SPSS Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery
By Ronald N. Forthofer, Eun Sul Lee, Mike Hernandez
Chapter 7: Interval Estimation

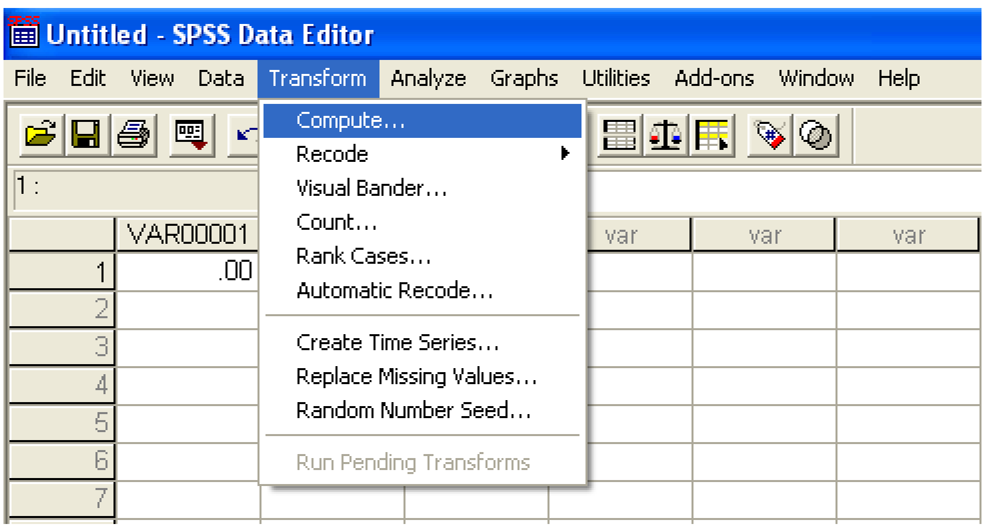
Chapter 7 Formulas

	Distribution	Confidence Interval
Upper and lower limits for a (1-α)*100% confidence interval for the population **mean** when the **variance is unknown**.	t-distribution with n-1 degrees of freedom	$\bar{x} \pm t_{n-1,1-\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$
Upper and lower limits for a (1-α)*100% confidence interval for a **population proportion**.	Standard normal distribution	$\hat{p} \pm \left(z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} + \frac{1}{2n}\right)$
Upper and lower limits for a (1-α)*100% confidence interval for the difference between population **means** when the **variance is unknown but equal**.	t-distribution with n-2 degrees of freedom Note: $n = n_1 + n_2$	$(\bar{x}_1 - \bar{x}_2) \pm t_{n-2,1-\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ where $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$
Upper and lower limits for a (1-α)*100% confidence interval for the difference between population **means** when the **variance is unknown and unequal**.	t-distribution distribution with degrees of freedom, $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{df,1-\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

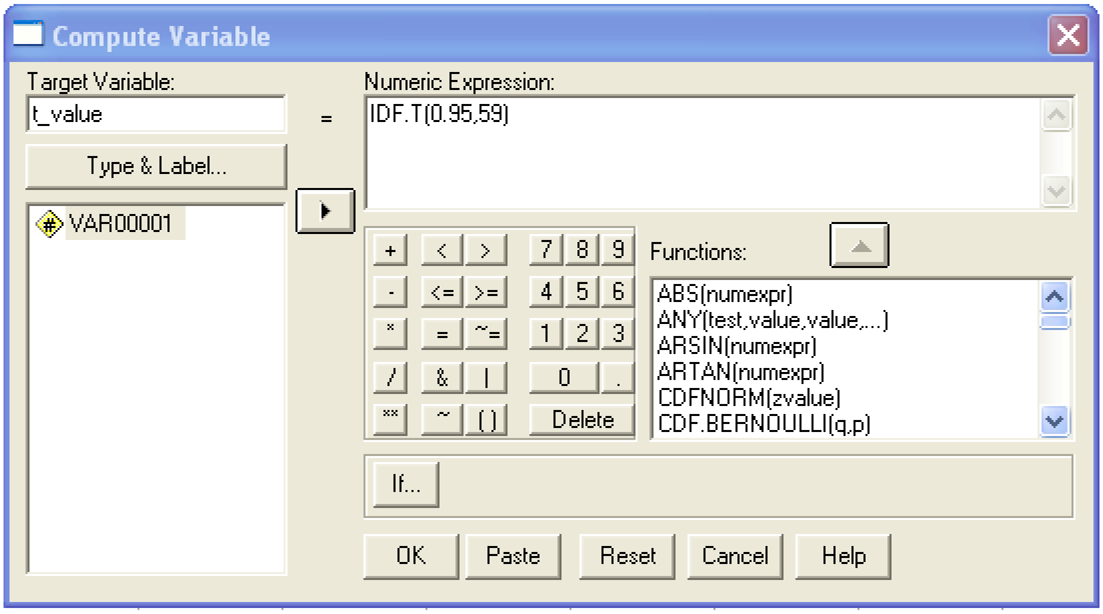
Upper and lower limits for a (1-α)*100% confidence interval for the **Pearson Correlation Coefficient**.	Requires Fisher's transformation	(See textbook Chapter 7 page 203)

Program Note 7.1 – Finding percentiles of a t-distribution

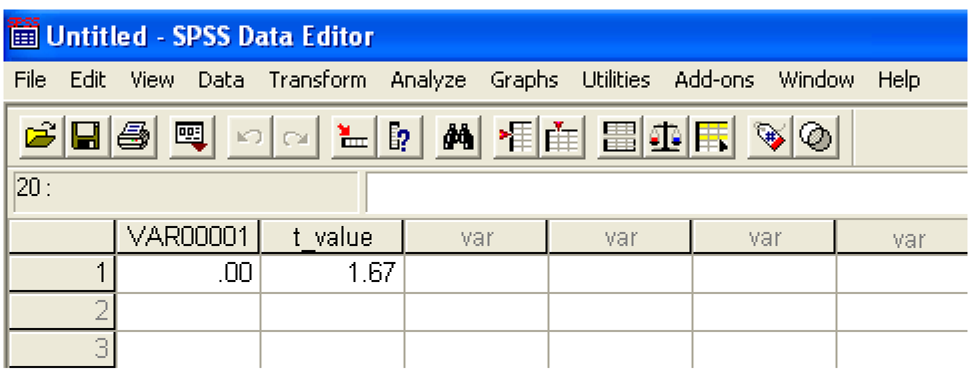
To find the value of the 95th percentile of for a t-distribution with 59 degrees of freedom, make the SPSS Data Editor active by typing a 0 in the cell corresponding to the first row - first column. Use the following SPSS procedure, **Transform**-> **Compute…** to obtain the **Compute Variable Window** shown below.



