Chapter. 14
Association for Ordinal

If your variables are "Group Ordinal Variables"…

- Strongly agree, agree, don't know, disagree, strongly disagree…
- The basic logic is PRE (Proportional Reduction of Error)
- Gamma, Somer's D, and Kendall Tau.
- The key idea, in addition to PRE, is the concept of "pairs."

The Concept of Pairs

- Nc: pairs that display the same order on both X and Y (concordant)
- Nd: pairs that display the inverse order on X and Y (discordant)
- Eg. The higher the social class of parents, the higher the social class of offspring: Nc
- The higher the social class of parents, the lower the social class of offspring: Nd.

Nc and Nd

POSITIVE RELATIONSHIP		INDEPENDENT VARIABLE		
		Low	Mid	High
DEPENDENT VARIABLE	Low			
	Mid			
	High			

NEGATIVE RELATIONSHIP		INDEPENDENT VARIABLE		
		Low	Mid	High
DEPENDENT VARIABLE	Low			
	Mid			
	High			

(+) ↑ ; (−) ↓

Cases	Parent	Offspring
1	Lower	Lower
2	Upper	Upper
3	Med	Upper
4	Upper	Lower
5	Med	Med

- (1,2) Nc
- (1,3) Nc
- (1,4) Tx (Pairs ties on X)
- (1,5) Nc
- (2,3) Ty (Pairs ties on Y)
- (2,4) Tx
- (2,5) Nc
- (3,4) Nd
- (3,5) Tx
- (4,5) Nd

L → U (+)
L → U (+)
same

L → M (+)
L → U (+)
same

L → M (+)
L → M (+)
same

U → M (−)
U → M (−)
same

M → U (+)
U → L (−)
opp.

U → M (−)
L → M (+)
opp.

Gamma

- Nc is the number of concordant pairs: 4
- Nd is the number of discordant pairs: 2

$$G = \frac{N_c - N_d}{N_c + N_d}$$

- Gamma is (4-2)/(4+2)=1/3=0.333…
- Positive relationship: It is more likely that two variables go with each other in the same direction.
- In terms of PRE, you can reduce the amount of error by 33% when predicting offspring's social class if you know about parents' social class than if you don't know.

Somer's D

- Gamma does not care about tied pairs. Therefore, it sometimes overestimates the association. The number of tied pairs increases, Gamma is not that reliable.
- Somer's D=(Nc-Nd)/(Nc+Nd+Ty)
- For the example above, Somer's D is (4-2)/(4+2+1)=0.286.
- When there are no tied pairs, gamma is the same with Somer's D.

Kendall Tau

- If there are many tied pairs, Kendall Tau is the most common because it considers both Tx and Ty. In other words, there are no tied pairs, then Gamma is the same with Kendall Tau.
- Kendall's Tau b is defined:

  [Nc-Nd]/sqrt[(Nc+Nd+Ty)(Nc+Nd+Tx)]

2. Do attitudes about capital punishment co-vary with formal education? Our researcher believes that, due to the "liberal bias" of educators, the more formal education one gets, the less likely one is to favor capital punishment for murder. Data from the 1998 General Social Survey are used to examine this question.

OPPOSE DEATH PENALTY FOR MURDER * RS HIGHEST DEGREE Crosstabulation

			RS HIGHEST DEGREE					Total
OPPOSE DEATH PENALTY FOR MURDER			LT HIGH SCHOOL	HIGH SCHOOL	JUNIOR COLLEGE	BACHELOR	GRADUATE	
	FAVOR	Count	34	114	17	26	10	201
	OPPOSE	Count	9	40	3	9	12	73
Total		Count	43	154	20	35	22	274

- Using SPSS is recommended.
- Caution: You have got the crosstabulation, not raw data like GSS. So, you have to command SPSS to read 34, 114, 17… as frequency. (Recall the homework question 2 in chapter 10!)

a. state a null and research hypothesis for this problem

b. which variable is independent, which is dependent? Why?

c. calculate: total number of pairs, pairs that are concordant, pairs that are discordant, pairs that are tied on X, pairs that are tied on Y.

