Crosstabs

Notes

Output Created		02-SEP-2019 16:42:30
Comments		
Input	Data	C:\Users\Bill\Desktop\current work\aaa Backup\Teaching\CLASSES\Psy 465-Adv Stats\SPSS-SAS-R\CHI-SQ_Fisherexact20.SAV
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	20
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		crosstabs / tables = gender by atthappy / format=avalue tables / statistics = all / cells = count expected sresid / count round cell.
Resources	Processor Time	00:00:00.08
	Elapsed Time	00:00:00.14
	Dimensions Requested	2
	Cells Available	349496

gender * atthappy Crosstabulation

			atthappy		
			1.00 no	2.00 yes	Total
gender	1.00 male	Count	7	3	10
		Expected Count	5.5	4.5	10.0
		Standardized Residual	.6	-.7	
	2.00 female	Count	4	6	10
		Expected Count	5.5	4.5	10.0
		Standardized Residual	-.6	.7	
Total		Count	11	9	20
		Expected Count	11.0	9.0	20.0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.818[a]	1	.178		
Continuity Correction[b]	.808	1	.369		
Likelihood Ratio	1.848	1	.174		
Fisher's Exact Test				.370	.185
Linear-by-Linear Association	1.727	1	.189		
McNemar Test				1.000[c]	
N of Valid Cases	20				

a. 2 cells (50.0%) have expected count less than 5. The minimum expected count is 4.50.

b. Computed only for a 2x2 table

c. Binomial distribution used.

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Nominal by Nominal	Phi	.302			.178
	Cramer's V	.302			.178
	Contingency Coefficient	.289			.178
Ordinal by Ordinal	Kendall's tau-b	.302	.213	1.414	.157
	Kendall's tau-c	.300	.212	1.414	.157
	Gamma	.556	.327	1.414	.157
	Spearman Correlation	.302	.213	1.342	.196[c]
Interval by Interval	Pearson's R	.302	.213	1.342	.196[c]
Measure of Agreement	Kappa	.300	.212	1.348	.178
N of Valid Cases		20			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Risk Estimate

	Value	95% Confidence Interval Lower	Upper
Odds Ratio for gender (1.00 male / 2.00 female)	3.500	.549	22.304
For cohort atthappy = 1.00 no	1.750	.740	4.139
For cohort atthappy = 2.00 yes	.500	.171	1.463
N of Valid Cases	20		

Tests of Conditional Independence

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	1.818	1	.178
Mantel-Haenszel	.768	1	.381

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			3.500
ln(Estimate)			1.253
Standard Error of ln(Estimate)			.945
Asymptotic Significance (2-sided)			.185
Asymptotic 95% Confidence Interval	Common Odds Ratio	Lower Bound	.549
		Upper Bound	22.304
	ln(Common Odds Ratio)	Lower Bound	-.599
		Upper Bound	3.105

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

```
DATASET ACTIVATE DataSet1.
weight by wt.
list / var all.
```

List

STATUS	HT	WT
1.00	1.00	31.00
2.00	1.00	12.00
1.00	2.00	20.00
2.00	2.00	32.00

Number of cases read: 4 Number of cases listed: 4

Crosstabs

Notes

Output Created		02-SEP-2019 16:43:58
Comments		
Input	Data	C:\Users\Bill\Desktop\Downloads\2011_andBACK_Downloads\chi-square-examp-465.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	WT
	Split File	<none>
	N of Rows in Working Data File	4
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		CROSSTABS /tables=status by ht /format=avalue tables /statistics=chisq corr risk cmh (1) /cells=count /count round cell.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.25
	Dimensions Requested	2
	Cells Available	349496