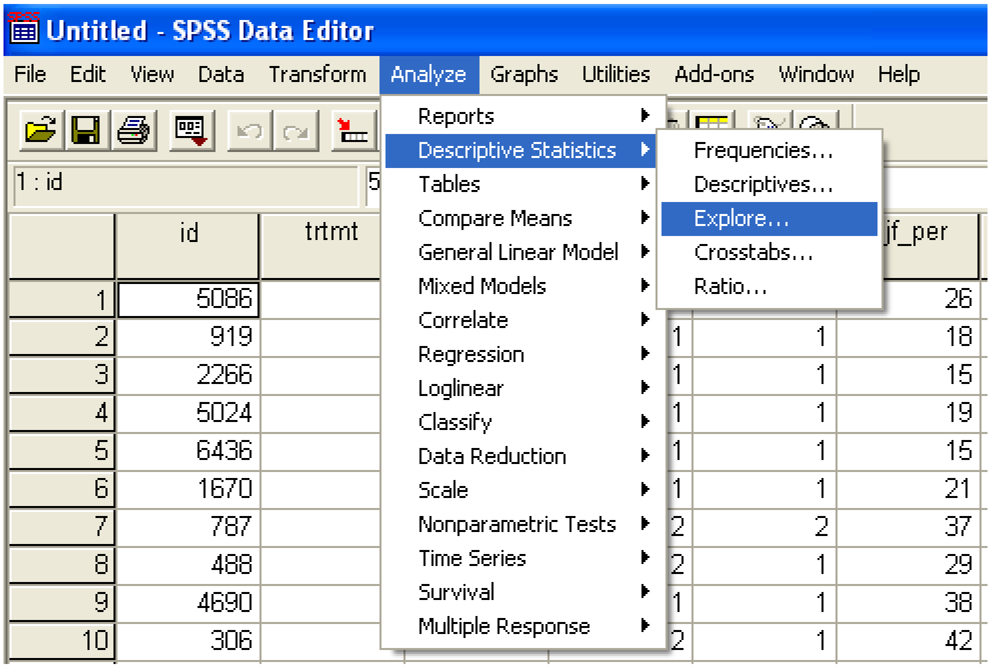
It doesn't matter what you call the **Target Variable:**, in this case we used t_value. Under functions, find **IDF.T(p,df)** which refers to the value from the t-distribution with degrees of freedom, **df**, for which the cumulative probability is **p**. Therefore, **IDF.T(0.95,59)** returns the value from the t-distribution with 59 degrees of freedom for which the cumulative probability is 0.95. Another way of stating this is to say that **IDF.T(0.95,59)** returns the 95th percentile from the t-distribution with 59 degrees of freedom.

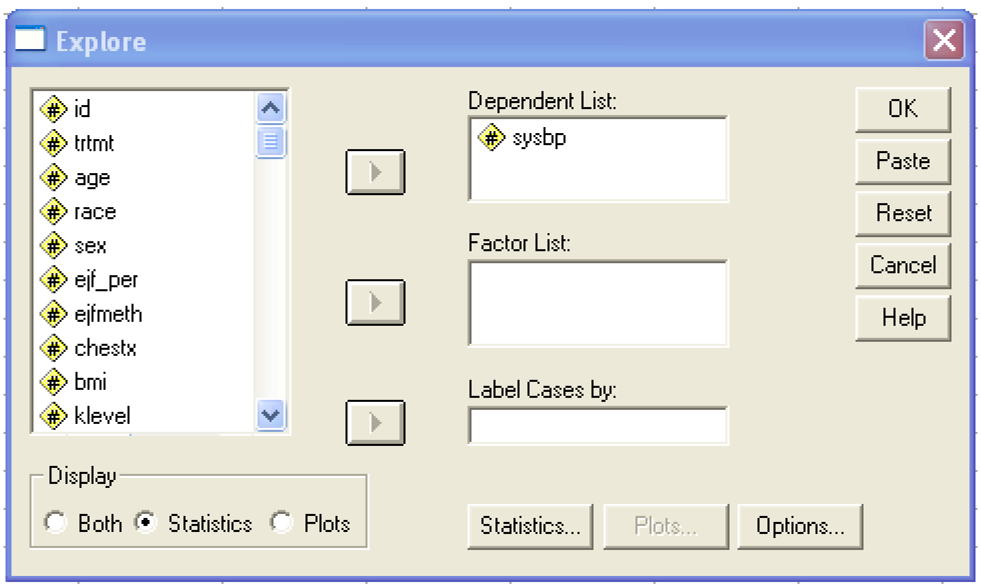


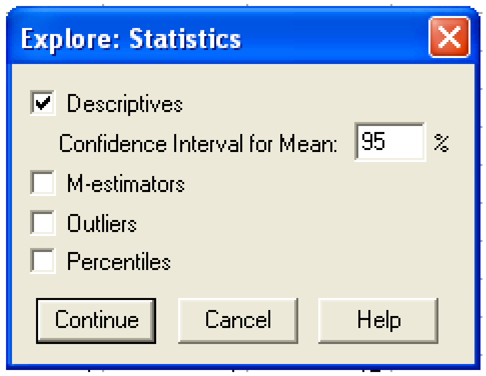
After clicking **OK**, you should find that the value is approximately 1.67 as shown below.



In **Example 7.1** on page 179, we calculate a 95% confidence interval for the mean systolic blood pressure using patient data from the **DIG200** dataset. However, for Example 7.1, we assume that the standard deviation for the population is 20mmHg. If the population variance is unknown, we can use the following SPSS procedure to calculate a 95% confidence interval using the t-distribution.


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : id
id
trtmt
jf_per
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
1 5086 26
2 919 1 1 18
3 2266 1 1 15
4 5024 1 1 19
5 6436 1 1 15
6 1670 1 1 21
7 787 2 2 37
8 488 2 1 29
9 4690 1 1 38
10 306 2 1 42


Explore
id
trtmt
age
race
sex
ejf_per
ejfmeth
chestx
bmi
klevel
Dependent List:
sysbp
Factor List:
Label Cases by:
OK
Paste
Reset
Cancel
Help
Display
Both
Statistics
Plots
Statistics...
Plots...
Options...