d. calculate gamma and Somer's D

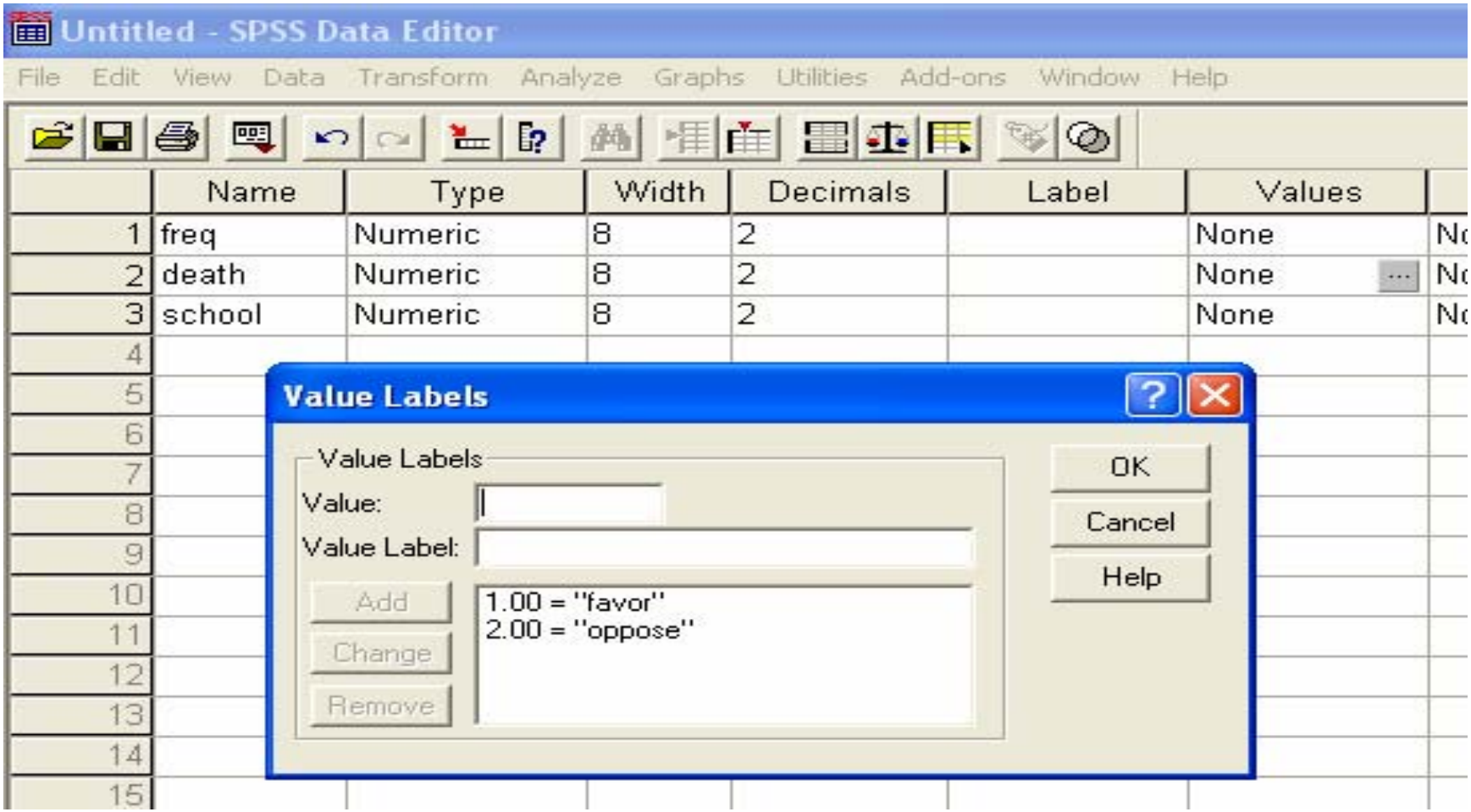
e. comment on the strength of association, and it's form. Are the data consistent with the researcher's hypothesis?

- It is unavoidable to solve the subquestion (c) by hand! Again, if both increase in the same direction, then Nc; if both move in the opposite direction, then Nd; if X changes but Y does not, then called tied on Y; if Y changes but X does not, then called tied on X.

