Issues on regression*

William Revelle
Northwestern University

October 31, 2020

Abstract

Contents

1 Overview of this and related documents

To do basic and advanced personality and psychological research using R is not as complicated as some think. This is one of a set of "How To" to do various things using R (R Core Team, 2020), particularly using the *psych* (Revelle, 2020) package.

The current list of How To's includes:

1. An introduction (vignette) of the *psych* package
2. An overview (vignette) of the *psych* package
3. Installing R and some useful packages
4. Using R and the *psych* package to find $omega_h$ and ω_t.
5. Using R and the *psych* for factor analysis and principal components analysis.
6. Using the `scoreItems` function to find scale scores and scale statistics (this document).
7. Using `mediate` and `setCor` to do mediation, moderation and regression analysis

R code

```
> library(psych)
> library(psychTools)
```

*Part of a set of tutorials for the psych package.

This HowTo is partly a debugging tool for the `setCor` and `mediate` functions. It is also meant as a short tutorial on regression with cross product terms.

2 Attitude data set

The `attitude` data set is built into core-R and has 7 variables for 30 subjects.

Always first describe your data:

R code

```
> describe(attitude)
```

```
           vars  n  mean    sd median trimmed   mad min max range  skew kurtosis   se
rating        1 30 64.63 12.17   65.5   65.21 10.38  40  85    45 -0.36    -0.77 2.22
complaints    2 30 66.60 13.31   65.0   67.08 14.83  37  90    53 -0.22    -0.68 2.43
privileges    3 30 53.13 12.24   51.5   52.75 10.38  30  83    53  0.38    -0.41 2.23
learning      4 30 56.37 11.74   56.5   56.58 14.83  34  75    41 -0.05    -1.22 2.14
raises        5 30 64.63 10.40   63.5   64.50 11.12  43  88    45  0.20    -0.60 1.90
critical      6 30 74.77  9.89   77.5   75.83  7.41  49  92    43 -0.87     0.17 1.81
advance       7 30 42.93 10.29   41.0   41.83  8.90  25  72    47  0.85     0.47 1.88
```

2.1 linear modeling

The `lm` function is the workhorse for linear regression. It works on raw data and provides very useful diagnostics.

We first do a simple multiple regression with three predictors and compare that to the `setCor` results. By default, `setCor` standardizes and zero centers the data. We turn this off for this first example.

R code

```
> summary(lm(rating ~ complaints  + privileges + learning, data=attitude))
```

```
Call:
lm(formula = rating ~ complaints + privileges + learning, data = attitude)

Residuals:
     Min       1Q   Median       3Q      Max
-11.2012  -5.7478   0.5599   5.8226  11.3241

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.2583     7.3183   1.538   0.1360
complaints    0.6824     0.1288   5.296 1.54e-05 ***
privileges   -0.1033     0.1293  -0.799   0.4318
learning      0.2380     0.1394   1.707   0.0997 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.863 on 26 degrees of freedom
Multiple R-squared:  0.715,        Adjusted R-squared:  0.6821
F-statistic: 21.74 on 3 and 26 DF,  p-value: 2.936e-07
```

R code

```
> print(setCor(rating ~ complaints + privileges + learning, data=attitude,std=FALSE,zero=FALSE),digits=
```

```
Call: setCor(y = rating ~ complaints + privileges + learning, data = attitude,
    std = FALSE, zero = FALSE)
```

```
Multiple Regression from raw data

 DV =  rating
               slope      se        t          p lower.ci upper.ci      VIF
(Intercept) 11.25831 7.32204  1.53759 1.3623e-01 -3.79236 26.30897 33.00922
complaints   0.68242 0.12884  5.29643 1.5395e-05  0.41757  0.94726 47.14893
privileges  -0.10328 0.12935 -0.79851 4.3181e-01 -0.36916  0.16259 30.62298
learning     0.23798 0.13941  1.70702 9.9735e-02 -0.04859  0.52454 39.66788

Residual Standard Error =  6.98031  with  26  degrees of freedom

 Multiple Regression
             R    R2    Ruw    R2uw Shrunken R2 SE of R2 overall F df1 df2           p
rating 0.84558 0.715 0.6942 0.48192     0.68212  0.07275  21.74316   3  26 2.93597e-07
```

To find interactions, we take the product terms.