2x2

OR = 4.13

r = .34

STATUS * HT Crosstabulation

Count

		HT		
		1.00	2.00	Total
STATUS	1.00	31	20	51
	2.00	12	32	44
Total		43	52	95

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10.707[a]	1	.001		
Continuity Correction[b]	9.397	1	.002		
Likelihood Ratio	10.971	1	.001		
Fisher's Exact Test				.002	.001
Linear-by-Linear Association	10.594	1	.001		
N of Valid Cases	95				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 19.92.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Interval by Interval	Pearson's R	.336	.096	3.437	.001[c]
Ordinal by Ordinal	Spearman Correlation	.336	.096	3.437	.001[c]
N of Valid Cases		95			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

Risk Estimate

	Value	95% Confidence Interval Lower	Upper
Odds Ratio for STATUS (1.00 / 2.00)	4.133	1.732	9.862
For cohort HT = 1.00	2.229	1.311	3.788
For cohort HT = 2.00	.539	.366	.794
N of Valid Cases	95		

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymptotic Significance (2-sided)
Breslow-Day	.000	0	.
Tarone's	.000	0	.

Tests of Conditional Independence

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	10.707	1	.001
Mantel-Haenszel	9.298	1	.002

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			4.133
ln(Estimate)			1.419
Standard Error of ln(Estimate)			.444
Asymptotic Significance (2-sided)			.001
Asymptotic 95% Confidence Interval	Common Odds Ratio	Lower Bound	1.732
		Upper Bound	9.862
	ln(Common Odds Ratio)	Lower Bound	.550
		Upper Bound	2.289

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

```
data one;
input STATUS HT WT;
datalines;
1 1 31
2 1 12
1 2 20
2 2 32
;
run;

proc freq data=one;
weight wt;
tables status*ht / expected chisq;
run;
```

The FREQ Procedure

Table of STATUS by HT

STATUS HT

Frequency Expected Percent Row Pct Col Pct	1	2	Total
1	31 23.084 32.63 60.78 72.09	20 27.916 21.05 39.22 38.46	51 53.68
2	12 19.916 12.63 27.27 27.91	32 24.084 33.68 72.73 61.54	44 46.32
Total	43 45.26	52 54.74	95 100.00

The FREQ Procedure

Statistics for Table of STATUS by HT

Statistic	DF	Value	Prob
Chi-Square	1	10.7069	0.0011
Likelihood Ratio Chi-Square	1	10.9705	0.0009
Continuity Adj. Chi-Square	1	9.3970	0.0022
Mantel-Haenszel Chi-Square	1	10.5942	0.0011
Phi Coefficient		0.3357	
Contingency Coefficient		0.3183	
Cramer's V		0.3357	

Fisher's Exact Test

Cell (1,1) Frequency (F)	31
Left-sided Pr <= F	0.9998
Right-sided Pr >= F	9.850E-04
Table Probability (P)	7.714E-04
Two-sided Pr <= P	0.0018

Sample Size = 95

465 class example - chi-square 2 x 2

```
> dataone <- matrix(c(31, 12, 20, 32), nr=2)
> dataone
     [,1] [,2]
[1,]   31   20
[2,]   12   32
> chisq.test(dataone)

        Pearson's Chi-squared test with Yates' continuity correction

data:  dataone
X-squared = 9.397, df = 1, p-value = 0.002173
```

```
> dataone <- matrix(c(31, 12, 20, 32), nr=2)
> dataone
     [,1] [,2]
[1,]   31   20
[2,]   12   32
>  chisq.test(dataone,y=NULL, correct=FALSE)

        Pearson's Chi-squared test

data:  dataone
X-squared = 10.7069, df = 1, p-value = 0.001067
```

```
CROSSTABS
  /TABLES=status  BY ht
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ CORR CMH(1)
  /CELLS= COUNT EXPECTED .
```

w/ O.R.