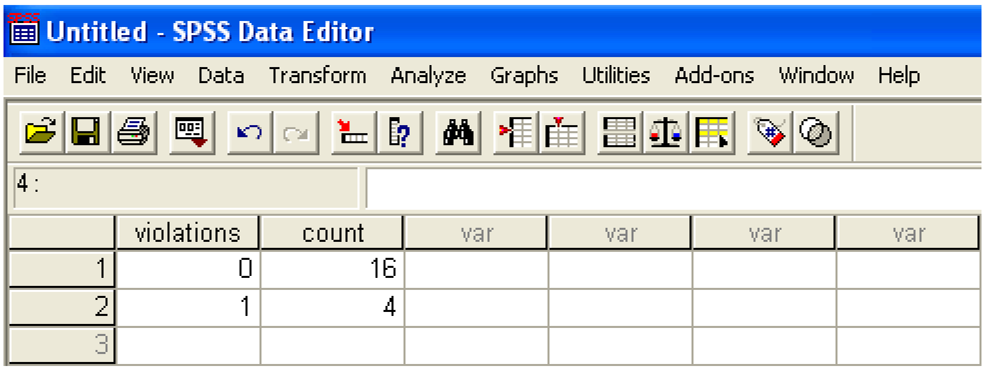
The SPSS output is shown below:

Descriptives

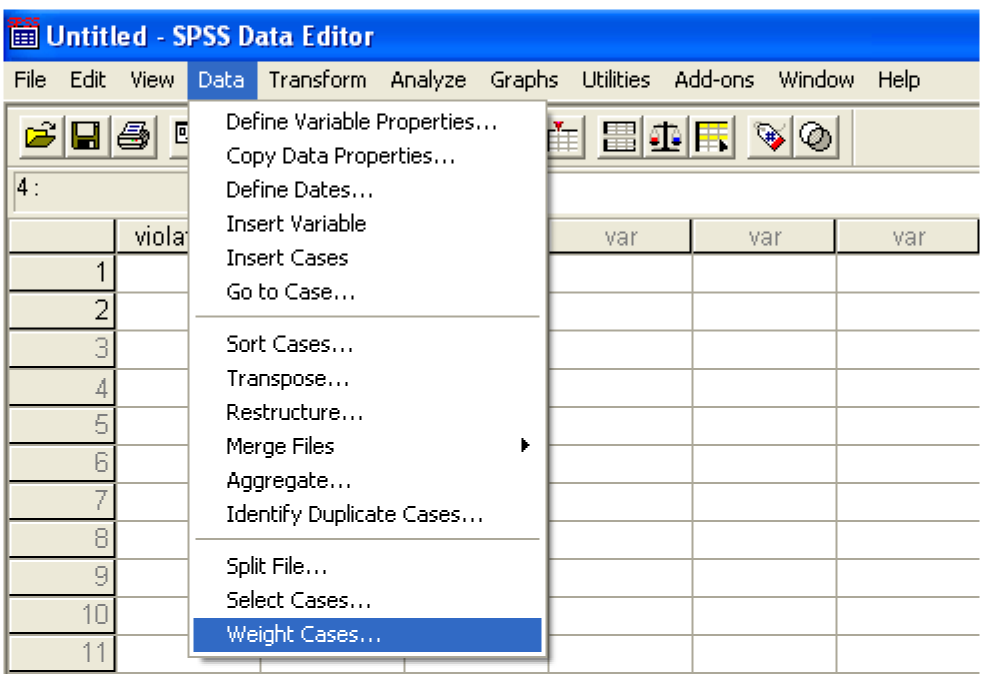
			Statistic	Std. Error
sysbp	Mean		125.82	1.289
	95% Confidence Interval for Mean	Lower Bound	123.28	
		Upper Bound	128.37	
	5% Trimmed Mean		125.48	
	Median		124.00	
	Variance		330.459	
	Std. Deviation		18.179	
	Minimum		85	
	Maximum		170	
	Range		85	
	Interquartile Range		26	
	Skewness		.209	.172
	Kurtosis		-.282	.343

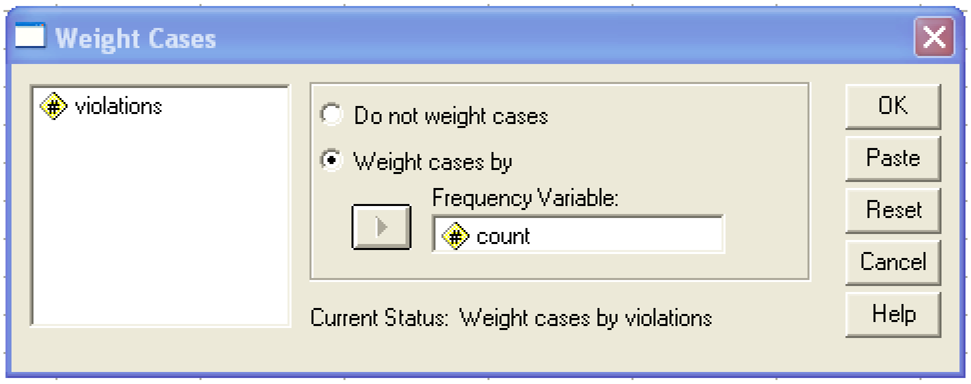
Program Note 7.2 – Binomial confidence intervals

In **Example 7.3** on page 183, we wish to find the 90 percent confidence interval for the population proportion of restaurants in violation of a health code. A sample of twenty restaurants are taken with four violations observed. We can use the SPSS procedure below to obtain a 90% confidence interval based on a t-distribution with 19 degrees of freedom, however, we recommend using exact confidence intervals (see Stata Program Notes for Chapter 7).

