Answer Key to Problem Set #1

Two notes:

q#4e: Please disregard

q#5e: The frequency tables of the total CESD scales of 94, 96 and 98 in question 5e should sum up to 328 observation not 924 (the student solving this PS forgot to reshape wide before doing the math)

Problem Set 1



3a)Three variables have missing values: mari1594, mari1596, mari1598
b)

Figure: Histogram of one measure of self esteem stratified by the 1994 smoked marijuana before age 15 report (1=before 15, 0=all others)

4a)We performed one split-half reliability test by randomly dividing the self-esteem scale. We chose to divide the 10-item scale into evens and odds. We then obtained the correlation of the even and odd variables and used the Spearman-Brown Prophecy Formula to compute one split-half reliability and obtained a correlation of 0.87063075.

. corr spliteven splitodd
(obs=328)

Graphs by mari1594

Pearson correlation for spliteven and slplitodd=0.7709

Calculate one split-half reliability test using the Spearman-Brown Prophecy Formula: display (2*0.7709)/(1+0.7709)=0.87063075

b)We computed the average of all possible split-half reliabilities for this scale by calculating Cronbach's alpha and obtained an overall apha = 0.8551

alpha slf941-slf9410

Test scale = mean(unstandardized items)

Average inte	ritem covaria	nce: .1491089
Number of it	ems in the so	ale: 10
Scale reliab.	ility coeffic	ient: 0.8551

c)We reshaped the data to use the anova to answer the following questions. We computed the ICC for the total scale assuming that each item is fixed (fixed rater design). The ICC for the fixed rater design = 0.85511819

anova slf94 id item

	Number of ob Root MSE	s = = .5	3280 R-s 02627 Adj	quared R-squared	= 0.4733 = 0.4131
Source	Partial SS	df	MS	F	Prob > F
Model	667.996341	336	1.98808435	7.87	0.0000
id item	570.197256 97.7990854	327 9	1.7437225 10.866565	6.90 43.01	0.0000 0.0000
Residual	743.500915	2943	.252633678		
Total	1411.49726	3279	.430465769		

Fixed rater design ICC =

(Between Mean Square Variance-(Error Variance))/Between Mean Square Variance

display (1.7437225-.252633678) / 1.7437225 = 0.85511819

d)The following is the formula for the average ICC for a single item on a scale:

(Between Mean Square Variance - Error Mean Square Variance)/(Between Mean Square Variance +(R-1)*Error Mean Square) where R=number of raters

display (1.7437225-.252633678)/(1.7437225 +((10-1)* .252633678))

The average reliability using the ICC for a single item on the self-esteem scale = 0.3711553

 $\mathsf{e}) \, \mathsf{W}\mathsf{e}$ used the ICC for the random rater design to calculate the reliability of the scale

(Between Mean Square-Error Mean Square)/(Between Mean Square +((Rater Mean Square-Error Mean Square)/n))

display (1.7437225 -.252633678)/(1.7437225 + ((10.866565-.252633678)/328))

The ICC using the random rater design = 0.83953825

slft94	Freq.	Percent	Cum.
16	1	0.30	0.30
22	1	0.30	0.61
24	8	2.44	3.05
25	7	2.13	5.18
26	10	3.05	8.23
27	17	5.18	13.41
28	21	6.40	19.82
29	30	9.15	28.96
30	55	16.77	45.73
31	22	6.71	52.44
32	37	11.28	63.72
33	18	5.49	69.21
34	20	6.10	75.30
35	12	3.66	78.96
36	15	4.57	83.54
37	13	3.96	87.50
38	11	3.35	90.85
39	14	4.27	95.12
40	16	4.88	100.00
Total	328	100.00	

