Psychology 405: Psychometric Theory
Course summary and review

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

NORTHWESTERN
UNIVERSITY

June, 2025

Outline

What is psychometrics?

We hardly recognize a subject as scientific if measurement is not one of its tools (Boring, 1929)

There is yet another [method] so vital that, if lacking it, any study is thought ... not be scientific in the full sense of the word. This further an crucial method is that of measurement. (Spearman, 1937)

One's knowledge of science begins when he can measure what he is speaking about and express in numbers (Eysenck, 1973)

Psychometrics: the assigning of numbers to observed psychological phenomena and to unobserved concepts.
Evaluation of the fit of theoretical models to empirical data.

Psychometric Theory: Data = Model + Residual

1. A summary of models of measurement
   - Statistics are smooths (models) of data
   - Models are idealized representations of data.
   - Models are projections of data from a higher order space into a lower order space.
   - Lack of model fit (Data-Model) is Error (for that model) and is Residual that needs to be explained
   - Models differ in complexity and fit
2. Review the models presented
   - Data = Model + Error = Model + Residual
   - Error = Data - Model
   - Residual = Data - Model
   - Observations = f(Model) + error
   - Model = $f^{-1}(Observations) + residual$
   - The problem is to find the inverse operator!

Data = Model + Residual

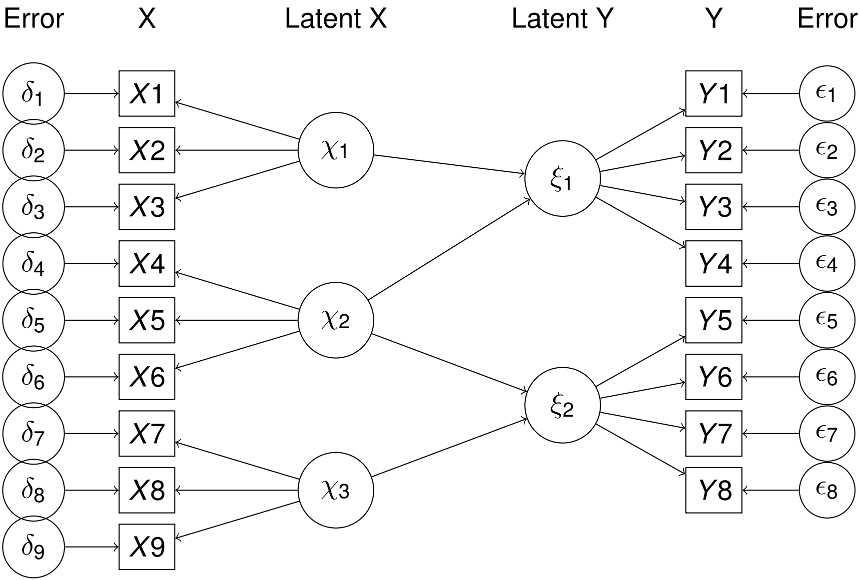
- The fundamental equations of statistics are that
  - Data = Model + Residual
  - Residual = Data - Model
- The problem is to specify the model and then evaluate the fit of the model to the data as compared to other models
  - Fit = f(Data, Residual)
  - Typically: $Fit = f(1 - \frac{Residual^2}{Data^2})$
  - $Fit = f(\frac{(Data-Model)^2}{Data^2})$
- Even for something as simple as the mean is a model of the data. The residual left over after we remove the mean is the variance.

Psychometrics as model estimation and model fitting

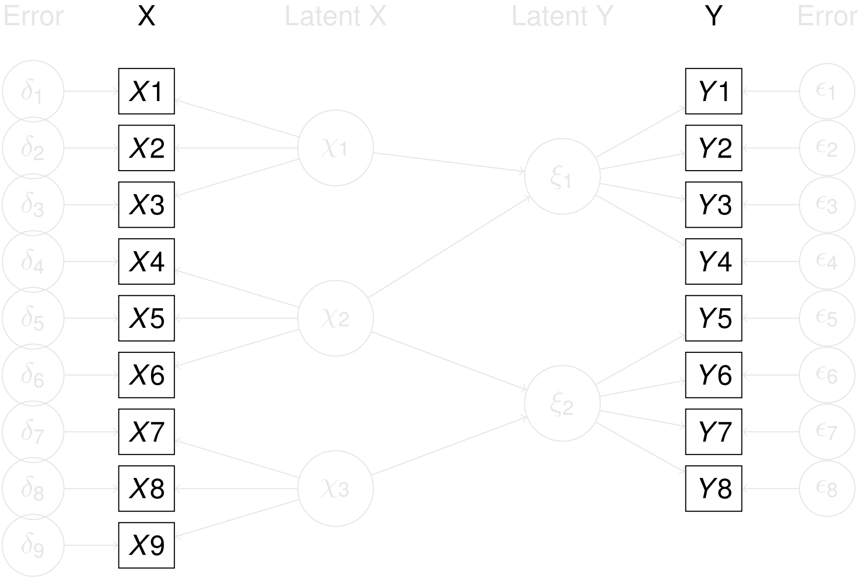
We explored a number of models

1. Modeling the process of data collection and of scaling
   - $X = f(\theta)$
   - How to measure X, properties of the function f.
2. Correlation and Regression
   - $Y = \beta X$
   - $R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
3. Factor Analysis and Principal Components Analysis
   - $R = FF' + U^2 \quad R = CC'$
4. Reliability $\rho_{xx} = \frac{\sigma^2_\theta}{\sigma^2_X}$
5. Item Response Theory
   - $p(X|\theta, \delta) = f(\theta - \delta)$
6. Structural Equation Modeling
   - $\rho_{yy} Y = \beta \rho_{xx} X$

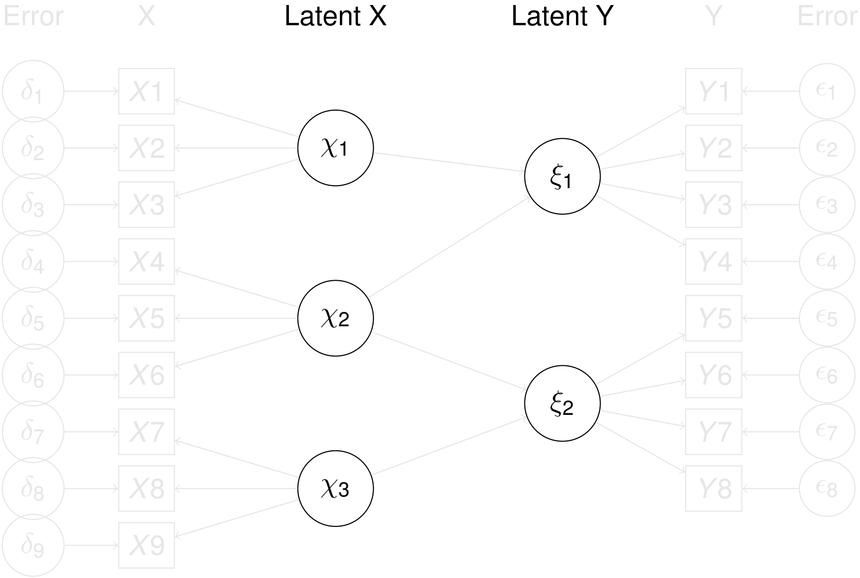
Psychometric Theory: A conceptual Syllabus



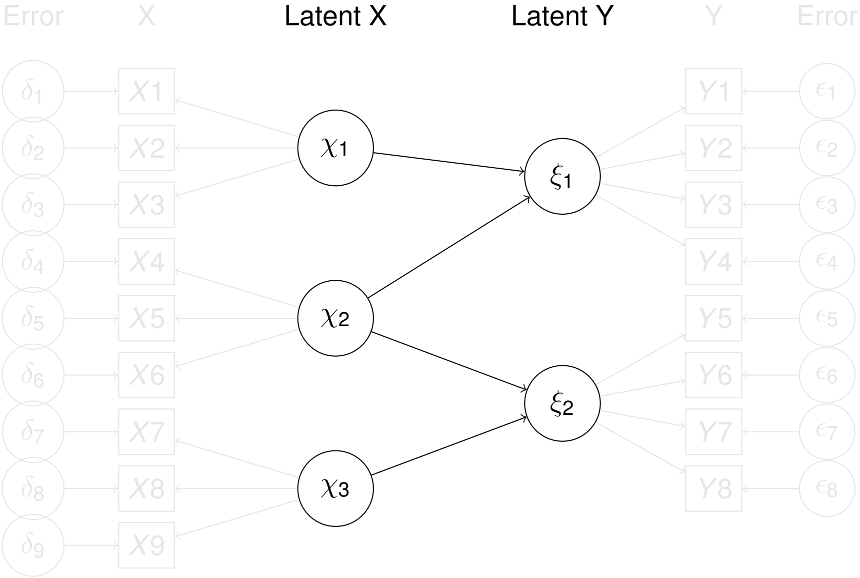
Observed Variables



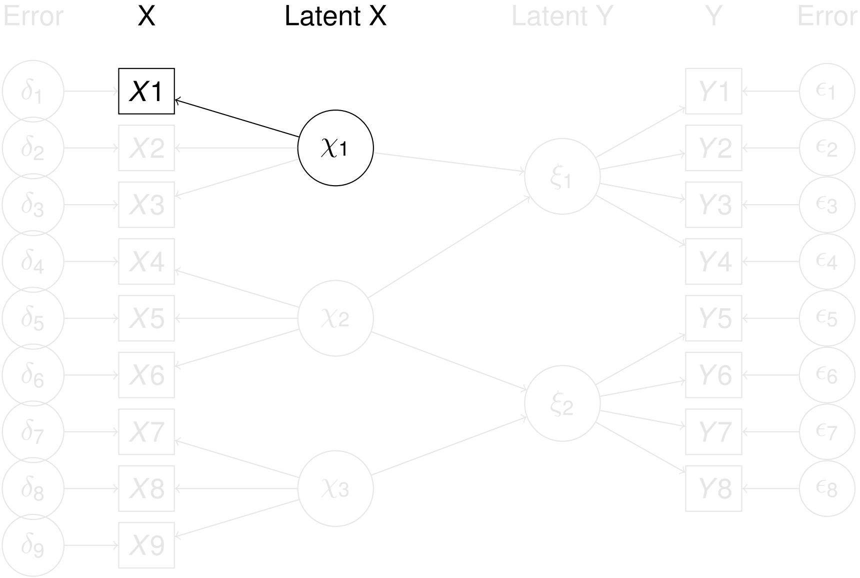
Latent Variables



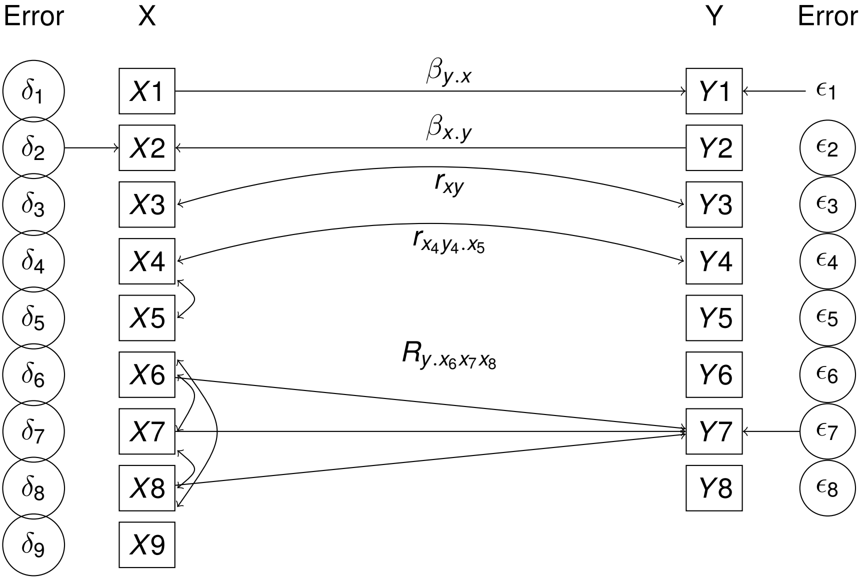
Theory


Error
X
Latent X
Latent Y
Y
Error
δ_1
δ_2
δ_3
δ_4
δ_5
δ_6
δ_7
δ_8
δ_9
X1
X2
X3
X4
X5
X6
X7
X8
X9
χ_1
χ_2
χ_3
ξ_1
ξ_2
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
ϵ_1
ϵ_2
ϵ_3
ϵ_4
ϵ_5
ϵ_6
ϵ_7
ϵ_8

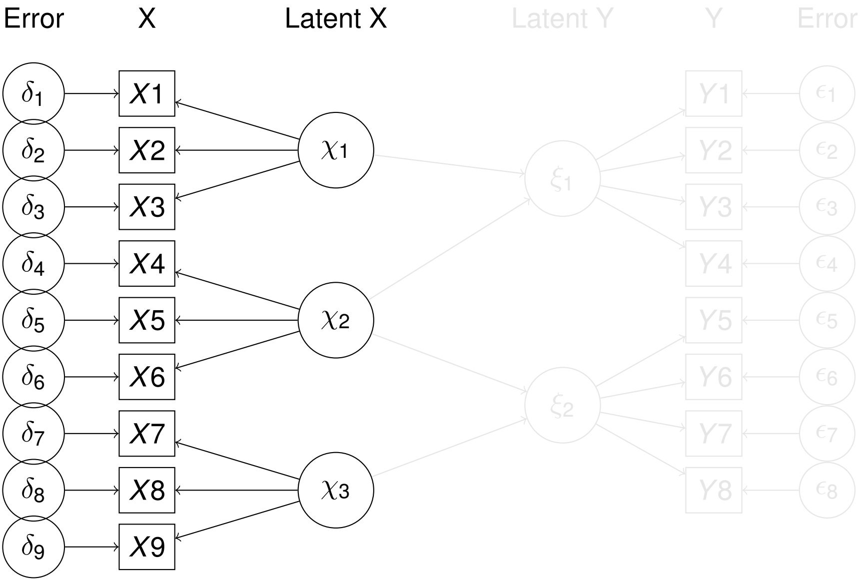
A theory of data and fundamentals of scaling



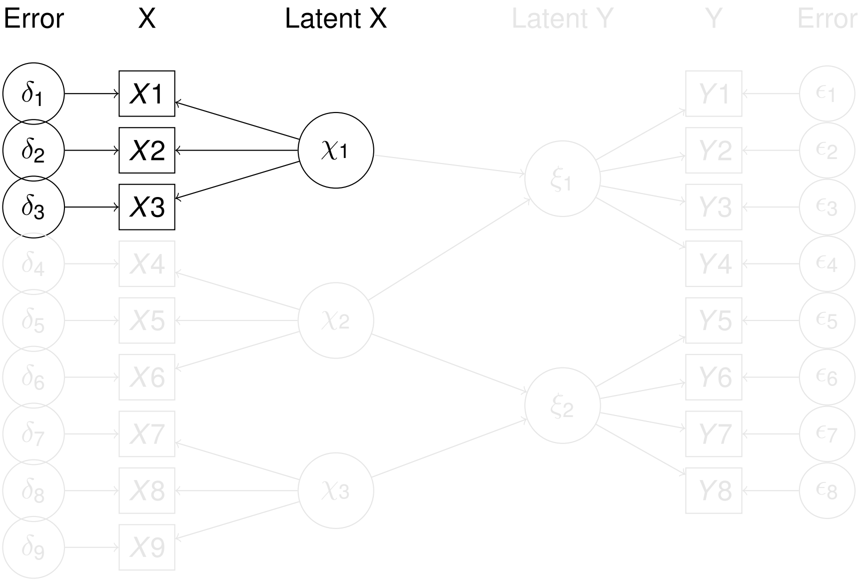
Correlation, Regression, Partial Correlation, Multiple Regression



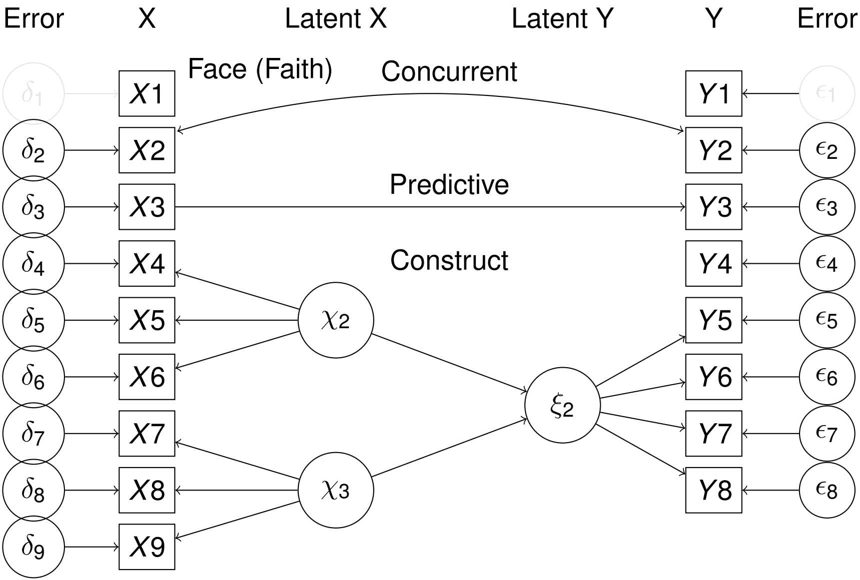
Measurement: A latent variable approach.



