330.657
Psychosocial Statistics
Reliability

Monday, September 11, 2006

Bill Eaton

1) Review of reliability theory so far

•

2) Different types of reliability coefficients

Correlation
Split half measures
Alpha coefficient
Kuder Richardson Coefficient
Kappa

After this class you will be able to:

- Define reliability in two ways
- Estimate reliability in five ways

Classical test theory:

$$x = T_x + e$$

Assumptions:

1) $E(e) = 0$
2) $cov(T_x, e) = 0$
3) $cov(e_i, e_j) = 0$

N.B.:

$$
\begin{aligned}
Var\ (X) &= Var\ (T_x + e) \\
&= Var\ (T_x) + 2\ COV\ (T_x, e) + Var\ (e) \\
&= Var\ (T_x) + Var\ (e)
\end{aligned}
$$

Reliability is the consistency of measurement

- 1) The correlation between parallel measures
  - $\rho_{xx} = r_{x1x2}$
- 2) The ratio of True score to Total score variance
  - $\rho_{xx} = \dfrac{V\,(Tx)}{V\,(Ox)}$

- Parallel measures
  - $T_{x1} = T_{x2}$ [= E(x)]
  - Cov(e1,e2) = 0
  - Var (e_1) = Var (e_2)
- Tau equivalent measures
  - $T_{x1} = T_{x2}$
  - Var $(e_1) \neq$ Var (e_2)
- Congeneric measures
  - $T_{x1} = \beta_{1T}$; $T_{x2} = \beta_{2T}$; etc. (factor model)

Correlation, *r*

Correlation (i.e. Pearson correlation) is a scaled version of covariance

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^{\ 2} s_y^{\ 2}}}$$

$-1 \leq r \leq 1$

$r = 1$	perfect positive correlation
$r = -1$	perfect negative correlation
$r = 0$	uncorrelated

Correlation of parallel tests equals the reliability of each test

$$r_{x_1x_2} = \frac{Cov\ (x_1x_2)}{\sqrt{(s_{x_1}{}^2 s_{x_2}{}^2)}} = \frac{\mathrm{cov}(x_1, x_2)}{\sqrt{s_{x1}^2 s_{x1}^2}}$$

$$= \frac{\mathrm{cov}(t + e_1, t + e_2)}{\mathrm{var}(x)}$$

$$= \frac{\mathrm{var}(t)}{\mathrm{var}(x)}$$

Measures to Assess Reliability

	Continuous	Categorical
Test-retest	R or ICC	Kappa or ICC
Inter-rater	R or ICC	Kappa or ICC
Internal Consistency	Alpha or Split-half or ICC	KR-20 or ICC (dichotomous)

Internal consistency: How well are three or more scale items measuring a single underlying characteristic?

Two requirements:

1) items should be moderately correlated with each other

2) each item should correlate with the total score

Two techniques to assess internal consistency:

1) split-half reliability

2) Cronbach's alpha/KR-20

Split-half estimates

Three-step procedure:

1) arbitrarily divide the scale into two halves and create total scores for two halves

2) correlate the two total scales

3) adjust the correlation upwards with the Spearman-Brown prophecy formula

G1_3 BEEN ABLE TO CONCENTRATE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	139	7.2	7.6	7.6
	2.00	1469	76.5	80.2	87.8
	3.00	192	10.0	10.5	98.3
	4.00	23	1.2	1.3	99.6
	8.00	5	.3	.3	99.8
	9.00	3	.2	.2	100.0
	Total	1831	95.4	100.0	
Missing	System	89	4.6		
Total		1920	100.0		

