Psychology 405: Psychometric Theory
More on regression and partial correlations

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2020

Outline

More on Regression, partial correlations, and mediation

1. Regression and multiple regression is just the solution to $\beta = \beta R R^{-1} = r_{xy} R^{-1}$
2. $\hat{y} = \beta_{xy} x$
3. We can do this for one, two, or many x variables
4. Lets examine what happens when we increase the predictors
5. Also, lets examine what if we don't use optimal weights
6. We will use the `Thurstone` data set of 9 cognitive measures and
7. The `Tal.Or` data set of 6 measures of behavior

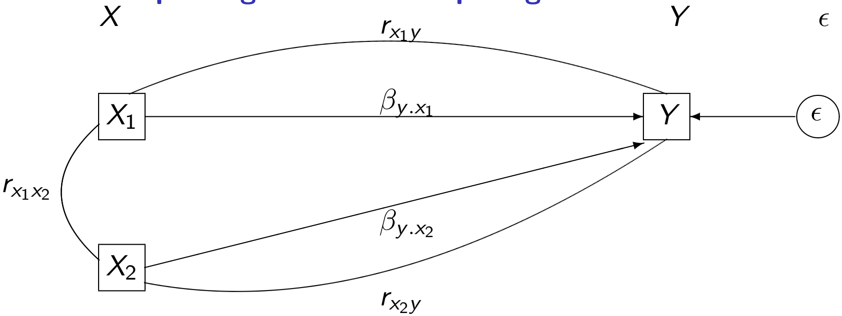
Thurstone correlation matrix (a classic demonstration set)

R code

```
lowerMat(Thurstone)
```

```
                Sntnc Vcblr Snt.C Frs.L 4.L.W Sffxs Ltt.S Pdgrs Ltt.G
Sentences       1.00
Vocabulary      0.83  1.00
Sent.Completion 0.78  0.78  1.00
First.Letters   0.44  0.49  0.46  1.00
4.Letter.Words  0.43  0.46  0.42  0.67  1.00
Suffixes        0.45  0.49  0.44  0.59  0.54  1.00
Letter.Series   0.45  0.43  0.40  0.38  0.40  0.29  1.00
Pedigrees       0.54  0.54  0.53  0.35  0.37  0.32  0.56  1.00
Letter.Group    0.38  0.36  0.36  0.42  0.45  0.32  0.60  0.45  1.00
```

Multiple Regression: decomposing correlations



$$r_{x_1y} = \overbrace{\beta_{y.x_1}}^{\color{blue}{direct}} + \overbrace{r_{x_1x_2}\beta_{y.x_2}}^{\color{red}{indirect}}$$

$$\beta_{y.x_1} = \frac{\color{blue}{r_{x_1y}} - \color{red}{r_{x_1x_2}r_{x_2y}}}{1 - \color{red}{r^2_{x_1x_2}}}$$

$$r_{x_2y} = \underbrace{\beta_{y.x_2}}_{\color{blue}{direct}} + \underbrace{r_{x_1x_2}\beta_{y.x_1}}_{\color{red}{indirect}}$$

$$\beta_{y.x_2} = \frac{\color{blue}{r_{x_2y}} - \color{red}{r_{x_1x_2}r_{x_1y}}}{1 - \color{red}{r^2_{x_1x_2}}}$$

$$R^2 = r_{x_1y}\beta_{y.x_1} + r_{x_2y}\beta_{y.x_2}$$

Beta weights are merely one possible way to weight a predictor set

Consider a matrix, **R** of predictors (x_i) with the criterion (y).

$$R = \left[\begin{array}{ccccc|c} r_{11} & r_{12} & r_{13} & \dots & r_{1n} & r_{1y} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2n} & r_{2y} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & r_{nn} & r_{ny} \\ \hline r_{y1} & r_{y2} & r_{y3} & \dots & r_{yn} & r_{yy} \end{array}\right] \quad R_{(x_1 x_2 \dots x_n)y} = \frac{\Sigma r_{x_i y}}{\sqrt{(\Sigma\Sigma r_{ij}) r_{yy}}}$$