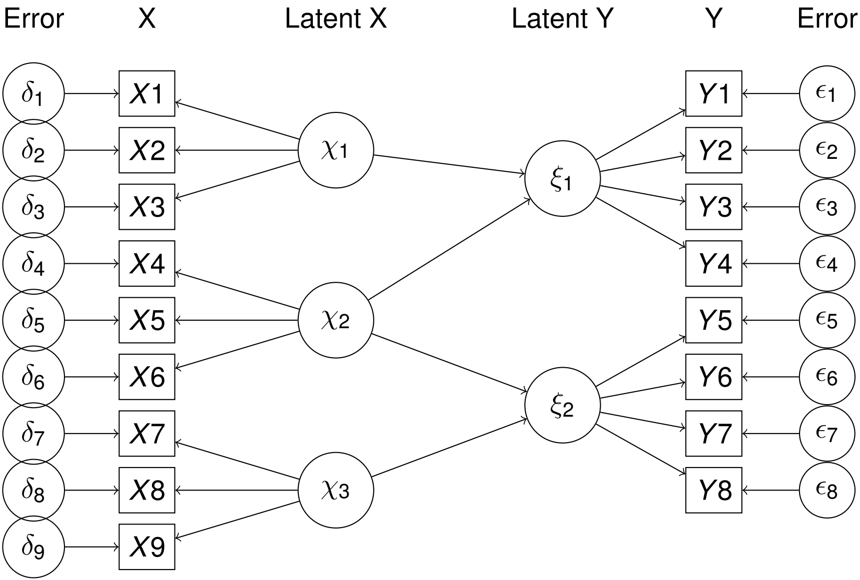
Reliability: How well does a test reflect one latent trait?



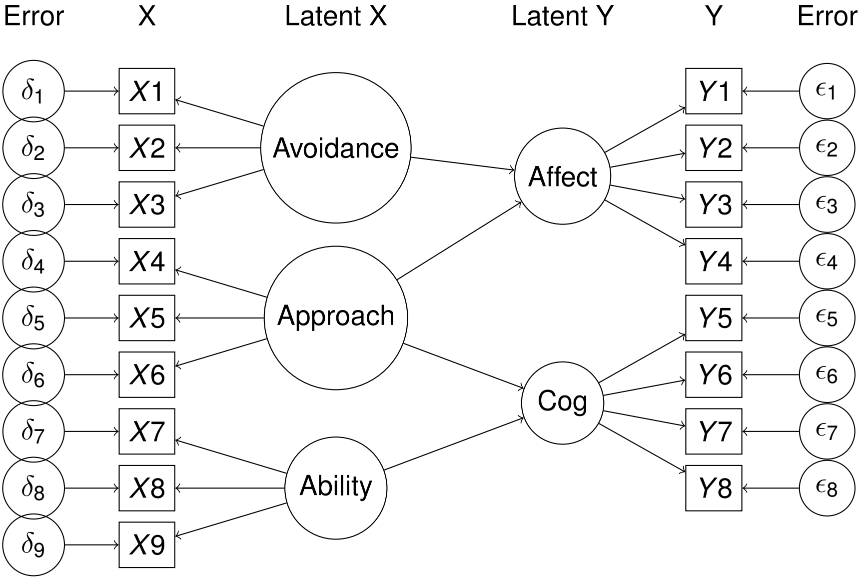
Face, Concurrent, Predictive, Consruct



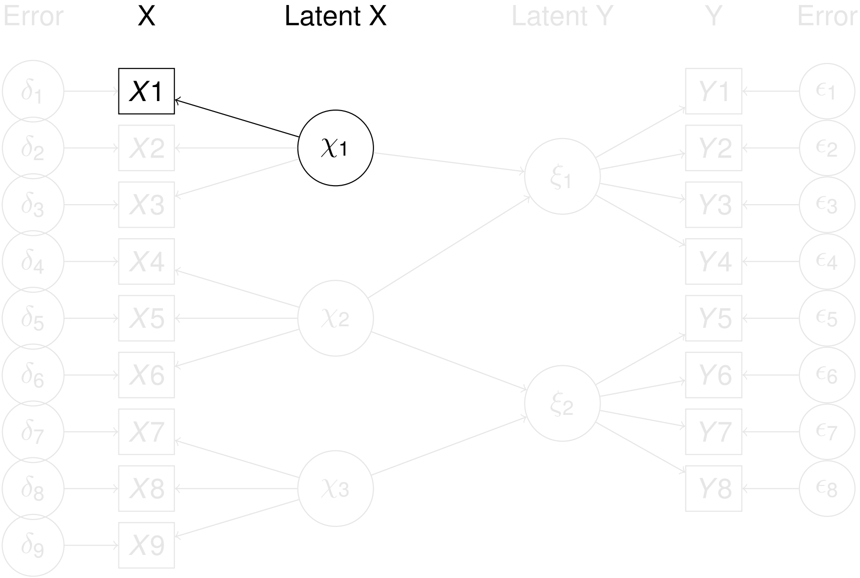
Psychometric Theory: Data, Measurement, Theory



Psychometric Theory: Data, Measurement, Theory



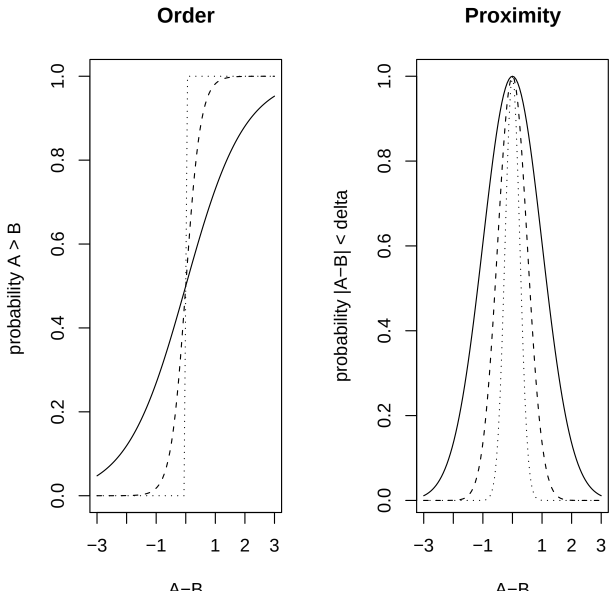
A theory of data and fundamentals of scaling



Clyde Coombs and the Theory of Data

1. O = the set of objects
   - $O = \{o_i, o_j \dots o_n\}$
2. S = the set of Individuals
   - $S = \{s_i, s_j \dots s_n\}$
3. Two comparison operations
   - order ($x > y$)
   - proximity ($|x - y| < \epsilon$
4. Two types of comparisons
   - Single dyads
     - (s_i, s_j) (s_i, o_j) (o_i, o_j)
   - Pairs of dyads
     - $(s_i, s_j)(s_k, s_l)$ $(s_i, o_j)(s_k, o_l)$ $(o_i, o_j)(o_k, o_l)$

2 types of comparisons: Monotone ordering and single peak proximity



Theory of Data and types of measures

Table: The theory of data provides a 3 x 2 x 2 taxonomy for various types of measures

Elements of Dyad	Number of Dyads	Comparison	Name
People x People	1	Order	Tournament rankings
People x People	1	Proximity	Social Networks
Objects x Objects	1	Order	Scaling
Objects x Objects	1	Proximity	Similarities
People x Objects	1	Order	Ability Measurement
People x Objects	1	Proximity	Attitude Measurement
People x People	2	Order	Tournament rankings
People x People	2	Proximity	Social Networks
Objects x Objects	2	Order	Scaling
Objects x Objects	2	Proximity	Multidimensional scaling
People x Objects	2	Order	Ability Comparisons
People x Objects	2	Proximity	Preferential Choice
People x Objects x Objects	2	Proximity	Individual Differences in Multidimensional Scaling

Thurstone's Vegetable data as an example of one dimensional scaling

Table: Consider the likelihood of liking a vegetable. Numbers reflect probability that the column is preferred to the row. Can we turn this into a scale?

The `veg` data set from the psych package in R

Variable	Turn	Cab	Beet	Asp	Car	Spin	S.Beans	Peas	Corn
Turn	0.50	0.82	0.77	0.81	0.88	0.89	0.90	0.89	0.93
Cab	0.18	0.50	0.60	0.72	0.74	0.74	0.81	0.84	0.86
Beet	0.23	0.40	0.50	0.56	0.74	0.68	0.84	0.80	0.82
Asp	0.19	0.28	0.44	0.50	0.56	0.59	0.68	0.60	0.73
Car	0.12	0.26	0.26	0.44	0.50	0.49	0.57	0.71	0.76
Spin	0.11	0.26	0.32	0.41	0.51	0.50	0.63	0.68	0.63
S.Beans	0.10	0.19	0.16	0.32	0.43	0.37	0.50	0.53	0.64
Peas	0.11	0.16	0.20	0.40	0.29	0.32	0.47	0.50	0.63
Corn	0.07	0.14	0.18	0.27	0.24	0.37	0.36	0.37	0.50

#show the data from the veg data set from the psych package
veg

Data = Model + Residual

1. What is the model?
   - Pref = Mean (preference)
   - $p(A > B) = f(A, B)$
   - what is f?
2. Possible functions
   - f = A - B (simple difference)
   - $\frac{A}{A+B}$ Luce choice rule
   - Thurstonian scaling
   - logistic scaling
3. Evaluating functions – Goodness of fit
   - Residual = Model - Data
   - Minimize residual
   - Minimize $residual^2$

Multidimensional Scaling: $(|o_i - o_j| < |o_k - o_l|)$

$$Distance_{xy} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \quad (1)$$

Consider the `cities` data set of airline distances.

```
> cities
```

	ATL	BOS	ORD	DCA	DEN	LAX	MIA	JFK	SEA	SFO	MSY
ATL	0	934	585	542	1209	1942	605	751	2181	2139	424
BOS	934	0	853	392	1769	2601	1252	183	2492	2700	1356
ORD	585	853	0	598	918	1748	1187	720	1736	1857	830
DCA	542	392	598	0	1493	2305	922	209	2328	2442	964
DEN	1209	1769	918	1493	0	836	1723	1636	1023	951	1079
LAX	1942	2601	1748	2305	836	0	2345	2461	957	341	1679
MIA	605	1252	1187	922	1723	2345	0	1092	2733	2594	669
JFK	751	183	720	209	1636	2461	1092	0	2412	2577	1173
SEA	2181	2492	1736	2328	1023	957	2733	2412	0	681	2101
SFO	2139	2700	1857	2442	951	341	2594	2577	681	0	1925
MSY	424	1356	830	964	1079	1679	669	1173	2101	1925	0

A two dimensional solution of the airline distances

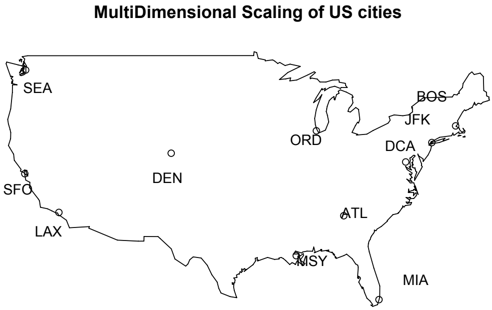
```
> city.location <- cmdscale(cities, k=2)
> plot(city.location,type="n", xlab="Dimension␣1",
   ylab="Dimension␣2",main ="cmdscale(cities)")
> text(city.location,labels=names(cities))
> round(city.location,0)
```

	[,1]	[,2]
ATL	-571	248
BOS	-1061	-548
ORD	-264	-251
DCA	-861	-211
DEN	616	10
LAX	1370	376
MIA	-959	708
JFK	-970	-389
SEA	1438	-607
SFO	1563	88
MSY	-301	577

1. Use the `cmdscale` function to do multidimensional scaling, ask for a 2 dimensional solution
2. Plot the results (don't actually show the points)
3. Add the names of the cities
4. Show the numeric results

Revised solution for 11 US cities after making `city.location <- -city.location` and adding a US map.

The correct locations of the cities are shown with circles. The MDS solution is the center of each label. The central cities (Chicago, Atlanta, and New Orleans are located very precisely, but Boston, New York and Washington, DC are north and west of their correct locations.



Four types of scales and their associated statistics

Table: Four types of scales and their associated statistics (Rossi, 2007; Stevens, 1946) The statistics listed for a scale are invariant for that type of transformation.

Scale	Basic operations	Transformations	Invariant statistic	Examples
Nominal	equality $x_i = x_j$	Permutations	Counts Mode χ^2 and (ϕ) correlation	Detection Species classification Taxons
Ordinal	order $x_i > x_j$	Monotonic (homeomorphic) x' =f(x) f is monotonic	Median Percentiles Spearman correlations*	Mhos Hardness scale Beaufort Wind (intensity) Richter earthquake scale
Interval	differences $(x_i - x_j) > (x_k - x_l)$	Linear (Affine) x' = a + bx	Mean (μ) Standard Deviation (σ) Pearson correlation (r) Regression (β)	Temperature (°F, °C) Beaufort Wind (velocity)
Ratio	ratios $\frac{x_i}{x_j} > \frac{x_k}{x_l}$	Multiplication (Similiarity) x' = bx	Coefficient of variation ($\frac{\sigma}{\mu}$)	Length, mass, time Temperature (°K) Heating degree days

The Beaufort wind speed scale is interval with respect to the velocity of the wind, but only ordinal with respect to the effect of the wind. The Richter scale of earthquake intensity is a logarithmic scale of the energy released but linear measure of

Multiple measures of central tendency

mode The most frequent observation. Not a very stable measure, depends upon grouping. Can be used for categorical data.

median The number with 50% above and 50%below. A powerful, if underused, measure. Not sensitive to transforms of the shape of the distribution, nor outliers. Appropriate for ordinal data, and useful for interval data.

mean One of at least seven measures that assume interval properties of the data.

Multiple ways to estimate the mean

Arithmetic mean $\bar{X} = X_{.} = (\sum_{i=1}^{N} X_i)/N$ `mean`(x)

Trimmed mean throws away the top and bottom t% of observations. This follows the principle that all data are normal at the middle. `mean`(x,trim=.1)

Winsorized mean Find the arithmetic mean after replacing the n lowest observations with the nth value, and the N largest values with the Nth largest. `winsor`(x,trim=.2)

Geometric Mean $\bar{X}_{geometric} = \sqrt[N]{\prod_{i=1}^{N} X_i} = e^{\Sigma(ln(x))/N}$ (The anti-log of the mean log score). `geometric.mean`(x)

Harmonic Mean $\bar{X}_{harmonic} = \frac{N}{\sum_{i=1}^{N} 1/X_i}$ (The reciprocal of the mean reciprocal). `harmonic.mean`(x)

Circular Mean $\bar{x}_{circular} = tan^{-1}\left(\frac{\sum cos(x)}{\sum sin(x)}\right)$ `circular.mean`(x) (where x is in radians)

circadian.mean `circular.mean`(x) (where x is in hours)

Circular statistics

Table: Hypothetical mood data from six subjects for four mood variables. The values reflect the time of day that each scale achieves its maximum value for each subject. Each mood variable is just the previous one shifted by 5 hours. Note how this structure is preserved for the *circular mean* but not for the arithmetic mean.