5a) Frequency of the total self-esteem scores for the 1994 scale

Frequency of the total self-esteem scores for the 1996 scale

slft96			Freq.	Percent		Cum	•		
20 21	-+ 		1 3	0.30 0.91		0.30	-) 2		
23	İ		1	0.30		1.52	2		
24			3	0.91		2.44	1		
25			3	0.91		3.35	5		
26			4	1.22		4.5	7		
27			5	1.52		6.10)		
28			18	5.49	1	1.59	9		
29			40	12.20	2	23.78	3		
30			63	19.21	4	12.99	9		
31			24	7.32		50.30	C		
32			12	3.66	۲ ۲	53.90	5		
33			18	5.49	۲ ۲	59.45	5		
34			26	7.93	6	57.38	3		
35			18	5.49	5	72.8	7		
36			16	4.88	5	77.74	1		
37			17	5.18	8	32.93	3		
38			12	3.66	8	36.59	9		
39			13	3.96	0	90.55	5		
40			31	9.45	10	0.00	C		
Total	-+		328	100.00			-		
Frequencies	of	the	total	self-esteem	scores	for	the	1998	scale

slft98		Freq.	Percent	Cum.
	+			

19	1	0.30	0.30
20	1	0.30	0.61
23	2	0.61	1.22
24	6	1.83	3.05
25	4	1.22	4.27
26	4	1.22	5.49
27	5	1.52	7.01
28	12	3.66	10.67
29	30	9.15	19.82
30	60	18.29	38.11
31	23	7.01	45.12
32	15	4.57	49.70
33	25	7.62	57.32
34	17	5.18	62.50
35	15	4.57	67.07
36	16	4.88	71.95
37	17	5.18	77.13
38	17	5.18	82.32
39	22	6.71	89.02
40	36	10.98	100.00
Total	328	100.00	

b)To assess the over-time reliability of the total self-esteem scale across 1994 and 1996 we could use Pearson's correlation or the ICC for a fixed rater design to obtain the test-retest reliability for continuous data. Cronbach's alpha can also be used to assess test-retest reliability.

c)Using Pearson's correlation, we get r= 0.4670 for the test-retest reliability for the self-worth scales for 1994 and 1996.

. corr slft94 slft96 (obs=328)

	slft94	slft96
slft94 slft96	1.0000 0.4670	1.0000

d) The ICC for a fixed rater design to assess the reliability of all the items asked at each of the three time points (30 items total) is 0.68242111.

anova slft id wave

	Number of obs Root MSE	= = 3.	984 R-s .29756 Adj	quared R-squared	= 0.6175 = 0.4251
Source	Partial SS	df	MS	F	Prob > F
Model	11481.6148	329	34.8985253	3.21	0.0000
id	11196.4787	327	34.2399959	3.15	0.0000
wave	285.136179	2	142.568089	13.11	0.0000
Residual	7111.53049	654	10.8738998		

Total | 18593.1453 983 18.9146951

display (34.2399959-10.8738998)/34.2399959

The ICC for the scale reliability for the three years combined = 0.68242111

6a)Reliability of repeated measures of a trait would be higher than the reliability of repeated measures of a state due to minimal within variance. The within variance of the true score would be lower for a trait because traits, such as self-esteem, are relatively constant over time. Therefore, people are more likely to give consistent responses. States, on the other hand, have higher within variance because they are more likely to fluctuate between two time points. To achieve high reliability, a researcher ideally would like to have high variance between subjects, while minimizing the within variance for each individual.

b) Frequency of the total depression score for the 1994 scale

cesdt94	Freq.	Percent	Cum.
0 1	+ 96 81	9.76 8.23	9.76 17.99
2	66	6.71	24.70
3	108	10.98	35.67
4	102	10.37	46.04
5	87	8.84	54.88
6	99	10.06	64.94
7	60	6.10	71.04
8	84	8.54	79.57
9	51	5.18	84.76
10	54	5.49	90.24
11	24	2.44	92.68
12	15	1.52	94.21
13	21	2.13	96.34
14	3	0.30	96.65
15	12	1.22	97.87
16	6	0.61	98.48
17	15	1.52	100.00
Total	984	100.00	

Frequency of the total depression score for the 1996 scale

cesdt96		Freq.	Percent	Cum.
	-+			
0		72	7.32	7.32
1		60	6.10	13.41
2		111	11.28	24.70
3		132	13.41	38.11
4		129	13.11	51.22
5		90	9.15	60.37
6		81	8.23	68.60
7		75	7.62	76.22
8		57	5.79	82.01
9		48	4.88	86.89
10		30	3.05	89.94
11		24	2.44	92.38

