Mediation Effects: Comparing different methods

Seunghun Lee

r sys.Date()

Import necessary libraries

```
libraries <- c("dplyr", "ggplot2", "kableExtra", "tidyverse", "lavaan", "MBESS")
lapply(libraries, require, character.only = TRUE)
```

```
## Loading required package: 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

## Loading required package: ggplot2

## Loading required package: kableExtra

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':
##
##     group_rows

## Loading required package: tidyverse

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.0.4     v purrr   0.3.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter()          masks stats::filter()
## x kableExtra::group_rows() masks dplyr::group_rows()
## x dplyr::lag()             masks stats::lag()

## Loading required package: lavaan

## This is lavaan 0.6-7

## lavaan is BETA software! Please report any bugs.

## Loading required package: MBESS

##
## Attaching package: 'MBESS'

## The following object is masked from 'package:lavaan':
##
##     cor2cov

## [[1]]
## [1] TRUE
##
## [[2]]
## [1] TRUE
##
## [[3]]
## [1] TRUE
##
## [[4]]
## [1] TRUE
##
## [[5]]
## [1] TRUE
##
## [[6]]
## [1] TRUE
```

Import the data

```
dat <- read.csv("./data/student-mat.csv", sep = ';')
```

Descriptive statistics

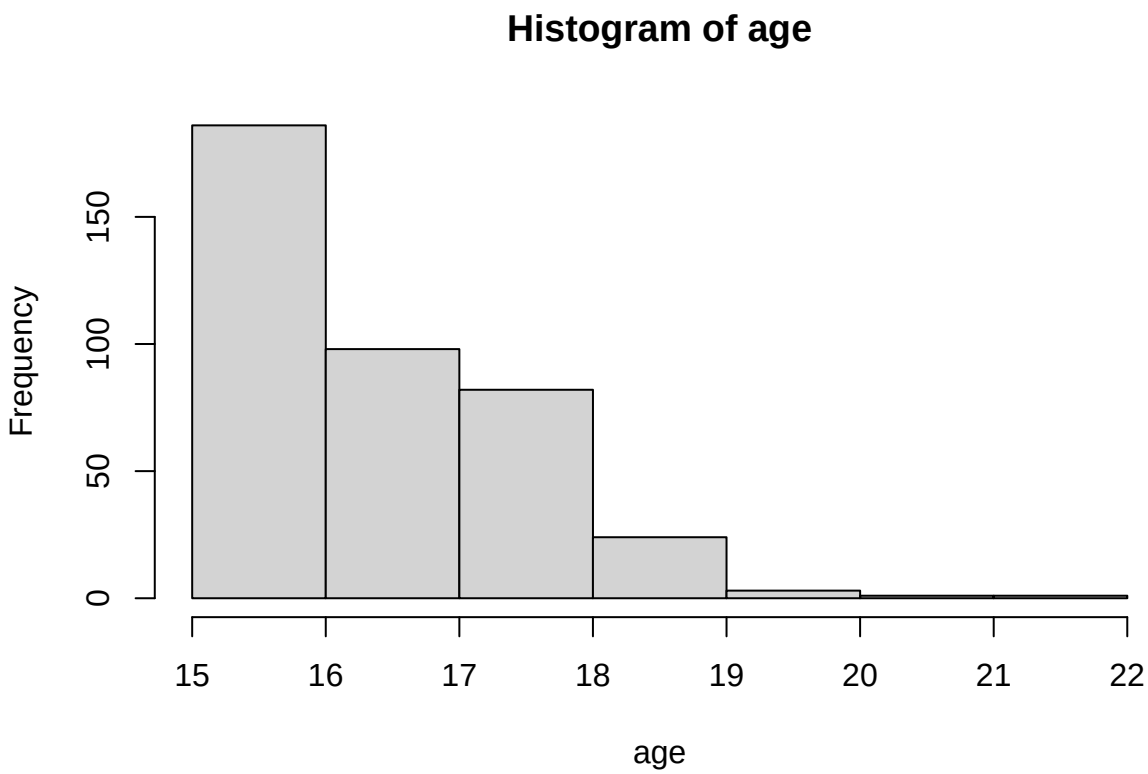
```
dat %>%
  is.na %>%
  apply(2, sum) # there is no missing value
```