G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES

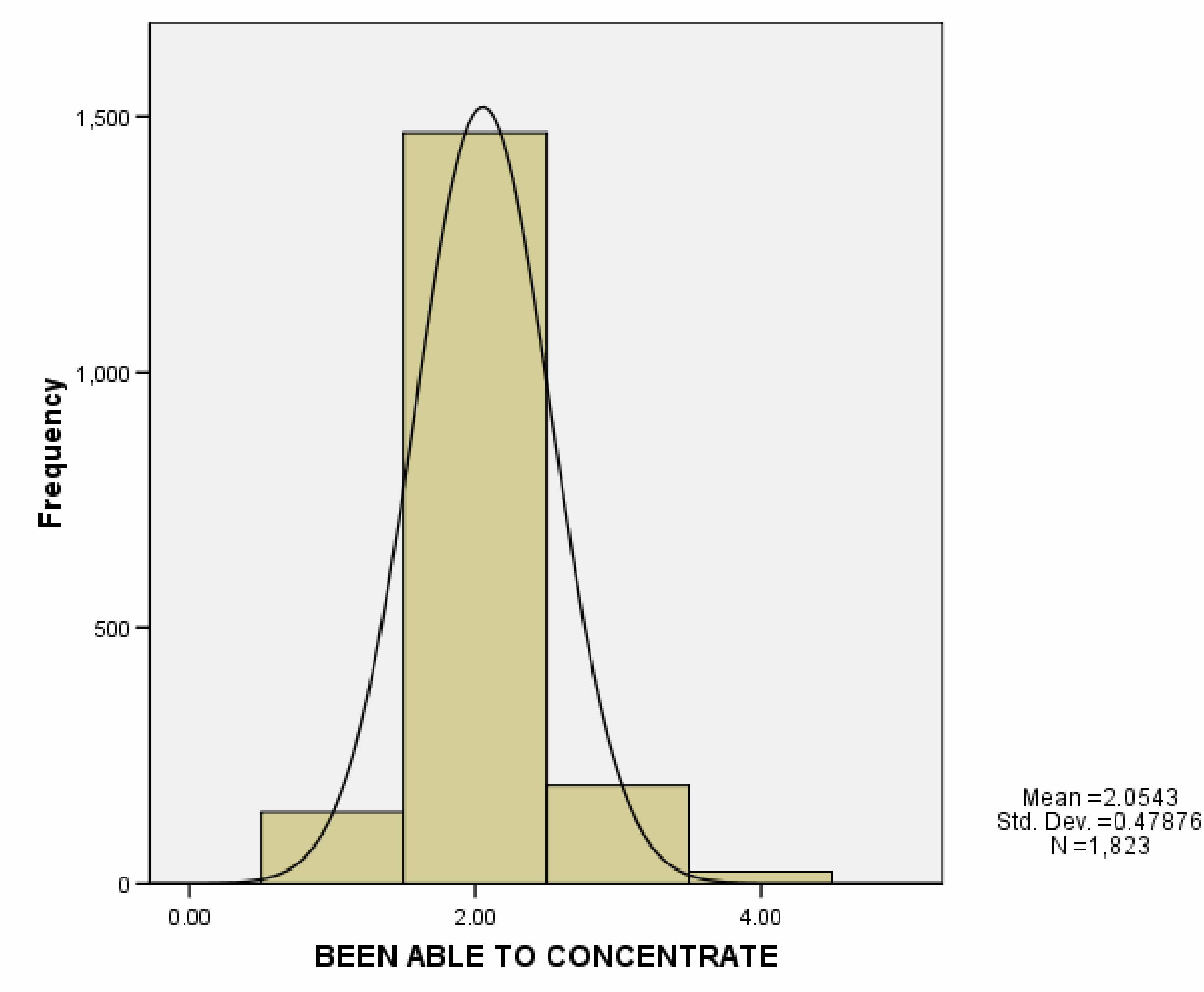
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	886	46.1	48.4	48.4
	2.00	808	42.1	44.1	92.5
	3.00	107	5.6	5.8	98.4
	4.00	17	.9	.9	99.3
	8.00	8	.4	.4	99.7
	9.00	5	.3	.3	100.0
	Total	1831	95.4	100.0	
Missing	System	89	4.6		
Total		1920	100.0		