12	27	2.74	95.12
13	12	1.22	96.34
14	15	1.52	97.87
15	9	0.91	98.78
16	6	0.61	99.39
18	3	0.30	99.70
20	3	0.30	100.00
Total	+ 984	100.00	

Frequency of the total depression score for the 1998 scale

cesdt98	Freq.	Percent	Cum.
0	99	10.06	10.06
1	75	7.62	17.68
2	93	9.45	27.13
3	132	13.41	40.55
4	90	9.15	49.70
5	108	10.98	60.67
6	99	10.06	70.73
7	63	6.40	77.13
8	69	7.01	84.15
9	57	5.79	89.94
10	18	1.83	91.77
11	33	3.35	95.12
12	18	1.83	96.95
13	6	0.61	97.56
14	12	1.22	98.78
15	6	0.61	99.39
16	6	0.61	100.00
Total	984	100.00	

c)The ICC for the fixed rater design is an appropriate reliability test to examine the reliability of the total CESD score using all three time points. The fixed rater design ICC for overall reliability of the CESD scale=0.64370603

anova cesdt id repeat

	Number of obs Root MSE	= 2.9	984 R-s 95345 Adj	quared R-squared	= 0.5852 = 0.3766
Source	Partial SS	df	MS	F	Prob > F
Model	8049.59858	329	24.466865	2.80	0.0000
id repeat	8005.69004 43.9085366	327 2	24.4822325 21.9542683	2.81 2.52	0.0000 0.0815
Residual	5704.75813	654	8.72287176		
Total	13754.3567	983	13.9922245		

display (24.4822325-8.72287176)/24.4822325= 0.64370603

d)The difference between the over-time reliability of the CESD (r=0.64370603) and self-esteem (r=0.68242111) scales was in the expected direction. The trait should have higher reliability due to minimal within variance.

e)A positive relationship exists between internal consistency and reliability: the higher the internal consistency, the higher the reliability. Because we did not observe a more distinct difference between the reliability for the state (depression) and trait (self-esteem), we hypothesize that the CESD scale has higher internal consistency than the self-esteem scale, narrowing the gap between their reliability scores.

f)Contrary to our hypothesis, the self-esteem scale has higher internal consistency (r=0.8551) than the depression scale (r=0.6925). The lower internal consistency for the depression scale may reflect that the items do not adequately measure the underlying latent construct (lower reliability). We have effectively ruled out higher internal consistency of the depression scale to explain our findings. Therefore, the between variance for depression must be greater than it is for self-esteem. If the state has more true variance between subjects, then it will result in a higher reliability.

7a) The overall agreement in the subject's report of sexual intercourse before age 15 across 1994 and 1996 is 0.80792683.

tab sex1594 sex1596

sex1594		sex1596 0	1	I	Total
0 1	+ - 	185 31	32 80		217 111
Total		216	112		328

Overall agreement=185+80/328=0.8079683

b) The chance-adjusted agreement in the subject's report of sexual intercourse before age 15 across 1994 and 1996 is 0.5720.

	Expected				
Agreement	Agreement	Kappa	Std. Err.	Z	Prob>Z
80.79%	55.12%	0.5720	0.0552	10.36	0.0000

kappa= 0.5720

c)Crosstabs

		sex1596			
sex1594	 +-	0	1	 -+-	Total
0	+ - 	185 85.25 85.65	32 14.75 28.57	-+- 	217 100.00 66.16

1		31 27.93 14.35	80 72.07 71.43		111 100.00 33.84
Total	-+	216 65.85 100.00	112 34.15 100.00	- + · 	328 100.00 100.00

	sex	1598	
sex1596	I 0	1	Total
0	184	32	216
	85.19	14.81	100.00
	83.64	29.63	65.85
1	36	76	112
	32.14	67.86	100.00
	16.36	70.37	34.15
Total	220	108	328
	67.07	32.93	100.00
	100.00	100.00	100.00

	sex15	98	
sex1594	0	1	Total
0	178	39	217
	82.03	17.97	100.00
	80.91	36.11	66.16
1	42	69	111
	37.84	62.16	100.00
	19.09	63.89	33.84
Total	220	108	328
	67.07	32.93	100.00
	100.00	100.00	100.00

d) The results suggest that over time, subject's report of sexual intercourse is subject to increasing non-systematic (random) error. There does not appear to be any trend to explain the changing values of the concordant and discordant cells, suggesting that some people are acknowledging sexual experiences later in life, whereas others are forgetting early sexual intercourse.