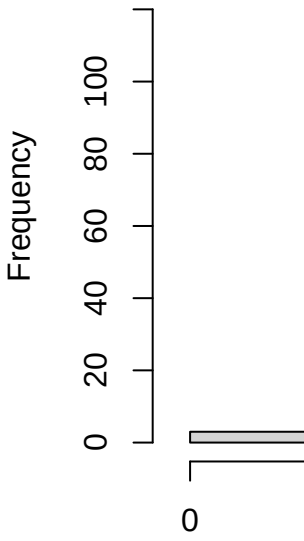
```
##     school        sex        age    address    famsize    Pstatus       Medu
##          0          0          0          0          0          0          0
##       Fedu       Mjob       Fjob     reason   guardian traveltime  studytime
##          0          0          0          0          0          0          0
##   failures  schoolsup     famsup       paid activities    nursery     higher
##          0          0          0          0          0          0          0
##   internet   romantic     famrel   freetime      goout       Dalc       Walc
##          0          0          0          0          0          0          0
##     health   absences         G1         G2         G3
##          0          0          0          0          0
```

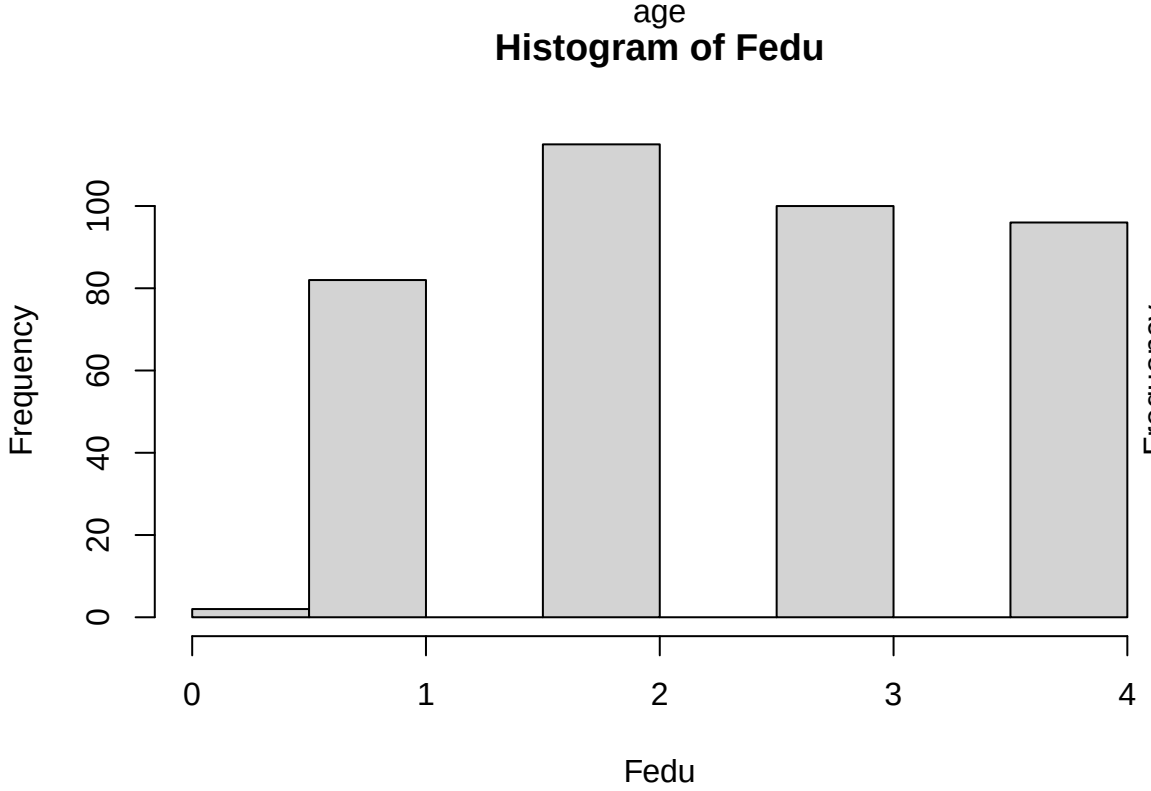
```
test <- dat %>%
  mutate(across(where(is.character), as.factor))
```