Multiply each x_i by a weight β_i and then find $R_{(x_1 x_2 \dots x_n)y}$

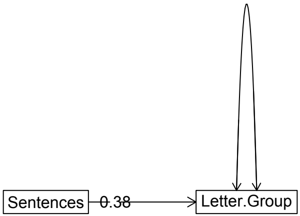
$$R = \left[\begin{array}{ccccc|c} \beta_1 r_{11}\beta_1 & \beta_1 r_{12}\beta_2 & \beta_1 r_{13}\beta_3 & \dots & \beta_1 r_{1n}\beta_n & \beta_1 r_{1y} \\ \beta_2 r_{21}\beta_1 & \beta_2 r_{22}\beta_1 & \beta_3 r_{23}\beta_3 & \dots & \beta_2 r_{2n}\beta_n & \beta_2 r_{2y} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \beta_n r_{n1}\beta_1 & \beta_n r_{n2}\beta_1 & \beta_n r_{n3}\beta_3 & \dots & \beta_n r_{nn}\beta_n & \beta_n r_{ny} \\ \hline r_{y1}\beta_1 & r_{y2}\beta_2 & r_{y3}\beta_3 & \dots & r_{yn}\beta_n & r_{yy} \end{array}\right]$$

R code

```
setCor(Letter.Group ~Sentences,data=Thurstone)
```

```
setCor(Letter.Group ~Sentences,data=Thurstone)

Call: setCor(y = "Letter.Group", x = "Sentences", data = Thurstone)

Multiple Regression from matrix input

Beta weights
          Letter.Group
Sentences         0.38

Multiple R
             Letter.Group
Letter.Group         0.38
multiple R2
             Letter.Group
Letter.Group         0.14

 Unweighted multiple R
Letter.Group
        0.38
 Unweighted multiple R2
Letter.Group
        0.14
```

Regression Models



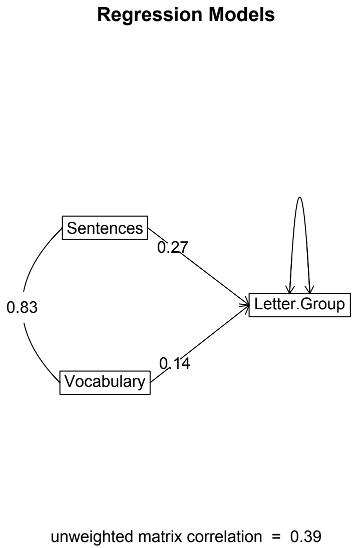
unweighted matrix correlation = 0.38

R code

```
> setCor(Letter.Group ~Sentences+Vocabulary,data=Thurstone)
```

```
>setCor(Letter.Group ~Sentences+Vocabulary, data=Thurstone)

Call: setCor(y = Letter.Group ~ Sentences + Vocabulary,
  data = Thurstone)

Multiple Regression from matrix input

Beta weights
           Letter.Group
Sentences          0.27
Vocabulary         0.14

Multiple R
             Letter.Group
Letter.Group         0.39
multiple R2
             Letter.Group
Letter.Group         0.15

 Unweighted multiple R
Letter.Group
        0.39
 Unweighted multiple R2
Letter.Group
        0.15
```

Two predictors – note the reduction of the beta weight for Sentences

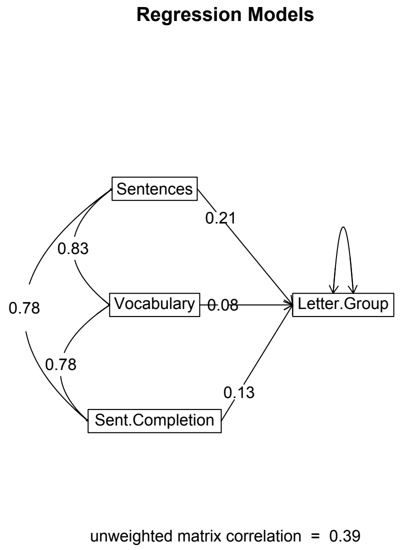


R code

```
setCor(Letter.Group ~Sentences + Vocabulary + Sent.Completion,
    data=Thurstone)
```

```
Call: setCor(y = Letter.Group ~ Sentences + Vocabulary + Sent.Completion,
    data = Thurstone)

Multiple Regression from matrix input

Beta weights
                Letter.Group
Sentences               0.21
Vocabulary              0.08
Sent.Completion         0.13

Multiple R
             Letter.Group
Letter.Group          0.4
multiple R2
             Letter.Group
Letter.Group         0.16

 Unweighted multiple R
Letter.Group
        0.39
 Unweighted multiple R2
Letter.Group
        0.15
```