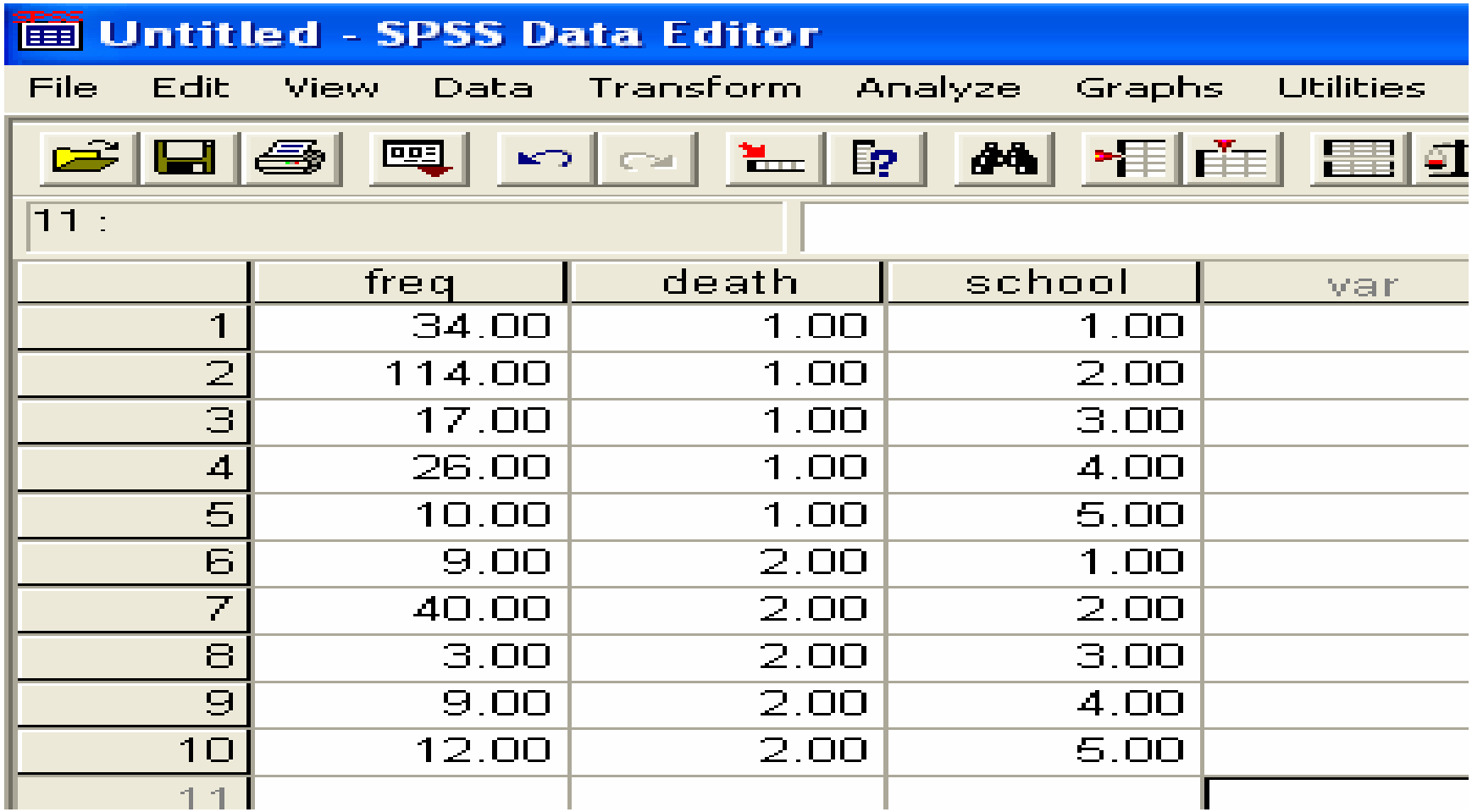
Basic Steps

- V1 is freq; V2 is death; V3 is school.
- For V2, favor is 1, and oppose is 2.
- For V3, from LT high school is 1… to graduate is 5.
- For example, in the data editor, 34 if a respondent is LT high school(1) and Favor(1). 12 if a respondent is graduate(5) and oppose(2).
- After inputting the data, don't forget to go to Data>Weight cases!