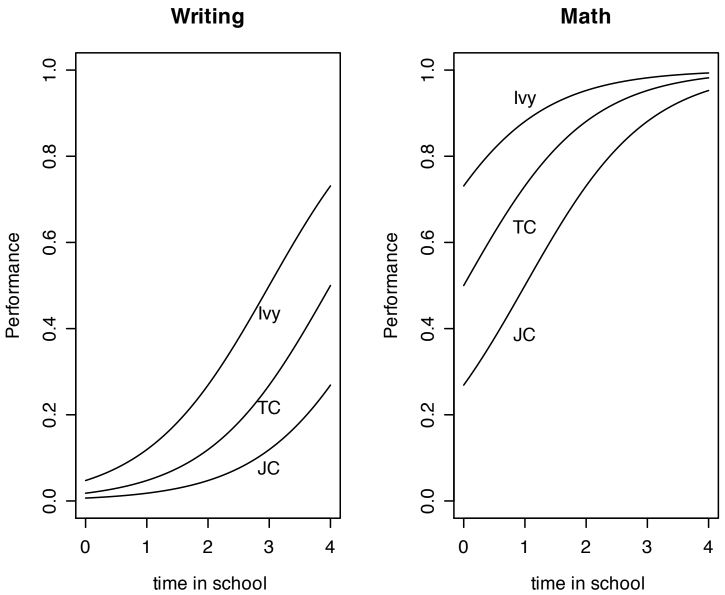
Subject	Energetic Arousal	Positive Affect	Tense Arousal	Negative Affect
1	9	14	19	24
2	11	16	21	2
3	13	18	23	4
4	15	20	1	6
5	17	22	3	8
6	19	24	5	10
Arithmetic Mean	14	19	12	9
Circular Mean	14	19	24	5

What is the “average" class size?

Table: Average class size depends upon point of view. For the faculty members, the median of 10 is very appealing. From the Dean’s perspective, the faculty members teach an average of 50 students per calls. But what about the students?

Faculty Member	Freshman/ Sophmore	Junior	Senior	Graduate	Mean	Median
A	20	10	10	10	12.5	10
B	20	10	10	10	12.5	10
C	20	10	10	10	12.5	10
D	20	100	10	10	35.0	15
E	200	100	400	10	177.5	150
Total						
Mean	56	46	110	10	50.0	39
Median	20	10	10	10	12.5	10

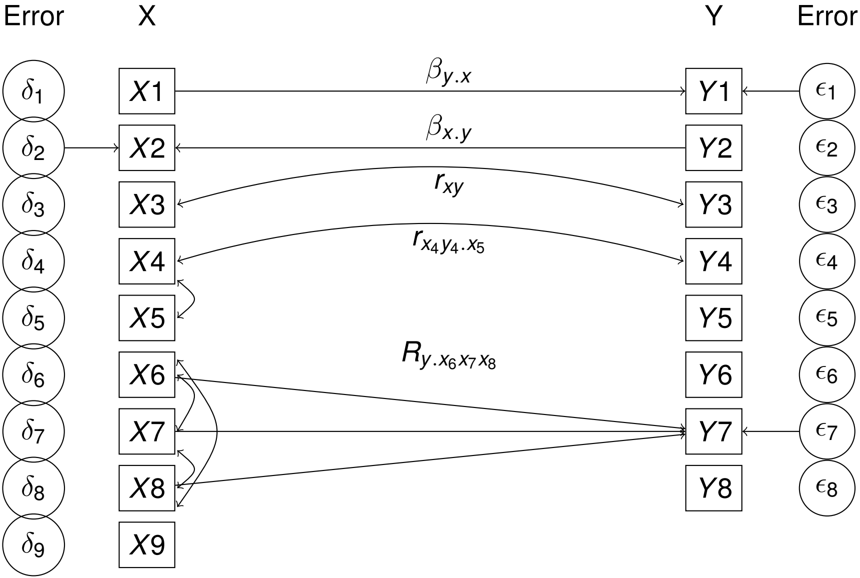
Effect of teaching, effect of students, or just scaling?



The best scale is the one that works best

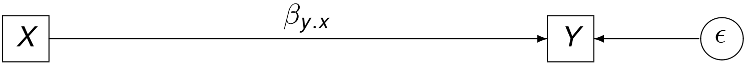
1. Money is linear but negatively accelerated with utility.
2. Perceived intensity is a log function of physical intensity.
3. Probabilty of being correct is a logistic or cumulative normal function of ability.
4. Energy used to heat a house is linear function of outdoor temperature.
5. Time to fall a particular distance varies as the square root of the distance.
6. Gravitational attraction varies as $1/distance^2$
7. Hull speed of sailboat varies as square root of length of boat.
8. Sound intensity in db is log(observed/reference)
9. pH of solutions is -log(concentration of hydrogen ions)

Correlation, Regression, Partial Correlation, Multiple Regression



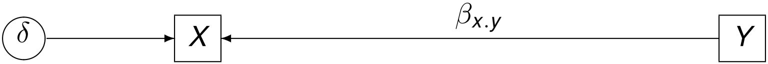
Bivariate Regression

δ X Y ϵ



$$\hat{y} = \beta_{y.x} x + \epsilon$$

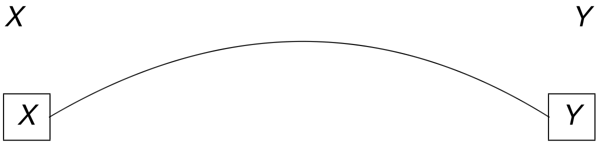
$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$$



$$\hat{x} = \beta_{x.y} y + \delta$$

$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_y^2}$$

Bivariate Correlation



$$\hat{x} = \beta_{x.y} y + \delta$$

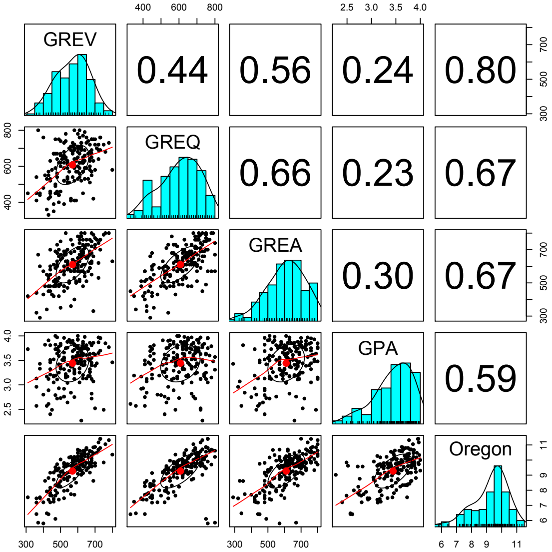
$$\hat{y} = \beta_{y.x} x + \epsilon$$

$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_y^2}$$

$$\beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma^{2x} \sigma_y^2}}$$

Scatter Plot Matrix showing correlation and LOESS regression

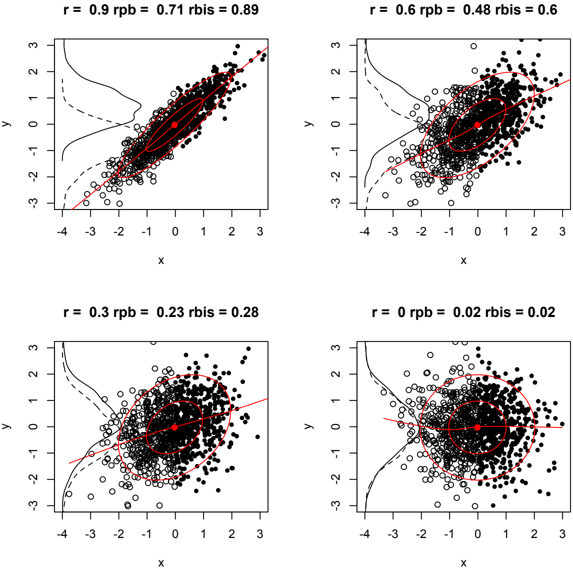


Alternative versions of the correlation coefficient

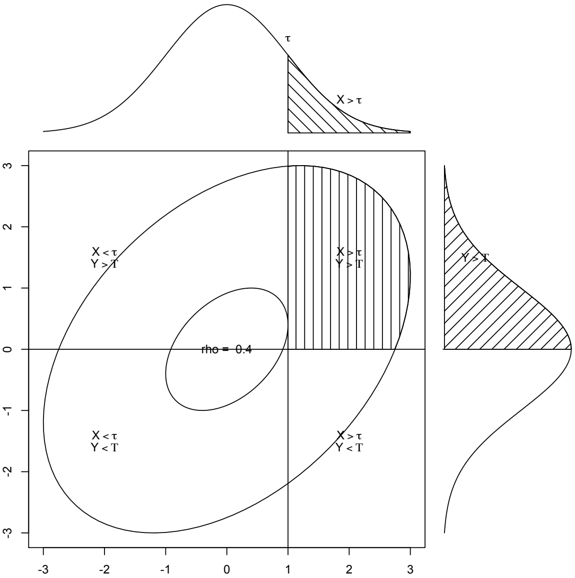
Table: A number of correlations are Pearson r in different forms, or with particular assumptions. If $r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$, then depending upon the type of data being analyzed, a variety of correlations are found.

Coefficient	symbol	X	Y	Assumptions
Pearson	r	continuous	continuous	
Spearman	rho (ρ)	ranks	ranks	
Point bi-serial	r_{pb}	dichotomous	continuous	
Phi	ϕ	dichotomous	dichotomous	
Bi-serial	r_{bis}	dichotomous	continuous	normality
Tetrachoric	r_{tet}	dichotomous	dichotomous	bivariate normality
Polychoric	r_{pc}	categorical	categorical	bivariate normality

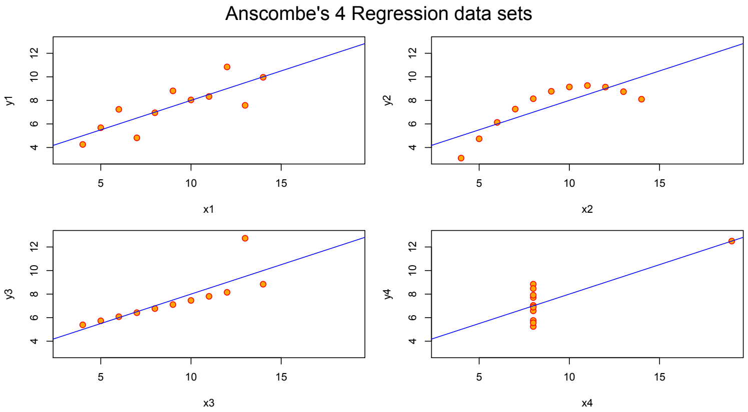
The biserial correlation estimates the latent correlation



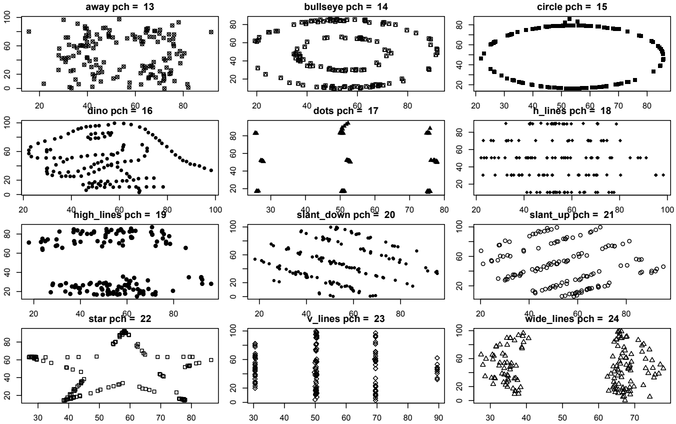
The tetrachoric correlation estimates the latent correlation



Cautions about correlations: Anscombe data set



Cautions about correlations: Always graph the data



The ubiquitous correlation coefficient

Table: Alternative Estimates of effect size. Using the correlation as a scale free estimate of effect size allows for combining experimental and correlational data in a metric that is directly interpretable as the effect of a standardized unit change in x leads to r change in standardized y.

Statistic	Estimate	r equivalent	as a function of r
Pearson correlation	$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$	r_{xy}	
Regression	$b_{y.x} = \frac{Cxy}{\sigma_x^2}$	$r = b_{y.x} \frac{\sigma_x}{\sigma_y}$	$b_{y.x} = r \frac{\sigma_y}{\sigma_x}$
Cohen's d	$d = \frac{X_1 - X_2}{\sigma_x}$	$r = \frac{d}{\sqrt{d^2+4}}$	$d = \frac{2r}{\sqrt{1-r^2}}$
Hedge's g	$g = \frac{X_1 - X_2}{s_x}$	$r = \frac{g}{\sqrt{g^2+4(df/N)}}$	$g = \frac{2r\sqrt{df/N}}{\sqrt{1-r^2}}$
t - test	$t = \frac{d\sqrt{df}}{2}$	$r = \sqrt{t^2/(t^2+df)}$	$t = \sqrt{\frac{r^2 df}{1-r^2}}$
F-test	$F = \frac{d^2 df}{4}$	$r = \sqrt{F/(F+df)}$	$F = \frac{r^2 df}{1-r^2}$
Chi Square		$r = \sqrt{\chi^2/n}$	$\chi^2 = r^2 n$
Odds ratio	$d = \frac{ln(OR)}{1.81}$	$r = \frac{ln(OR)}{1.81\sqrt{(ln(OR)/1.81)^2+4}}$	$ln(OR) = \frac{3.62r}{\sqrt{1-r^2}}$
$r_{equivalent}$	r with probability p	$r = r_{equivalent}$	

Partial R: Correlations without the effects of 3rd or 4th variables

1. We sometimes want to find the relationship between X and Y controlling for the effects of Z
2. This logically is the correlation of X predicted by Z with Y predicted by Z
3. The matrix of partial correlations is just the negative of the inverse with the diagonal replaced by the negative of the diagonal.
4. Or, just use `partial.r`

R code

```
R <- sim.congeneric()
R.inv <- -solve(R)
diag(R.inv) <- -diag(R.inv)
lowerMat(cov2cor(R.inv))
```

Partial R: Correlations without the effects of 3rd or 4th variables

R code

```
R <- sim.congeneric()
lowerMat(R)   #show it
R.inv <- -solve(R)
diag(R.inv) <- -diag(R.inv)
lowerMat(cov2cor(R.inv))   #show the solution
```

```
lowerMat(R)   #show it
   V1   V2   V3   V4
V1 1.00
V2 0.56 1.00
V3 0.48 0.42 1.00
V4 0.40 0.35 0.30 1.00
> R.inv <- -solve(R)
> diag(R.inv) <- -diag(R.inv)
> lowerMat(cov2cor(R.inv))   #show the solution
   V1   V2   V3   V4
V1 1.00
V2 0.40 1.00
V3 0.29 0.19 1.00
V4 0.22 0.14 0.10 1.00
```

A problem with partial R: reliability

1. Partial r corrects for the effect of the other variables, but does not correct for the reliability of the variables.
2. This leads to partial rs that are non-zero, even though the underlying model shows no relationship when the factor is removed.
3. We consider the example from the homework for reliability and correlation. (help page for esem)

R code

```
fx <-matrix(c( .9,.8,.6,rep(0,4),.6,.8,-.7),ncol=2)
 fy <- matrix(c(.6,.5,.4),ncol=1)
 rownames(fx) <- c("V","Q","A","nach","Anx")
 rownames(fy)<- c("gpa","Pre","MA")
 Phi <-matrix( c(1,0,.7,.0,1,.7,.7,.7,1),ncol=3)
 gre.gpa <- sim.structural(fx,Phi,fy)
 gre.gpa
partial.gre.gpa <- partial.r(gre.gpa$model)
lowerMat(partial.gre.gpa)
#compare to what happens if we correct for reliability
corrected<- gre.gpa$model
diag(corrected) <- gre.gpa$reliability
lowerMat(cov2cor(corrected))  #this divides by the square roots of the
```