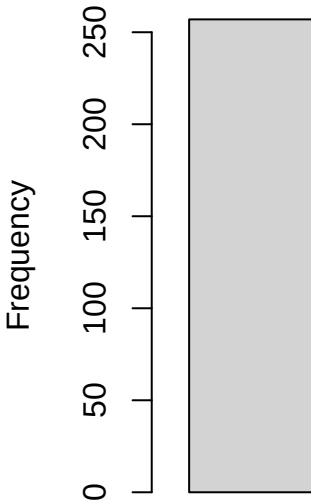
Assumptions

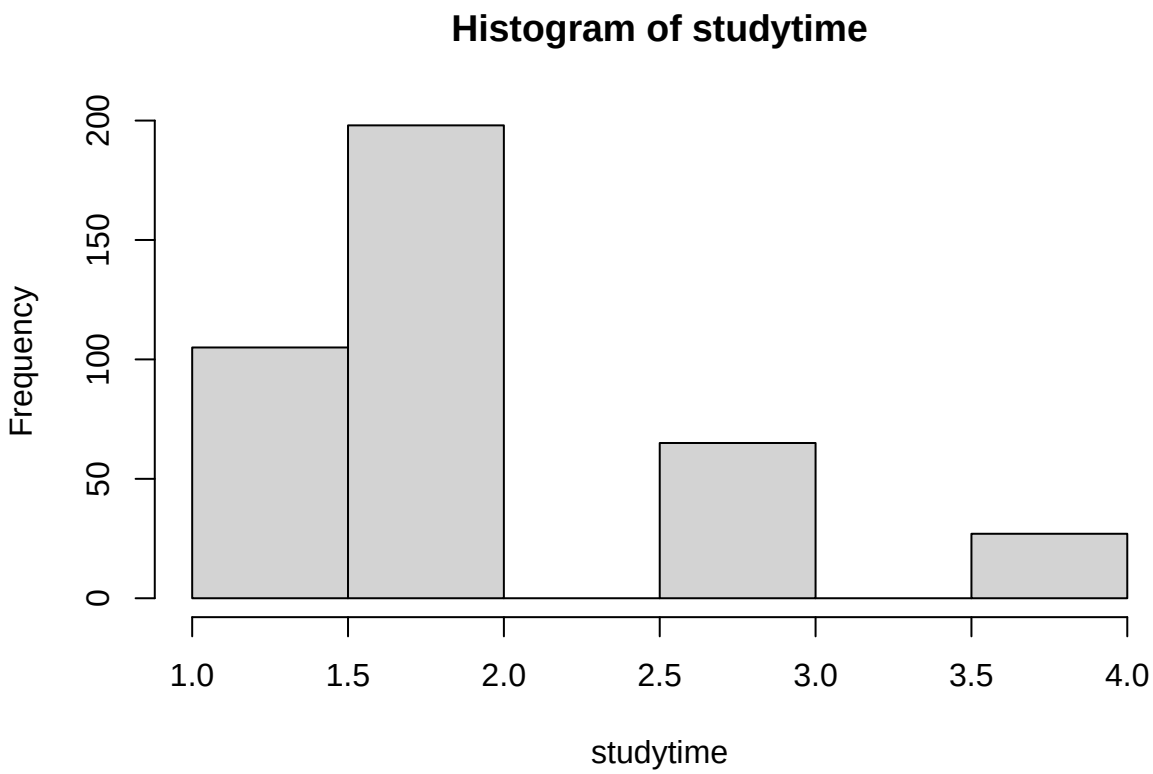
```
summary_for_all <- function(data){
  for(i in 1:ncol(data)){
    if(class(data[[i]]) == "integer"){
      hist(data[[i]], breaks = 7,
           main = paste("Histogram of", names(test)[i]),
           xlab = names(test[i]),
           ylab = "Frequency"
          )} else if (class(data[[i]]) == "factor"){
            kbl(summary(data[[i]]), caption = paste("Table of", names(data[i]))) %>%
                kable_classic(full_width = FALSE, html_font = "Cambria") %>%
                kable_styling(font_size = 12, position = "center")
          }
  }
}

summary_for_all(test)
```