8a)Frequency of the dichotomous variable for the 1994 self-esteem scale

self94	Freq.	Percent	Cum.
0 1	172 156	52.44 47.56	52.44 100.00
Total	328	100.00	

Frequency of the dichotomous variable for the 1996 self-esteem scale

self96	Freq.	Percent	Cum.
0 1	165 163	50.30 49.70	50.30 100.00
Total	328	100.00	

Frequency of the dichotomous variable for the 1998 self-esteem scale self98 | Freq. Percent Cum. 0 | 148 45.12 45.12 1 | 180 54.88 100.00 Total | 328 100.00

b) The kappa statistic is appropriate to assess the test-retest reliability for the dichotomous variables for the 1994 ad 1996 scales. The chanceadjusted reliability is 0.3718.

. kap self94 self96, tab

	self	96				
self94	0	1	 _+	Total		
0 1	117 48	55 108	' _+	172 156		
Total	165	163		328		
Agreement	Expected Agreement	Карра	Std.	Err.	Z	Prob>Z
68.60%	50.01%	0.3718	0.0	0552	6.74	0.0000

c)The reliability based on the dichotomized scales (r=0.3718) is lower than the reliability obtained when using the continuous version of the scale (r=0.4670).

d) The major reason for the difference in reliabilities is that when a continuous variable is categorized, in this case dichotomized, data is "thrown away." This grouping of data minimizes the between variance for subjects, thus lowering the reliability. The reliability calculated using the continuous scale preserves the between reliability, so it has higher reliability.

EXTRA CREDIT

A- The answer is 0.467, (corr selft94 selft96)

B- The answer is 0.637, (alpha selft94 selft96)

C- We use the Spearman Brown Prophecy, using the coefficient "r" from (a) to get an answer similar to (b) because when using the SB in this case, the end result is "alpha correlation" which is mathematically equivalent to ICC for Fixed Rater design (applied in (b)).

SB= N*r/ 1+ (N-1)*r → 2*(0.467)/1+0.467= 0.637

D- The correlation coefficient we use in (a) is pearson's correlation. But correlation assumes continuous variables, and when using "r" on "binary variables" would attenuate the relation. An alternative measure of correlation (as per our notes) that we could use on binary variables is the "tetrachoric correlation" that estimates what would the correlation between 2 binary variables be if they were measured on a continuous scale.

It equals to \rightarrow OR-1/OR+1 \rightarrow 4.79-1/4.79+1= 0.65

selft94br	selft96br 0	1	Total
0 1	117 48	55 108	172 156
Total		163	328

OR= ad/bc= 4.7863636

One might argue "but how come it is higher than the "r" which is quasi-continuous, and the answer is it is higher because of its assumption (a strong one), which is that the "underlying quantity of interest is truly continuous"!!

E-

Cut-off is equal or below 25:

selft94br1	Freq.	Percent	Cum.
0 1	17 311	5.18 94.82	5.18 100.00
Total	328	100.00	
selft96br1	Freq.	Percent	Cum.
0	11	3.35	3.35

1	1	317	96.65	100.00
Total		328	100.00	

Cut-off is equal or below 28:

selft94br2		Freq.	Percent	Cum.
0 1		65 263	19.82 80.18	19.82 100.00
Total		328	100.00	

Cum.	Percent	Freq.	selft96br2
11.59 100.00	11.59 88.41	38 290	0 1
	100.00	328	Total

F-

	Alpha-correlation
Cut-off <=25	0.41
Cut-off <=28	0.50

G- Alpha is getting lower as the cut-off point decreases. Alpha works better for continuous or quasi-continuous variables than binary variables (the estimate in the latter case is attenuated). When categorizing the variables into a dichotomous variable, some of the variability is lost.