Compare the results

See the help page for `esem`

```
 gre.gpa
Call: sim.structural(fx = fx, Phi = Phi, fy = fy)

 $model (Population correlation matrix)
        V    Q    A nach   Anx   gpa   Pre    MA
V    1.00 0.72 0.54 0.00  0.00  0.38  0.32  0.25
Q    0.72 1.00 0.48 0.00  0.00  0.34  0.28  0.22
A    0.54 0.48 1.00 0.48 -0.42  0.50  0.42  0.34
nach 0.00 0.00 0.48 1.00 -0.56  0.34  0.28  0.22
Anx  0.00 0.00 -0.42 -0.56 1.00 -0.29 -0.24 -0.20
gpa  0.38 0.34 0.50 0.34 -0.29  1.00  0.30  0.24
Pre  0.32 0.28 0.42 0.28 -0.24  0.30  1.00  0.20
MA   0.25 0.22 0.34 0.22 -0.20  0.24  0.20  1.00

$reliability (population reliability)
   V    Q    A nach  Anx  gpa  Pre   MA
0.81 0.64 0.72 0.64 0.49 0.36 0.25 0.16
```

```
lowerMat(cov2cor(corrected)[1:6,1:6])
     V     Q     A     nach  Anx   gpa
V    1.00
Q    1.00  1.00
A    0.71  0.71  1.00
nach 0.00  0.00  0.71  1.00
Anx  0.00  0.00 -0.71 -1.00  1.00
gpa  0.70  0.70  0.99  0.70 -0.70  1.00

lowerMat(partial.r(gre.gpa$model)[1:6,1:6])
     V     Q     A     nach  Anx   gpa
V    1.00
Q    0.54  1.00
A    0.33  0.17  1.00
nach -0.19 -0.10  0.34  1.00
Anx  0.13  0.06 -0.23 -0.35  1.00
gpa  0.13  0.07  0.17  0.14 -0.09  1.00
```

Adjusting for reliability

R code

```
lowerMat(R)
lowerMat(r.adj)
```

```
lowerMat(R)
     V    Q    A    nach  Anx   gpa  Pre  MA
V    1.00
Q    0.72 1.00
A    0.54 0.48 1.00
nach 0.00 0.00 0.48 1.00
Anx  0.00 0.00 -0.42 -0.56 1.00
gpa  0.38 0.34 0.50 0.34 -0.29 1.00
Pre  0.32 0.28 0.42 0.28 -0.24 0.30 1.00
MA   0.25 0.22 0.34 0.22 -0.20 0.24 0.20 1.00
> lowerMat(R.adj)
Error in lower.tri(R, diag = TRUE) : object 'R.adj' not found
> lowerMat(r.adj)
     V    Q    A    nach  Anx   gpa  Pre  MA
V    1.00
Q    1.00 1.00
A    0.71 0.71 1.00
nach 0.00 0.00 0.71 1.00
Anx  0.00 0.00 -0.71 -1.00 1.00
gpa  0.70 0.70 0.99 0.70 -0.70 1.00
Pre  0.70 0.70 0.99 0.70 -0.70 1.00 1.00
MA   0.70 0.70 0.99 0.70 -0.70 1.00 1.00 1.00
```

Adjusting correlations for reliability and covariance analysis

R code

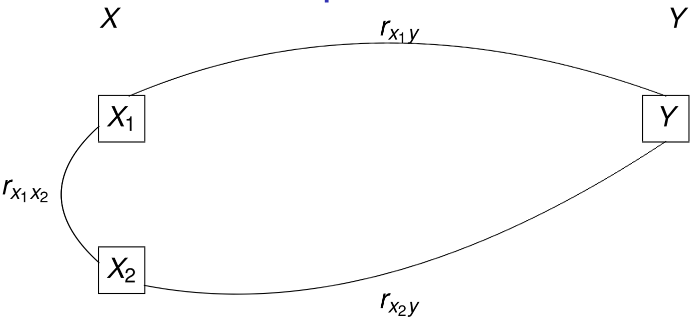
```
R <- gre.gpa$model
r <- R; diag(r) <- gre.gpa$reliability
r.adj <- cov2cor(r)
lowerMat(R)
```

```
     V     Q     A     nach  Anx   gpa   Pre   MA
V     1.00
Q     0.72  1.00
A     0.54  0.48  1.00
nach  0.00  0.00  0.48  1.00
Anx   0.00  0.00 -0.42 -0.56  1.00
gpa   0.38  0.34  0.50  0.34 -0.29  1.00
Pre   0.32  0.28  0.42  0.28 -0.24  0.30  1.00
MA    0.25  0.22  0.34  0.22 -0.20  0.24  0.20  1.00
```

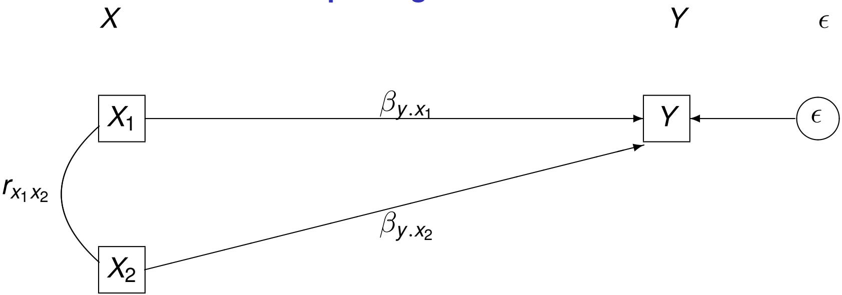
```
lowerMat(partial.r(R,1:5,6:8))
     V     Q     A     nach  Anx
V     1.00
Q     0.66  1.00
A     0.37  0.32  1.00
nach -0.23 -0.20  0.32  1.00
Anx   0.19  0.17 -0.28 -0.49  1.00
```

```
lowerMat(partial.r(r.adj,1:5,6:8))
     V     Q     A     nach  Anx
V     1.00
Q     1.00  1.00
A     0.11  0.11  1.00
nach -1.00 -1.00 -0.07  1.00
Anx   1.00  1.00  0.07 -1.00  1.00
```

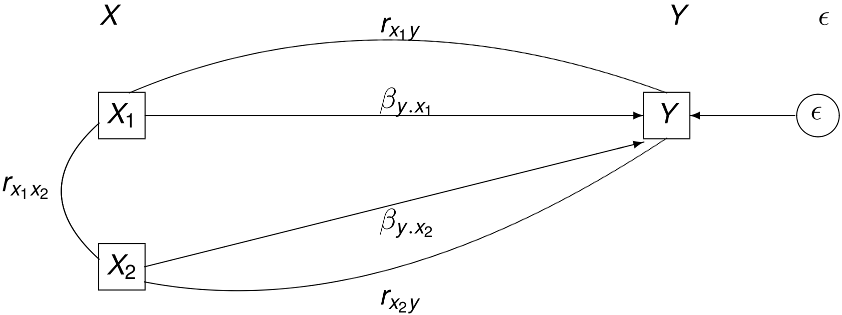
Multiple correlations



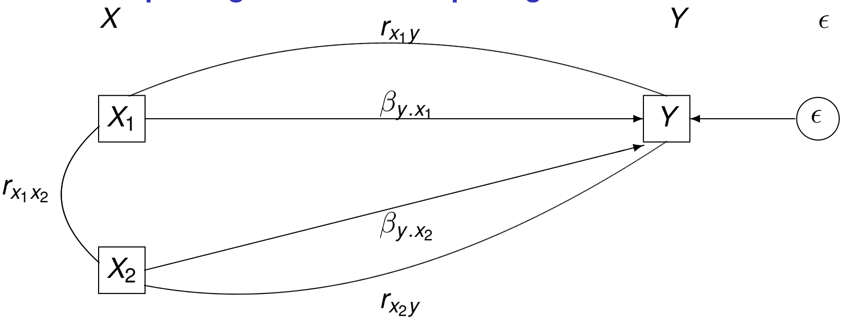
Multiple Regression



Multiple Regression: decomposing correlations



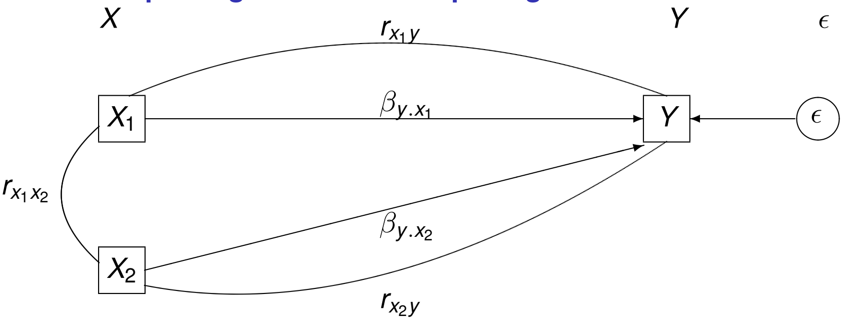
Multiple Regression: decomposing correlations



$$r_{x_1 y} = \overbrace{\beta_{y.x_1}}^{direct} + \overbrace{r_{x_1 x_2}\beta_{y.x_2}}^{indirect}$$

$$r_{x_2 y} = \underbrace{\beta_{y.x_2}}_{direct} + \underbrace{r_{x_1 x_2}\beta_{y.x_1}}_{indirect}$$

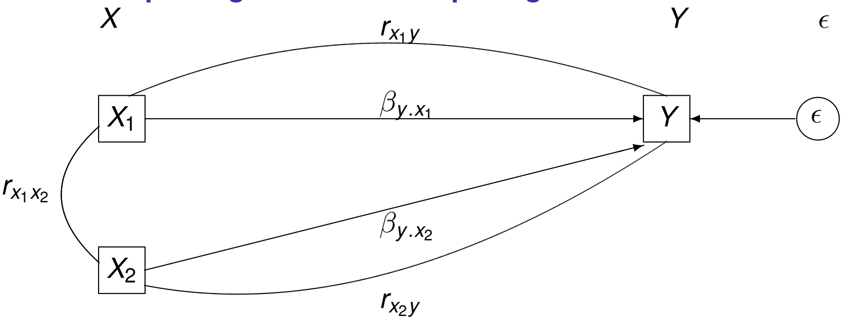
Multiple Regression: decomposing correlations



$$r_{x_1y} = \overbrace{\beta_{y.x_1}}^{\text{direct}} + \overbrace{r_{x_1x_2}\beta_{y.x_2}}^{\text{indirect}} \qquad \beta_{y.x_1} = \frac{r_{x_1y} - r_{x_1x_2}r_{x_2y}}{1 - r_{x_1x_2}^2}$$

$$r_{x_2y} = \underbrace{\beta_{y.x_2}}_{\text{direct}} + \underbrace{r_{x_1x_2}\beta_{y.x_1}}_{\text{indirect}} \qquad \beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2}r_{x_1y}}{1 - r_{x_1x_2}^2}$$

Multiple Regression: decomposing correlations



$$r_{x_1 y} = \overbrace{\beta_{y.x_1}}^{\text{direct}} + \overbrace{r_{x_1 x_2}\beta_{y.x_2}}^{\text{indirect}}$$

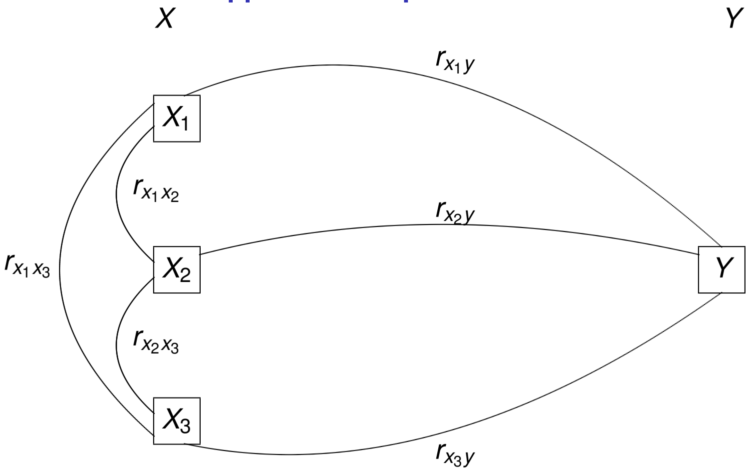
$$r_{x_2 y} = \underbrace{\beta_{y.x_2}}_{\text{direct}} + \underbrace{r_{x_1 x_2}\beta_{y.x_1}}_{\text{indirect}}$$

$$\beta_{y.x_1} = \frac{r_{x_1 y} - r_{x_1 x_2} r_{x_2 y}}{1 - r^2_{x_1 x_2}}$$

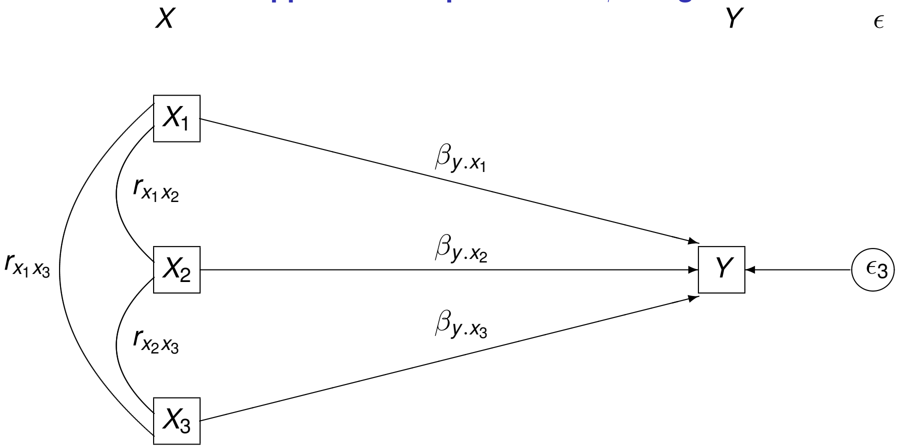
$$\beta_{y.x_2} = \frac{r_{x_2 y} - r_{x_1 x_2} r_{x_1 y}}{1 - r^2_{x_1 x_2}}$$

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

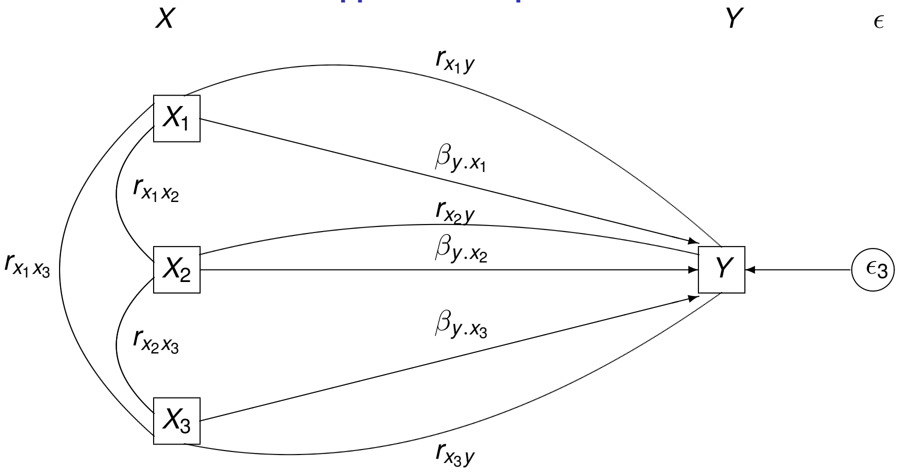
What happens with 3 predictors? The correlations



What happens with 3 predictors? β weights



What happens with 3 predictors?