A trivial increase in R, a reduction in beta weights



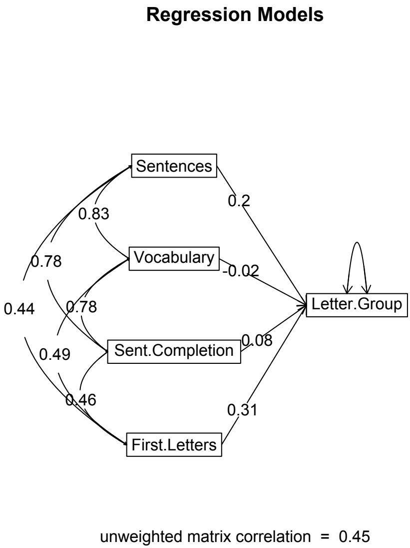
Yet one more predictor, slightly different content

R code

```
setCor(Letter.Group ~Sentences + Vocabulary + Sent.Completion + First.Letters,
    data=Thurstone)
```

```
Call: setCor(y = Letter.Group ~ Sentences + Vocabulary + Sent.Completion +
    First.Letters, data = Thurstone)
Multiple Regression from matrix input

Beta weights
                Letter.Group
Sentences               0.20
Vocabulary             -0.02
Sent.Completion         0.08
First.Letters           0.31

Multiple R
             Letter.Group
Letter.Group         0.48
multiple R2
             Letter.Group
Letter.Group         0.23

 Unweighted multiple R
Letter.Group
        0.45
 Unweighted multiple R2
Letter.Group
         0.2
```

A change in the pattern of the beta weights



Add a fifth variable

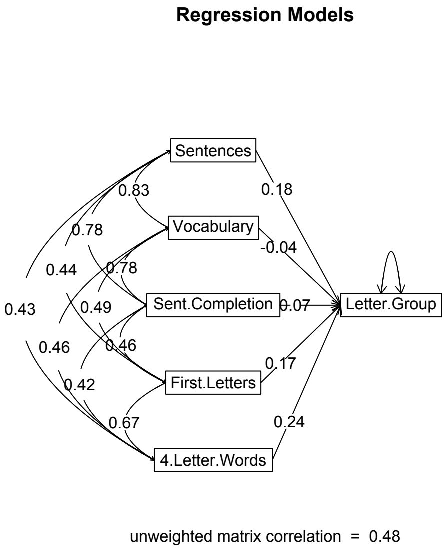
(Change the way we specify y and x)

R code

```
setCor(y=9,x = 1:5,data=Thurstone)
```

```
setCor(y=9,x = 1:5,data=Thurstone)
Call: setCor(y = 9, x = 1:5, data = Thurstone)

Multiple Regression from matrix input

Beta weights
                Letter.Group
Sentences               0.18
Vocabulary             -0.04
Sent.Completion         0.07
First.Letters           0.17
4.Letter.Words          0.24

Multiple R
             Letter.Group
Letter.Group         0.51
multiple R2
             Letter.Group
Letter.Group         0.26

 Unweighted multiple R
Letter.Group
        0.48
 Unweighted multiple R2
Letter.Group
        0.23
```

Yet another variable



Add a 6th – doesn't do anything!

R code

```
Call: setCor(y = 9, x = 1:6, data = Thurstone)
```

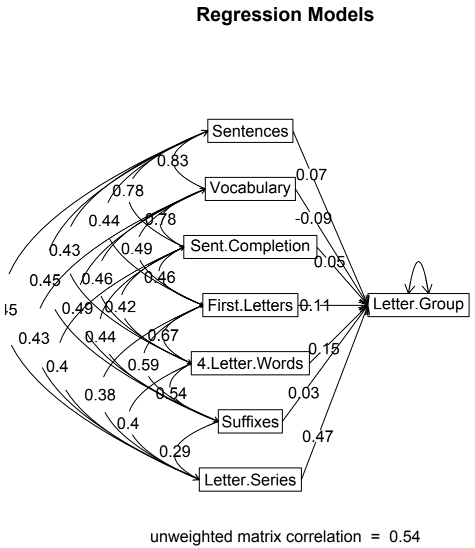
```
setCor(y=9,x = 1:6,data=Thurstone)

Multiple Regression from matrix input

Beta weights
                Letter.Group
Sentences               0.18
Vocabulary             -0.05
Sent.Completion         0.07
First.Letters           0.17
4.Letter.Words          0.24
Suffixes                0.00

Multiple R
             Letter.Group
Letter.Group         0.51
multiple R2
             Letter.Group
Letter.Group         0.26

 Unweighted multiple R
Letter.Group
        0.48
 Unweighted multiple R2
Letter.Group
        0.23
```