15

Crosstabs

Notes

Output Created		27-JAN-2003 15:35:59
Comments		
Input	Data	C:\My Documents\stats\chi-square-example.sav
	Filter	<none>
	Weight	WT wt
	Split File	<none>
	N of Rows in Working Data File	4
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		CROSSTABS /TABLES=status BY ht /FORMAT= AVALUE TABLES /STATISTIC=CHISQ CORR CMH(1) /CELLS= COUNT EXPECTED .
Resources	Elapsed Time	
Resources	Dimensions Requested	2
	Cells Available	116508
	Elapsed Time	0:00:00.16

O.R.

A	B
C	D

STATUS follower-und * HT ht Crosstabulation

			HT ht		
			1.00 short	2.00 tall	Total
STATUS follower-und	1.00 f or u	Count	31	20	51
		Expected Count	23.1	27.9	51.0
	2.00 leader	Count	12	32	44
		Expected Count	19.9	24.1	44.0
Total		Count	43	52	95
		Expected Count	43.0	52.0	95.0

A D / B C

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10.707[b]	1	.001		
Continuity Correction[a]	9.397	1	.002		
Likelihood Ratio	10.971	1	.001		
Fisher's Exact Test				.002	.001
Linear-by-Linear Association	10.594	1	.001		
N of Valid Cases	95				

a Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 19.92.

31	20
12	32

$$\frac{(31)(32)}{(20)(12)} = 4.13$$

Odds ratio...if you apply that nice little formula i gave in class to the numbers in the Howell text, you will get an odds ratio of 0.54. But Howell reports a value of 1.83. Yipes! Is there an error? Nope.

If you switch Howell's heart attack columns, you will get 1.83. Recall in class that i mentioned sometimes having to move columns or rows around to get an odds ratio that one can easily interpret. Well, this is a grand example. The value of 0.54 is actually fine and correct...but its more interpretable if you switch the columns. You can actually convert the 0.54 to 1.83 by dividing by 1 (1/.54)...well, you get something close to 1.83 (i rounded to get .54, so the division does not exactly yield 1.83)

-square and Fisher's exact test -- creating a raw matrix datafile...

```
> dataone <- matrix(c(7, 4, 3, 6), nr=2)

> dataone
     [,1] [,2]
[1,] 7    3
[2,] 4    6

> chisq.test(dataone)

     Pearson's Chi-squared test with Yates' continuity correction

data:  dataone
X-squared = 0.8081, df = 1, p-value = 0.3687

Warning message:
In chisq.test(dataone) : Chi-squared approximation may be incorrect

> fisher.test(dataone)

     Fisher's Exact Test for Count Data

data:  dataone
p-value = 0.3698
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
  0.4057961 33.2738448
     sample estimates:
     odds ratio
       3.274233
```

not Zero cell !!

		var1			expected	
		1	2			
	1	**20**	**0**	20	**10**	**10**
var2	2	**40**	**60**	100	**50**	**50**
		60	60	120		

Table 5: Chi-square critical value table (partial – from Appendix C)

	Level of Significance		
Df	0.05	0.01	0.001
1	3.84	6.63	10.83
2	5.99	9.21	13.82
3	7.81	11.34	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.51
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.12

partial Table

Chi-Square Example Write-up: Gender and reading:

A study was conducted to examine the association between gender and reading – do males and females differ in their reading in public places? A total of 40 individuals were observed in the student center. There were 15 females and 25 males seated alone. Of the 15 females, 10 were reading (67%). Only 5 of the 25 males were reading (20%). A chi-square test of independence revealed an association beyond the .05 significance level, $\chi^2(2, N = 40) = 8.82, p < .05$. The data indicate that females are more likely to be reading than are males.