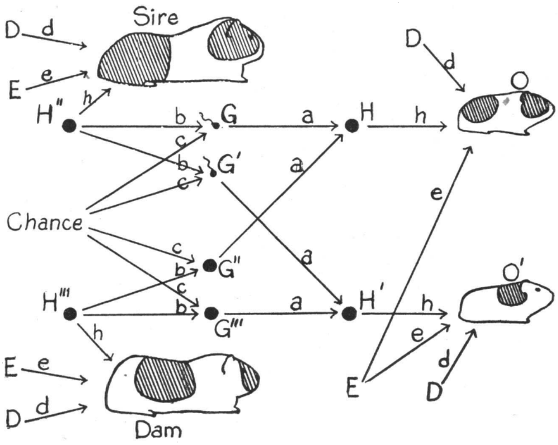
$$r_{x_1 y} = \underbrace{\beta_{y.x_1}}_{\color{blue}{direct}} + \underbrace{r_{x_1 x_2}\beta_{y.x_1} + r_{x_1 x_3}\beta_{y.x_3}}_{\color{red}{indirect}} \qquad r_{x_2 y} = \ldots \qquad r_{x_3 y} = \ldots$$

The math gets tedious

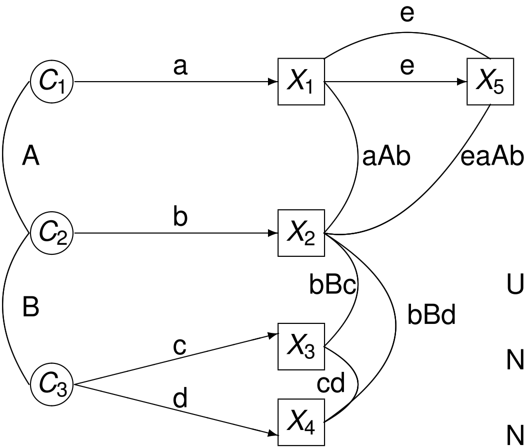
Multiple regression and matrix algebra

- Multiple regression requires solving multiple, simultaneous equations to estimate the direct and indirect effects.
    - Each equation is expressed as a $r_{x_i y}$ in terms of direct and indirect effects.
    - Direct effect is $\beta_{y.x_i}$
    - Indirect effect is $\sum_{j \neq i} beta_{y.x_j} r_{x_j y}$
- How to solve these equations?
- Tediously, or just use matrix algebra

Wright's Path model of inheritance in the Guinea Pig (Wright, 1921)



The basic rules of path analysis–think genetics



Up ... and over and down ...

No down and up

No double overs

Up ... and down ...

Parents cause children

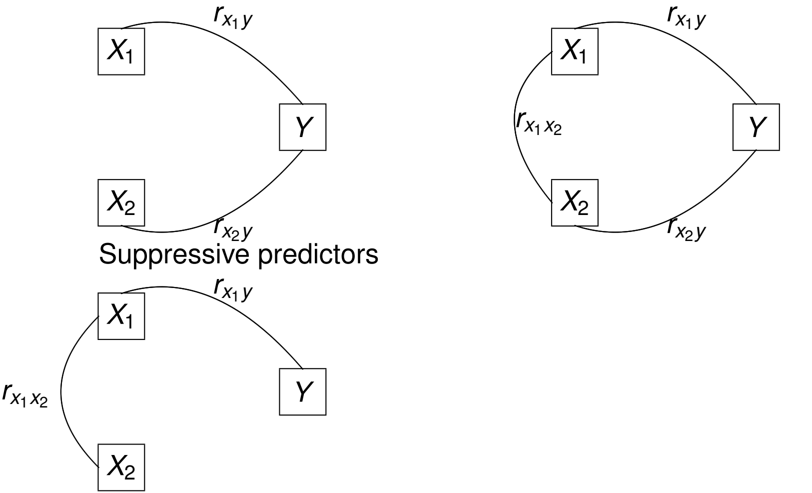
children do not cause parents

3 special cases of regression

Orthogonal predictors

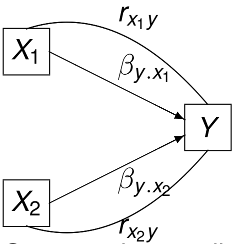
Correlated predictors

Suppressive predictors

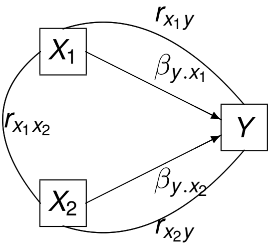


3 special cases of regression

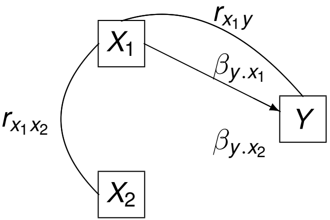
Orthogonal predictors



Correlated predictors



Suppressive predictors



$$\beta_{y.x_1} = \frac{r_{x_1 y} - r_{x_1 x_2} r_{x_2 y}}{1 - r^2_{x_1 x_2}}$$

$$\beta_{y.x_2} = \frac{r_{x_2 y} - r_{x_1 x_2} r_{x_1 y}}{1 - r^2_{x_1 x_2}}$$

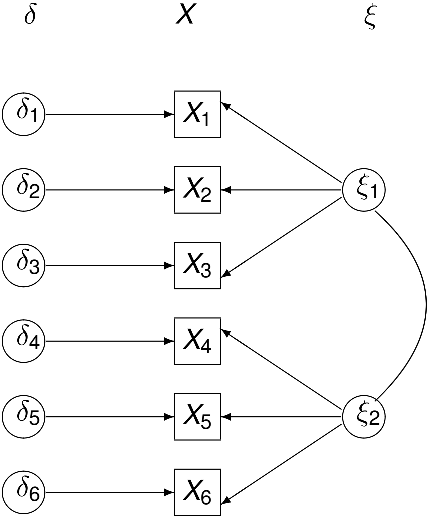
$$R^2 = r_{x_1 y}\beta_{y.x_1} + r_{x_2 y}\beta_{y.x_2}$$

Models of data

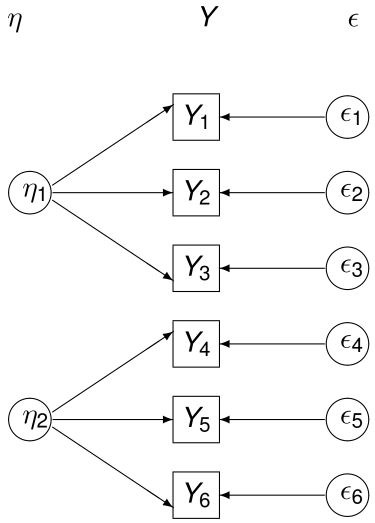
MacCallum et al. (2007) "A factor analysis model is not an exact representation of real-world phenomena.
Always wrong to some degree, even in population.
At best, model is an approximation of real world."
Box (1979): "Models, of course, are never true, but fortunately it is only necessary that they be useful. For this it is usually needful only that they not be grossly wrong."
Tukey (1961) "In a single sentence, the moral is: Admit that complexity always increases, first from the model you fit to the data, thence to the model you use to think and plan about the experiment and its analysis, and thence to the true situation."

(From MacCallum, 2004); http://www.fa100.info/maccallum2.pdf

A measurement model for X



A measurement model for Y



Various measurement models

1. Observed variables models
   - Singular Value Decompostion
   - Eigen Value – Eigen Vector decomposition
   - Principal Components
   - First k principal components as an approximation
2. Latent variable models
   - Factor analysis
3. Interpretation of models
   - Choosing the appropriate number of components/factors
   - Transforming/rotating towards interpretable structures
4. $R = FF' + U^2 \qquad R = CC'$

Eigen vector decomposition

Given a *nxn* matrix **R**, each eigenvector, $\mathbf{x_i}$, solves the equation

$$\mathbf{x_i R} = \lambda_i \mathbf{x_i}$$

and the set of n eigenvectors are solutions to the equation

$$\mathbf{XR} = \boldsymbol{\lambda}\mathbf{X}$$

where **X** is a matrix of orthogonal eigenvectors and $\boldsymbol{\lambda}$ is a diagonal matrix of the the eigenvalues, λ_i. Then

$$\mathbf{x_i R} - \lambda_i \mathbf{XI} = 0 <=> \mathbf{x_i}(\mathbf{R} - \lambda_i \mathbf{I}) = 0$$

Finding the eigenvectors and eigenvalues is computationally tedious, but may be done using the `eigen` function. That the vectors making up **X** are orthogonal means that

$$\mathbf{XX'} = \mathbf{I}$$

and because they form the *basis space* for **R** that

$$\mathbf{R} = \mathbf{X}\boldsymbol{\lambda}\mathbf{X'}.$$

From eigen vectors to Principal Components

1. For n variables, there are n eigen vectors
   - There is no parsimony in thinking of the eigen vectors
   - Except that the vectors provide the orthogonal basis for the variables
2. Principal components are formed from the eigen vectors and eigen values
   - $\mathbf{R} = \mathbf{V}\lambda\mathbf{V}' = \mathbf{C}\mathbf{C}'$
   - $\mathbf{C} = \mathbf{V}\sqrt{\lambda}$
3. But there will still be as many Principal Components as variables, so what is the point?
4. Take just the first k Principal Components and see how well this reduced model fits the data.

Factors vs. components

Originally developed by Spearman (1904) for the case of one common factor, and then later generalized by Thurstone (1947) and others to the case of multiple factors, factor analysis is probably the most frequently used and sometimes the most controversial psychometric procedure. The factor model, although seemingly very similar to the components model, is in fact very different. For rather than having components as linear sums of variables, in the factor model the variables are themselves linear sums of the unknown factors. That is, while components can be solved for by doing an *eigenvalue* or *singular value decomposition*, factors are estimated as best fitting solutions (Eckart and Young, 1936; Householder and Young, 1938), normally through iterative methods (Jöreskog, 1978; Lawley and Maxwell, 1963). Cattell (1965) referred to components analysis as a closed model and factor analysis as an open model, in that by explaining just the common variance, there was still more variance to explain.

The Thurstone 9 variable problem

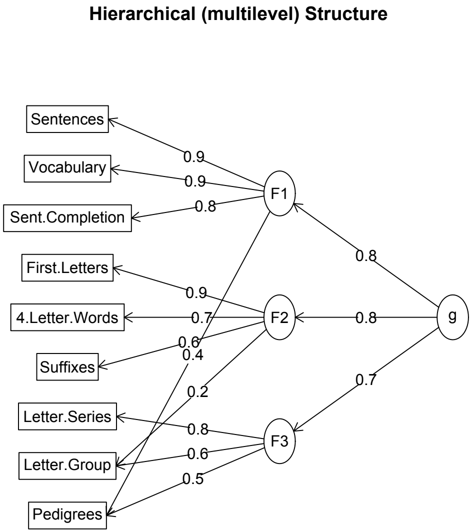
```
> lower.mat(Thurstone)
```

	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

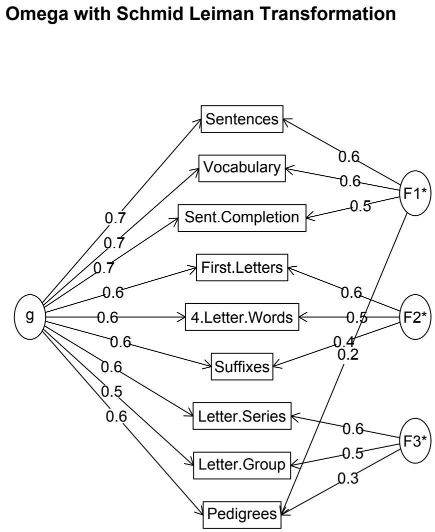
Three factors from Thurstone 9 variables

```
> f3 <- fa(Thurstone,3)
> f3

Factor Analysis using method =  minres
Call: fa(r = Thurstone, nfactors = 3)
Standardized loadings (pattern matrix) based upon correlation matrix
                   MR1   MR2   MR3   h2   u2 com
Sentences         0.91 -0.04  0.04 0.82 0.18 1.0
Vocabulary        0.89  0.06 -0.03 0.84 0.16 1.0
Sent.Completion   0.83  0.04  0.00 0.73 0.27 1.0
First.Letters     0.00  0.86  0.00 0.73 0.27 1.0
4.Letter.Words   -0.01  0.74  0.10 0.63 0.37 1.0
Suffixes          0.18  0.63 -0.08 0.50 0.50 1.2
Letter.Series     0.03 -0.01  0.84 0.72 0.28 1.0
Pedigrees         0.37 -0.05  0.47 0.50 0.50 1.9
Letter.Group     -0.06  0.21  0.64 0.53 0.47 1.2

                       MR1  MR2  MR3
SS loadings           2.64 1.86 1.50
Proportion Var        0.29 0.21 0.17
Cumulative Var        0.29 0.50 0.67
Proportion Explained  0.44 0.31 0.25
Cumulative Proportion 0.44 0.75 1.00

 With factor correlations of
     MR1  MR2  MR3
MR1 1.00 0.59 0.54
MR2 0.59 1.00 0.52
MR3 0.54 0.52 1.00

Mean item complexity =  1.2
```

A hierarchical/multilevel solution to the Thurstone 9 variables



A bifactor solution using the Schmid Leiman transformation



How many factors – no right answer, one wrong answer

1. Statistical
   - Extracting factors until the χ^2 of the residual matrix is not significant.
   - Extracting factors until the change in χ^2 from factor n to factor n+1 is not significant.
2. Rules of Thumb
   - Parallel Extracting factors until the eigenvalues of the real data are less than the corresponding eigenvalues of a random data set of the same size (*parallel analysis*)
   - Plotting the magnitude of the successive eigenvalues and applying the *scree test*.
3. Interpretability
   - Extracting factors as long as they are interpretable.
   - Using the *Very Simple Structure* Criterion (VSS)
   - Using the Minimum Average Partial criterion (MAP).
4. Eigen Value of 1 rule

Factor extraction is a model

1. Maximum likelihood weights residuals by unique variance
2. Maximum likelihood $F_{ML} \approx \Sigma\Sigma[\frac{(s_{jk}-\hat{\sigma}_{jk})^2}{u_j^2 u_k^2}]$
3. OLS weights all residuals equally
4. $F_{OLS} = \Sigma\Sigma(s_{jk} - \hat{\sigma}_{jk})^2$

(MacCallum et al., 2007)

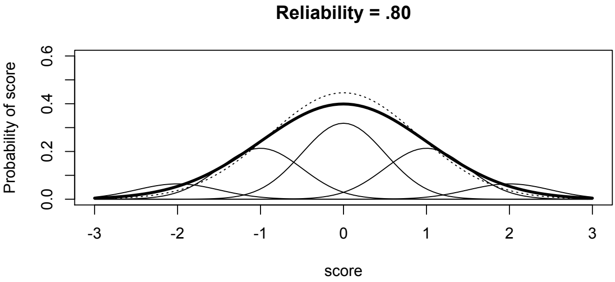
Factors are models

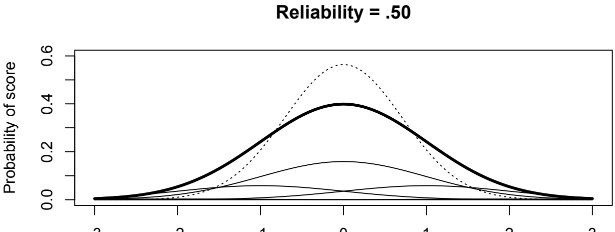
1. Like all other scientific models, factor analysis models are approximations of the real world.
2. Some early controversy due in part to fact that this principle was not yet developed or understood.
3. This principle must be kept in mind when we do FA and interpret our findings.
4. This principle may have important implications for how we do factor analysis.