Variable View Window

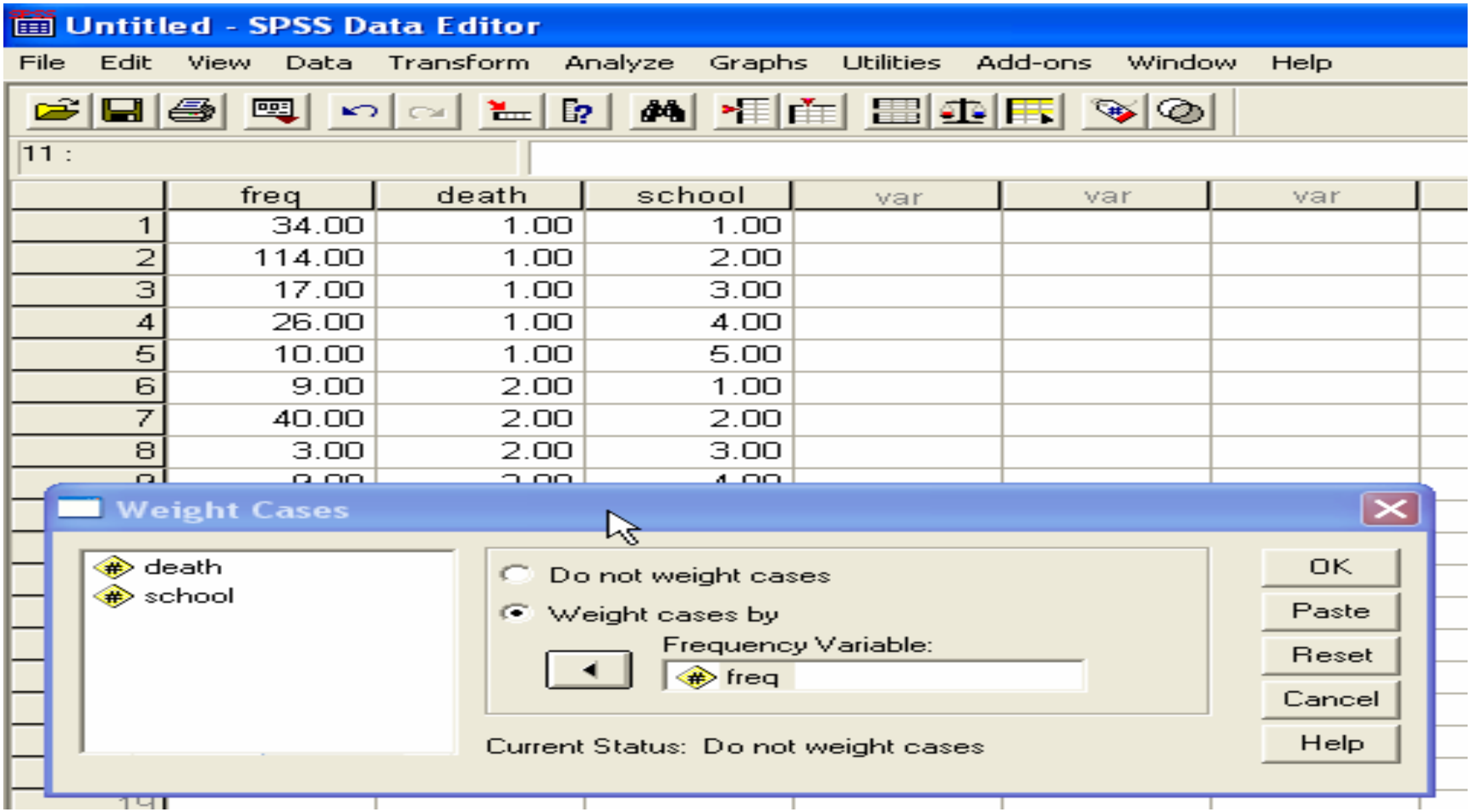


Input the data

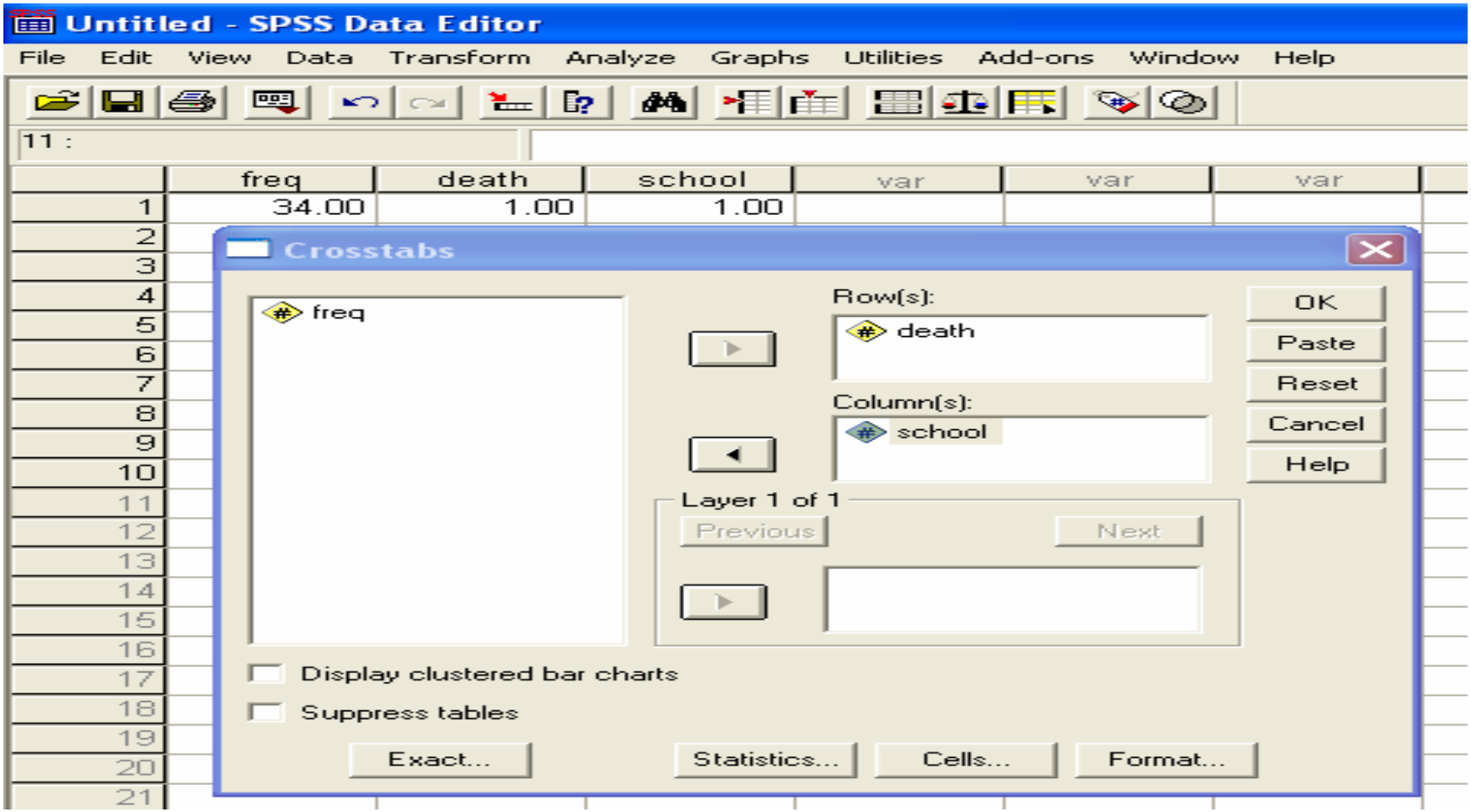


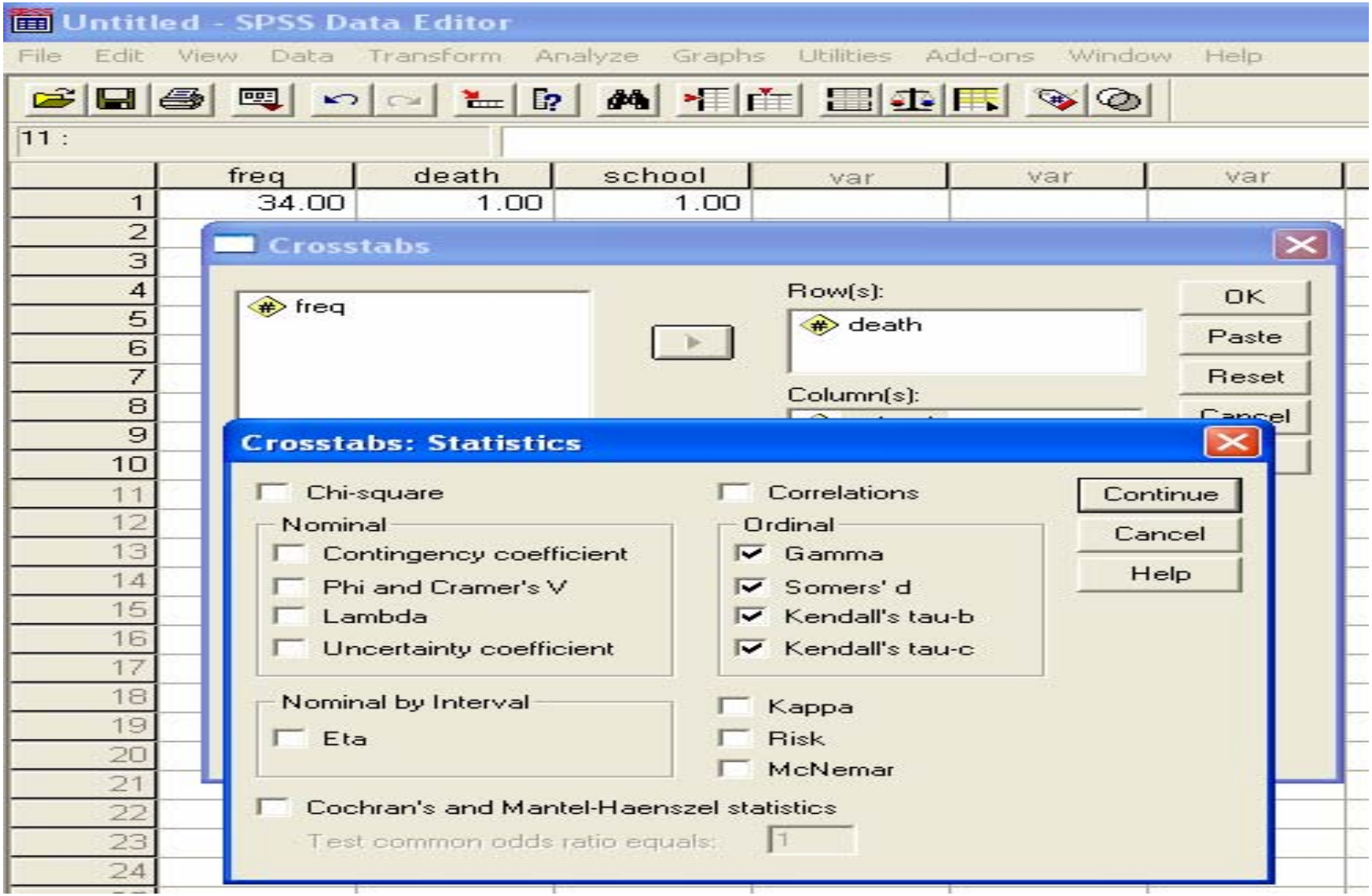
	freq	death	school	var
1	34.00	1.00	1.00	
2	114.00	1.00	2.00	
3	17.00	1.00	3.00	
4	26.00	1.00	4.00	
5	10.00	1.00	5.00	
6	9.00	2.00	1.00	
7	40.00	2.00	2.00	
8	3.00	2.00	3.00	
9	9.00	2.00	4.00	
10	12.00	2.00	5.00	
11				