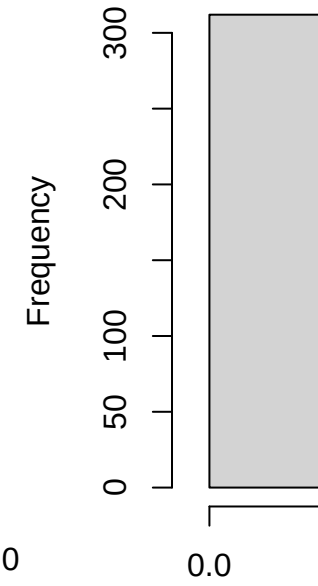

Histogram of age
Frequency
0
50
100
150
15
16
17
18
19
20
21
22
age

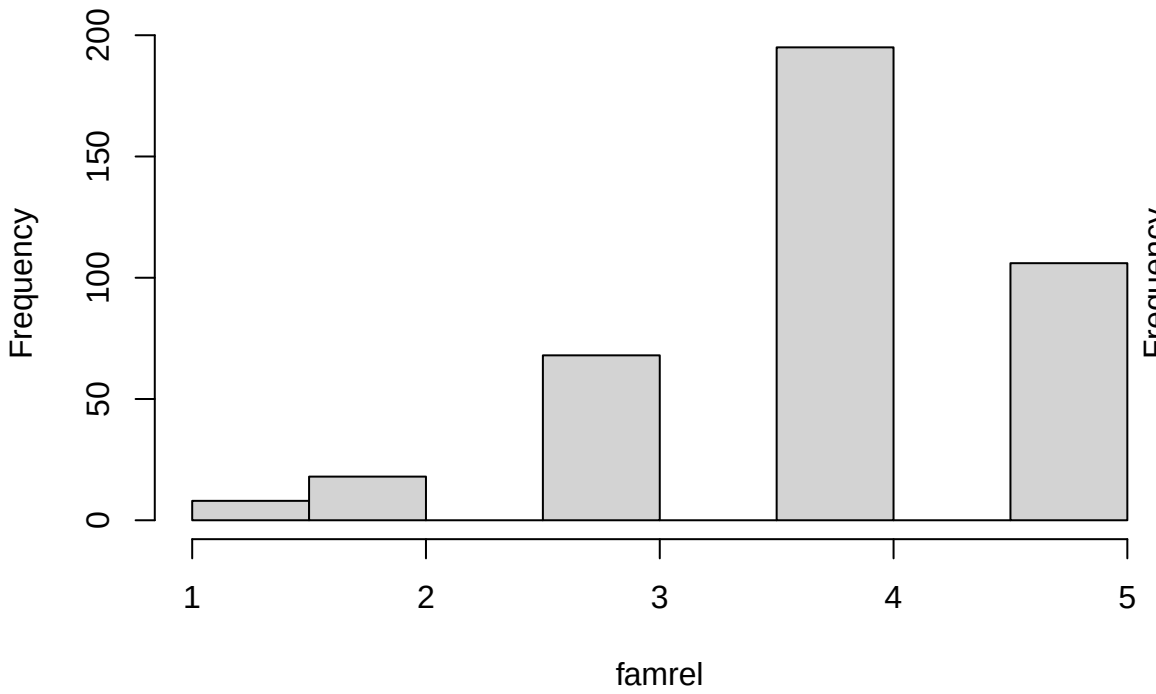

Frequency
0
20
40
60
80
100
0

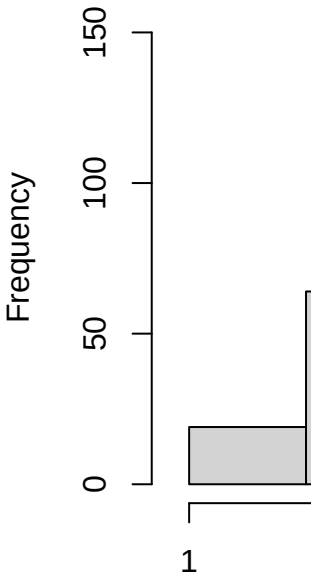

Histogram of Fedu