Chi-square Assumptions

For inferential tests, assumptions are made about the underlying distributions that the data form. Tests such as chi-square are carried out on data assuming that certain data properties are in place. First, we need to insure that our *data are counts or frequency data*, not interval or ratio-level scores. If you have such score-level data (sometimes interval and ratio data are called "score-level" data), consider other tests covered in this textbook such as correlation, or a t-test

Correlation Example (Marelich, Psy 465)

Variables:

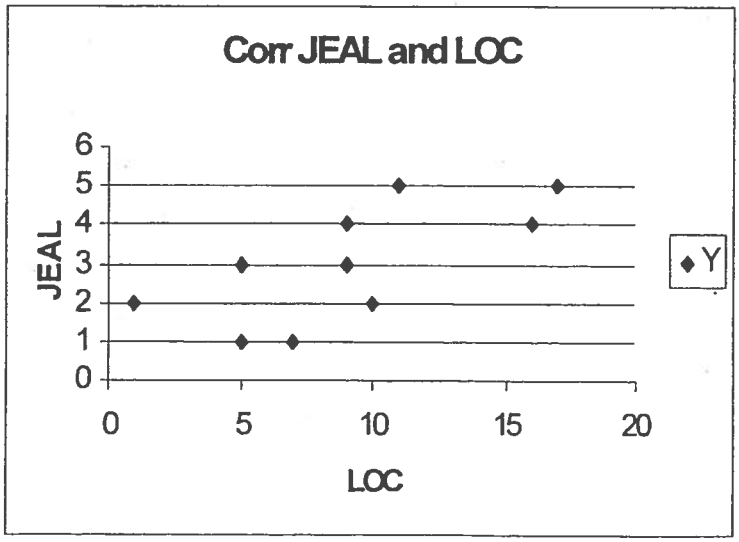
Locus of Control (LOC): higher levels = external control
Jealousy (JEAL): higher levels = greater jealousy

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Sub	LOC X	JEAL Y	X^2	Y^2	XY
1	9	3	81	9	27
2	10	2	100	4	20
3	16	4	256	16	64
4	5	1	25	1	5
5	9	4	81	16	36
6	7	1	49	1	7
7	5	3	25	9	15
8	11	5	121	25	55
9	1	2	1	4	2
10	17	5	289	25	85
Totals	**90**	**30**	**1028**	**110**	**316**

$$r = \frac{(10)(316) - (90)(30)}{\sqrt{[(10)(1028) - (90)^2)][(10)(110) - (30)^2]}}$$

r = .70 [the more externally controlled, the greater jealousy reported



```
EXAMINE VARIABLES=LOC Jealousy
  /PLOT BOXPLOT STEMLEAF
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Explore

Notes

Output Created		2009-07-17T13:27:40.712
Comments		
Input	Active Dataset	$DataSet
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	10
Missing Value Handling	Definition of Missing	User-defined missing values for dependent variables are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any dependent variable or factor used.
Syntax		EXAMINE VARIABLES=LOC Jealousy /PLOT BOXPLOT STEMLEAF /COMPARE GROUP /STATISTICS DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL.
ources	Processor Time	0:00:00.795
	Elapsed Time	0:00:00.889

[$DataSet]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
LOC	10	100.0%	0	.0%	10	100.0%
Jealousy	10	100.0%	0	.0%	10	100.0%

Descriptives

			Statistic	Std. Error
LOC	Mean		9.0000	1.55635
	95% Confidence Interval for Mean	Lower Bound	5.4793	
		Upper Bound	12.5207	
	5% Trimmed Mean		9.0000	
	Median		9.0000	
	Variance		24.222	
	Std. Deviation		4.92161	
	Minimum		1.00	
	Maximum		17.00	
	Range		16.00	
	Interquartile Range		7.25	
	Skewness		.252	.687
	Kurtosis		-.196	1.334
Jealousy	Mean		3.0000	.47140
	95% Confidence Interval for Mean	Lower Bound	1.9336	
		Upper Bound	4.0664	
	5% Trimmed Mean		3.0000	
	Median		3.0000	
	Variance		2.222	
	Std. Deviation		1.49071	

Descriptives

		Statistic	Std. Error
Jealousy	Minimum	1.00	
	Maximum	5.00	
	Range	4.00	
	Interquartile Range	2.50	
	Skewness	.000	.687
	Kurtosis	-1.334	1.334

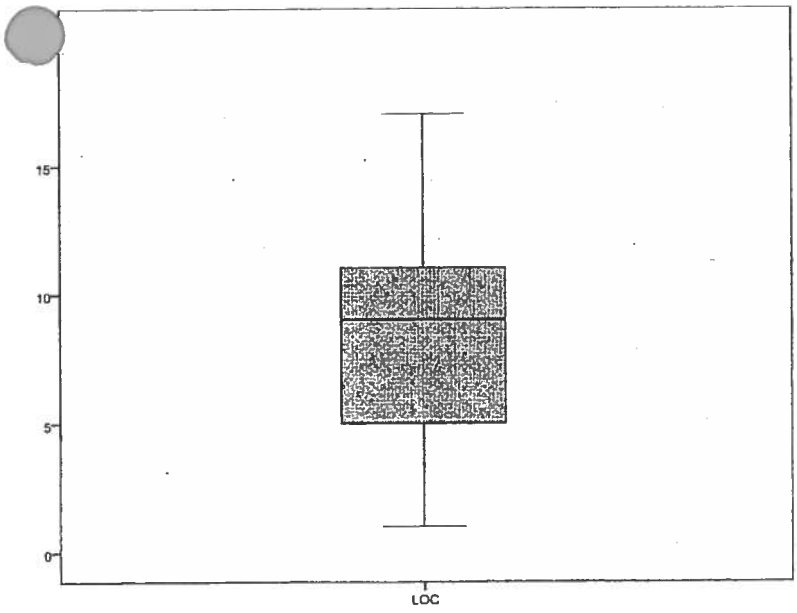
LOC

LOC Stem-and-Leaf Plot