(MacCallum et al., 2007)

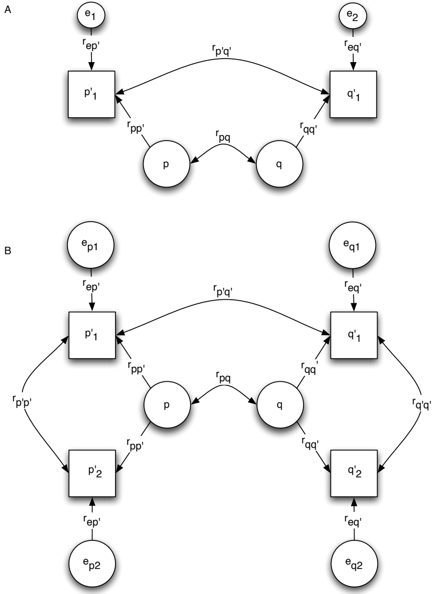
Factors are fictions (Revelle, 1983; Revelle and Ellman, 2016)

All data are befuddled by error: Observed Score = True score + Error score





Spearman's parallell test theory



Guttman's alternative estimates of reliability

Reliability is amount of test variance that is not error variance. But what is the error variance?

$$r_{xx} = \frac{V_x - V_e}{V_x} = 1 - \frac{V_e}{V_x}. \tag{2}$$

$$\lambda_1 = 1 - \frac{tr(\mathbf{V_x})}{V_x} = \frac{V_x - tr(\mathbf{V}_x)}{V_x}. \tag{3}$$

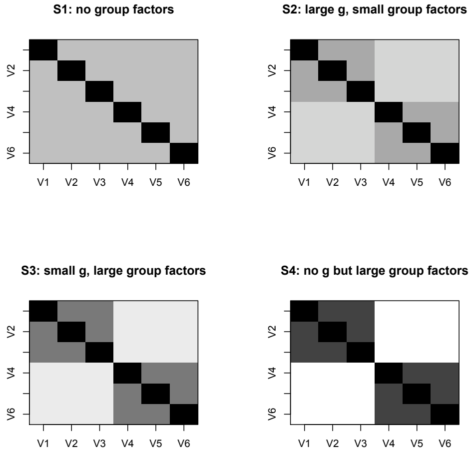
$$\lambda_2 = \lambda_1 + \frac{\sqrt{\frac{n}{n-1}C_2}}{V_x} = \frac{V_x - tr(\mathbf{V}_x) + \sqrt{\frac{n}{n-1}C_2}}{V_x}. \tag{4}$$

$$\lambda_3 = \lambda_1 + \frac{\frac{V_X - tr(\mathbf{V}_X)}{n(n-1)}}{V_X} = \frac{n\lambda_1}{n-1} = \frac{n}{n-1}\left(1 - \frac{tr(\mathbf{V})_x}{V_x}\right) = \frac{n}{n-1}\frac{V_x - tr(\mathbf{V}_x)}{V_x} = \alpha \tag{5}$$

$$\lambda_4 = 2\left(1 - \frac{V_{X_a} + V_{X_b}}{V_X}\right) = \frac{4c_{ab}}{V_x} = \frac{4c_{ab}}{V_{X_a} + V_{X_b} + 2c_{ab}V_{X_a}V_{X_b}}. \tag{6}$$

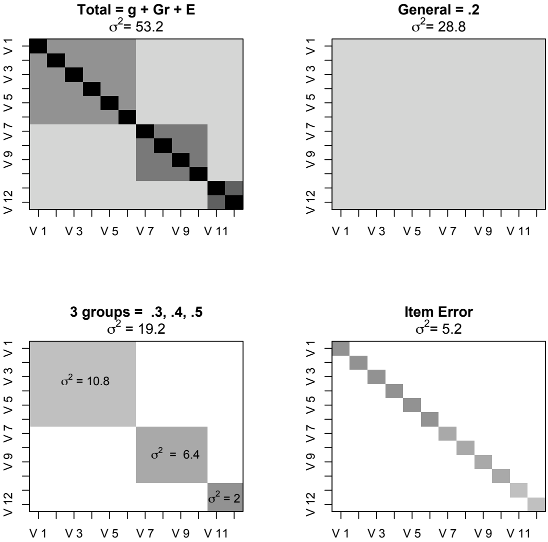
$$\lambda_6 = 1 - \frac{\sum e_j^2}{V_x} = 1 - \frac{\sum(1 - r_{smc}^2)}{V_x} \tag{7}$$

Four different correlation matrices, one value of α



1. The problem of group factors
2. If no groups, or many groups, α is ok

Decomposing a test into general, Group, and Error variance



1. Decompose total variance into general, group, specific, and error
2. $\alpha <$ total
3. $\alpha >$ general

Two additional alternatives to α: $\omega_{hierarchical}$ and ω_{total}

If a test is made up of a general, a set of group factors, and specific as well as error:

$$\mathbf{x} = \mathbf{cg} + \mathbf{Af} + \mathbf{Ds} + \mathbf{e} \tag{8}$$

then the communality of item$_j$, based upon general as well as group factors,

$$h_j^2 = c_j^2 + \sum f_{ij}^2 \tag{9}$$

and the unique variance for the item

$$u_j^2 = \sigma_j^2(1 - h_j^2) \tag{10}$$

may be used to estimate the test reliability.

$$\omega_t = \frac{\mathbf{1cc'1'} + \mathbf{1AA'1'}}{V_x} = 1 - \frac{\sum(1 - h_j^2)}{V_x} = 1 - \frac{\sum u^2}{V_x} \tag{11}$$

McDonald (1999) introduced two different forms for ω

$$\omega_t = \frac{\mathbf{1cc'1'} + \mathbf{1AA'1'}}{V_x} = 1 - \frac{\sum(1-h_j^2)}{V_x} = 1 - \frac{\sum u^2}{V_x} \tag{12}$$

and

$$\omega_h = \frac{\mathbf{1cc'1}}{V_x} = \frac{(\sum \Lambda_i)^2}{\sum\sum R_{ij}}. \tag{13}$$

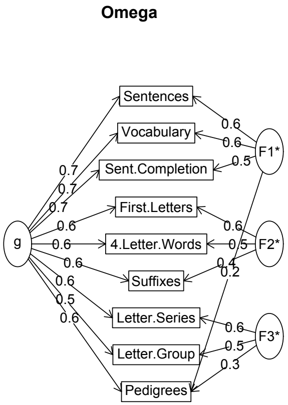
These may both be find by factoring the correlation matrix and finding the g and group factor loadings using the `omega` function.

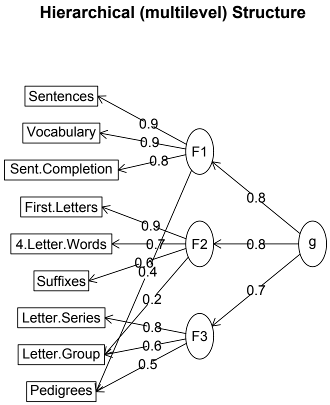
Using omega on the Thurstone data set to find alternative reliability estimates

```
> lower.mat(Thurstone)
> omega(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

Omega
Call: omega(m = Thurstone)
Alpha:                 0.89
G.6:                   0.91
Omega Hierarchical:    0.74
Omega H asymptotic:    0.79
Omega Total            0.93
```

Two ways of showing a general factor





omega function does a Schmid Leiman transformation

```
> omega(Thurstone, sl=FALSE)

Omega
Call: omega(m = Thurstone, sl = FALSE)
Alpha:                 0.89
G.6:                   0.91
Omega Hierarchical:    0.74
Omega H asymptotic:    0.79
Omega Total            0.93
Schmid Leiman Factor loadings greater than  0.2
                   g   F1*  F2*  F3*   h2   u2   p2
Sentences       0.71 0.57            0.82 0.18 0.61
Vocabulary      0.73 0.55            0.84 0.16 0.63
Sent.Completion 0.68 0.52            0.73 0.27 0.63
First.Letters   0.65      0.56       0.73 0.27 0.57
4.Letter.Words  0.62      0.49       0.63 0.37 0.61
Suffixes        0.56      0.41       0.50 0.50 0.63
Letter.Series   0.59           0.61  0.72 0.28 0.48
Pedigrees       0.58 0.23      0.34  0.50 0.50 0.66
Letter.Group    0.54           0.46  0.53 0.47 0.56
With eigenvalues of:
   g  F1*  F2*  F3*
3.58 0.96 0.74 0.71
```

Alpha and its alternatives

- Reliability = $\frac{\sigma_t^2}{\sigma_x^2} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$
- If there is another test, then $\sigma_t = \sigma_{t_1 t_2}$ (covariance of test X_1 with test $X_2 = C_{xx}$)
- But, if there is only one test, we can *estimate* σ_t^2 based upon the observed covariances within test 1
- How do we find σ_e^2 ?
- The worst case, (Guttman case 1) all of an item's variance is error and thus the error variance of a test X with variance-covariance C_x
  - $C_x = \sigma_e^2 = diag(C_x)$
  - $\lambda_1 = \frac{C_x - diag(C_x)}{C_x}$
- A better case (Guttman case 3, α) is that that the average covariance between the items on the test is the same as the average true score variance for each item.
  - $C_x = \sigma_e^2 = diag(C_x)$
  - $\lambda_3 = \alpha = \lambda_1 * \frac{n}{n-1} = \frac{(C_x - diag(C_x)) * n/(n-1)}{C_x}$

Reliability is an upper bound to validity

1. Validity can not exceed the square root of reliabiliity
2. But what is reliability?
3. Confusion of internal consistency and reliablity.
4. Interpretability $\neq$ internal consistency

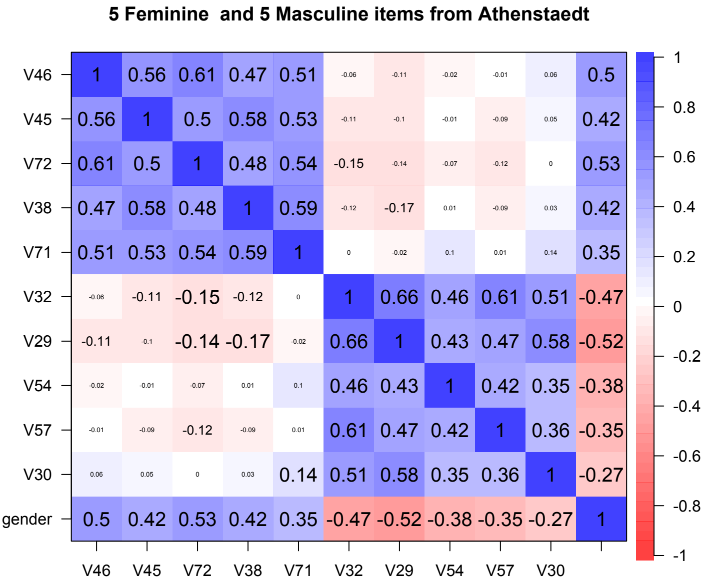
Comparing internal consistency and validity

Table: Estimates of internal consistency and validity for various scales from the Athenstaedt data set. Uni is an estimate of the unidimensionality of a scale, Max and Min splits are based upon split half reliabilities, validity is the correlation with gender.

Reliability estimates using the `reliability` function in *psych*

Variable	ω_h	α	ω_t	Uni	r.fit	fa.ft	max	min	$\bar{r}$	med r	N	Validity
Femininity	0.57	0.90	0.91	0.74	0.81	0.91	0.94	0.79	0.24	0.22	29	0.67
Masculinity	0.70	0.87	0.89	0.73	0.78	0.94	0.92	0.76	0.23	0.21	23	-0.53
MF	0.13	0.88	0.90	0.22	0.39	0.56	0.93	0.68	0.13	0.11	52	0.82
F10	0.78	0.89	0.90	0.94	0.96	0.99	0.92	0.85	0.44	0.43	10	0.62
M10	0.68	0.87	0.89	0.93	0.95	0.98	0.90	0.81	0.40	0.37	10	-0.52
MF20	0.14	0.85	0.89	0.29	0.52	0.57	0.92	0.34	0.22	0.22	20	0.77
F5	0.71	0.85	0.88	0.99	0.99	0.99	0.85	0.78	0.54	0.53	5	0.56
M5	0.69	0.82	0.85	0.95	0.96	0.99	0.83	0.78	0.48	0.47	5	-0.52
MF10	0.15	0.77	0.85	0.28	0.50	0.56	0.86	0.12	0.25	0.14	10	0.74

Consider 10 items: one scale or two?



One scale or two?

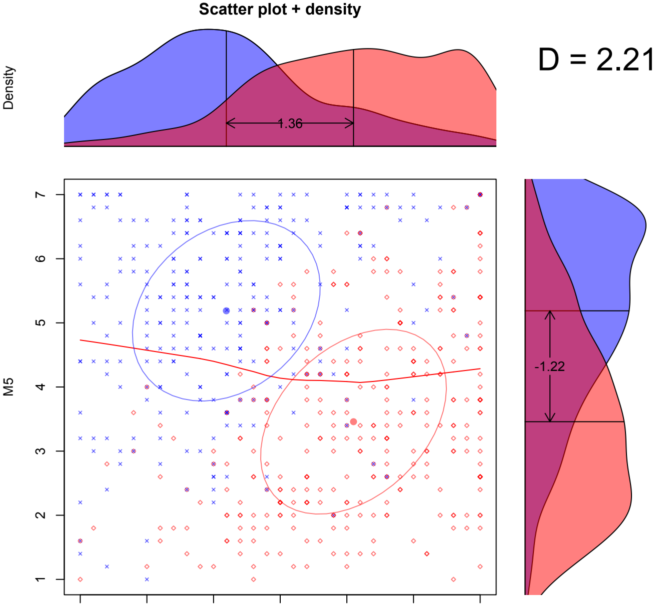
The correlations of ten items from this data set show a clear two factor structure, although the α estimate (.77) is reasonable for a 10 item scale. It is by finding ω_h for this scale (.15) that we recognize that we should not combine this scale into one.

Table: The items in the 10 item scale

Variable	Item label	Item content
V46	V46	Sew on a Button
V45	V45	Change Bed Sheets
V72	V72	Do the Ironing
V38	V38	Dust the Furniture
V71	V71	Wash Windows
V32-	V32	Do Repair Work
V29-	V29	Change Fuses
V54-	V54	Shovel Snow
V57-	V57	Do Home Improvement Jobs
V30-	V30	Clean a Drain

Validity is enhanced with unrelated scales

D = 2.21



Validity is enhanced with unrelated scales

1. As discussed by Eagly and Revelle (2022) it is useful to consider the joint effect of these scales when examining their effects.
2. The validity of the 10 item scale may be shown in the scatter plot/density distribution of the scores on the two 5 item scale.
3. Although two scales have average within gender correlations of .32 and Cohen d values of 1.36 and 1.22 they correlate -.07, and have a pooled Mahalobnis Distance of 2.21.