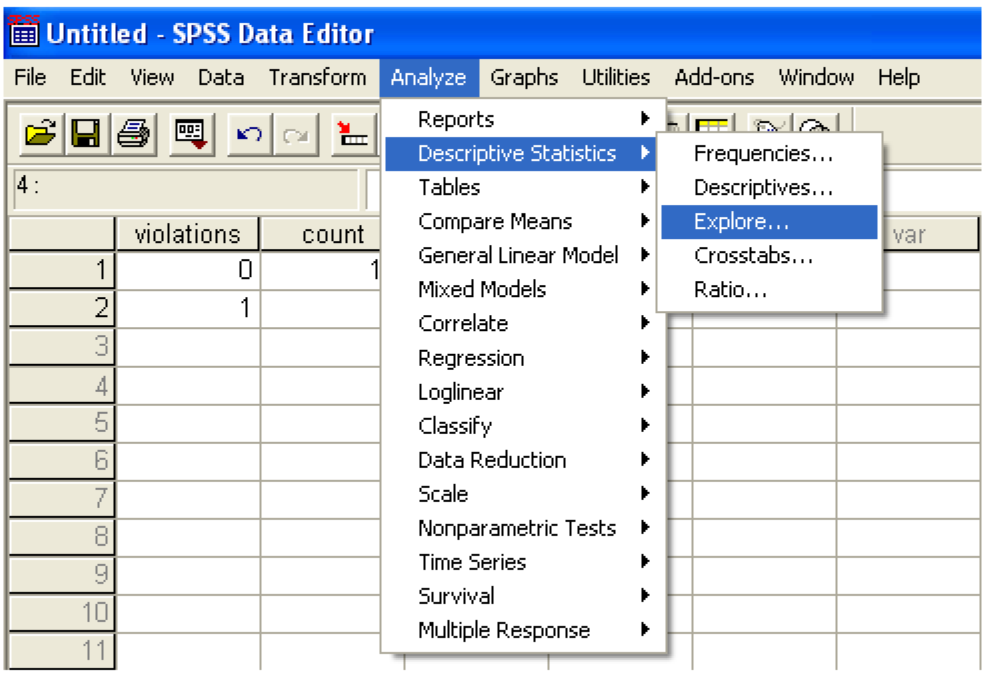


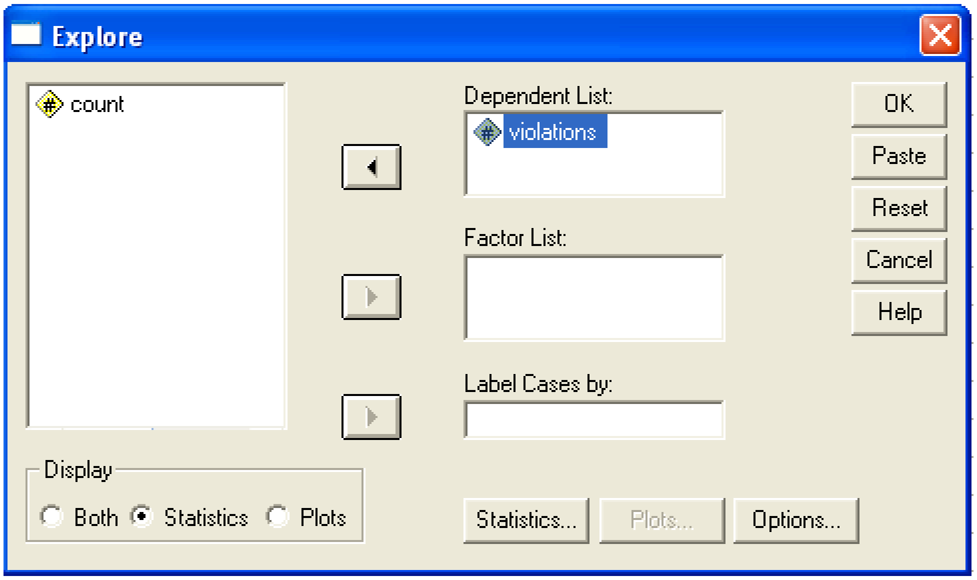
Rather than creating 16 zero's and 4 one's, we use **Data** -> **Weight Cases…** as shown below.

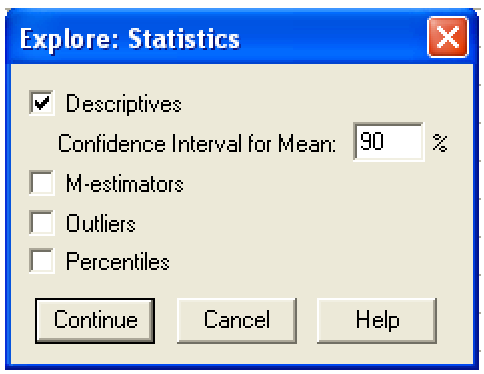




Now, we can use the SPSS procedure **Analyze** -> **Descriptive Statistics** -> **Explore...** to calculate the 90% confidence interval as shown below.







The SPSS output is shown below:

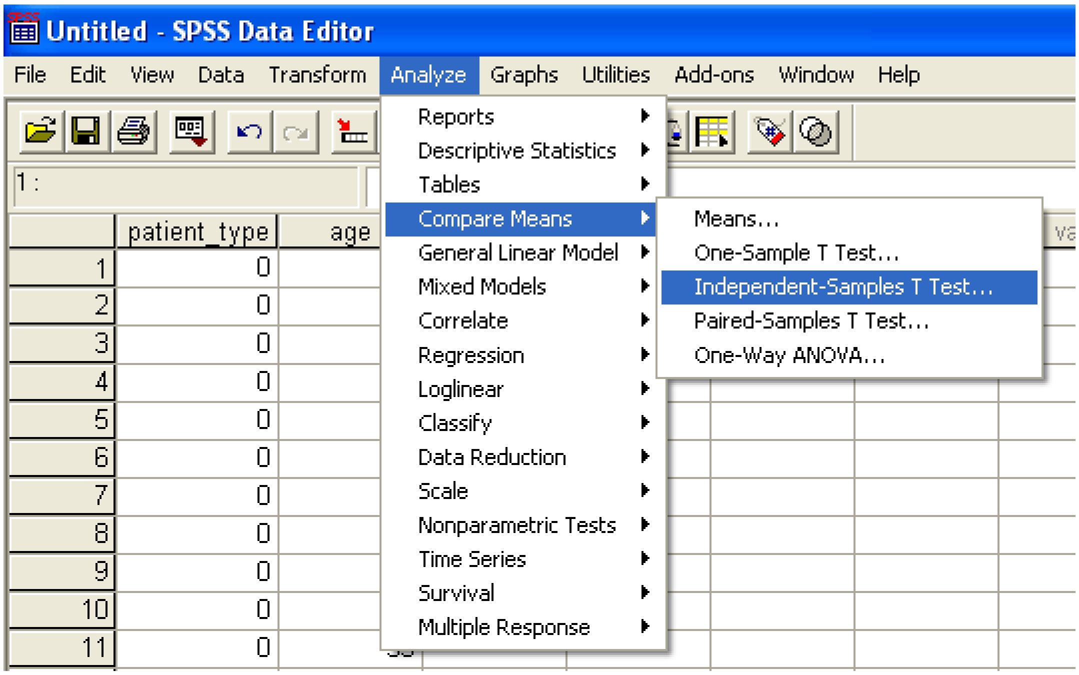
Descriptives

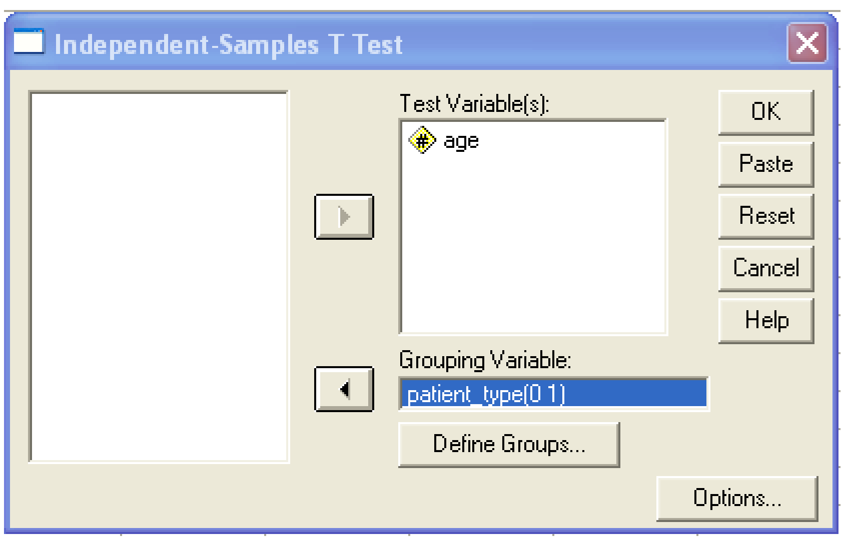
			Statistic	Std. Error
violations	Mean		.20	.092
	90% Confidence Interval for Mean	Lower Bound	.04	
		Upper Bound	.36	
	5% Trimmed Mean		.17	
	Median		.00	
	Variance		.168	
	Std. Deviation		.410	
	Minimum		0	
	Maximum		1	
	Range		1	
	Interquartile Range		0	
	Skewness		1.624	.512
	Kurtosis		.699	.992

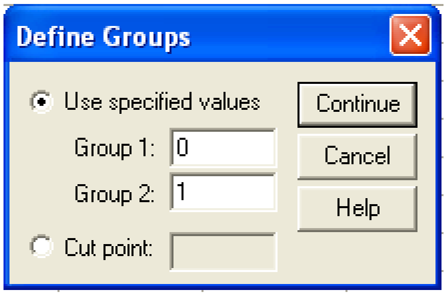
From the SPSS output, we see that the 90% confidence interval is (0.04, 0.36).

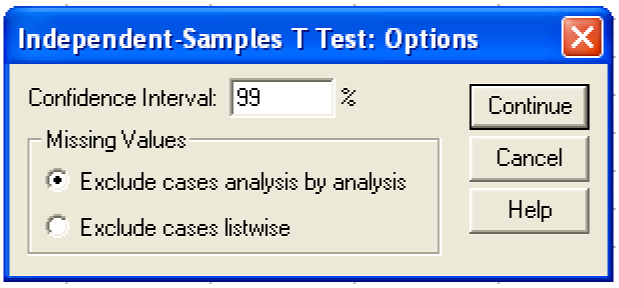
Program Note 7.3 – Confidence intervals using the t-distribution

In **Example 7.7** on page 192, we are comparing the mean ages between AML and ALL patients. Using the SPSS procedure below, we can calculate a 99% confidence interval for the difference between mean ages of AML and ALL patients as shown below.









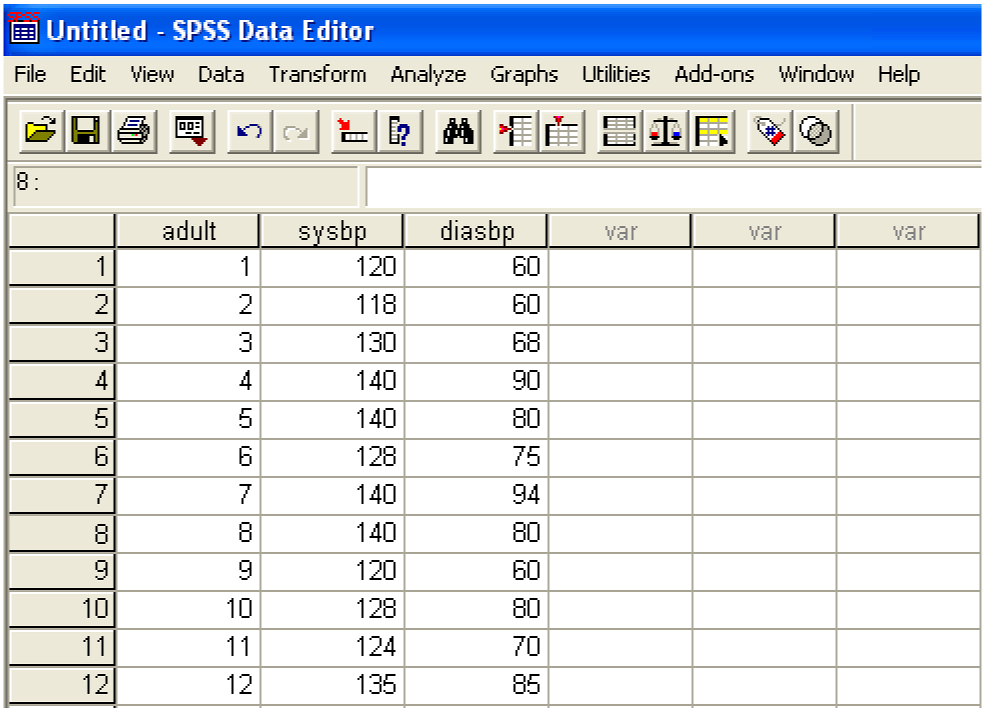
The SPSS output is shown below:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									99% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
age	Equal variances assumed	.019	.890	2.965	69	.004	13.213	4.456	1.409	25.017
	Equal variances not assumed			2.865	32.501	.007	13.213	4.612	.594	25.831