R code

```
setCor(y=9,x = 1:7,data=Thurstone)
```

```
Call: setCor(y = 9, x = 1:7, data = Thurstone)

Multiple Regression from matrix input

Beta weights
                Letter.Group
Sentences               0.07
Vocabulary             -0.09
Sent.Completion         0.05
First.Letters           0.11
4.Letter.Words          0.15
Suffixes                0.03
Letter.Series           0.47

Multiple R
             Letter.Group
Letter.Group         0.65
multiple R2
             Letter.Group
Letter.Group         0.42

 Unweighted multiple R
Letter.Group
        0.54
 Unweighted multiple R2
Letter.Group
         0.3
```

Adding a 7th actually makes the 6th variable predicting



Partial regression coefficients - compare to using x = 1:3

R code

```
setCor(Letter.Group ~ Sentences - Vocabulary - Sent.Completion ,
  data=Thurstone)
```

```
Call: setCor(y = Letter.Group ~ Sentences - Vocabulary - Sent.Completion,
    data = Thurstone)

    Multiple Regression from matrix input

Beta weights
          Letter.Group
Sentences         0.21

Multiple R
             Letter.Group
Letter.Group         0.12
multiple R2
             Letter.Group
Letter.Group        0.014

 Unweighted multiple R
Letter.Group
        0.11
 Unweighted multiple R2
Letter.Group
        0.01
```

Partial multiple R – compare with regression

R code

```
setCor(y=9,x = 1:2,z=3:7,data=Thurstone)
```

```
Call: setCor(y = 9, x = 1:2, data = Thurstone, z = 3:7)

Multiple Regression from matrix input

Beta weights
           Letter.Group
Sentences          0.07
Vocabulary        -0.09

Multiple R
             Letter.Group
Letter.Group         0.06
multiple R2
             Letter.Group
Letter.Group       0.0037

 Unweighted multiple R
Letter.Group
        0.03
 Unweighted multiple R2
Letter.Group
           0
```

Mediation and partial correlations

1. "Mediation" models have become popular to "explain" effects of X on Y as mediated through M
2. This makes particular sense if time is a variable because then X can effect M which in turn can effect Y
3. Mediation models are just regression models expressed somewhat differently
4. Total direct effect of x on y is the c path
5. Direct paths (x -> y) (the c' path)
6. Indirect path (x -> M) (the a path) (M -> y) (the b path)
7. indirect effect is ab
8. Compare c to c' where c' = c - ab
9. Test of mediation (ab) is found through bootstrapping

The Tal.Or data set

Tal-Or, Cohen, Tsfati & Gunther (2010) examined the presumed effect of the media in two experimental studies. These data are from study 2. '... perceptions regarding the influence of a news story about an expected shortage in sugar were manipulated indirectly, by manipulating the perceived exposure to the news story, and behavioral intentions resulting from the story were consequently measured." (p 801).
The data were downloaded from the webpages of Andrew Hayes (https://www.afhayes.com/public/hayes2018data.zip)webpagessupporting the first and second edition of his book (Hayes, 2013) The name of the original data set was pmi. (Gender was recoded to reflect the number of X chromosomes).

The Tal.Or data set

- **cond** Experimental Condition: 0 low media importance, 1 high media importance
- **pmi** Presumed media influence (based upon the mean of two items
- **import** Importance of the issue
- **reaction** Subjects rated agreement about possible reactions to the story (mean of 4 items).
- **gender** 1 = male, 2 = female
- **age** a numeric vector

The Tal.Or data set

R code

```
describe(Tal_Or)
```

```
> describe(Tal_Or)
         vars   n  mean   sd median trimmed  mad min max range  skew kurtosis   se
cond        1 123  0.47 0.50   0.00    0.46 0.00   0   1     1  0.11    -2.00 0.05
pmi         2 123  5.60 1.32   6.00    5.78 1.48   1   7     6 -1.17     1.30 0.12
import      3 123  4.20 1.74   4.00    4.26 1.48   1   7     6 -0.26    -0.89 0.16
reaction    4 123  3.48 1.55   3.25    3.44 1.85   1   7     6  0.21    -0.90 0.14
gender      5 123  1.65 0.48   2.00    1.69 0.00   1   2     1 -0.62    -1.62 0.04
age         6 123 24.63 5.80  24.00   23.76 1.48  18  61    43  4.71    24.76 0.52
```