R code

```
> summary(lm(rating ~ complaints  * privileges + learning, data=attitude))

Call:
lm(formula = rating ~ complaints * privileges + learning, data = attitude)

Residuals:
     Min       1Q   Median       3Q      Max
-11.5795  -4.9937   0.2601   5.9985  10.5996

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          0.118124  27.312048   0.004   0.9966
complaints           0.855545   0.428889   1.995   0.0571 .
privileges           0.141025   0.591132   0.239   0.8134
learning             0.219815   0.147999   1.485   0.1500
complaints:privileges -0.003405   0.008032  -0.424   0.6753
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.974 on 25 degrees of freedom
Multiple R-squared:  0.717,        Adjusted R-squared:  0.6718
F-statistic: 15.84 on 4 and 25 DF,  p-value: 1.396e-06
```

R code

```
> print(setCor(rating ~ complaints * privileges + learning, data=attitude,std=FALSE,zero=TRUE),digits=5
Call: setCor(y = rating ~ complaints * privileges + learning, data = attitude,
    std = FALSE, zero = TRUE)

Multiple Regression from raw data

 DV =  rating
                        slope      se        t          p lower.ci upper.ci     VIF
(Intercept)           0.29934 1.47504  0.20294 8.4083e-01 -2.73855  3.33723 1.29734
complaints            0.67464 0.13220  5.10301 2.8557e-05  0.40236  0.94692 1.84761
privileges           -0.08573 0.13781 -0.62209 5.3952e-01 -0.36955  0.19809 1.69522
learning              0.21982 0.14800  1.48525 1.4998e-01 -0.08500  0.52463 1.79922
complaints*privileges -0.00340 0.00803 -0.42390 6.7526e-01 -0.01995  0.01314 1.49186

Residual Standard Error =  7.09311  with  25  degrees of freedom

 Multiple Regression
             R      R2     Ruw    R2uw Shrunken R2 SE of R2 overall F df1 df2           p
```

```
rating 0.84678 0.71704 0.70709 0.49998     0.67176  0.06956  15.83779   4  25 1.39632e-06
```

But, this form of interaction gives inappropriate estimates for the linear terms. We can solve this problem by zero centering the data.

We do this explicitly in `lm` and by default in `setCor`

R code

```
> summary(lm(rating ~ complaints  * privileges + learning, data=data.frame(scale(attitude,scale=FALSE))
```

```
Call:
lm(formula = rating ~ complaints * privileges + learning, data = data.frame(scale(attitude,
    scale = FALSE)))

Residuals:
     Min       1Q   Median       3Q      Max
-11.5795  -4.9937   0.2601   5.9985  10.5996

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.299341   1.455962   0.206    0.839
complaints          0.674642   0.132205   5.103 2.86e-05 ***
privileges         -0.085728   0.137807  -0.622    0.540
learning            0.219815   0.147999   1.485    0.150
complaints:privileges -0.003405   0.008032  -0.424    0.675
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.974 on 25 degrees of freedom
Multiple R-squared:  0.717,        Adjusted R-squared:  0.6718
F-statistic: 15.84 on 4 and 25 DF,  p-value: 1.396e-06
```

R code

```
> print(setCor(rating ~ complaints * privileges + learning, data=attitude,std=FALSE),digits=6)
```

```
Call: setCor(y = rating ~ complaints * privileges + learning, data = attitude,
    std = FALSE)

Multiple Regression from raw data

 DV =  rating
                        slope       se         t           p  lower.ci upper.ci      VIF
(Intercept)          0.299341 1.475035  0.202938 0.840826000 -2.738551 3.337232 1.297338
complaints           0.674642 0.132205  5.103007 0.000028557  0.402361 0.946923 1.847605
privileges          -0.085728 0.137807 -0.622086 0.539518000 -0.369545 0.198090 1.695222
learning             0.219815 0.147999  1.485245 0.149981000 -0.084995 0.524626 1.799217
complaints*privileges -0.003405 0.008032 -0.423901 0.675263000 -0.019947 0.013137 1.491865

Residual Standard Error =  7.093106  with  25  degrees of freedom

 Multiple Regression
             R       R2     Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2           p
rating 0.846781 0.717038 0.70709 0.499976    0.671764 0.069556  15.83779   4  25 1.39632e-06
```

Note that the se estimate for the intercept does not agree with `lm`. I am investigating this problem.