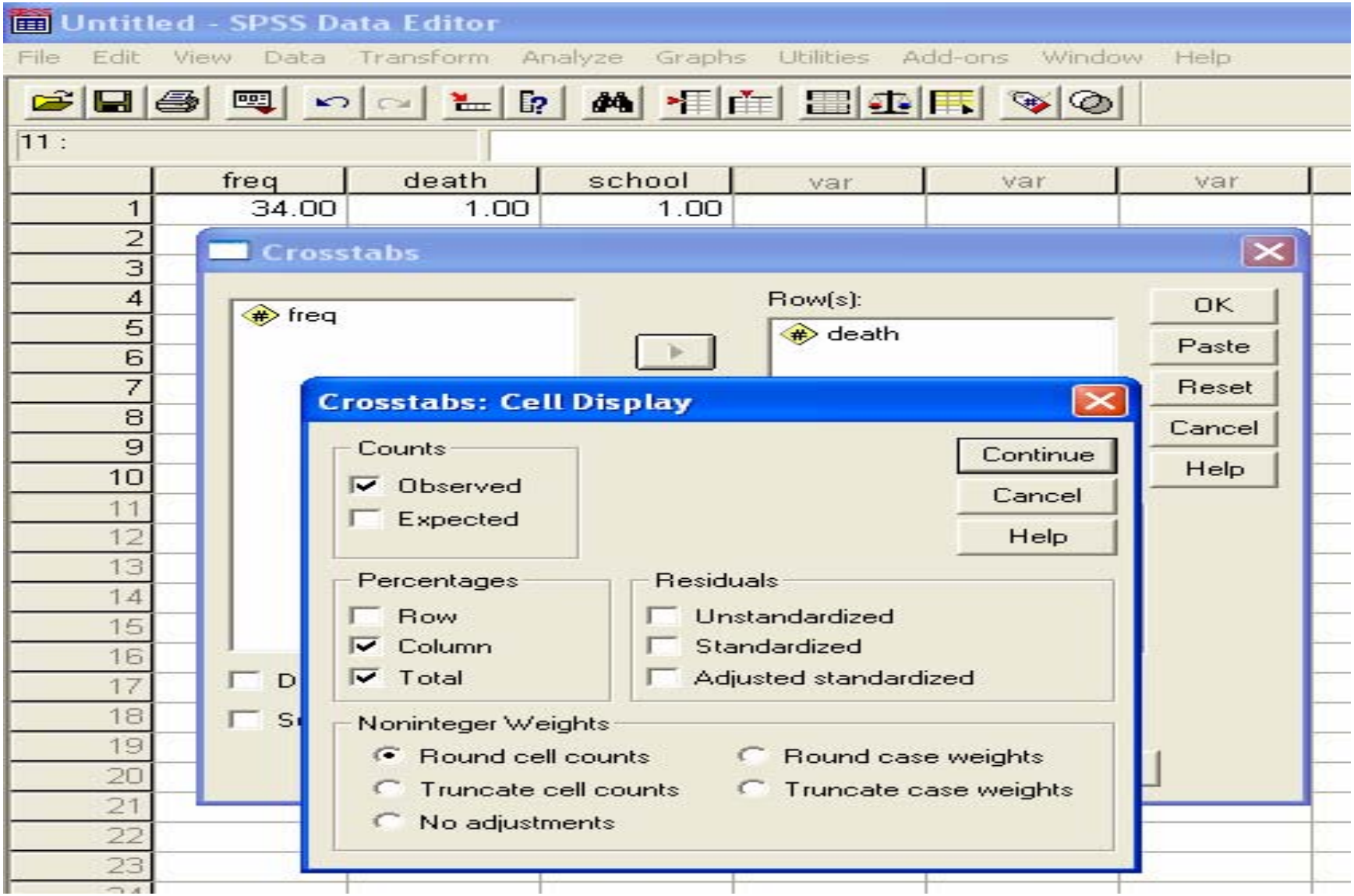
Weight Cases



Analyze>Descriptive Statistics> Crosstabs




Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
11 :
freq death school var var var
1 34.00 1.00 1.00
Crosstabs
freq
Row(s):
death
Column(s):
OK
Paste
Reset
Cancel
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Continue
Cancel
Help
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
11 :
freq death school var var var
1 34.00 1.00 1.00
Crosstabs
freq
Row(s):
death
OK
Paste
Reset
Cancel
Help
Crosstabs: Cell Display
Counts
Observed
Expected
Continue
Cancel
Help
Percentages
Row
Column
Total
Residuals
Unstandardized
Standardized
Adjusted standardized
Noninteger Weights
Round cell counts
Truncate cell counts
No adjustments
Round case weights
Truncate case weights

If your variables are "Full Rank Ordered"...

- Rank all of cases from highest to lowest.
- Spearman's Rho. (from -1 to +1)
- This is not based on PRE principle: Unlike the first three, for example, even if Rho is 0.4, you cannot say that you are able to reduce the error by 40% with the help of independent variable when predicting dependent variable.

Strength and pattern

- If rho is +1, this means that high ranks on X are perfectly matched with high ranks on Y.
- If rho is -1, this means that high ranks on X are perfectly matched with low ranks on Y. (Inverse match)
- If rho is 0… there is no consistent agreement between cases in terms of their ranks.

Spearman rank-order correlation

- The equation: $r_s = 1 - \frac{6\sum D^2}{N(N^2-1)}$
- D is the difference between corresponding ranks on X and Y; N is the total number of cases.
- H0: The correlation (or association) you observed is just because of sampling errors.
- H1: The correlation is not that small. Therefore, it is more likely that two variables are associated significantly with each other in the population.

Example.

X	Y
38	19
35	10
40	24
30	17
33	13

Rank on X	Rank on Y	D	D^2
4	4	0	0
3	1	2	4
5	5	0	0
1	3	-2	4
2	2	0	0

Calculation and Interpretation

- What is N? 5 (not 10!) $r_s = 1 - \frac{6\sum D^2}{N(N^2-1)}$
- Rho=1-[6*(0+4+0+4+0)]/[5*(25-1)]=0.6
- What does this imply?
- Again, three points: 1) statistical significance, 2) the strength of association, 3) direction of association.
- 1) You don't know yet. 2) Relatively strong. 3) Positive (The rank of X goes with the rank of Y in the same direction)

APPENDIX 4: SPEARMAN'S RHO TEST

	Level of significance for two-tailed test			
	0.10	0.05	0.02	0.01
	Level of significance for one-tailed test			
	0.05	0.025	0.01	0.005
$N = 4$	1.000			
5	0.900	1.000	1.000	
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.671	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.566	0.675
15	0.443	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.485	0.566	0.615
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.520	0.570

How to know statistical significance?

- There is no table in your textbook.
- Rho is 0.6 and N is 5. If your confidence level is 95% and you are interested in two-tailed test, then CV is 1.00.
- Your test statistic falls outside the rejection region.
- In conclusion, the positive association we have got is not that large to be considered reliable.
- Caution: All of your test statistic except 1 will not help to reject the null. In this way, if your sample size is so small, the hypothesis test becomes very strict.

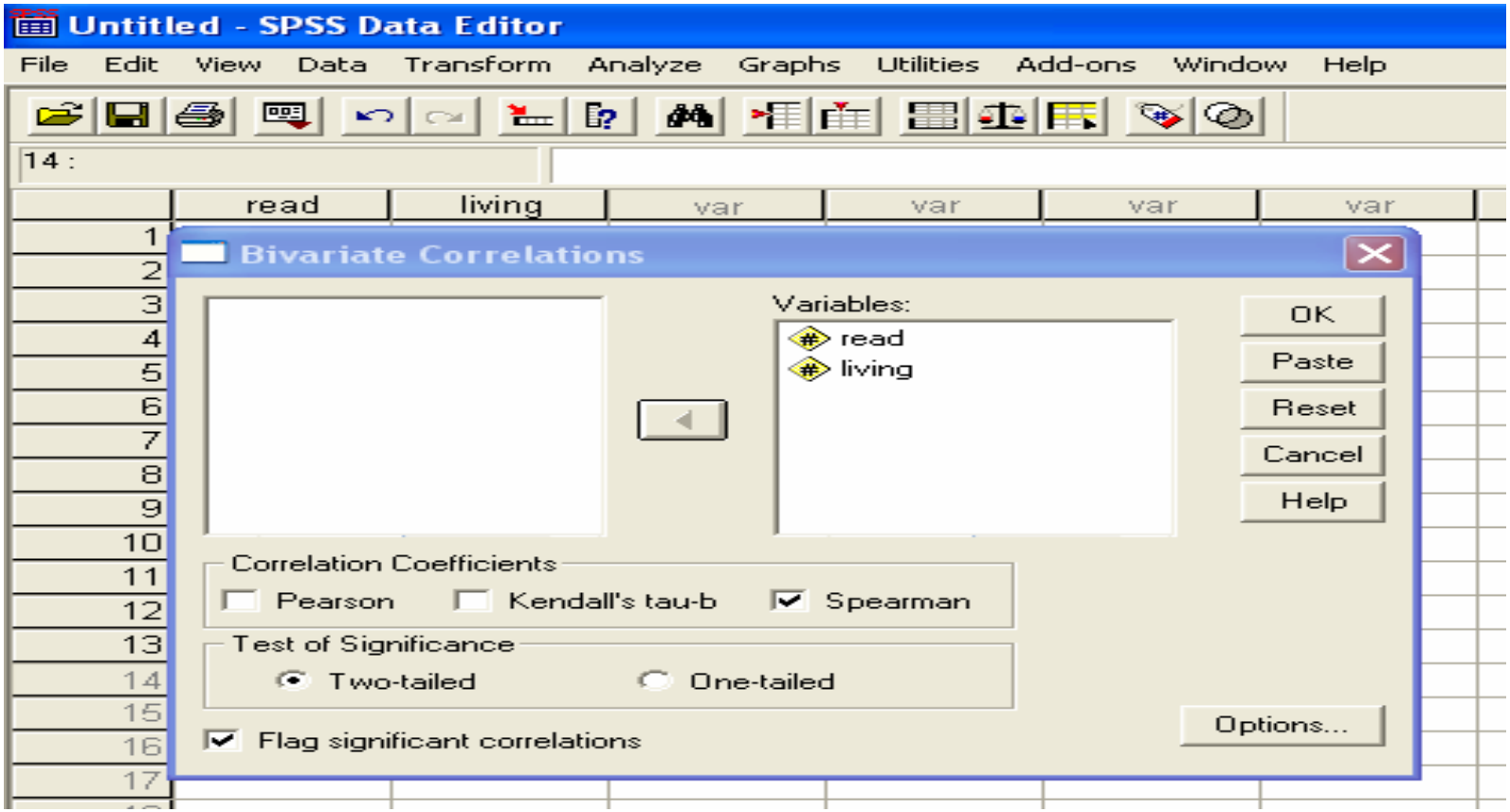
How to use SPSS
(Input ranks instead of scores!)