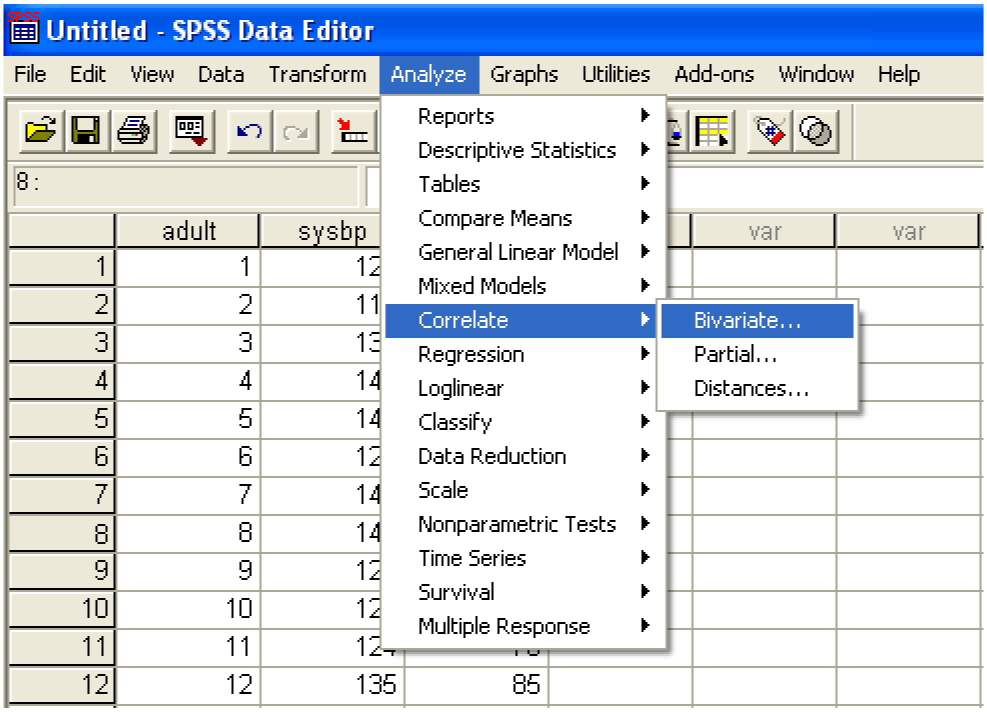
Program Note 7.4 – Confidence intervals for the Pearson correlation coefficient

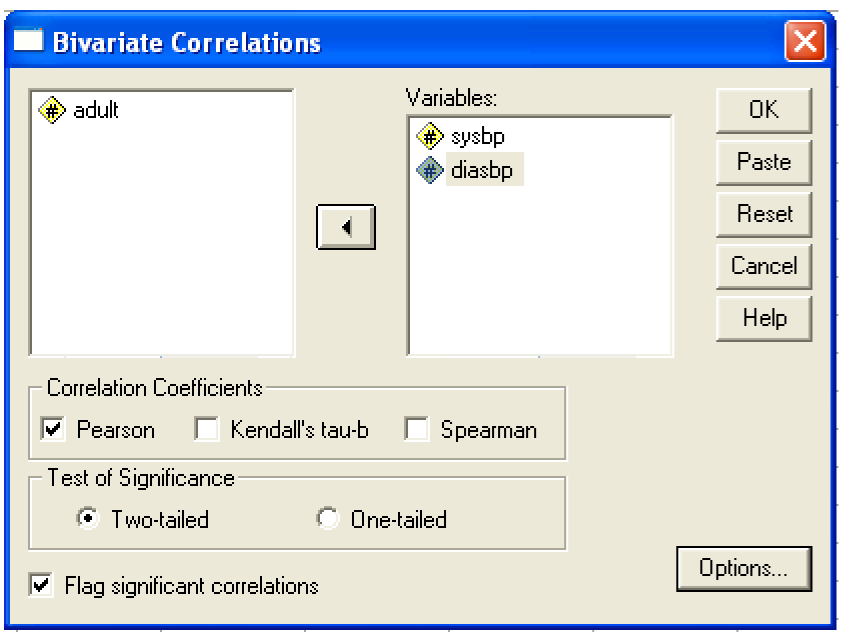
The data below correspond to **Example 3.18** on page 61. In **Example 3.18**, we calculated the Pearson correlation coefficient for the data on systolic and diastolic blood pressure. Using the data on the same 12 adults, the following SPSS procedure can be used to calculate confidence interval limits for the correlation coefficient.



	adult	sysbp	diasbp	var	var	var
1	1	120	60			
2	2	118	60			
3	3	130	68			
4	4	140	90			
5	5	140	80			
6	6	128	75			
7	7	140	94			
8	8	140	80			
9	9	120	60			
10	10	128	80			
11	11	124	70			
12	12	135	85			

First, calculate the Pearson correlation coefficient using **Analyze** -> **Correlate** -> **Bivariate…**.