It is also possible to standardize the data and then do the regressions.

R code

```
> summary(lm(rating ~ complaints  * privileges + learning, data=data.frame(scale(attitude))))
```

```
Call:
lm(formula = rating ~ complaints * privileges + learning, data = data.frame(scale(attitude)))
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.95128 -0.41025  0.02137  0.49279  0.87078

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.02459    0.11961   0.206    0.839
complaints            0.73795    0.14461   5.103 2.86e-05 ***
privileges           -0.08617    0.13852  -0.622    0.540
learning              0.21195    0.14270   1.485    0.150
complaints:privileges -0.04557    0.10749  -0.424    0.675
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5729 on 25 degrees of freedom
Multiple R-squared:  0.717,        Adjusted R-squared:  0.6718
F-statistic: 15.84 on 4 and 25 DF,  p-value: 1.396e-06
```

R code

```
> print(setCor(rating ~ complaints * privileges + learning, data=attitude),digits=6)
Call: setCor(y = rating ~ complaints * privileges + learning, data = attitude)

Multiple Regression from raw data

 DV =  rating
                          slope       se         t           p  lower.ci upper.ci      VIF
(Intercept)            0.000000 0.106388  0.000000 1.000000000 -0.219111 0.219111 1.000000
complaints             0.737946 0.144610  5.103007 0.000028557  0.440116 1.035776 1.847605
privileges            -0.086170 0.138518 -0.622086 0.539518000 -0.371454 0.199113 1.695222
learning               0.211950 0.142704  1.485245 0.149981000 -0.081954 0.505854 1.799217
complaints*privileges -0.049290 0.116276 -0.423901 0.675263000 -0.288766 0.190186 1.194527

Residual Standard Error =  0.582713  with  25  degrees of freedom

 Multiple Regression
              R       R2      Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2           p
rating 0.846781 0.717038 0.749691 0.562037    0.671764 0.069556  15.83779   4  25 1.39632e-06
```

These disagree. Lets examine why. It is due to the way `lm` and `setCor` standardize. We can see this if we do the standardization outside of `lm`

We can form the standard scores and then find the product of the standard scores.

R code

```
>  att1 <- scale(attitude)   #standardize the data
>  cp <- att1[,"complaints"] * att1[,"privileges"] #take the products
>  att1 <- data.frame(cbind(att1,cp))  #put them back into the data frame
>  describe(att1)  #note that the products have non-zero means
           vars  n mean   sd median trimmed  mad   min  max range  skew kurtosis   se
rating        1 30 0.00 1.00   0.07    0.05 0.85 -2.02 1.67  3.70 -0.36    -0.77 0.18
complaints    2 30 0.00 1.00  -0.12    0.04 1.11 -2.22 1.76  3.98 -0.22    -0.68 0.18
privileges    3 30 0.00 1.00  -0.13   -0.03 0.85 -1.89 2.44  4.33  0.38    -0.41 0.18
learning      4 30 0.00 1.00   0.01    0.02 1.26 -1.91 1.59  3.49 -0.05    -1.22 0.18
raises        5 30 0.00 1.00  -0.11   -0.01 1.07 -2.08 2.25  4.33  0.20    -0.60 0.18
critical      6 30 0.00 1.00   0.28    0.11 0.75 -2.60 1.74  4.35 -0.87     0.17 0.18
advance       7 30 0.00 1.00  -0.19   -0.11 0.86 -1.74 2.83  4.57  0.85     0.47 0.18
cp            8 30 0.54 1.08   0.22    0.43 0.52 -1.34 3.29  4.62  0.96     0.23 0.20
```

Now do the regressions on these standardized scores.