The Tal.Or data set

R code

```
lowerCor(Tal_Or)
```

```
lowerCor(Tal_Or)
         cond  pmi   imprt rectn gendr age
cond      1.00
pmi       0.18  1.00
import    0.18  0.28  1.00
reaction  0.16  0.45  0.46  1.00
gender   -0.13 -0.02  0.03  0.01  1.00
age       0.03  0.00  0.07 -0.08 -0.32  1.00
```

A simple regression

R code

```
setCor(reaction ~ import + cond,data=Tal_Or,std=FALSE)
```

```
Call: setCor(y = reaction ~ import + cond, data = Tal_Or, std = FALSE)

Multiple Regression from raw data

 DV =  reaction
            slope   se    t       p lower.ci upper.ci  VIF
(Intercept)  1.68 0.33 5.04 1.7e-06     1.02     2.34 7.17
import       0.40 0.07 5.50 2.1e-07     0.26     0.55 7.14
cond         0.24 0.25 0.96 3.4e-01    -0.26     0.74 1.96

Residual Standard Error =  1.38  with  120  degrees of freedom

 Multiple Regression
           R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2        p
reaction 0.47 0.22 0.34 0.12        0.21     0.06     17.12   2 120 2.87e-07
```

Use the mediate function and graphics

R code

```
 A simple mediation example is the Tal_Or data set (pmi for Hayes)
#The pmi data set from Hayes is available as the Tal_Or data set.
mod4 <- mediate(reaction ~ cond + (pmi), data =Tal_Or,n.iter=50)
summary(mod4)
```

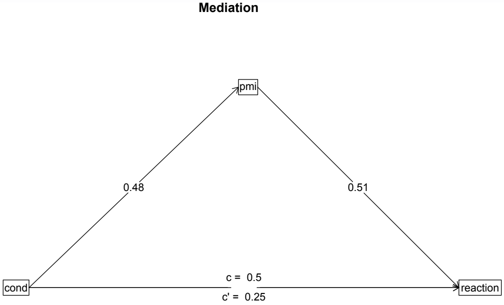
```
 summary(mod4)
Call: mediate(y = reaction ~ cond + (pmi), data = Tal_Or, n.iter = 50)

Direct effect estimates (traditional regression)    (c')
          reaction   se    t  df     Prob
Intercept     0.53 0.55 0.96 120 3.40e-01
cond          0.25 0.26 0.99 120 3.22e-01
pmi           0.51 0.10 5.22 120 7.66e-07

R = 0.45 R2 = 0.21   F = 15.56 on 2 and 120 DF   p-value:  9.83e-07

 Total effect estimates (c)
     reaction   se    t  df   Prob
cond      0.5 0.28 1.79 122 0.0754

 'a'  effect estimates
           pmi   se     t  df     Prob
Intercept 5.38 0.16 33.22 121 1.16e-62
cond      0.48 0.24  2.02 121 4.54e-02
(continued)
```

1 mediator (continued

R code

```
 A simple mediation example is the Tal_Or data set (pmi for Hayes)
#The pmi data set from Hayes is available as the Tal_Or data set.
mod4 <- mediate(reaction ~ cond + (pmi), data =Tal_Or,n.iter=50)
summary(mod4)
```

```
(continued from previous slide)

'b'  effect estimates
    reaction  se    t  df     Prob
pmi     0.51 0.1 5.24 121 6.88e-07

 'ab'  effect estimates (through mediators)
     reaction boot   sd lower upper
cond     0.24 0.22 0.12  0.06  0.46
```