Guttman 6: estimating using the Squared Multiple Correlation

- Reliability = $\frac{\sigma_t^2}{\sigma_x^2} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$
- Estimate true item variance as squared multiple correlation with other items
- $\lambda_6 = \frac{(C_x - diag(C_x) + \Sigma(smc_i)}{C_x}$
  - This takes observed covariance, subtracts the diagonal, and replaces with the squared multiple correlation
  - Similar to α which replaces with average inter-item covariance
- Squared Multiple Correlation is found by `smc` and is just $smc_i = 1 - 1/R_{ii}^{-1}$

Classical Reliability

1. Classical model of reliability
   - Observed = True + Error
   - Reliability = $1 - \frac{\sigma^2_{error}}{\sigma^2_{observed}}$
   - Reliability = $r_{xx} = r^2_{x_{domain}}$
   - Reliability as correlation of a test with a test just like it
2. Reliability requires variance in observed score
   - As σ^2_x decreases so will $r_{xx} = 1 - \frac{\sigma^2_{error}}{\sigma^2_{observed}}$
3. Alternate estimates of reliability all share this need for variance
   3.1 Internal Consistency
   3.2 Alternate Form
   3.3 Test-retest
   3.4 Between rater
4. Item difficulty is ignored, items assumed to be sampled at random

The "new psychometrics"

1. Model the person as well as the item
    - People differ in some latent score
    - Items differ in difficulty and discriminability
2. Original model is a model of ability tests
    - $p(correct|ability, difficulty, ...) = f(ability - difficulty)$
    - What is the appropriate function?
3. Extensions to polytomous items, particularly rating scale models

FA and IRT

IRT parameters from FA

$$\delta_j = \frac{D\tau}{\sqrt{1-\lambda_j^2}}, \qquad \alpha_j = \frac{\lambda_j}{\sqrt{1-\lambda_j^2}} \tag{14}$$

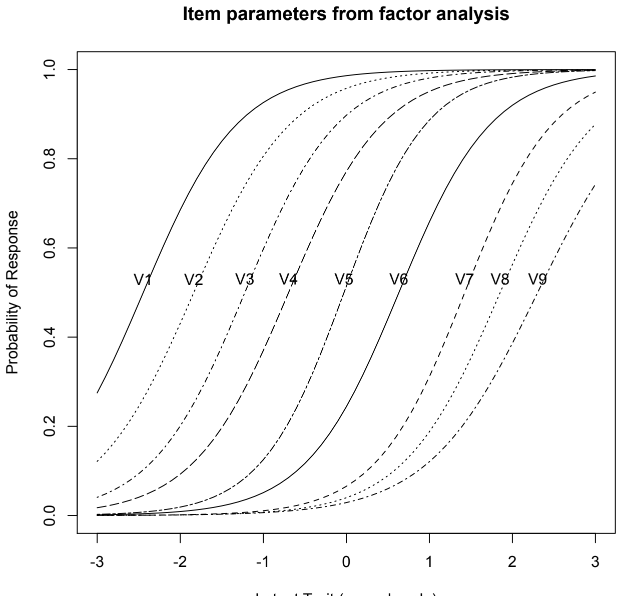
FA parameters from IRT

$$\lambda_j = \frac{\alpha_j}{\sqrt{1+\alpha_j^2}}, \qquad \tau_j = \frac{\delta_j}{\sqrt{1+\alpha_j^2}}.$$

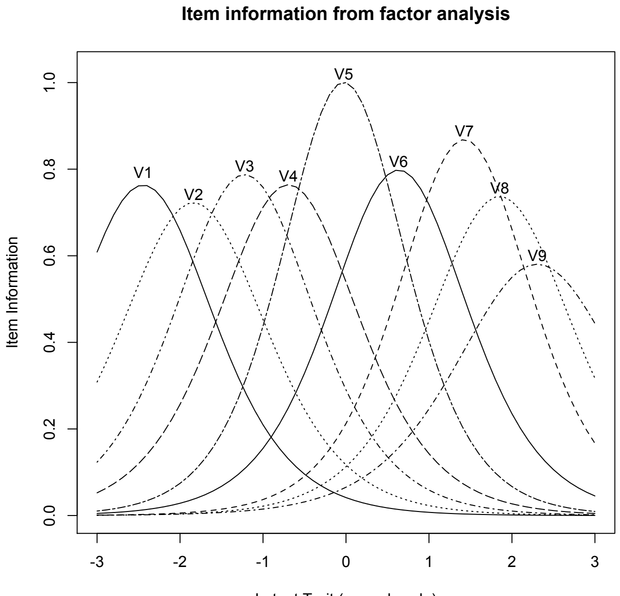
the irt.fa function

```
> set.seed(17)
> items <- sim.npn(9,1000,low=-2.5,high=2.5)$items
> p.fa <-irt.fa(items)

Summary information by factor and item
 Factor =  1
               -3   -2   -1    0    1    2     3
V1           0.61 0.66 0.21 0.04 0.01 0.00  0.00
V2           0.31 0.71 0.45 0.12 0.02 0.00  0.00
V3           0.12 0.51 0.76 0.29 0.06 0.01  0.00
V4           0.05 0.26 0.71 0.54 0.14 0.03  0.00
V5           0.01 0.07 0.44 1.00 0.40 0.07  0.01
V6           0.00 0.03 0.16 0.59 0.72 0.24  0.05
V7           0.00 0.01 0.04 0.21 0.74 0.66  0.17
V8           0.00 0.00 0.02 0.11 0.45 0.73  0.32
V9           0.00 0.00 0.01 0.07 0.25 0.55  0.44
Test Info    1.11 2.25 2.80 2.97 2.79 2.28  0.99
SEM          0.95 0.67 0.60 0.58 0.60 0.66  1.01
 Reliability 0.10 0.55 0.64 0.66 0.64 0.56 -0.01
```

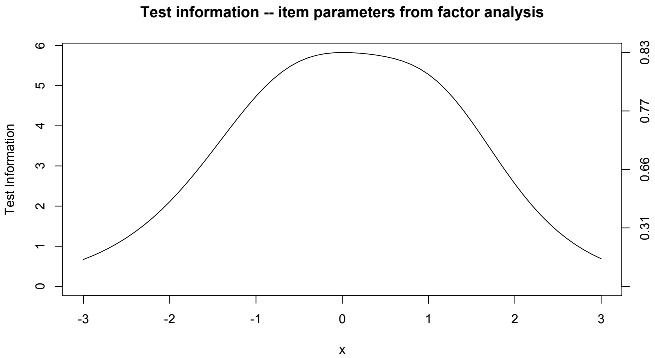
Item Characteristic Curves from FA



Item information from FA



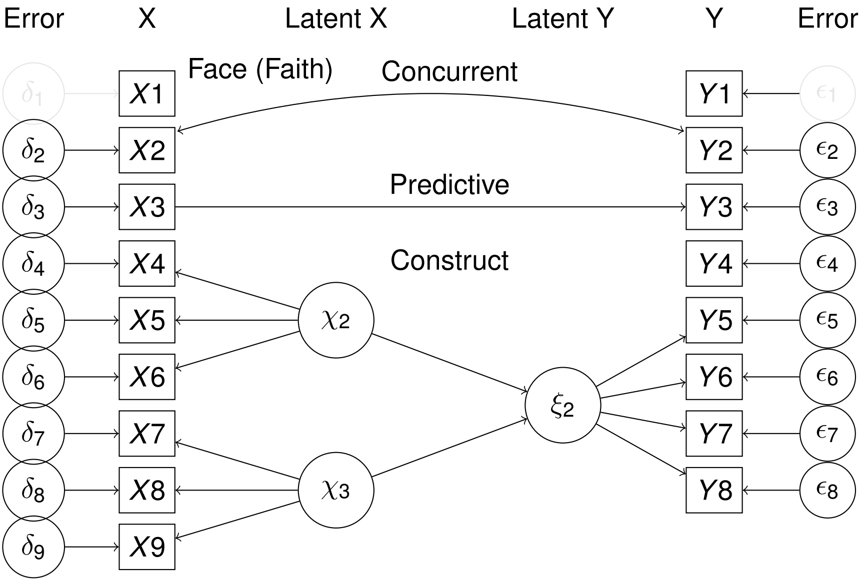
Test Information Curve



IRT and CTT don't really differ except

1. Correlation of classic test scores and IRT scores $> .98$.
2. Test information for the person doesn't require people to vary
3. Possible to item bank with IRT
   - Make up tests with parallel items based upon difficulty and discrimination
   - Detect poor items
4. Adaptive testing
   - No need to give a person an item that they will almost certainly pass (or fail)
   - Can tailor the test to the person
   - (Problem with anxiety and item failure)

Face, Concurrent, Predictive, Consruct



Validity

1. Face/Faith (does it look right)
2. Concurrent (does it correlate with a criterion?)
3. Predictive (Does it correlate with a later criterion?)
4. Construct
   - Convergent (do measures that should go together, go together)
   - Discriminant (do measures that should not go together not go together)
   - Incremental (does it actual make a difference?)
   - Multi-Method, Multi Trait Matrix
5. Validity for what?
   - for the institution
   - for the individual
6. Validity and decision making

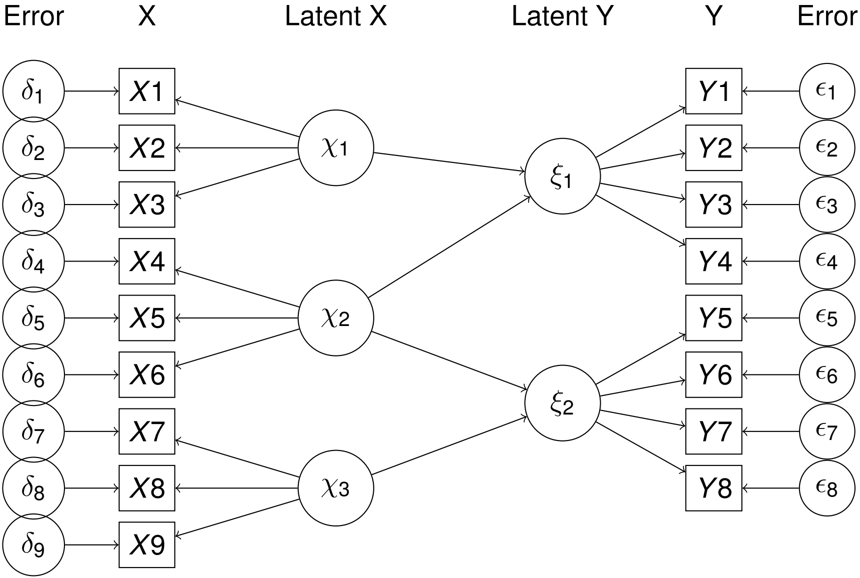
Zola peer rating and self report as example of MTMMs

R code

```
scores <- psych::scoreOverlap(zola.keys[c(1:5,33:37)],zola) #MTMM of I
```

```
lowerMat(scores$cor)
                   Agrbl Cnscn Nrtcs Extrv Opnnn Agrbl Cnscn Stblt Extrv IntlO
Agreeableness       1.00
Conscientiousness   0.28  1.00
Neuroticism        -0.12 -0.18  1.00
Extraversion        0.25  0.12 -0.25  1.00
Opennness           0.08  0.05 -0.09  0.13  1.00
Agreeableness       0.47  0.10 -0.01  0.00 -0.09  1.00
Conscientiousness   0.15  0.55 -0.12 -0.01 -0.04  0.18  1.00
Stability           0.13  0.16 -0.58  0.05  0.07  0.25  0.25  1.00
Extraversion        0.23  0.28 -0.27  0.49  0.11  0.07  0.23  0.22  1.00
IntellectOpenness   0.14  0.08 -0.15  0.09  0.30  0.19  0.24  0.27  0.15  1.00
```

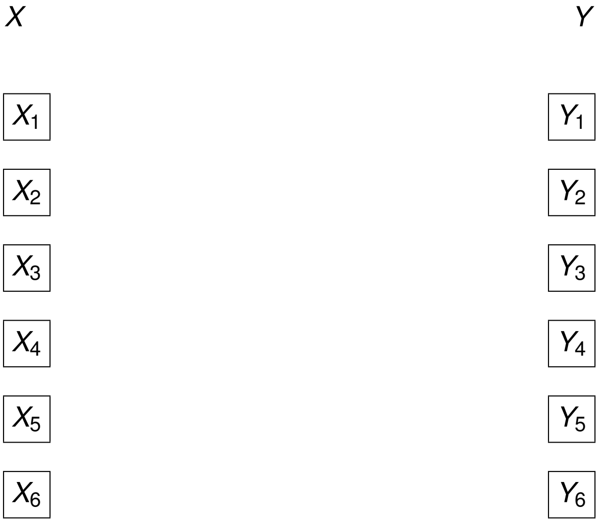
Psychometric Theory: Data, Measurement, Theory



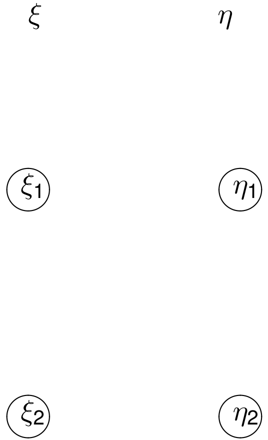
Two types of variables, three types of relationships

1. Variables
   1.1 Observed Variables (X, Y)
   1.2 Latent Variables (ξ η ϵ ζ)
2. Three kinds of variance/covariances
   2.1 Observed with Observed C_{xy} or σ_{xy}
   2.2 Observed with Latent λ
   2.3 Latent with Latent ϕ
3. Direction
   - Bidirectional (correlation)
   - Directional (regression)

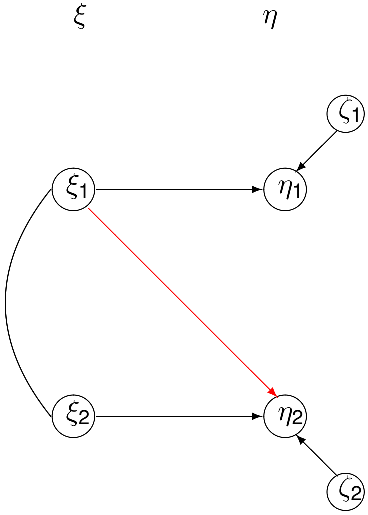
Observed Variables



Latent Variables

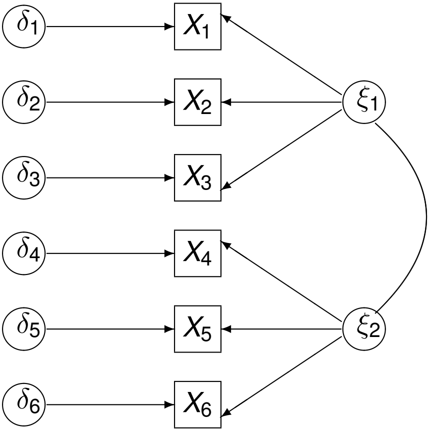


Theory: A regression model of latent variables

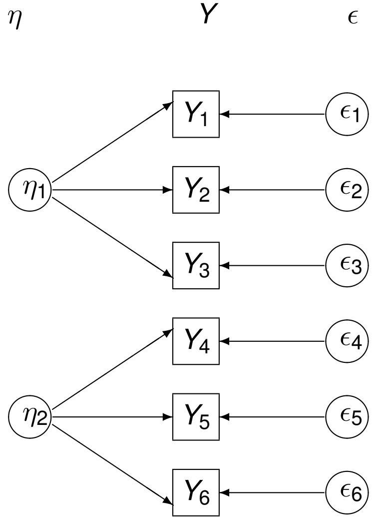


A measurement model for X

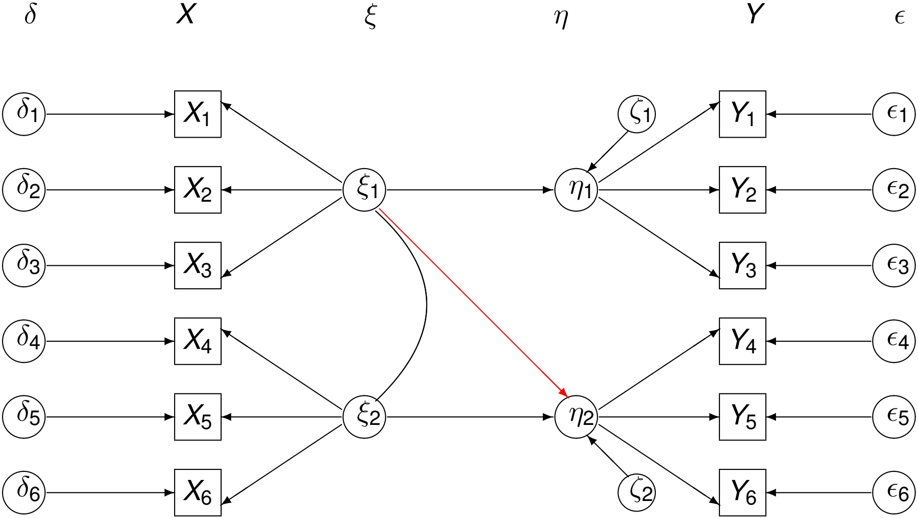
δ X ξ



A measurement model for Y



A complete structural model



Latent Variable Modeling

1. Requires measuring observed variables
   - Requires defining what is relevant and irrelevant to our theory.
   - Issues in quality of scale information, levels of measurement.
2. Formulating a measurement model of the data: estimating latent constructs
   - Perhaps based upon exploratory and then confirmatory factor analysis, definitely based upon theory.
   - Includes understanding the reliability of the measures.
3. Modeling the structure of the constructs
   - This is a combination of theory and fitting. Do the data fit the theory.
   - Comparison of models. Does one model fit better than alternative models?

Two fundamentally different types of observed variables

1. Observed variables can be "reflective" of the latent variables. They are "effect indicators".
   - Variables are caused by the latent variables.
   - Covariation between the variables are explained by the latent variables
2. Observed variables can be "causal indicators" or formative indicators that can directly effect the latent variable
   - Variables cause the "latent" variable
   - Covariation of the the observed variables is not modeled

Formative indicators (Bollen, 2002; Loevinger, 1957)

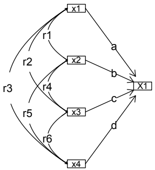
1. The correlational structure of formative indicators is independent of the loadings on a factor.
   - They are not locally independent
2. Examples of formative indicators Time spent in social interaction
   - Time spent with family, time spent with friends, time spent with coworkers.
   - These might in fact be negatively correlated even though total score is important.