R code

```
>  summary(lm(rating ~ complaints + privileges + learning + cp , data=att1))  #standardized data
```

```
Call:
lm(formula = rating ~ complaints + privileges + learning + cp,
    data = att1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.95128 -0.41025  0.02137  0.49279  0.87078

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.02459    0.11961   0.206    0.839
complaints   0.73795    0.14461   5.103 2.86e-05 ***
privileges  -0.08617    0.13852  -0.622    0.540
learning     0.21195    0.14270   1.485    0.150
cp          -0.04557    0.10749  -0.424    0.675
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5729 on 25 degrees of freedom
Multiple R-squared:  0.717,        Adjusted R-squared:  0.6718
F-statistic: 15.84 on 4 and 25 DF,  p-value: 1.396e-06
```

R code

```
>   summary(lm(rating ~ complaints * privileges + learning , data=att1))
```

```
Call:
lm(formula = rating ~ complaints * privileges + learning, data = att1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.95128 -0.41025  0.02137  0.49279  0.87078

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.02459    0.11961   0.206    0.839
complaints            0.73795    0.14461   5.103 2.86e-05 ***
privileges           -0.08617    0.13852  -0.622    0.540
learning              0.21195    0.14270   1.485    0.150
complaints:privileges -0.04557    0.10749  -0.424    0.675
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5729 on 25 degrees of freedom
Multiple R-squared:  0.717,        Adjusted R-squared:  0.6718
F-statistic: 15.84 on 4 and 25 DF,  p-value: 1.396e-06
```

R code

```
>  print(setCor(rating ~ complaints + privileges + learning + cp , data=att1,std=FALSE,zero=FALSE),digi
```

```
Call: setCor(y = rating ~ complaints + privileges + learning + cp,
    data = att1, std = FALSE, zero = FALSE)

Multiple Regression from raw data

 DV =  rating
             slope      se        t          p lower.ci upper.ci     VIF
(Intercept)  0.02459 0.12118  0.20294 8.4083e-01 -0.22498  0.27416 1.29734
```

```
complaints   0.73795 0.14461  5.10301 2.8557e-05  0.44012  1.03578 1.84761
privileges  -0.08617 0.13852 -0.62209 5.3952e-01 -0.37145  0.19911 1.69522
learning     0.21195 0.14270  1.48525 1.4998e-01 -0.08195  0.50585 1.79922
cp          -0.04557 0.10749 -0.42390 6.7526e-01 -0.26695  0.17582 1.49186

Residual Standard Error =  0.58271  with  25  degrees of freedom

 Multiple Regression
             R      R2     Ruw    R2uw Shrunken R2 SE of R2 overall F df1 df2           p
rating 0.84678 0.71704 0.70709 0.49998     0.67176  0.06956  15.83779   4  25 1.39632e-06
```

R code

```
>  print(setCor(rating ~ complaints * privileges + learning  , data=att1,std=FALSE,zero=FALSE),digits=5
```

```
Call: setCor(y = rating ~ complaints * privileges + learning, data = att1,
    std = FALSE, zero = FALSE)

Multiple Regression from raw data

 DV =  rating
                          slope      se        t          p lower.ci upper.ci     VIF
(Intercept)             0.02459 0.12118  0.20294 8.4083e-01 -0.22498  0.27416 1.29734
complaints              0.73795 0.14461  5.10301 2.8557e-05  0.44012  1.03578 1.84761
privileges             -0.08617 0.13852 -0.62209 5.3952e-01 -0.37145  0.19911 1.69522
learning                0.21195 0.14270  1.48525 1.4998e-01 -0.08195  0.50585 1.79922
complaints*privileges -0.04557 0.10749 -0.42390 6.7526e-01 -0.26695  0.17582 1.49186

Residual Standard Error =  0.58271  with  25  degrees of freedom

 Multiple Regression
             R      R2     Ruw    R2uw Shrunken R2 SE of R2 overall F df1 df2           p
rating 0.84678 0.71704 0.70709 0.49998     0.67176  0.06956  15.83779   4  25 1.39632e-06
```