G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES

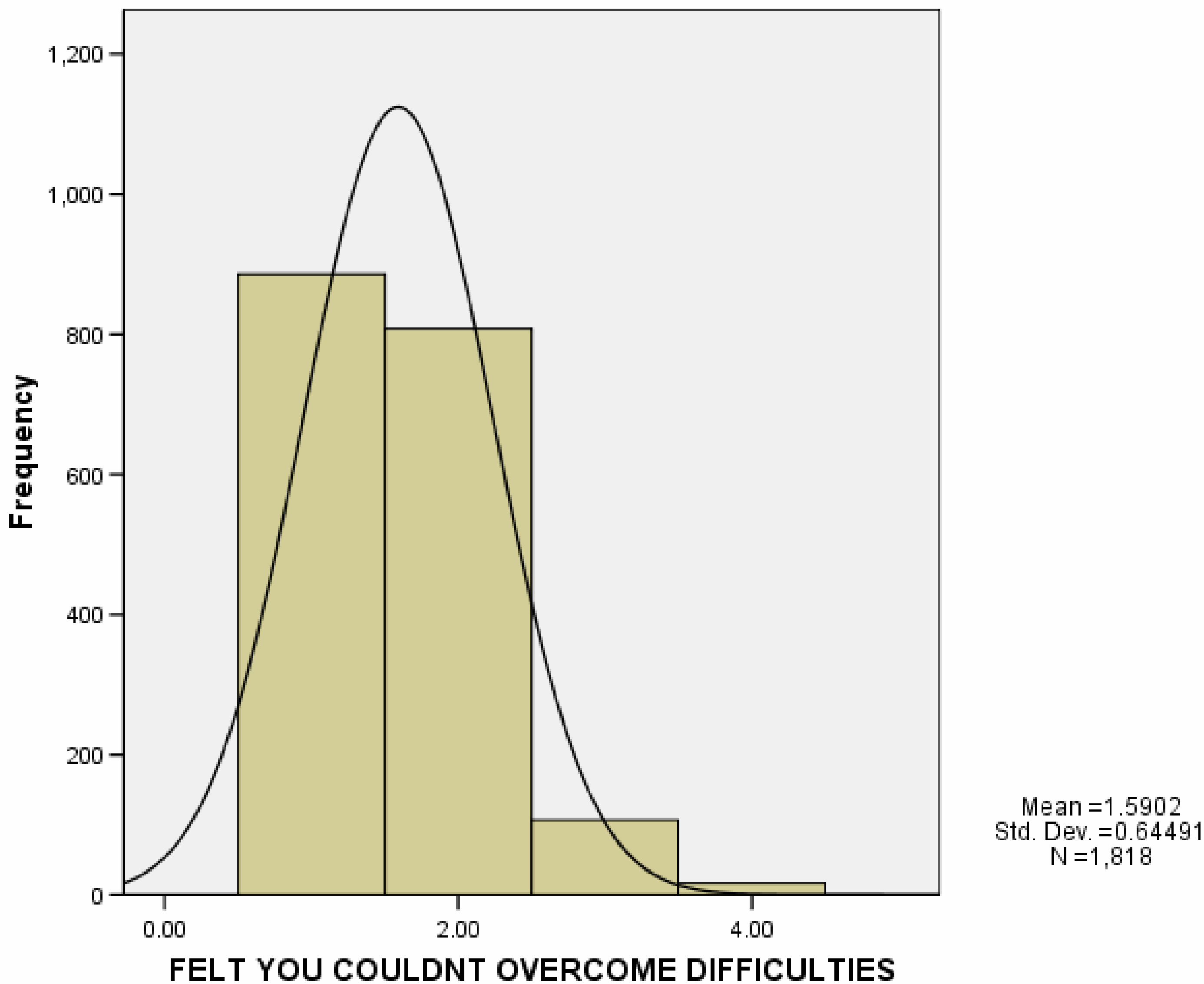
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	93	4.8	5.1	5.1
	2.00	1439	74.9	78.6	83.7
	3.00	247	12.9	13.5	97.2
	4.00	48	2.5	2.6	99.8
	8.00	2	.1	.1	99.9
	9.00	2	.1	.1	100.0
	Total	1831	95.4	100.0	
Missing	System	89	4.6		
Total		1920	100.0		

G12_3 BEEN TAKING THINGS HARD

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	619	32.2	33.8	33.8
	2.00	942	49.1	51.4	85.3
	3.00	229	11.9	12.5	97.8
	4.00	32	1.7	1.7	99.5
	8.00	4	.2	.2	99.7
	9.00	5	.3	.3	100.0
	Total	1831	95.4	100.0	
Missing	System	89	4.6		
Total		1920	100.0		



Frequency
1,500
1,000
500
0
0.00
2.00
4.00
BEEN ABLE TO CONCENTRATE
Mean =2.0543
Std. Dev. =0.47876
N =1,823


Frequency
1,200
1,000
800
600
400
200
0
0.00
2.00
4.00
FELT YOU COULDNT OVERCOME DIFFICULTIES
Mean =1.5902
Std. Dev. =0.64491
N =1,818