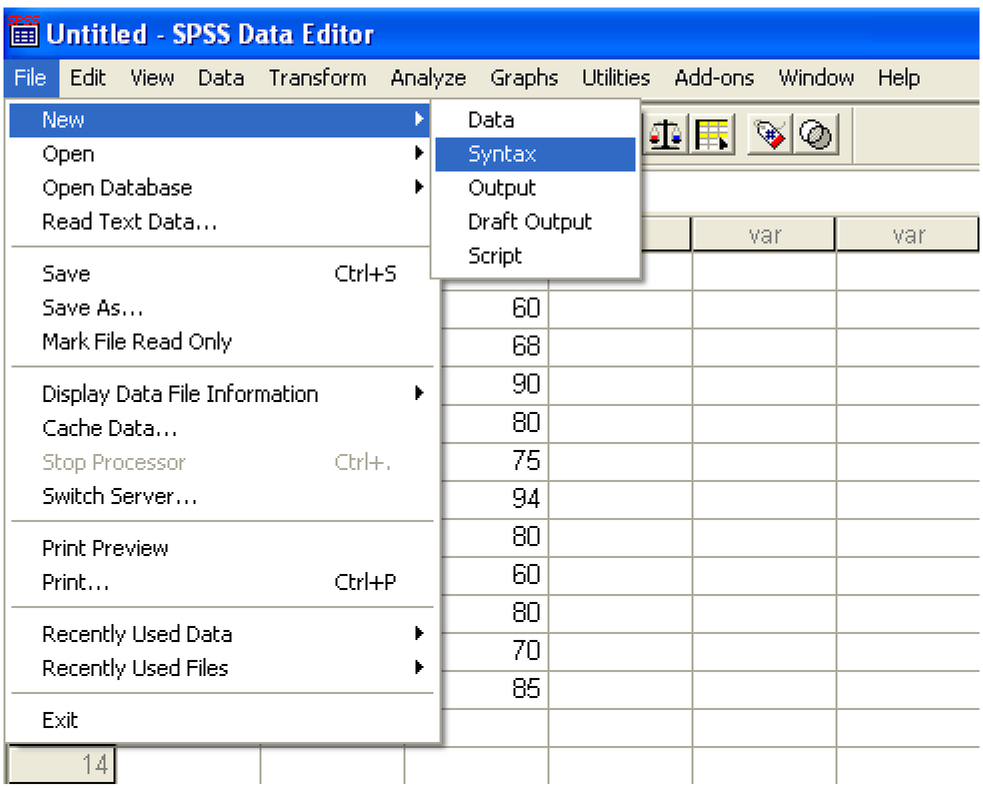
From the SPSS output, you should note that a confidence interval is not provided directly. However, the value of the correlation coefficient will be used to calculate confidence intervals using the SPSS Syntax Editor.

Correlations

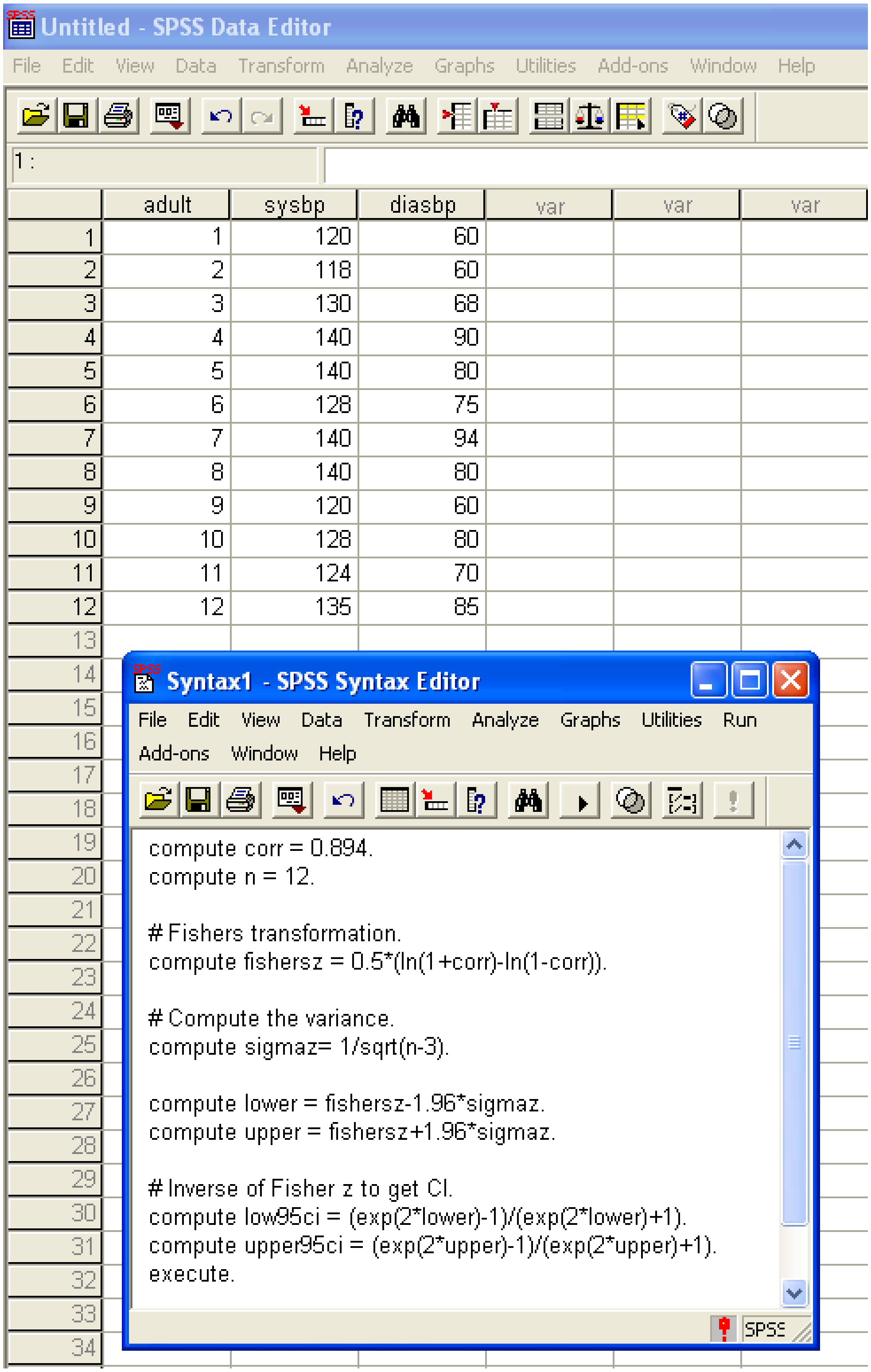
		sysbp	diasbp
sysbp	Pearson Correlation	1	.894(**)
	Sig. (2-tailed)	.	.000
	N	12	12
diasbp	Pearson Correlation	.894(**)	1
	Sig. (2-tailed)	.000	.
	N	12	12