R code

```
>  print(setCor(rating ~ complaints * privileges + learning, data=attitude,std=TRUE,zero=TRUE),digits=
```

```
Call: setCor(y = rating ~ complaints * privileges + learning, data = attitude,
    std = TRUE, zero = TRUE)

Multiple Regression from raw data

 DV =  rating
                            slope       se         t           p  lower.ci upper.ci      VIF
(Intercept)              0.000000 0.106388  0.000000 1.000000000 -0.219111 0.219111 1.000000
complaints               0.737946 0.144610  5.103007 0.000028557  0.440116 1.035776 1.847605
privileges              -0.086170 0.138518 -0.622086 0.539518000 -0.371454 0.199113 1.695222
learning                 0.211950 0.142704  1.485245 0.149981000 -0.081954 0.505854 1.799217
complaints*privileges -0.049290 0.116276 -0.423901 0.675263000 -0.288766 0.190186 1.194527

Residual Standard Error =  0.582713  with  25  degrees of freedom

 Multiple Regression
              R       R2      Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2           p
rating 0.846781 0.717038 0.749691 0.562037    0.671764 0.069556  15.83779   4  25 1.39632e-06
```

3 sat.act data set

Built into *psychTools* is the sat.act data set.

Because `lm` works only on complete data sets, we first form the complete data set and then do some regressions to compare the two procedures.

There is an error in the print function in mediate. This is shown here to remind me to fix it.

R code

```
> complete <- sat.act[complete.cases(sat.act),]
> summary(lm(SATQ ~ SATV*gender + ACT,data=complete))  #the uncentered model
```

```
Call:
lm(formula = SATQ ~ SATV * gender + ACT, data = complete)

Residuals:
     Min       1Q   Median       3Q      Max
-296.210  -45.738    4.323   52.355  252.306

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 138.52395   61.18770   2.264   0.0239 *
SATV          0.50280    0.10030   5.013 6.84e-07 ***
gender      -22.24995   35.59228  -0.625   0.5321
ACT           7.71702    0.77707   9.931  < 2e-16 ***
SATV:gender  -0.01984    0.05706  -0.348   0.7281
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81.18 on 682 degrees of freedom
Multiple R-squared:   0.51,        Adjusted R-squared:  0.5071
F-statistic: 177.5 on 4 and 682 DF,  p-value: < 2.2e-16
```

R code

```
> print(setCor(SATQ ~ SATV*gender + ACT,data=complete,std=FALSE,zero=FALSE),digits=6) #the base model
```

```
Call: setCor(y = SATQ ~ SATV * gender + ACT, data = complete, std = FALSE,
    zero = FALSE)

Multiple Regression from raw data

 DV =  SATQ
                slope        se         t           p   lower.ci   upper.ci      VIF
(Intercept) 138.523946 61.187814  2.263914 2.38927e-02  18.384826 258.663065 389.6774
SATV          0.502799  0.100300  5.012950 6.83987e-07   0.305865   0.699733 406.0444
gender      -22.249948 35.592276 -0.625134 5.32092e-01 -92.133547  47.633651 386.3878
ACT           7.717021  0.777071  9.930903 8.46915e-22   6.191282   9.242761  52.6958
SATV*gender  -0.019842  0.057060 -0.347747 7.28137e-01  -0.131877   0.092192 383.6756

Residual Standard Error =  81.24389  with  682  degrees of freedom

 Multiple Regression
           R       R2      Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2             p
SATQ 0.714141 0.509998 0.417056 0.173936    0.507124 0.026449  177.4575   4 682 3.984723e-104
```