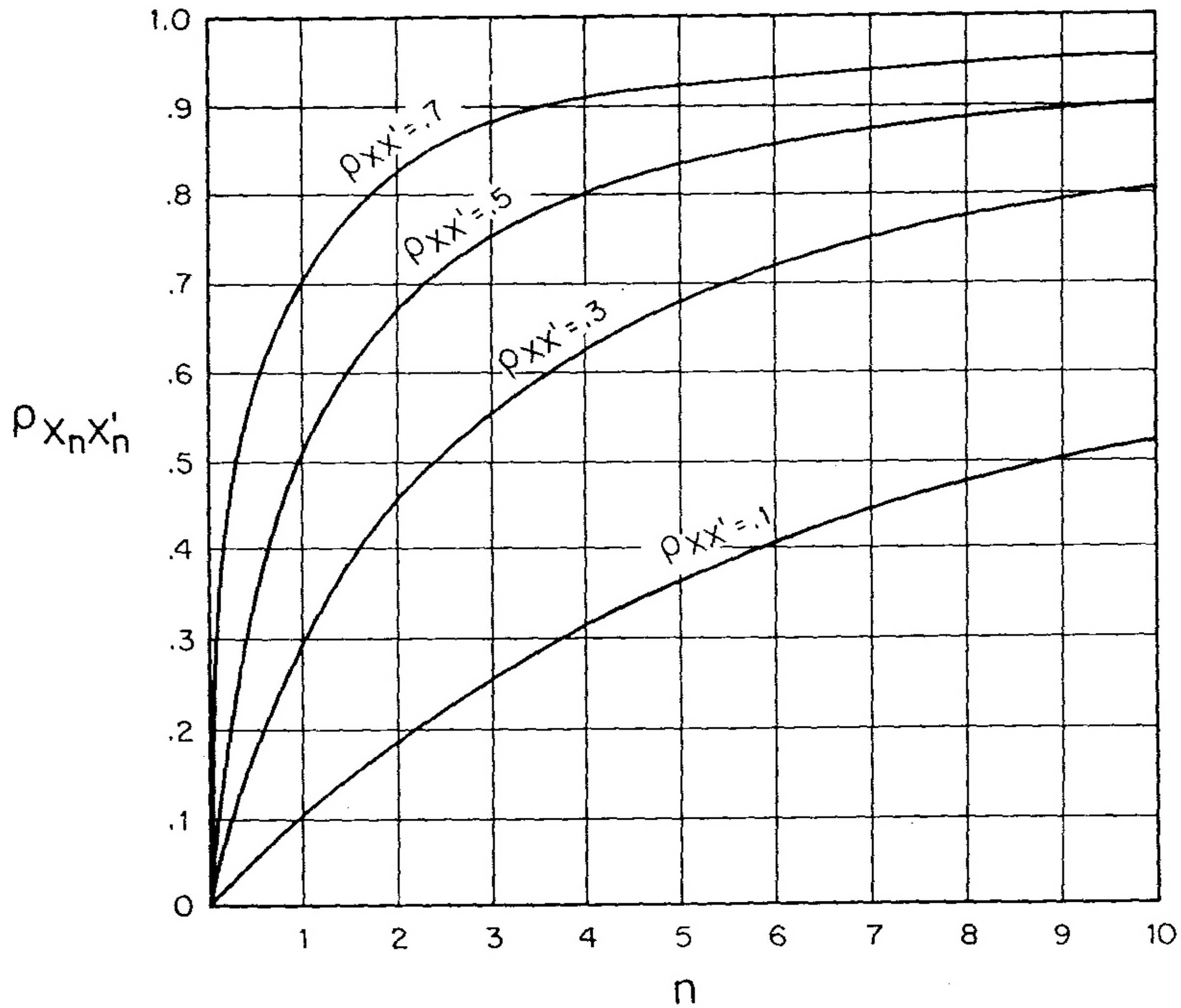


FIGURE 3.1 Diagram for showing increase in reliability as a function of increase in the length of scale (*n*) for different initial reliabilities.

Bohrnstedt, in Rossi et al, Handbook of Survey Research 1983, p. 78.

Spearman-Brown Prophecy formula

$$r_{SB} = \frac{Nr}{1 + (N-1)r}.$$

Where r = reliability of observed scale

$$N = \frac{\text{\# items in to-be-formed theoretical scale}}{\text{\# items in current, observed scale}}$$

Reliability Statistics

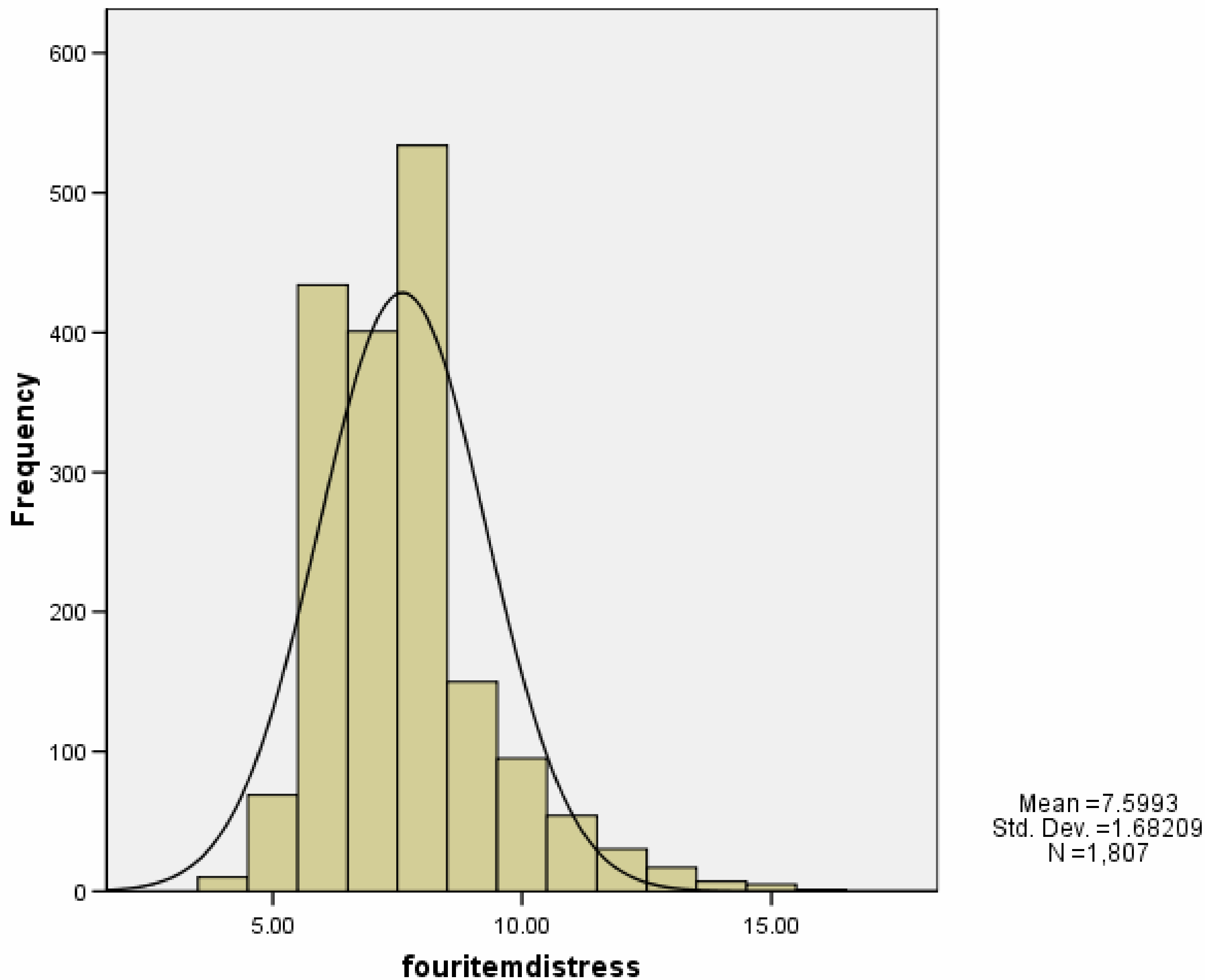
Cronbach's Alpha	Part 1	Value	.346
		N of Items	2[a]
	Part 2	Value	.523
		N of Items	2[b]
	Total N of Items		4
Correlation Between Forms			.555
Spearman-Brown Coefficient	Equal Length		.714
	Unequal Length		.714
Guttman Split-Half Coefficient			.709