** Correlation is significant at the 0.01 level (2-tailed).

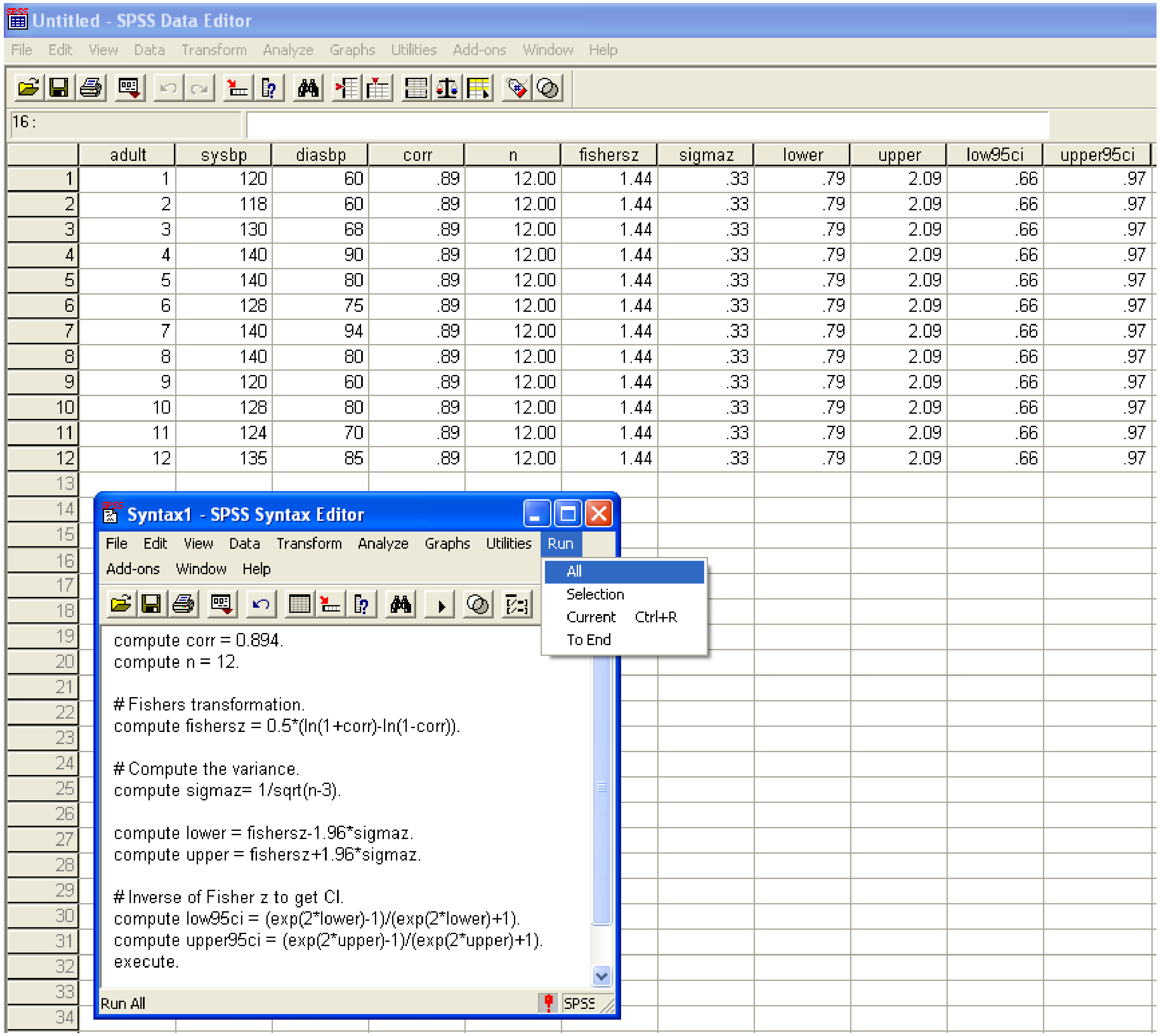
Next, use the SPSS procedure **File** -> **New** -> **Syntax**, to open an **SPSS Syntax Editor**.



Enter commands directly into the **SPSS Syntax Editor** as shown below.



In the **SPSS Syntax Editor** window, use the SPSS procedure: **Run** -> **All**. Finally, all of the variables computed in the **Syntax Editor** should be displayed in the columns next to diasbp as shown below. Note that the value for the lower 95% confidence interval is 0.66, and the value for the upper 95% confidence interval is 0.97.



	adult	sysbp	diasbp	corr	n	fishersz	sigmaz	lower	upper	low95ci	upper95ci
1	1	120	60	.89	12.00	1.44	.33	.79	2.09	.66	.97
2	2	118	60	.89	12.00	1.44	.33	.79	2.09	.66	.97
3	3	130	68	.89	12.00	1.44	.33	.79	2.09	.66	.97
4	4	140	90	.89	12.00	1.44	.33	.79	2.09	.66	.97
5	5	140	80	.89	12.00	1.44	.33	.79	2.09	.66	.97
6	6	128	75	.89	12.00	1.44	.33	.79	2.09	.66	.97
7	7	140	94	.89	12.00	1.44	.33	.79	2.09	.66	.97
8	8	140	80	.89	12.00	1.44	.33	.79	2.09	.66	.97
9	9	120	60	.89	12.00	1.44	.33	.79	2.09	.66	.97
10	10	128	80	.89	12.00	1.44	.33	.79	2.09	.66	.97
11	11	124	70	.89	12.00	1.44	.33	.79	2.09	.66	.97
12	12	135	85	.89	12.00	1.44	.33	.79	2.09	.66	.97

```
compute corr = 0.894.
compute n = 12.

# Fishers transformation.
compute fishersz = 0.5*(ln(1+corr)-ln(1-corr)).

# Compute the variance.
compute sigmaz= 1/sqrt(n-3).

compute lower = fishersz-1.96*sigmaz.
compute upper = fishersz+1.96*sigmaz.

# Inverse of Fisher z to get CI.
compute low95ci = (exp(2*lower)-1)/(exp(2*lower)+1).
compute upper95ci = (exp(2*upper)-1)/(exp(2*upper)+1).
execute.
```

SPSS Syntax is displayed again below. You should be able to copy and paste the text below directly into the **SPSS Syntax Editor**.

```
compute corr = 0.894.
compute n = 12.

# Fishers transformation.
compute fishersz = 0.5*(ln(1+corr)-ln(1-corr)).

# Compute the variance.
compute sigmaz= 1/sqrt(n-3).

compute lower = fishersz-1.96*sigmaz.
compute upper = fishersz+1.96*sigmaz.

# Inverse of Fisher z to get CI.
compute low95ci = (exp(2*lower)-1)/(exp(2*lower)+1).
compute upper95ci = (exp(2*upper)-1)/(exp(2*upper)+1).
execute.
```

For more information on using statistical software to calculate a confidence interval for the Pearson correlation coefficient download **Paper 170-31** by David Shen and Zaizai Lu found on the web at http://www2.sas.com/proceedings/sugi31/170-31.pdf .