R code

```
> mod0 <- mediate(SATQ ~ SATV*gender + ACT,data=complete,std=FALSE,zero=FALSE)
> mod0        ##<- error in print
```

```
Mediation/Moderation Analysis
Call: mediate(y = SATQ ~ SATV * gender + ACT, data = complete, std = FALSE,
    zero = FALSE)
```

```
The DV (Y) was  SATQ . The IV (X) was  SATV gender ACT SATV*gender . The mediating variable(s) =  .
 DV =  SATQ
             slope    se     t       p
Intercept   138.52 61.19  2.26 2.4e-02
SATV          0.50  0.10  5.01 6.8e-07
gender      -22.25 35.59 -0.63 5.3e-01
ACT           7.72  0.78  9.93 8.5e-22
SATV*gender  -0.02  0.06 -0.35 7.3e-01

With R2 =  NA
R = NA R2 = NA   F = NA on 5 and 682 DF   p-value:  NA
```

R code

```
> summary(mod0,digits=5)

Call: mediate(y = SATQ ~ SATV * gender + ACT, data = complete, std = FALSE,
    zero = FALSE)

No mediator specified leads to traditional regression
               SATQ       se         t  df        Prob
Intercept   138.52395 61.18770  2.26392 682 2.38924e-02
SATV          0.50280  0.10030  5.01295 682 6.83987e-07
gender      -22.24995 35.59228 -0.62513 682 5.32092e-01
ACT           7.71702  0.77707  9.93090 682 8.46915e-22
SATV*gender  -0.01984  0.05706 -0.34775 682 7.28137e-01

R = 0.71414 R2 = 0.51   F = 177.4575 on 4 and 682 DF   p-value:  3.98472e-104
```

Now do the regressions on the interaction terms for a variety of options.

R code

```
> summary(lm(SATQ ~ SATV*gender + ACT,data=data.frame(scale(complete,scale=FALSE))))  #the centered mod

Call:
lm(formula = SATQ ~ SATV * gender + ACT, data = data.frame(scale(complete,
    scale = FALSE)))

Residuals:
     Min       1Q   Median       3Q      Max
-296.210  -45.738    4.323   52.355  252.306

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.02140    3.09800  -0.007    0.994
SATV          0.47019    0.03305  14.227  < 2e-16 ***
gender      -34.40016    6.47229  -5.315 1.45e-07 ***
ACT           7.71702    0.77707   9.931  < 2e-16 ***
SATV:gender  -0.01984    0.05706  -0.348    0.728
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81.18 on 682 degrees of freedom
Multiple R-squared:   0.51,        Adjusted R-squared:  0.5071
F-statistic: 177.5 on 4 and 682 DF,  p-value: < 2.2e-16
```

R code

```
> print(setCor(SATQ ~ SATV*gender + ACT,data=complete,std=FALSE,zero=TRUE),digits=6) #the centered mode

Call: setCor(y = SATQ ~ SATV * gender + ACT, data = complete, std = FALSE,
    zero = TRUE)

Multiple Regression from raw data
```

```
 DV =  SATQ
                slope       se         t           p   lower.ci   upper.ci      VIF
(Intercept)  -0.021401 3.100259 -0.006903 9.94494e-01  -6.108600   6.065797 1.000394
SATV          0.470190 0.033049 14.226972 2.06592e-40   0.405300   0.535081 1.459210
gender      -34.400163 6.472289 -5.314992 1.44657e-07 -47.108169 -21.692157 1.001841
ACT           7.717021 0.777071  9.930903 8.46915e-22   6.191282   9.242761 1.466780
SATV*gender  -0.019842 0.057060 -0.347747 7.28137e-01  -0.131877   0.092192 1.005319

Residual Standard Error =  81.24389  with  682  degrees of freedom

 Multiple Regression
           R       R2      Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2             p
SATQ 0.714141 0.509998 0.565925 0.320271    0.507124 0.026449  177.4575   4 682 3.984723e-104
```