a. The items are: G1_3 BEEN ABLE TO CONCENTRATE, G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES.

b. The items are: G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES, G12_3 BEEN TAKING THINGS HARD.

From 2 to 4 items, that is, n = 2

$$\rho_{sb} = \frac{2 * .555}{1 + 1 * .555} = .714$$



Frequency
600
500
400
300
200
100
0
5.00
10.00
15.00
fouritemdistress
Mean =7.5993
Std. Dev. =1.68209
N =1,807

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.346
		N of Items	2[a]
	Part 2	Value	.523
		N of Items	2[b]
	Total N of Items		4
Correlation Between Forms			.555
Spearman-Brown Coefficient	Equal Length		.714
	Unequal Length		.714
Guttman Split-Half Coefficient			.709

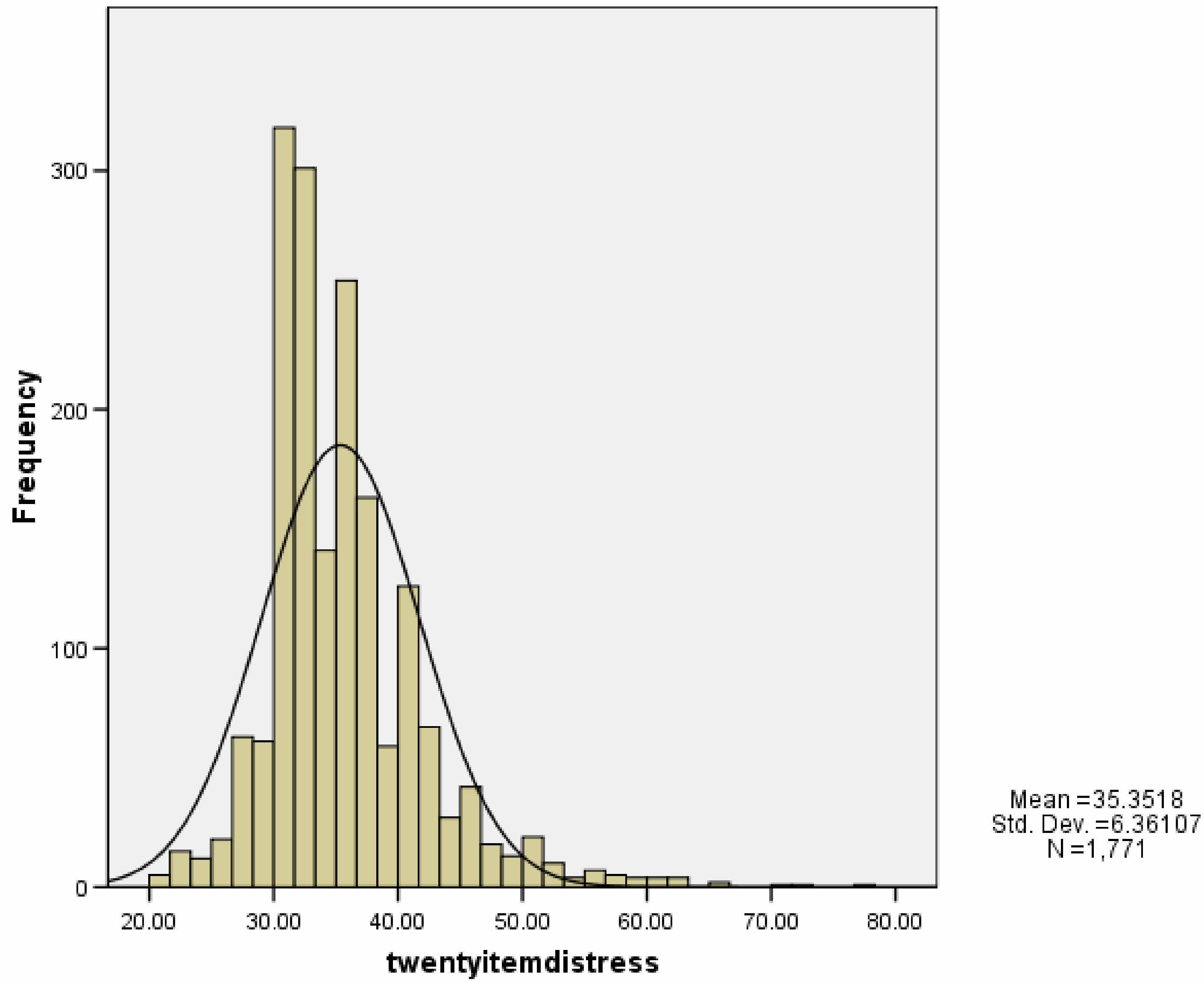
a. The items are: G1_3 BEEN ABLE TO CONCENTRATE, G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES.