Frequency
0
20
40
60
80
100
0
1
2
3
4
Fedu

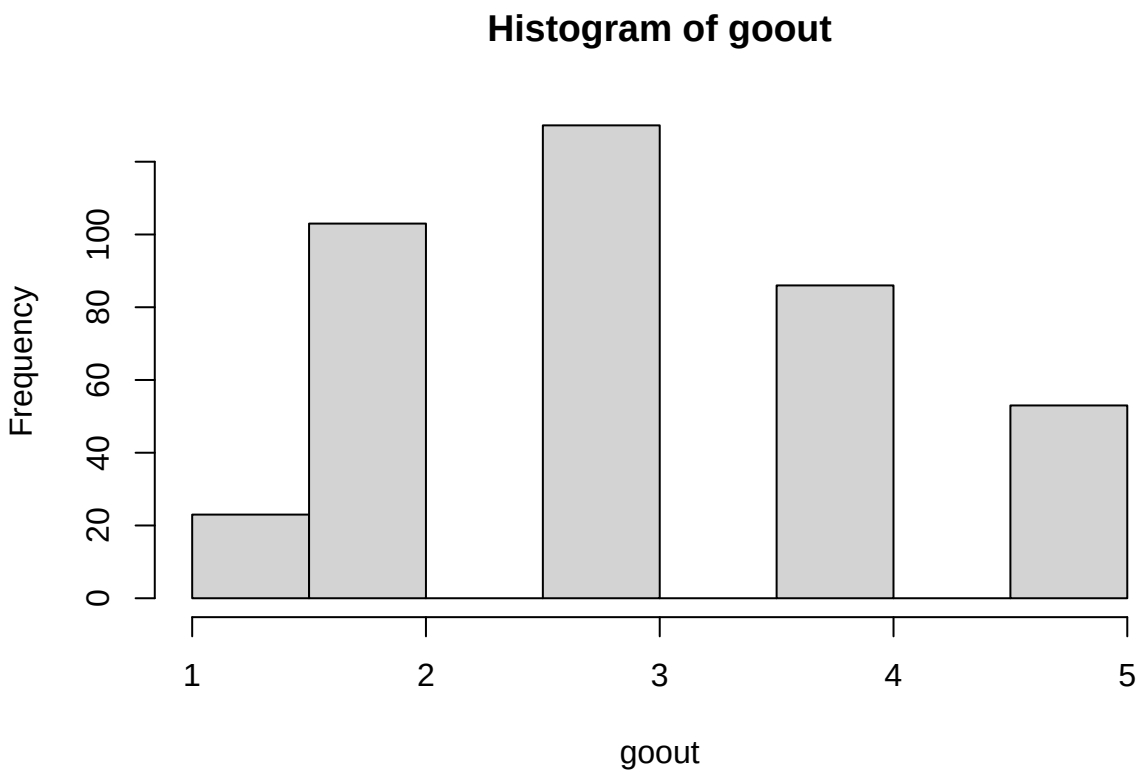

Frequency
0
50
100
150
200
250
1.0

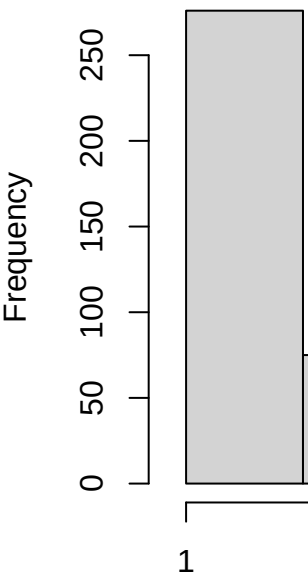

Histogram of studytime
Frequency
0
50
100
150
200
1.0
1.5
2.0
2.5
3.0
3.5
4.0
studytime