R code

```
> mod1 <- mediate(SATQ ~ SATV*gender + ACT,data=complete,std=FALSE,zero=TRUE)
> mod1
```

```
Mediation/Moderation Analysis
Call: mediate(y = SATQ ~ SATV * gender + ACT, data = complete, std = FALSE,
    zero = TRUE)

The DV (Y) was  SATQ . The IV (X) was  SATV gender ACT SATV*gender . The mediating variable(s) =  .
 DV =  SATQ
             slope   se     t       p
Intercept    -0.02 3.10 -0.01 9.9e-01
SATV          0.47 0.03 14.23 2.1e-40
gender      -34.40 6.47 -5.31 1.4e-07
ACT           7.72 0.78  9.93 8.5e-22
SATV*gender  -0.02 0.06 -0.35 7.3e-01

With R2 =  NA
R = NA R2 = NA   F = NA on 5 and 682 DF   p-value:  NA
```

R code

```
> summary(mod1,digits=6)
```

```
Call: mediate(y = SATQ ~ SATV * gender + ACT, data = complete, std = FALSE,
    zero = TRUE)

No mediator specified leads to traditional regression
                 SATQ       se         t  df         Prob
Intercept    -0.021401 3.098002 -0.006908 682 9.944902e-01
SATV          0.470190 0.033049 14.226972 682 2.065924e-40
gender      -34.400163 6.472289 -5.314992 682 1.446572e-07
ACT           7.717021 0.777071  9.930903 682 8.469147e-22
SATV*gender  -0.019842 0.057060 -0.347747 682 7.281372e-01

R = 0.714141 R2 = 0.509998   F = 177.4575 on 4 and 682 DF   p-value:  3.984723e-104
```

R code

```
> summary(lm(SATQ ~ SATV*gender + ACT,data=data.frame(scale(complete,scale=TRUE))))  #the Standardized
```

```
Call:
lm(formula = SATQ ~ SATV * gender + ACT, data = data.frame(scale(complete,
    scale = TRUE)))

Residuals:
     Min       1Q   Median       3Q      Max
-2.56150 -0.39552  0.03738  0.45275  2.18183
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0001851  0.0267902  -0.007    0.994
SATV         0.4606576  0.0323792  14.227  < 2e-16 ***
gender      -0.1425965  0.0268291  -5.315 1.45e-07 ***
ACT          0.3223874  0.0324631   9.931  < 2e-16 ***
SATV:gender -0.0093187  0.0267972  -0.348    0.728
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7021 on 682 degrees of freedom
Multiple R-squared:   0.51,        Adjusted R-squared:  0.5071
F-statistic: 177.5 on 4 and 682 DF,  p-value: < 2.2e-16
```

R code

```
> print(setCor(SATQ ~ SATV*gender + ACT,data=complete),digits=6) #the standardized and centered model
```

```
Call: setCor(y = SATQ ~ SATV * gender + ACT, data = complete)

Multiple Regression from raw data

 DV =  SATQ
                slope       se         t           p  lower.ci  upper.ci      VIF
(Intercept)  0.000000 0.026804  0.000000 1.00000e+00 -0.052629  0.052629 1.000000
SATV         0.460658 0.032379 14.226972 2.06592e-40  0.397083  0.524232 1.459210
gender      -0.142596 0.026829 -5.314992 1.44657e-07 -0.195274 -0.089919 1.001841
ACT          0.322387 0.032463  9.930903 8.46915e-22  0.258648  0.386127 1.466780
SATV*gender -0.009344 0.026870 -0.347747 7.28137e-01 -0.062103  0.043414 1.004924

Residual Standard Error =  0.702563  with  682  degrees of freedom

 Multiple Regression
           R       R2      Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2             p
SATQ 0.714141 0.509998 0.615931 0.379371    0.507124 0.026449  177.4575   4 682 3.984723e-104
```

R code

```
> mod2 <- mediate(SATQ ~ SATV*gender + ACT,data=complete,std=TRUE,zero=TRUE)
> mod2
```

```
Mediation/Moderation Analysis
Call: mediate(y = SATQ ~ SATV * gender + ACT, data = complete, std = TRUE,
    zero = TRUE)

The DV (Y) was  SATQ . The IV (X) was  SATV gender ACT SATV*gender . The mediating variable(s) =  .
 DV =  SATQ
            slope   se     t       p
Intercept    0.00 0.03  0.00 1.0e+00
SATV         0.46 0.03 14.23 2.1e-40
gender      -0.14 0.03 -5.31 1.4e-07
ACT          0.32 0.03  9.93 8.5e-22
SATV*gender -0.01 0.03 -0.35 7.3e-01

With R2 =  0.51
R = 0.71 R2 = 0.51   F = 141.97 on 5 and 682 DF   p-value:  3.97e-103
```