b. The items are: G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES, G12_3 BEEN TAKING THINGS HARD.

From 2 to 20 items, that is, n = 10

$$\rho_{sb} = \frac{10 * .555}{1 + 9 * .555} = .925$$

(Alpha from 20 item scale = .87)



Frequency
300
200
100
0
20.00
30.00
40.00
50.00
60.00
70.00
80.00
twentyitemdistress
Mean =35.3518
Std. Dev. =6.36107
N =1,771

number of possible split halves = $\dfrac{N!/2}{[(N/2)!]^2}$

$$\frac{\binom{n}{r}}{2}$$

$(r = n/2)$

Number of questions	Possible split-halves
2	1 (A vs. B)
4	3 (AB, AC, AD vs. …)
6	10
8	35
10	126
12	462
14	1716
16	6435
18	24310
20	92378
30	77558760
50	63205303218876

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} V(y_i)}{\sum_{i=1}^{n} V(y_i) + 2\sum_{i<j}^{n}\sum^{n} C(y_i, y_j)}\right]$$

$$= \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} V(y_i)}{\sigma_x^2}\right]$$

n = number of items
y = individual items
x = total scale score $= \sum_{i=1}^{n} y_i$
i and j index items
$V(y_i)$ *= Variance of item* y_i
$C(y_i, y_j)$ *= Covariance of item* y_i *with item* y_j

Inter-Item Covariance Matrix

	G1_3 BEEN ABLE TO CONCENTRATE	G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES	G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES	G12_3 BEEN TAKING THINGS HARD
G1_3 BEEN ABLE TO CONCENTRATE	.229	.067	.090	.081
G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES	.067	.416	.099	.231
G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES	.090	.099	.269	.137
G12_3 BEEN TAKING THINGS HARD	.081	.231	.137	.504

$$\Sigma_{i=1}^{n} V(y_i) = .229+.416+.269+.504 = 1.418$$

$$\Sigma_{i<j}^{n}\Sigma^{n} C(y_i, y_j) = .067+.090+.081+.099+.231+.137 = .707$$

$$\alpha = \frac{4}{3}\left[1 - \frac{1.418}{1.418 + 2\,(.707)}\right] = .66$$

Inter-Item Correlation Matrix

	G1_3 BEEN ABLE TO CONCENTRATE	G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES	G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES	G12_3 BEEN TAKING THINGS HARD
G1_3 BEEN ABLE TO CONCENTRATE	1.000	.218	.361	.238
G10_3 FELT YOU COULDNT OVERCOME DIFFICULTIES	.218	1.000	.297	.505
G11_3 BEEN ABLE TO ENJOY DAILY ACTIVITIES	.361	.297	1.000	.372
G12_3 BEEN TAKING THINGS HARD	.238	.505	.372	1.000

Alpha in terms of average inter-item correlations
(Assume all items have equal variances)

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	1.900	1.590	2.134	.544	1.342	.060	4
Item Variances	.355	.229	.504	.275	2.201	.016	4
Inter-Item Covariances	.118	.067	.231	.164	3.432	.003	4
Inter-Item Correlations	.332	.218	.505	.287	2.314	.010	4

$$\alpha = \frac{n * r_{ij}}{1 + (n-1) * r_{ij}}$$

$$\alpha = \frac{4 * .332}{1 + 3 * .332} \quad \sim= .66$$

Values of Cronbach's alpha for various combinations of different number of items and different average interitem correlations

