##          X
## 0.09748237
```

R package of "mediation"

```r
set.seed(123)

#install.packages("mediation") #in case not installed "mediation" yet
library(mediation)

results = mediate(a_path, b_path ,sims=5000, treat='X', mediator='M', boot=T)
summary(results)
```

```
##
## Causal Mediation Analysis
##
## Nonparametric Bootstrap Confidence Intervals with the Percentile Method
##
##                 Estimate 95% CI Lower 95% CI Upper p-value
## ACME             0.09748      0.03819         0.16  0.0028 **
## ADE              0.00229     -0.02167         0.03  0.8424
## Total Effect     0.09977      0.03157         0.16  0.0052 **
## Prop. Mediated  0.97709      0.75227         1.42  0.0032 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 100
##
##
## Simulations: 5000
```

ACME: Average Causal Mediation Effects = Indirect Effect = a*b

ADE: Average Direct Effects = c'

Total Effect: c

Prop. Mediated: ACME/Total Effect

Method 2 - Mediation Analysis in R from Scratch

```r
library(boot)
set.seed(123)


Mediation_function<-function(data_used,i)
{
 # Sample a data
   data_temp=data_used[i,]
 # a path
    result_a<-lm(M~X, data = data_temp)
    a<-result_a$coefficients[2]

 # b path
    result_b<-lm(Y~M+X, data = data_temp)
    b<-result_b$coefficients[2]

 #calculating the indirect effect
    indirect_effect<-a*b
 return(indirect_effect)
}


# use boot() to do bootstrapping mediation analysis
boot_mediation <- boot(Mediation_data, Mediation_function, R=5000)
boot_mediation
```

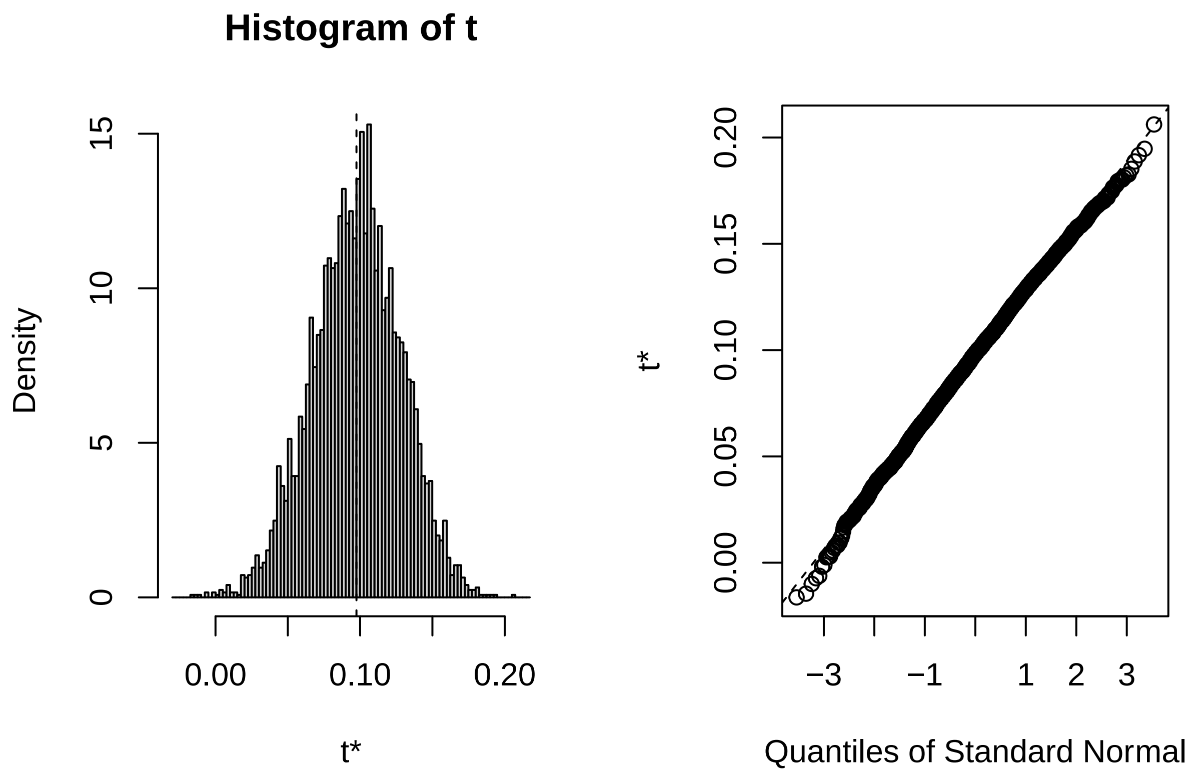
```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Mediation_data, statistic = Mediation_function, R = 5000)
##
##
## Bootstrap Statistics :
##       original        bias    std. error
## t1* 0.09748237 0.0001266305  0.03044034
```

Confidence Intervals

```r
# print out confidence intervals
boot.ci(boot.out = boot_mediation, type = c("norm", "perc"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 5000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_mediation, type = c("norm", "perc"))
##
## Intervals :
## Level      Normal             Percentile
## 95%   ( 0.0377,  0.1570 )   ( 0.0379,  0.1559 )
## Calculations and Intervals on Original Scale
```

```r
# plot the 5000 indirect effects
plot(boot_mediation)
```



```r
head(boot_mediation$t)
```

```
##            [,1]
## [1,] 0.11422596
## [2,] 0.14378983
## [3,] 0.05868453
## [4,] 0.09819629
## [5,] 0.07659858
## [6,] 0.08390458
```

Understand SE and SD in boostrapping

```r
# calculate the standard deviation
sd(boot_mediation$t)
```

```
## [1] 0.03044034
```

```r
# print out the bootstrapping output again
boot_mediation
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Mediation_data, statistic = Mediation_function, R = 5000)
##
##
## Bootstrap Statistics :
##       original       bias    std. error
## t1* 0.09748237 0.0001266305  0.03044034
```

Understand "bias" in the output

The difference between the mean of the bootstrap estimates and the original sample estimate is the bias.

```
# the mean of 5000 indirect effect
print(mean(boot_mediation$t))
```

```
## [1] 0.097609
```

```
#the original sample estimate
print(boot_mediation$t0)
```

```
##          X
## 0.09748237
```

```
# the difference = the bias
bias=mean(boot_mediation$t)-boot_mediation$t0
print(bias)
```

```
##            X
## 0.0001266305
```

A better understanding of confidence interval

```r
# calculate the confidence interval from scratch
print(boot_mediation$t0-bias + c(-1, 1) * 1.96 * sd(boot_mediation$t))
```

```
## [1] 0.03769268 0.15701880
```

```r
# print out confidence interval based on normal distribution
boot.ci(boot.out = boot_mediation, type = c("norm"))
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 5000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot_mediation, type = c("norm"))
##
## Intervals :
## Level      Normal
## 95%   ( 0.0377,  0.1570 )
## Calculations and Intervals on Original Scale
```