Mediation
pmi
0.48
0.51
cond
c = 0.5
c' = 0.25
reaction

Two mediating variables

R code

```
#Two mediators (from Hayes model 6 (chapter 5))
mod6 <- mediate(reaction ~ cond + (pmi) + (import), data =Tal_Or,n.iter=50)
summary(mod6)
```

```
summary(mod6)
Call: mediate(y = reaction ~ cond + (pmi) + (import), data = Tal_Or,
    n.iter = 50)

Direct effect estimates (traditional regression)    (c')
          reaction   se     t  df     Prob
Intercept    -0.15 0.53 -0.28 119 7.78e-01
cond          0.10 0.24  0.43 119 6.66e-01
pmi           0.40 0.09  4.26 119 4.04e-05
import        0.32 0.07  4.59 119 1.13e-05

R = 0.57 R2 = 0.33   F = 19.11 on 3 and 119 DF   p-value:  3.5e-10

 Total effect estimates (c)
     reaction   se    t  df   Prob
cond      0.5 0.28 1.79 122 0.0754

 'a'  effect estimates
           pmi   se     t  df     Prob
Intercept 5.38 0.16 33.22 121 1.16e-62
cond      0.48 0.24  2.02 121 4.54e-02
          import   se     t  df     Prob
Intercept   3.91 0.21 18.37 121 8.39e-37
cond        0.63 0.31  2.02 121 4.52e-02
(continued)
```

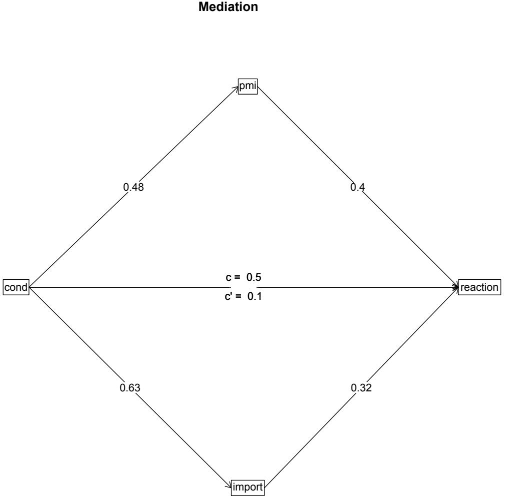
Two mediating variables (continued)

R code

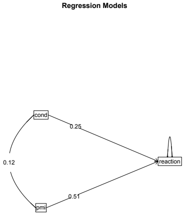
```
#Two mediators (from Hayes model 6 (chapter 5))
mod6 <- mediate(reaction ~ cond + (pmi) + (import), data =Tal_Or,n.iter=50)
summary(mod6)
```

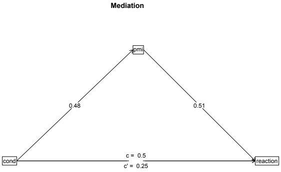
```
 'b'  effect estimates
       reaction   se    t  df      Prob
pmi         0.40 0.09 4.28 120 3.75e-05
import      0.32 0.07 4.60 120 1.03e-05

 'ab'  effect estimates (through mediators)
     reaction boot   sd lower upper
cond     0.39 0.38 0.18   0.1  0.73

 'ab' effects estimates for each mediator for reaction
      pmi   sd lower upper
cond 0.24 0.13 -0.04  0.48
     import   sd lower upper
cond    0.2 0.11 -0.03  0.43
```

Two “mediating” variables

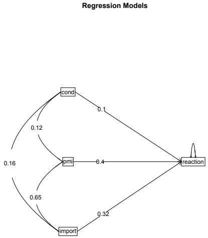


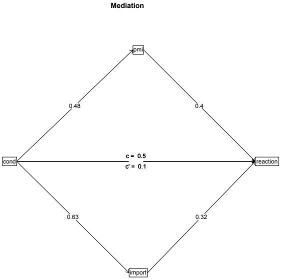
Compare regression versus mediation results





Compare regression versus mediation results





Squared Multiple Correlations

1. What is the predictability of one variable from a set of variables?
2. Do the standard multiple R from the correlation matrix **R**
3. But what if we want to know how much each variable is predictable by all the other variables?
4. $SMC = 1 - \frac{1}{diag(\mathbf{R}^{-1})}$