```
 Frequency    Stem &  Leaf

     1.00      0 .  1
     5.00      0 .  55799
     2.00      1 .  01
     2.00      1 .  67

 Stem width:     10.00
 Each leaf:       1 case(s)
```



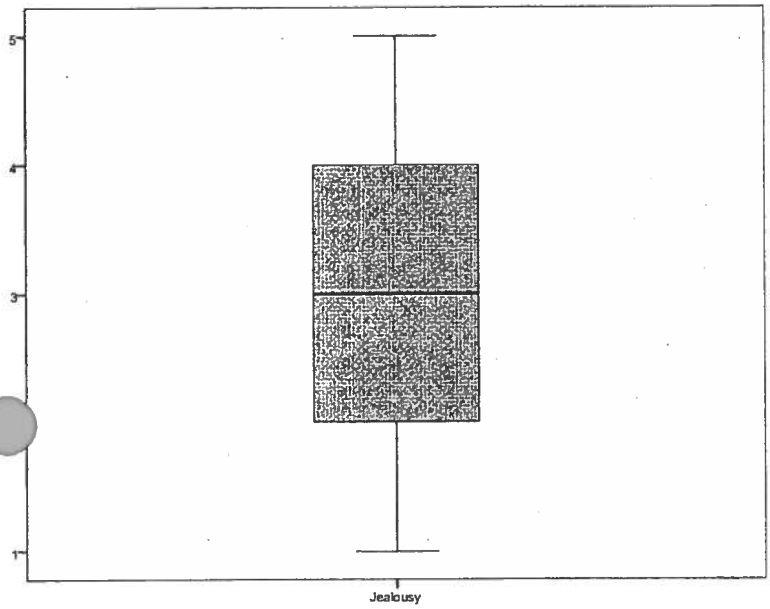
Jealousy

Jealousy Stem-and-Leaf Plot

```
 Frequency    Stem &  Leaf
```

```
    2.00        1 .  00
    2.00        2 .  00
    2.00        3 .  00
    2.00        4 .  00
    2.00        5 .  00

Stem width:     1.00
Each leaf:      1 case(s)
```




```
CORRELATIONS
  /VARIABLES=LOC Jealousy
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

Notes

Output Created		2009-07-17T13:28:45.857
Comments		
Input	Active Dataset	$DataSet
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	10
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.