# items	average interitem correlation					
	.0	.2	.4	.6	.8	1.0
2	.0	.333	.572	.750	.889	1.0
4	.0	.500	.727	.857	.941	1.0
6	.0	.600	.800	.900	.960	1.0
8	.0	.666	.842	.924	.970	1.0
10	.0	.714	.870	.938	.976	1.0

Notes on Cronbach's Alpha:

1) It is the same as the average of all split-half reliabilities

2) It is mathematically equivalent to the ICC for the mean of multiple observations with fixed raters/items

3) The most common measure of reliability in the social sciences

Kuder-Richardson 20:

$$KR20 = \left(\frac{N}{N-1} \right) \left(1 - \frac{\Sigma p_i q_i}{S_x^2} \right)$$

N is the number of dichotomous items
p_i is the proportion responding positively to the ith item
q_i equals 1 - p_i
S_x^2 is the variance of the total composite

note: in Stata, the alpha command, when given dichotomous items as arguments, will produce the KR20 coefficient

Kappa Coefficient

Contingency table for two observers

		Observer 2		
		Present	Absent	Total
	Present	20	15	35
Observer 1	Absent	10	55	65
	Total	30	70	100

Overall agreement is:

Kappa Coefficient

		Observer 2		
		Present	Absent	Total
	Present	20	15	35
Observer 1	Absent	10	55	65
	Total	30	70	100

$$\text{kappa} = \frac{P_o - P_e}{1.0 - P_e}$$

P_o = observed proportion of agreements

P_e = expected proportion of agreements

Expected agreement in top left cell is:

Expected agreement in bottom right cell:

$$\text{kappa} = \frac{(75/100) - ((10.5+45.5)/100)}{1.0 - (10.5+45.5)/100} = .43$$

Kappa Coefficient

		Observer 2		
		Present	Absent	Total
	Present	10	5	15
Observer 1	Absent	5	80	85
	Total	15	85	100

Overall agreement is:

Expected agreement in top left cell is:
Expected agreement in bottom right cell:

kappa =

1) Best interpretation of kappa is to compare its values on other, similar scales

2) Another suggested kappa interpretation scale:

Kappa Value	Interpretation
Below 0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

(source: Landis, J. R. and Koch, G. G. 1977. *Biometrics* 33: 159-174)

Discrepancy Between the DIS and SCAN for the Lifetime Occurrence of Depressive Disorder in the Baltimore ECA follow-up

	Psychiatrist Using SCAN		
Interview Using DIS	**Never a case**	**Positive Diagnosis**	**Total**
Never a case	260	55	315
Positive diagnosis	11	23	34
Total	**271**	**78**	**349**

Kappa = 0.20

Source: Eaton et al. Arch Gen Psychiatry, 2000

Weighted kappa

1) arbitrary weights

2) linear weights

3) quadratic weights

disagreement weights based on the square of the amount of discrepancy

Baltimore ECA Follow-Up

349 subjects											
		Number of DSM Symptom Groups with SCAN									
		0	1	2	3	4	5	6	7	8	9
Number of DSM Symptom Groups With DIS	0	119	51	28	12	11	22	12	8	3	0
	1	0	1	2	3	0	0	0	0	0	0
	2	2	3	0	0	0	1	0	0	0	0
	3	0	2	2	2	1	0	1	0	1	0
	4	1	4	4	2	0	4	3	2	2	1
	5	1	1	0	1	1	1	2	3	4	0
	6	0	1	1	1	0	1	3	2	0	0
	7	0	1	2	0	1	0	0	6	0	1
	8	0	0	0	2	0	0	0	0	0	0
	9	0	0	0	0	2	0	0	0	2	0