Finding the SMC

R code

```
R.inv <- solve(Thurstone)  #Find the inverse
lowerMat(R.inv)  #show the inverse (lower diagonal)
 smc(Thurstone)  #find the SMC by the smc function
  1- 1/diag(R.inv)  #find the SMC by the formula
```

```
                Sntnc Vcblr Snt.C Frs.L 4.L.W Sffxs Ltt.S Pdgrs Ltt.G
Sentences        3.82
Vocabulary      -2.02  3.99
Sent.Completion -1.16 -1.14  3.07
First.Letters    0.14 -0.27 -0.20  2.24
4.Letter.Words  -0.06 -0.11  0.01 -0.94  2.09
Suffixes        -0.10 -0.27 -0.07 -0.58 -0.36  1.74
Letter.Series   -0.20 -0.11  0.09 -0.09 -0.12  0.09  1.91
Pedigrees       -0.20 -0.24 -0.34  0.06 -0.05 -0.01 -0.56  1.82
Letter.Group    -0.10  0.17 -0.05 -0.20 -0.26 -0.05 -0.75 -0.21  1.75

> smc(Thurstone)
      Sentences      Vocabulary Sent.Completion   First.Letters  4.Letter.Words        Suffixes   Letter.
      0.7380810       0.7495701       0.6746377       0.5531690       0.5209648       0.4262465       0.4
>  1- 1/diag(R.inv)
      Sentences      Vocabulary Sent.Completion   First.Letters  4.Letter.Words        Suffixes   Letter.
      0.7380810       0.7495701       0.6746377       0.5531690       0.5209648       0.4262465       0.4
```

Beta weights are merely one possible way to weight a predictor set

Consider a matrix, **R** of predictors (x_i) with the criterion (y).

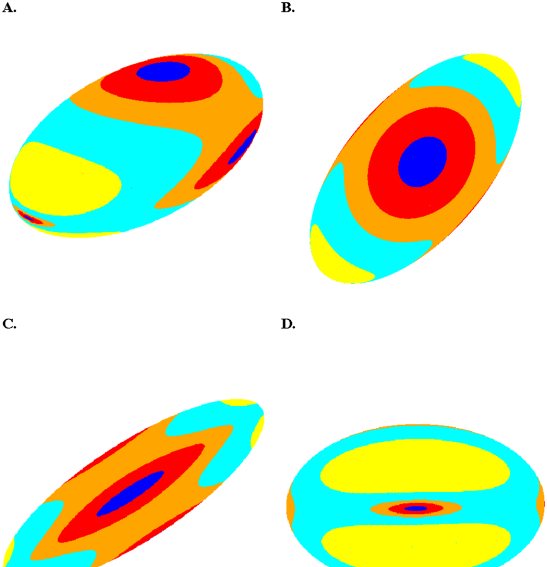
$$R = \left[\begin{array}{ccccc|c} r_{11} & r_{12} & r_{13} & \dots & r_{1n} & r_{1y} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2n} & r_{2y} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & r_{nn} & r_{ny} \\ \hline r_{y1} & r_{y2} & r_{y3} & \dots & r_{yn} & r_{yy} \end{array}\right] \quad R_{(x_1 x_2 \dots x_n)y} = \frac{\sum r_{x_i y}}{\sqrt{(\sum\sum r_{ij}) r_{yy}}}$$

Multiply each x_i by a weight β_i and then find $R_{(x_1 x_2 \dots x_n)y}$

$$R = \left[\begin{array}{ccccc|c} \beta_1 r_{11} \beta_1 & \beta_1 r_{12} \beta_2 & \beta_1 r_{13} \beta_3 & \dots & \beta_1 r_{1n} \beta_n & \beta_1 r_{1y} \\ \beta_2 r_{21} \beta_1 & \beta_2 r_{22} \beta_1 & \beta_3 r_{23} \beta_3 & \dots & \beta_2 r_{2n} \beta_n & \beta_2 r_{2y} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \beta_n r_{n1} \beta_1 & \beta_n r_{n2} \beta_1 & \beta_n r_{n3} \beta_3 & \dots & \beta_n r_{nn} \beta_n & \beta_n r_{ny} \\ \hline r_{y1} \beta_1 & r_{y2} \beta_2 & r_{y3} \beta_3 & \dots & r_{yn} \beta_n & r_{yy} \end{array}\right]$$

Four views of effect of Fungible weights (Waller, 2008; Waller & Jones, 2010)

blue: $R^2_\Delta \leq .01$ Red: $.01 < R^2_\Delta \leq .05$ Orange: $.05 \leq R^2_\Delta \leq .10$



Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Tal-Or, N., Cohen, J., Tsfati, Y., & Gunther, A. C. (2010). Testing causal direction in the influence of presumed media influence. *Communication Research*, *37*(6), 801–824.

Waller, N. G. (2008). Fungible weights in multiple regression. *Psychometrika*, *73*(4), 691–703.

Waller, N. G. & Jones, J. A. (2010). Correlation weights in multiple regression. *Psychometrika*, *75*(1), 58–69.