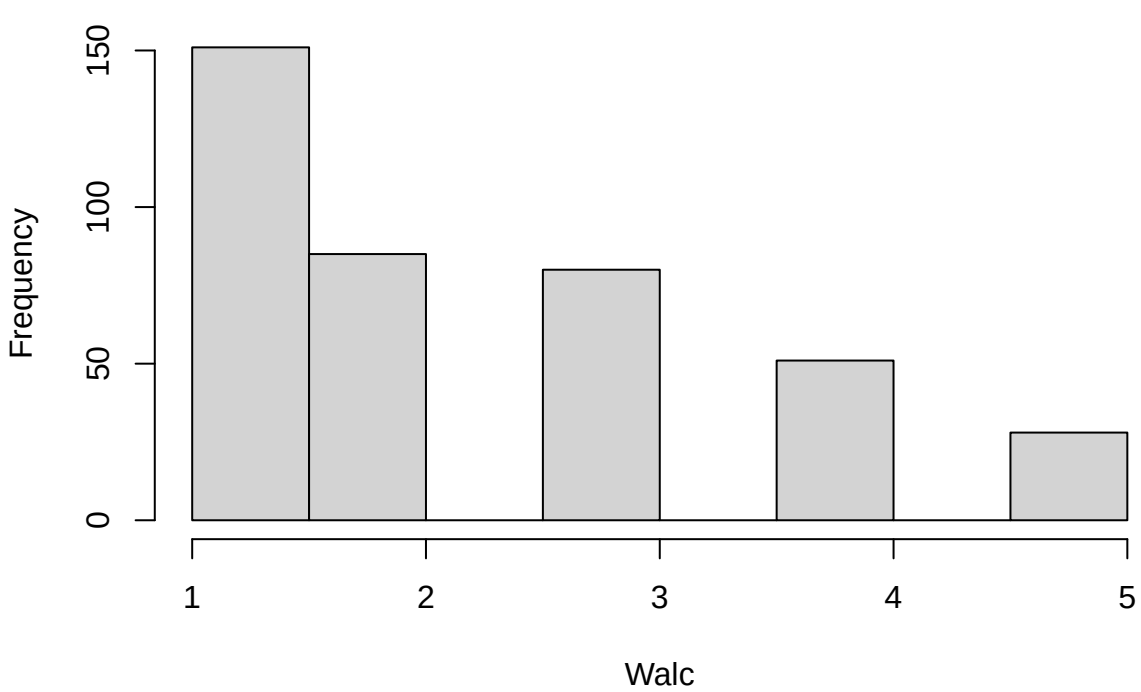

Frequency
0
50
100
200
300
0.0

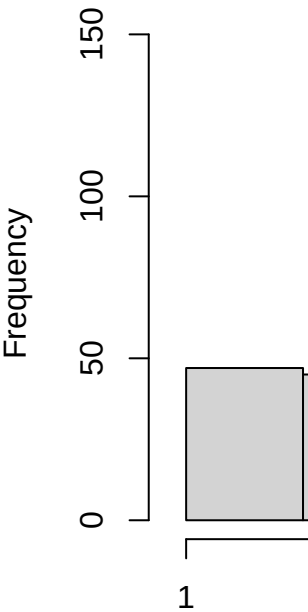

Histogram of famrel
Frequency
0
50
100
150
200
1
2
3
4
5
famrel

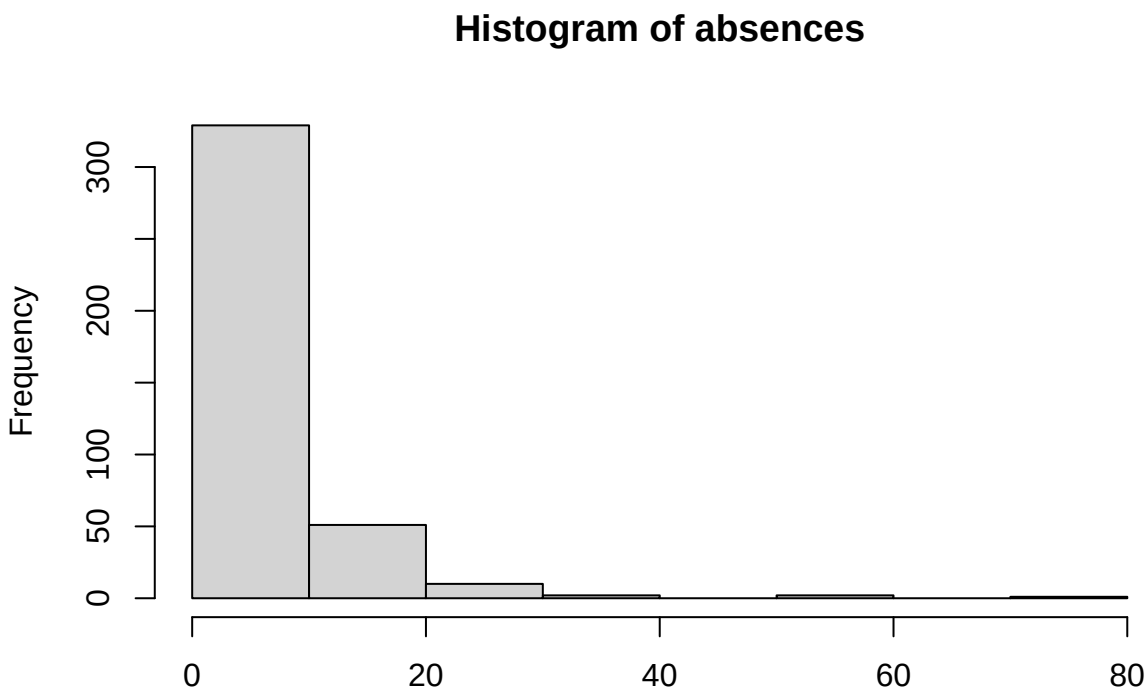

Frequency
0
50
100
150
1

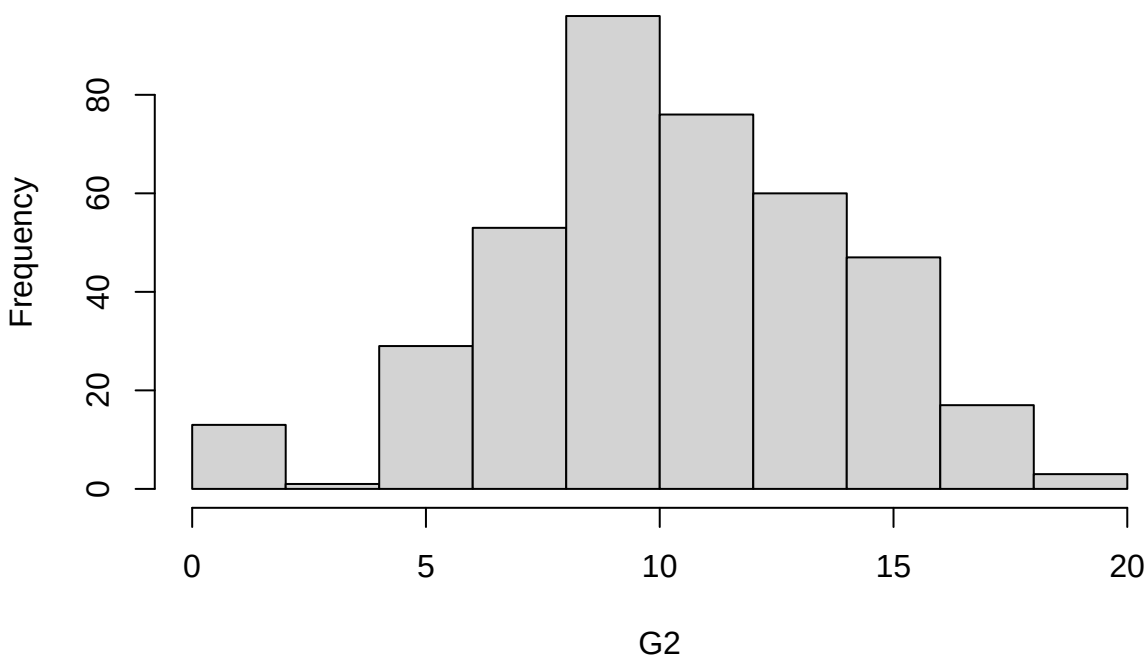

Histogram of goout
Frequency
0
20
40
60
80
100
1
2
3
4
5
goout


Frequency
0
50
100
150
200
250
1


Histogram of Walc
Frequency
0
50
100
150
1
2
3
4
5
Walc


Frequency
0
50
100
150
1






```
my_test <- test %>%
  select(G1, G2, G3) %>%
  as.data.frame()
```

SEM

```
testmod <- "
# Path c' (direct effect)
G3 ~ c*G1

# Path a
G2 ~ a*G1

# Path b
G3 ~ b*G2

# Indreict effect (a*b): Sobel test (Delta Method)
ab := a*b
"

# Fit estimate
fitmod <- sem(testmod, data = my_test)
```

```
summary(fitmod, fit.measures = TRUE, rsquare = TRUE)
```

```
## lavaan 0.6-7 ended normally after 18 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          5
##
##   Number of observations                           395
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Model Test Baseline Model:
##
##   Test statistic                              1193.651
##   Degrees of freedom                                 3
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    1.000
##   Tucker-Lewis Index (TLI)                       1.000
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1647.634
##   Loglikelihood unrestricted model (H1)      -1647.634
##
##   Akaike (AIC)                                3305.268
##   Bayesian (BIC)                              3325.162
```

```
##   Sample-size adjusted Bayesian (BIC)         3309.297
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.000
##   90 Percent confidence interval - lower         0.000
##   90 Percent confidence interval - upper         0.000
##   P-value RMSEA <= 0.05                             NA
##
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Standard
##   Information                                 Expected
##   Information saturated (h1) model          Structured
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   G3 ~
##     G1         (c)    0.153    0.056    2.739    0.006
##   G2 ~
##     G1         (a)    0.966    0.030   32.360    0.000
##   G3 ~
##     G2         (b)    0.987    0.049   19.985    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .G3                3.723    0.265   14.053    0.000
##    .G2                3.866    0.275   14.053    0.000
##
## R-Square:
##                    Estimate
##     G3                 0.822
##     G2                 0.726
##
## Defined Parameters:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     ab                 0.953    0.056   17.003    0.000
```