Effect (reflective) indicators

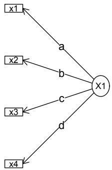
1. Test scores on various quantitative tests as effect indicators of trait
   - Feelings of self worth as effect indicators of self esteem
   - Ability items as indicators of ability
2. Correlational structure is a function of the path coefficients with latent variables
3. Variables are locally independent
   - (uncorrelated with each other when latent variable is partialled out)

Formative vs. Effect variables as regression versus factors

Formative Variables



Effect Variables



MIMIC models: Multiple Indicators, Multiple Causes

MIMIC models Jöreskog and Goldberger (1975)

1. Formative indicators as causes
   - income
   - occupation
   - education
2. Latent Variables as organizing constructs (e.g. social participation)
3. Effect indicators as measures of latent variables
   - church attendance
   - memberships
   - friends seen

Final Comments on structural models

1. Theory First
   - What are the alternative theories?
   - Are there specific differences in the theories that are testable?
2. Measurement Model
   - Comparison of measurement models?
   - How many latent variables? How do you know?
   - Measurement Invariance?
3. Structural Model
   - Comparison of multiple models?
   - What happens if the arrows are reversed?
4. Theory Last
   - What do we know now that we did not know before?
   - What do we have shown is not correct?

Cattell's data box (Cattell, 1946, 1966, 1978)

1. Person by Variables
   - Variables over People, fixed Occasion (R)
   - People over Variables, fixed Occasion (Q)
2. Person by Occasions
   - Occasions over People, fixed Variable (T)
   - People over Occasions, fixed Variable (S)
3. Variables by Occasions
   - Variables over Occasions, fixed People (P)
   - Occasions over Variables, fixed People (O)

Now, with multilevel modeling techniques, we can integrate 3 modes at once. See also 3 mode factor analysis and INDSCAL.

Note that Cattell changed the abbreviations for the dimensions across his various papers (e.g., (Cattell, 1946, 1966, 1978).

Traditional measures

1. Individuals across items
    - Correlations of items taken over people to identify dimensions of items which are in turn used to describe dimensions of individual differences
    - Ability
    - Non-cognitive measures of individual differences
        - Stable over time: Traits
        - Variable over time: States
2. INDSCAL type comparisons of differences in structure of items across people
3. 3 Mode factor analyis
4. Dynamic factor analysis (over time)
5. Factoring by individual over time (is the structure the same, or are the factors the same?)

Scale construction: A 10 steps program

1. Personality scales are not created in a theoretical vacuum. Perhaps the most important step in developing a new scale is a consideration of what is the construct of interest. What is it, what are manifestations of it, what is it not, and what should it not relate to.
2. Then, what is the population of interest? Are they young or old, highly literate, or somewhat challenged by literacy. Write items suitable for the population of interest.
3. Give the items to the participants. Make sure that they are engaged in the task.

Scale construction: A 10 steps program (continued)

4. To analyze the data, it is necessary to enter the data into a machine readable form.
    - This is a source of error. Double check for data entry errors.
    - Double entry (two different people enter the data and then the two files are automatically compared) is recommended.
    - Even better is automatic data entry (but then you need to check and double check the program).
    - my.data <- read.file() #go find the file on your computer
    - my.data <- read.file(myfile) #if you have the file name some
    - my.data <- read.clipboard() #if you have already copied the data to the clipboard
5. Run basic descriptive statistics to do one more check for errors. Graphically check as well.
    - describe(my.data)
    - pairs.panels(my.data)
6. Form the variance/covariance matrix from the items and examine the dimensionality of the resulting space.

Scale construction: A 10 steps program (continued)

7. Apply various data reduction techniques (factor analysis, principal components analysis, cluster analysis).
   - `fa` For most factor analysis and rotation/transformation algorithms (need to specify the number of factors)
   - `irt.fa` If you have polytomous or dichotomous items and want to take an IRT approach.
   - `principal` aka `pca` for principal components analysis
   - `fa.parallel` For parallel analysis and scree tests
   - `vss` The Very Simple Structure criterion as well as the Minimum Average Parcel test
   - `nfactors` Combine a number of different tests
   - `iclust` Cluster analysis of items shows structure pretty well.
   - `bestScales` Apply an empirical scale construction procedure.

Scale construction: A 10 steps program (continued)

8. Form composite scales of the selected items. Check these scales for various measures of internal consistency.
    - Form a list of items to score. Can use `make.keys` if desired.
    - `scoreItems` wlll give reliability statistics and raw scores. If given correlation matrix input, will return reliability statistics.
    - `scoreOverlap`: Given a correlation matrix, will find scale statistics correcting for item overlap.
    - `scoreFast` and `scoreVeryFast` will give just scores given data and a keys.list
    - `bestScales` (For empirical scale construction)
    - `alpha` – just one scale at a time – Do you really want to do that?

Scale construction: A 10 steps program (continued)

9. Validity
   - Discriminant validity requires that the scales not correlate with other, unrelated traits.
   - Convergent validity requires that the scale do correlate with other, alternative measures of the same trait.
10. Cross validation to show the results are not sample dependent.
    - Validate on a new sample
    - Cross vaidate on a hold sample
    - KFold cross validation
    - Bootstrap validation

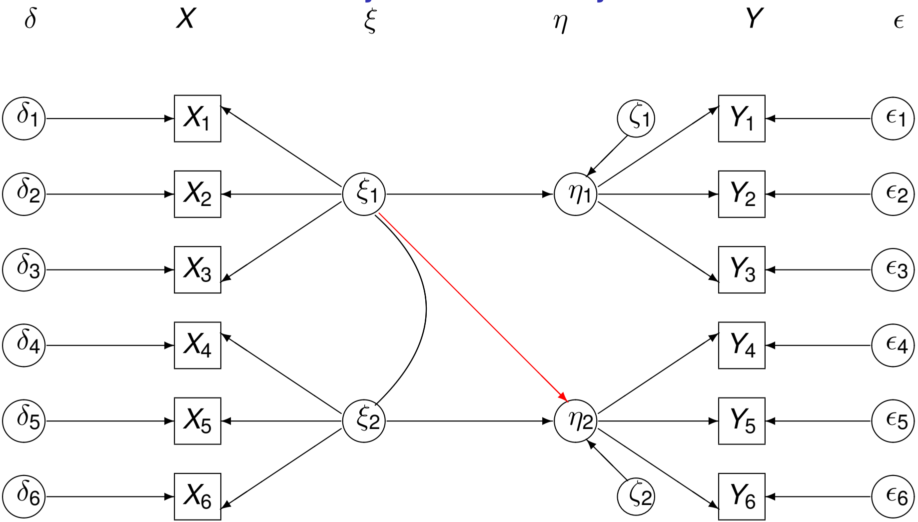
Multiple ways to construct scales

1. Rational/Theoretical
   - Learn Theory
   - Write good items
2. Homogeneous keying
   - Write good items
   - Factor/Cluster analyze
3. Empirical Keys
   - Write good items
   - Select those items that correlate with the criteria

Putting it all together

1. Theory first
2. Good item design
3. Data collection from appropriate population
4. Practical help using the *psych* package (Revelle, 2023) in R (R Core Team, 2025). See `http://personality-project.org/r`
5. Data cleaning: `pairs.panel`, `describe` and `scrub`.
6. Number of dimensions: `nfactors`, `fa.parallel`
7. Dimensional reduction: `fa`, `principal`, `iclust`
8. Scale reliability and scoring: `scoreItems`, `scoreOverlap`, `omega`
9. Scale validity: basic 0 order correlations: `lowerCor`, `corr.test`, `cor.plot`
10. Structural relations: *lavaan* (Rosseel, 2012) `lmCor`, `mediate`
11. Theory last (revise and iterate)

Psychometric Theory



A few of the most useful data manipulations functions (adapted from Rpad-refcard). Use ? for details

file.choose	() find a file
file.choose	(new=TRUE) create a new file
read.table	(filename)
read.csv	(filename) reads a comma separated file
read.delim	(filename) reads a tab delimited file
c	(...) combine arguments
from:to	e.g., 4:8
seq	(from,to, by)
rep	(x,times,each) repeat x
gl	(n,k,...) generate factor levels
matrix	(x,nrow=,ncol=) create a matrix
data.frame	(...) create a data frame

dim	(x) dimensions of x
str	(x) Structure of an object
list	(...) create a list
colnames	(x) set or find column names
rownames	(x) set or find row names
ncol(x), nrow(z)	number of row, columns
rbind	(...) combine by rows
cbind	(...) combine by columns
is.na	(x) also is.null(x), is...
na.omit	(x) ignore missing data
table	(x)
merge	(x,y)
apply	(x,rc,FUNCTION)
ls	() show workspace
rm	() remove variables from workspace

More useful statistical functions, Use ? for details

mean	(x,na.rm=TRUE) *
is.na	(x) also is.null(x), is...
na.omit	(x) ignore missing data
sum	(x)
rowSums	(x) see also colSums(x)
colSums	(x) see also rowSums(x)
min	(x,na.rm=TRUE)*
max	(x) *ignores NA values
range	(x)
table	(x)
summary	(x) depends upon x
sd	(x) standard deviation
cor	(x,use="pairwise") correlation
cov	(x) covariance
solve	(x) inverse of x
lm	`(y~x) linear model`
aov	`(y~x) ANOVA`

Selected functions from *psych* package

describe	(x) descriptive stats
describeBy	(x,y) descriptives by group
pairs.panels	(x) SPLOM
error.bars	(x) means + error bars
error.bars.by	(x) Error bars by groups
fa	(x,n) Factor analysis
principal	(x,n) Principal components
iclust	(x) Item cluster analysis
scoreItems	(x) score multiple scales
score.multiple.choice	(x) score multiple choice scales
alpha	(x) Cronbach's alpha
omega	(x) MacDonald's omega
irt.fa	(x) Item response theory through factor analysis
mediate	(y,x,m,data) Mediation/moderation

More help

1. An introduction to R as HTML, PDF or EPUB from http://cran.r-project.org/manuals.html (many different links on this page
2. FAQ General and then Mac and PC specific
3. R reference card http://cran.r-project.org/doc/contrib/Baggott-refcard-v2.pdf
4. Various “cheat sheets" from RStudio http://www.rstudio.com/resources/cheatsheets/
5. Using R for psychology http://personality-project.org/r/
6. Package vignettes (e.g., http://personality-project.org/r/psych/vignettes/overview.pdf)
7. R listserve, StackOverflow, your students and colleagues

Steps to data analysis (in R)

See the various vignettes and HowTos for *psych* and *psychTools*

1. Read in the data from a file
   - A small text file or Excel file: Copy to the clipboard and then

     R code
     ```
     my.data <- read.clipboard()
     ```
   - A SPSS or SAS file: Find it, read it

     R code
     ```
     library(foreign)
     my.file <- file.choose()   #use your normal directory  search
     my.data <- read.spss(my.file,use.value.labels=FALSE,to.data.frame=TRUE)
     ```
2. Check to make sure you did it right

   R code
   ```
   dim(my.data)
   describe(my.data)
   ```

Steps (continued)

1. Graphically display the data

R code
```
pairs.panels(my.data)  #if not too many variables
R <- lowerCor(my.data)
cor.plot(R,numbers=TRUE,upper=FALSE)
```

Dimension reduction

1. How many dimensions?

R code
```
nfactors(my.data)
fa.parallel(my.data)
```

2. Extract factors

R code
```
       # specify the number of factors  you want
my.factors <- fa(my.data,nfactors = 5)
my.scores <- my.factors$scores #in case you want them
```

Reliability of a scale

1. α (but why do you want this)

R code
```
alpha(my.data)
```

2. Better to find ω_h

R code
```
om <- omega(my.data)
omega.diagram(om,sl=FALSE) # to show a hierarchical structure
```

Score multiple scales from one inventory, find reliabilities

Make up a keys matrix and then find the scores. (Using the bfi example)

R code

```
keys.list <-
 list(agree=c("-A1","A2","A3","A4","A5"),
   conscientious=c("C1","C2","C3","-C4","-C5"),
   extraversion=c("-E1","-E2","E3","E4","E5"),
   neuroticism=c("N1","N2","N3","N4","N5"),
   openness = c("O1","-O2","O3","O4","-O5"))
 keys <- make.keys(bfi,keys.list)

scores <- scoreItems(keys[1:27,],bfi[1:27]) #don't score age
scores
#show the use of the fa.lookup with a dictionary
fa.lookup(keys,bfi.dictionary[,1:4])

#multiple reliability estimates
reliability(bfi.keys,bfi)
```

The vignettes and Howtos

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.
7. Using `mediate` and `lmCor` to do mediation, moderation and regression analysis.

For a step by step tutorial in the use of the psych package and the base functions in R for basic personality research, see the guide for using R for personality research at https://personalitytheory.org/r/r.short.html. For an *introduction to psychometric theory with applications in R*, see the draft chapters at https://personality-project.org/r/book).

Psychometrics as model estimation and model fitting

We explored a number of models

1. Modeling the process of data collection and of scaling
   - $X = f(\theta)$
   - How to measure X, properties of the function f.
2. Correlation and Regression
   - $Y = \beta X$
   - $R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
3. Factor Analysis and Principal Components Analysis
   - $R = FF' + U^2 \quad R = CC'$
4. Reliability $\rho_{xx} = \frac{\sigma^2_\theta}{\sigma^2_X}$
5. Item Response Theory
   - $p(X|\theta, \delta) = f(\theta - \delta)$
6. Structural Equation Modeling
   - $\rho_{yy} Y = \beta \rho_{xx} X$

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53:605–634.

Box, G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, 74(365):1–4.

Cattell, R. B. (1946). *Description and measurement of personality*. World Book Company, Oxford, England.

Cattell, R. B. (1965). A biometrics invited paper. factor analysis: An introduction to essentials i. the purpose and underlying models. *Biometrics*, 21(1):190–215.

Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of possible relational systems. In Cattell, R. B., editor, *Handbook of multivariate experimental psychology*, pages 67–128. Rand-McNally, Chicago.

Cattell, R. B. (1978). *The scientific use of factor analysis*. Plenum Press, New York.

Eagly, A. H. and Revelle, W. (2022). Understanding the Magnitude of Psychological Differences Between Women and Men

Requires Seeing the Forest and the Trees. *Perspectives on Psychological Science*, 17(5):1339–1358.

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.

Householder, A. S. and Young, G. (1938). Matrix approximation and latent roots. *The American Mathematical Monthly*, 45(3):165–171.

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4):443–477.

Jöreskog, K. G. and Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable,. *Journal of the American Statistical Association*, 70(351a):631–639.

Lawley, D. N. and Maxwell, A. E. (1963). *Factor analysis as a statistical method*. Butterworths, London.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports Monograph Supplement 9*, 3:635–694.

MacCallum, R. C., Browne, M. W., and Cai, L. (2007). Factor analysis models as approximations. In Cudeck, R. and MacCallum, R. C., editors, *Factor analysis at 100: Historical developments and future directions*, pages 153–175. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

McDonald, R. P. (1999). *Test theory: A unified treatment*. L. Erlbaum Associates, Mahwah, N.J.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revelle, W. (1983). Factors are fictions, and other comments on individuality theory. *Journal of Personality*, 51(4):707–714.

Revelle, W. (2023). *psych:Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, https://CRAN.r-project.org/package=psych, 2.3.9 edition. R package version 2.3.9.

Revelle, W. and Ellman, L. G. (2016). Factors are still fictions [peer commentary on "towards more rigorous personality trait–outcome research," by R. Mõttus]. *European Journal of Personality*, 30:324–325.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36.

Rossi, G. B. (2007). Measurability. *Measurement*, 40(6):545 – 562.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, 15(2):201–292.

Stevens, S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

Thurstone, L. L. (1947). *Multiple-factor analysis: a development and expansion of The vectors of the mind*. The University of Chicago Press, Chicago, Ill.

Tukey, J. W. (1961). Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics*, 3(2):191–219.