R code

```
> summary(mod2,digits=6)
```

```
Call: mediate(y = SATQ ~ SATV * gender + ACT, data = complete, std = TRUE,
    zero = TRUE)
```

```
No mediator specified leads to traditional regression
                SATQ       se          t  df         Prob
Intercept    0.000000 0.026804  0.000000 682 1.000000e+00
SATV         0.460658 0.032379 14.226972 682 2.065924e-40
gender      -0.142596 0.026829 -5.314992 682 1.446572e-07
ACT          0.322387 0.032463  9.930903 682 8.469147e-22
SATV*gender -0.009344 0.026870 -0.347747 682 7.281372e-01

R = 0.714141 R2 = 0.509998   F = 177.4575 on 4 and 682 DF   p-value:  3.984723e-104
```

R code

```
> print(setCor(SATQ ~ SATV*gender + ACT,data=complete,std=TRUE,zero=FALSE),digits=6) #the std uncenterc
```

```
Call: setCor(y = SATQ ~ SATV * gender + ACT, data = complete, std = TRUE,
    zero = FALSE)

Multiple Regression from raw data

 DV =  SATQ
                slope       se          t           p  lower.ci upper.ci      VIF
(Intercept)  0.000000 0.026804  0.000000 1.00000e+00 -0.052629 0.052629  1.00000
SATV         0.492605 0.098267  5.012950 6.83987e-07  0.299664 0.685546 13.43994
gender      -0.092231 0.147538 -0.625134 5.32092e-01 -0.381914 0.197452 30.29663
ACT          0.322387 0.032463  9.930903 8.46915e-22  0.258648 0.386127  1.46678
SATV*gender -0.059871 0.172167 -0.347747 7.28137e-01 -0.397912 0.278171 41.25606

Residual Standard Error =  0.702563  with  682  degrees of freedom

 Multiple Regression
           R       R2      Ruw     R2uw Shrunken R2 SE of R2 overall F df1 df2             p
SATQ 0.714141 0.509998 0.705377 0.497557    0.507124 0.026449  177.4575   4 682 3.984723e-104
```

R code

```
>
```

4 Session Information

R code

```
> sessionInfo()
```

```
R version 4.0.2 (2020-06-22)
Platform: x86_64-apple-darwin17.0 (64-bit)
Running under: macOS Catalina 10.15.7

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib

Random number generation:
 RNG:     Mersenne-Twister
 Normal:  Inversion
 Sample:  Rounding

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
```

```
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] psychTools_2.0.9 psych_2.0.10

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.5      lattice_0.20-41 MASS_7.3-51.6   grid_4.0.2      nlme_3.1-148    minqa_1.2.4     nl
 [8] Matrix_1.2-18   boot_1.3-25     splines_4.0.2   statmod_1.4.34  lme4_1.1-23     tools_4.0.2     fo
[15] parallel_4.0.2  compiler_4.0.2  mnormt_2.0.1    tmvnsim_1.0-2
```

References

Bashaw, W. and Anderson Jr, H. E. (1967). A correction for replicated error in correlation coefficients. *Psychometrika*, 32(4):435–441.

Cureton, E. (1966). Corrected item-test correlations. *Psychometrika*, 31(1):93–96.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rafaeli, E. and Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*, 30(1):1–12.

Revelle, W. (2020). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, https://CRAN.r-project.org/package=psych. R package version 2.0.8.

Revelle, W. and Anderson, K. J. (1997). Personality, motivation and cognitive performance. final report to the Army Research Institute on contract MDA 903-93-K-0008. Technical report, Northwestern University:.