	COUNTRY	People who read (%)	People living in cities (%)
1	Afghanistan	29	18
2	Argentina	95	86
3	Armenia	98	68
4	Australia	100	85
5	Austria	99	58
6	Azerbaijan	98	54
7	Bahrain	77	83
8	Bangladesh	35	16
9	Barbados	99	45
10	Belarus	99	65
11	Belgium	99	96
12	Bolivia	78	51
13	Bosnia	86	36
14	Botswana	72	25
15	Brazil	81	75
16	Bulgaria	93	68
17	Burkina Faso	18	15
18	Burundi	50	5
19	Cambodia	35	12
20	Cameroon	54	40
Total N		20	20	20

a. Calculate rho.

b. Briefly interpret rho to inform us about the strength and form of the association between fertility and infant mortality.

- Analyze>Correlate>Bivariate



How to calculate pairs by hand?

This table is the same as your homework in terms of the size (2 by 5) Just do the same way.

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

Look for Nc! (Concordant pairs)

	freshman	sophomore	junior	senior	graduates
Like (+)	30	30	20	50	40
Dislike (-)	60	70	60	80	120

(-) (+)

- Caution: Education level increases to the right side, from lower rank to higher rank.
- Favor decreases to the down, from higher rank (like) to lower rank (dislike)!

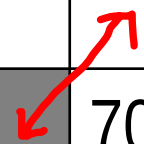
Look for Nc! (Continued)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- What is concordant pairs? The same order on both variables.
- Let's start with the grey cell (60).

Look for Nc! (Continued)

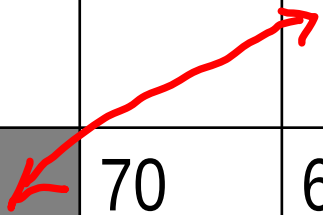
	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- Focusing on the red cell, this and the starting cell are concordant since X increases from freshman to sophomore, and Y also from dislike to like.
- Therefore, the number of pairs is 60X30.

Look for Nc! (Continued)

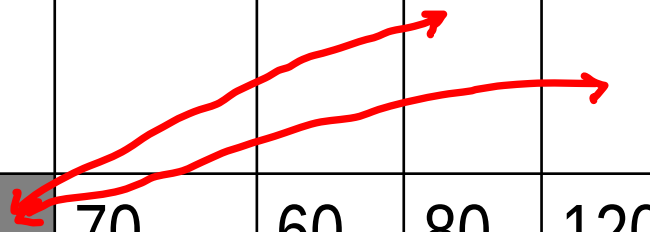
	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- Focusing on another red cell, this and the starting cell are again concordant since X increases from freshman to junior, and Y also from dislike to like.
- In this case, the number of pairs is 60X20.

Look for Nc! (Continued)

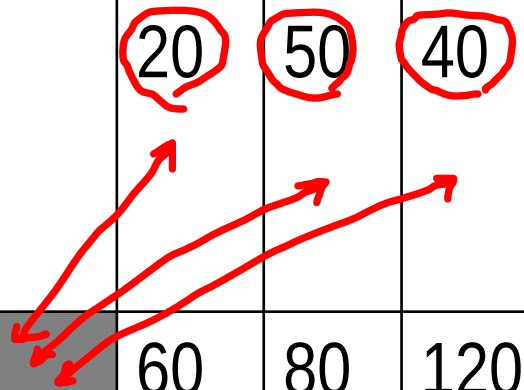
	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- In the same way, we can know another two cells and the starting cell are concordant since X increases from freshman to senior (graduates), and Y also from dislike to like.
- In this case, the number of pairs is 60X50 and 60X40.

Look for Nc! (Continued)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- What is the next victim? Let's start with the next cell (70).
- In the same way above, you can find three cells (red) that are concordant.
- Therefore, the number of pairs is 70X20, 70X50, 70X40.
- What else? No.