Continued

Agreement between DIS and SCAN for Lifetime Depressive Disorder

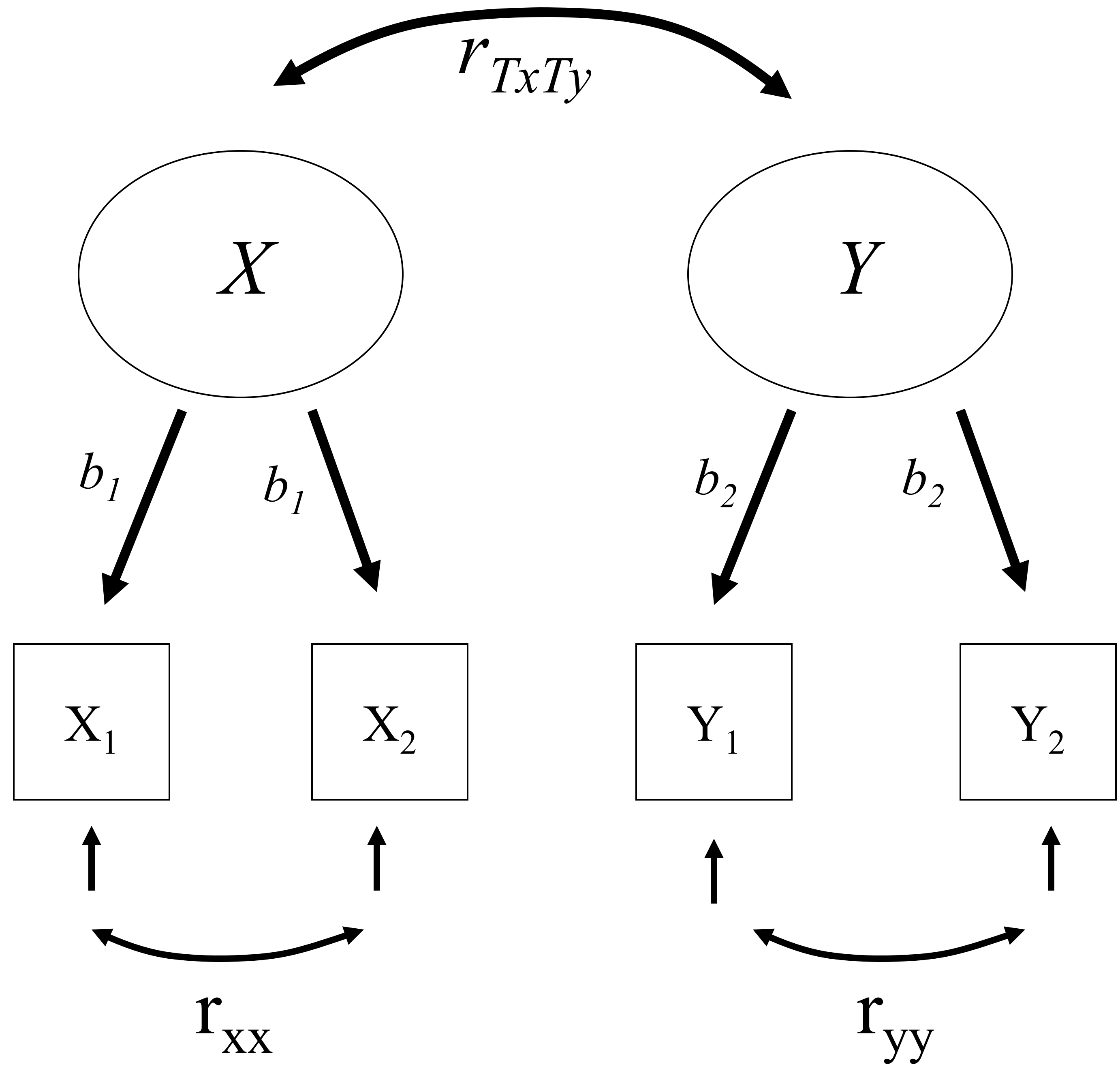
- Kappa values
  - Two by two table: 0.32
  - Nine by nine table
    - Unweighted: 0.20
    - Linear weights: 0.31
    - Squared weights: 0.43
    - Pearson correlation: 0.49

Effects of reliability on statistical estimates:

Correction for attenuation:

$$r_{TxTy} = \frac{r(x, y)}{\sqrt{r_{xx} r_{yy}}}$$

Variables with low reliability will have low observed correlations, even if the true correlation between them is high.


r TxTy
X
Y
b1
b1
b2
b2
X1
X2
Y1
Y2
rxx
ryy

$$r_{xy} = b_1 * r_{TxTy} * b_2$$

$$r_{xx} = b_1 * b_1 \ \ and \ \ r_{yy} = b_2 * b_2$$

For tau equivalent measures:

$$r_{TxTy} = \frac{r(x,y)}{\sqrt{r_{xx} r_{yy}}}$$

"The correlation of the true scores is equal to the correlation of the observed scores divided by the square root of the product of the reliabilities."

Examples of correction for attenuation:

observed correlation of .3

	.2	.4	.6	.8	1.0
.2	–	–	.87	.75	.67
.4	–	.75	.61	.53	.47
.6	.87	.61	.50	.43	.39
.8	.75	.53	.43	.38	.33
1.0	.67	.47	.39	.33	.30

observed correlation of .5

	.2	.4	.6	.8	1.0
.2	–	–	–	–	–
.4	–	–	–	.88	.79
.6	–	–	.83	.72	.65
.8	–	.88	.72	.63	.56
1.0	–	.79	.65	.56	.50

Examples of correction for attenuation (continued):

true correlation of .5

	.2	.4	.6	.8	1.0
.2	.10	.14	.17	.20	.22
.4	.14	.20	.24	.28	.32
.6	.17	.24	.30	.35	.39
.8	.20	.28	.35	.40	.45
1.0	.22	.32	.39	.45	.50

true correlation of .7

	.2	.4	.6	.8	1.0
.2	.14	.20	.24	.28	.31
.4	.20	.28	.34	.40	.44
.6	.24	.34	.42	.48	.54
.8	.28	.40	.48	.46	.63
1.0	.31	.44	.54	.63	.70