Notes

Syntax		CORRELATIONS /VARIABLES=LOC Jealousy /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE.
Resources	Processor Time	0:00:00.093
	Elapsed Time	0:00:00.046

[$DataSet]

Correlations

		LOC	Jealousy
LOC	Pearson Correlation	1.000	.697*
	Sig. (2-tailed)		.025
	N	10.000	10
Jealousy	Pearson Correlation	.697*	1.000
	Sig. (2-tailed)	.025	
	N	10	10.000

*. Correlation is significant at the 0.05 level (2-tailed).

Table R – significance values for r (partial table from Appendix R)

df = *n* -2			
Read - /+		Alpha	
Df	0.100	0.050	0.010
1	0.988	0.997	1.000
2	0.900	0.950	0.990
3	0.805	0.878	0.959
4	0.729	0.811	0.917
5	0.669	0.754	0.874
6	0.622	0.707	0.834
7	0.582	0.666	0.798
8	0.549	0.632	0.765
9	0.521	0.602	0.735
10	0.497	0.576	0.708
11	0.476	0.553	0.684
12	0.458	0.532	0.661
13	0.441	0.514	0.641
14	0.426	0.497	0.623
15	0.412	0.482	0.606
16	0.400	0.468	0.590
17	0.389	0.456	0.575
18	0.378	0.444	0.561
19	0.369	0.433	0.549
20	0.360	0.432	0.537

A second way to proceed is to analyze the data using a computer program such as SPSS or SAS. These programs calculate the correlation value that is derived from the obtained data and provide an exact probability. With our data, the computer program tells us that $p(-0.55|H_o) =$ .012. This tells us that the probability of getting our obtained correlation if the null hypothesis is true is exactly .012. Since our typical probability cutoff is .05, we can conclude as we did with the hand calculations that it is very unlikely that we could obtain our findings if the null hypothesis is true. This is why we now reject the null hypothesis – it seems like a much better decision to conclude that the alternative hypothesis is true. Therefore, our conclusion is that there is as association between Trust and Jealousy, with greater Trust leading to lower levels of Jealousy.

#2C

```
data three;
input LOC JEAL;
datalines;
    9.00        3.00
   10.00        2.00
   16.00        4.00
    5.00        1.00
    9.00        4.00
    7.00        1.00
    5.00        3.00
   11.00        5.00
    1.00        2.00
   17.00        5.00
;
run;

proc corr data=three;
var LOC JEAL;
run;
```

The SAS System 12:56 Saturday, January 5, 2008

The CORR Procedure

2 Variables: LOC JEAL

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
LOC	10	9.00000	4.92161	90.00000	1.00000	17.00000
JEAL	10	3.00000	1.49071	30.00000	1.00000	5.00000

Pearson Correlation Coefficients, N = 10
Prob > |r| under H0: Rho=0

	LOC	JEAL
LOC	1.00000	0.69665 0.0252
JEAL	0.69665 0.0252	1.00000

```
> library(foreign)

> corr10.spss<- read.spss("c:/data/corr_465_N10.sav", use.value.labels=TRUE)

Warning message:
In read.spss("c:/data/corr_465_N10.sav", use.value.labels = TRUE) :
  c:/data/corr_465_N10.sav: Unrecognized record type 7, subtype 18
encountered in system file

> head(corr10.spss)
$LOC
 [1]  9 10 16  5  9  7  5 11  1 17

$Jeal
 [1] 3 2 4 1 4 1 3 5 2 5

> attach(corr10.spss)
The following object(s) are masked from 'corr10.spss (position 3)':

    Jeal, LOC

> cor.test (LOC, Jeal)

        Pearson's product-moment correlation

data:  LOC and Jeal
t = 2.7466, df = 8, p-value = 0.02519
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1193927 0.9219028
sample estimates:
      cor
0.6966499
```