Look for Nc! (Continued)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- Let's move on to another next cell (60).
- In the same way above, you can find two cells (red) that are concordant in this time.
- The number of pairs is 60X50, 60X40.
- What else? No.

Look for Nc! (Continued)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- Let's move on to the last cell (80).
- In this time, there is only one cell (red) that is concordant.
- The number of pairs is 80X40.
- In this way, you finished the number of Nc. To get the total number, just add up all of numbers so far.

Look for Nd! (Discordant)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- What is discordant pairs? The opposite order on both variables.
- Let's start with the grey cell (30).

Look for Nd! (Discordant)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- Looking at the red cell, X increases from freshman to sophomore, but Y decreases from like to dislike. This pair is none other than discordant!
- The number of pairs is 30X70.

Look for Nd! (Discordant)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- Now, you can find out there are three more cells that are discordant. (red)
- The number of pairs is 30X60, 30X80, 30X120.

Look for Nd! (Discordant)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- Let's move on to the next cell. For the new grey cell, you can realize there are three cells that are discordant. (red)
- The number of pairs is 30X60, 30X80, 30X120.

Look for Nd! (Discordant)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- For the new grey cell, in the same way above, the number of pairs is 20X80, 20X120.

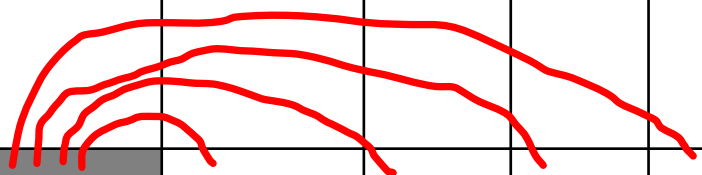
Look for Nd! (Discordant)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- For the new grey cell (50), in the same way above, the number of pairs is 50X120.
- This is the last cell.
- To get the total number of discordant cells, just sum up.

What else remains?

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- Let's think about any other combinations we did not deal with so far.
- For the grey cell, we did not consider the following pairs (red) where X increases but Y remains as the same.

Ty (pairs tied on Y)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- These pairs are called “Pairs tied on Y”.
- The number of these pairs is 60X70, 60X60, 60X80, 60X120.

What else remains? (Continued)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- What else? There must be another relationship we did not care about so far.
- What about the relationship between the grey cell and the red? Y increases from dislike to like, but X is the same.

Tx (Pairs tied on X)

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- These are called pairs tied on X.
- The number of Tx is 60X30.

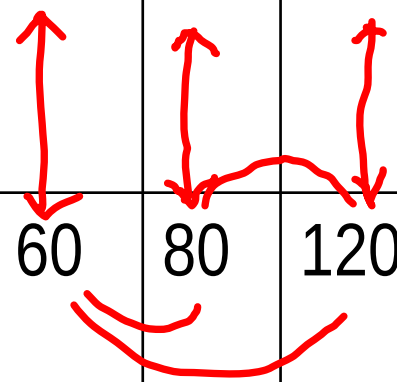
In the same way, look for Tx and Ty

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120

- For 70 in the second row, there are three pairs tied on Y. The number of pairs is 70X60, 70X80, 70X120.
- What is the number of pairs tied on X? 70X30.

Continued...

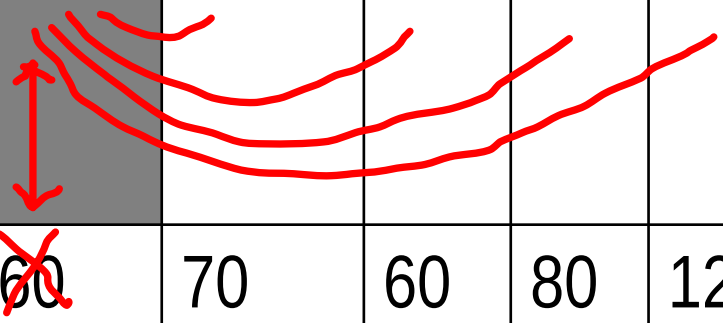
	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- Just do the same with the cell (60), the cell (80).
- How about the last cell (120)? Because we have moved toward the right, there is no pair tied on Y. However, there is one pair tied on X. So, (40X120).

Continued...

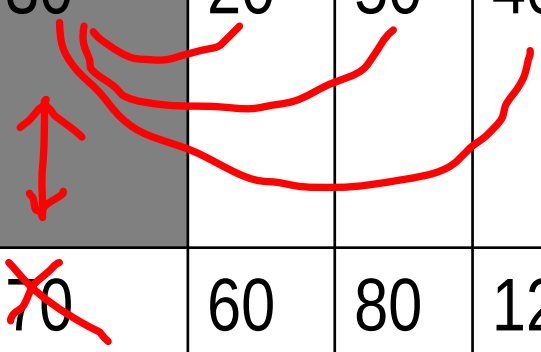
	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	60	80	120



- Wow! Sooo boring. The next victim is the grey cell (30) again.
- However, you counted the relationship between this cell and the red just before.
- Therefore, just focus on four pairs tied on Y here.

Continued...

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	~~70~~	60	80	120



- In the same way, calculate only the number of three pairs tied on Y again.
- Why? You already counted the relationship between this cell and the red before.

Continued...

	freshman	sophomore	junior	senior	graduates
Like	30	30	20	50	40
Dislike	60	70	~~60~~	~~80~~	~~120~~

- Do the same with the cell (20), the cell (50).
- How about the last cell (40)?
- Since we have moved right, there is no pair tied on Y.
- You don't need to count the number of pair tied on X because you already counted that.