Resampling

```
testmod <- "
# Path c' (direct effect)
G3 ~ c*G1

# Path a
G2 ~ a*G1
```

```
# Path b
G3 ~ b*G2

# Indreict effect (a*b): Sobel test (Delta Method)
ab := a*b
"

set.seed(1)
testmod2 <- sem(testmod, data = my_test,
                se = "bootstrap", bootstrap = 100)
```

```
summary(testmod2,  fit.measures = TRUE, rsquare = TRUE)
```

```
## lavaan 0.6-7 ended normally after 18 iterations
##
##   Estimator                                         ML
##   Optimization method                           NLMINB
##   Number of free parameters                          5
##
##   Number of observations                           395
##
## Model Test User Model:
##
##   Test statistic                                 0.000
##   Degrees of freedom                                 0
##
## Model Test Baseline Model:
##
##   Test statistic                              1193.651
##   Degrees of freedom                                 3
##   P-value                                        0.000
##
## User Model versus Baseline Model:
##
##   Comparative Fit Index (CFI)                    1.000
##   Tucker-Lewis Index (TLI)                       1.000
##
## Loglikelihood and Information Criteria:
##
##   Loglikelihood user model (H0)              -1647.634
##   Loglikelihood unrestricted model (H1)      -1647.634
##
##   Akaike (AIC)                                3305.268
##   Bayesian (BIC)                              3325.162
##   Sample-size adjusted Bayesian (BIC)         3309.297
##
## Root Mean Square Error of Approximation:
##
##   RMSEA                                          0.000
##   90 Percent confidence interval - lower         0.000
##   90 Percent confidence interval - upper         0.000
##   P-value RMSEA <= 0.05                             NA
##
```

```
## Standardized Root Mean Square Residual:
##
##   SRMR                                           0.000
##
## Parameter Estimates:
##
##   Standard errors                             Bootstrap
##   Number of requested bootstrap draws              100
##   Number of successful bootstrap draws             100
##
## Regressions:
##                    Estimate  Std.Err  z-value  P(>|z|)
##   G3 ~
##     G1         (c)    0.153    0.044    3.521    0.000
##   G2 ~
##     G1         (a)    0.966    0.030   32.326    0.000
##   G3 ~
##     G2         (b)    0.987    0.032   30.830    0.000
##
## Variances:
##                    Estimate  Std.Err  z-value  P(>|z|)
##    .G3                3.723    0.527    7.069    0.000
##    .G2                3.866    0.598    6.459    0.000
##
## R-Square:
##                    Estimate
##     G3                0.822
##     G2                0.726
##
## Defined Parameters:
##                    Estimate  Std.Err  z-value  P(>|z|)
##     ab                0.953    0.036   26.452    0.000
```

```
parameterEstimates(testmod2, ci = TRUE, level = 0.95, boot.ci.type = "perc")
```

```
##   lhs op rhs label    est    se      z pvalue ci.lower ci.upper
## 1  G3  ~  G1     c  0.153 0.044  3.521      0    0.062    0.248
## 2  G2  ~  G1     a  0.966 0.030 32.326      0    0.899    1.019
## 3  G3  ~  G2     b  0.987 0.032 30.830      0    0.929    1.066
## 4  G3 ~~  G3       3.723 0.527  7.069      0    2.714    4.893
## 5  G2 ~~  G2       3.866 0.598  6.459      0    2.842    5.441
## 6  G1 ~~  G1      10.989 0.000     NA     NA   10.989   10.989
## 7  ab := a*b    ab  0.953 0.036 26.452      0    0.870    1.021
```

```
mediation(x = my_test$G1,
          mediator = my_test$G2,
          dv = my_test$G3,
          bootstrap = TRUE, B = 100)
```

```
## [1] "Bootstrap resampling has begun. This process may take a considerable amount of time if the numbe
```

```
##                                       Estimate CI.Lower_Percentile
```

```
## Indirect.Effect                          0.9529875           0.8755722
## Indirect.Effect.Partially.Standardized 0.2080104           0.1893937
## Index.of.Mediation                       0.6904269           0.6341650
## R2_4.5                                   0.6389737           0.5805129
## R2_4.6                                   0.3650581           0.2906729
## R2_4.7                                   0.4440214           0.3808944
## Ratio.of.Indirect.to.Total.Effect        0.8614529           0.7860602
## Ratio.of.Indirect.to.Direct.Effect       6.2177615           3.6742202
## Success.of.Surrogate.Endpoint            1.1455842           1.0881683
## Residual.Based_Gamma                     0.5201373           0.4703380
## Residual.Based.Standardized_gamma        0.5235306           0.4757046
## SOS                                      0.9947425           0.9874601
##                                          CI.Upper_Percentile CI.Lower_BCa
## Indirect.Effect                                    1.0287243           NA
## Indirect.Effect.Partially.Standardized             0.2315293           NA
## Index.of.Mediation                                 0.7586671           NA
## R2_4.5                                             0.6985184           NA
## R2_4.6                                             0.4623102           NA
## R2_4.7                                             0.5252615           NA
## Ratio.of.Indirect.to.Total.Effect                  0.9379093           NA
## Ratio.of.Indirect.to.Direct.Effect                15.1982279           NA
## Success.of.Surrogate.Endpoint                      1.2175729           NA
## Residual.Based_Gamma                               0.5634116           NA
## Residual.Based.Standardized_gamma                  0.5659546           NA
## SOS                                                0.9989668           NA
##                                          CI.Upper_BCa
## Indirect.Effect                                    NA
## Indirect.Effect.Partially.Standardized             NA
## Index.of.Mediation                                 NA
## R2_4.5                                             NA
## R2_4.6                                             NA
## R2_4.7                                             NA
## Ratio.of.Indirect.to.Total.Effect                  NA
## Ratio.of.Indirect.to.Direct.Effect                 NA
## Success.of.Surrogate.Endpoint                      NA
## Residual.Based_Gamma                               NA
## Residual.Based.Standardized_gamma                  NA
## SOS                                                NA
```