SPSS Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery Second Edition
by Ronald N. Forthofer, Eun Sul Lee, Mike Hernandez

Chapter 8: Hypothesis Tests

(Note: these program notes were developed under an older version of SPSS. The current version of IBM® SPSS® Statistics is version 22)

Program Notes Outline

Note 8.1 – Testing a hypothesis about the mean assuming the variance is unknown
Note 8.2 – Testing the hypothesis about a population proportion
Note 8.3 – Correlation coefficients and their p-values
Note 8.4 – Testing the hypothesis of no difference in two population means assuming equal variances
Note 8.5 – Testing the hypothesis of no difference in two population means assuming unequal variances
Note 8.6 – Paired t-test
Note 8.7 – Testing a hypothesis about the difference of two proportions

Chapter 8 Formulas

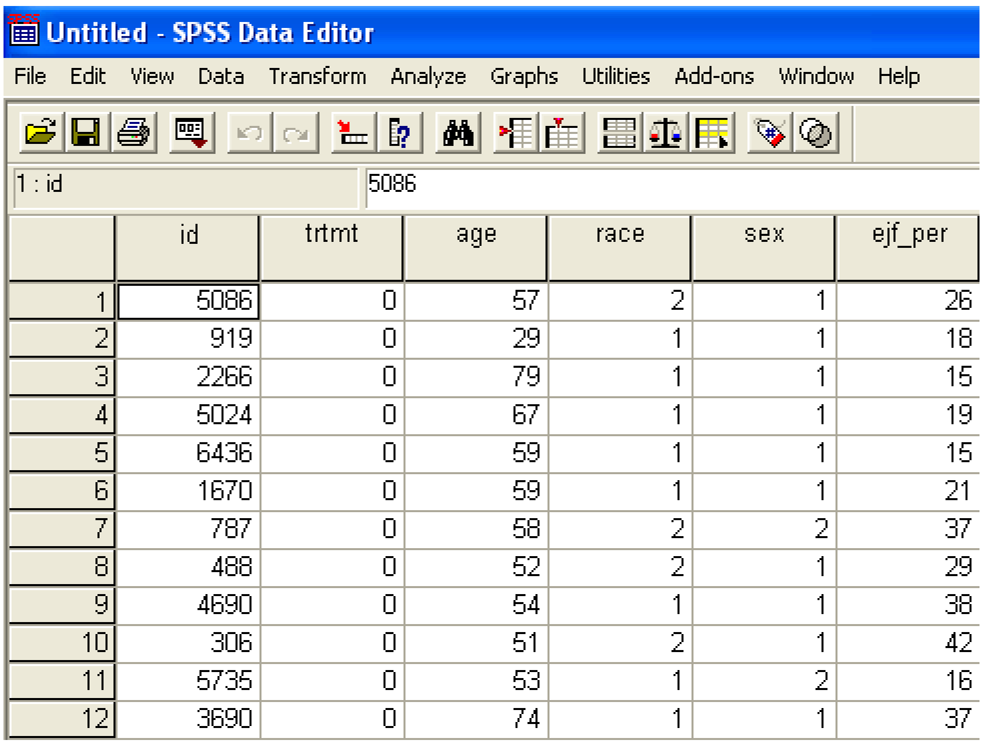
Test	Test Statistic	Distribution of Test Statistic
One sample z-test	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	standard normal distribution
One sample t-test	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	t-distribution with n-1 degrees of freedom
Independent samples t-test assuming equal variances	$\frac{(\bar{x}_1-\bar{x}_2)-\Delta_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	t-distribution distribution with degrees of freedom, df = $n_1 + n_2 - 2$, and s_p = $\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$
Independent samples t-test assuming unequal variances	$\frac{(\bar{x}_1-\bar{x}_2)-\Delta_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	t-distribution distribution with degrees of freedom,

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}$$

Paired t-test	$\frac{\bar{x}_d - \mu_d}{s_d / \sqrt{n}}$	t-distribution with n-1 degrees of freedom

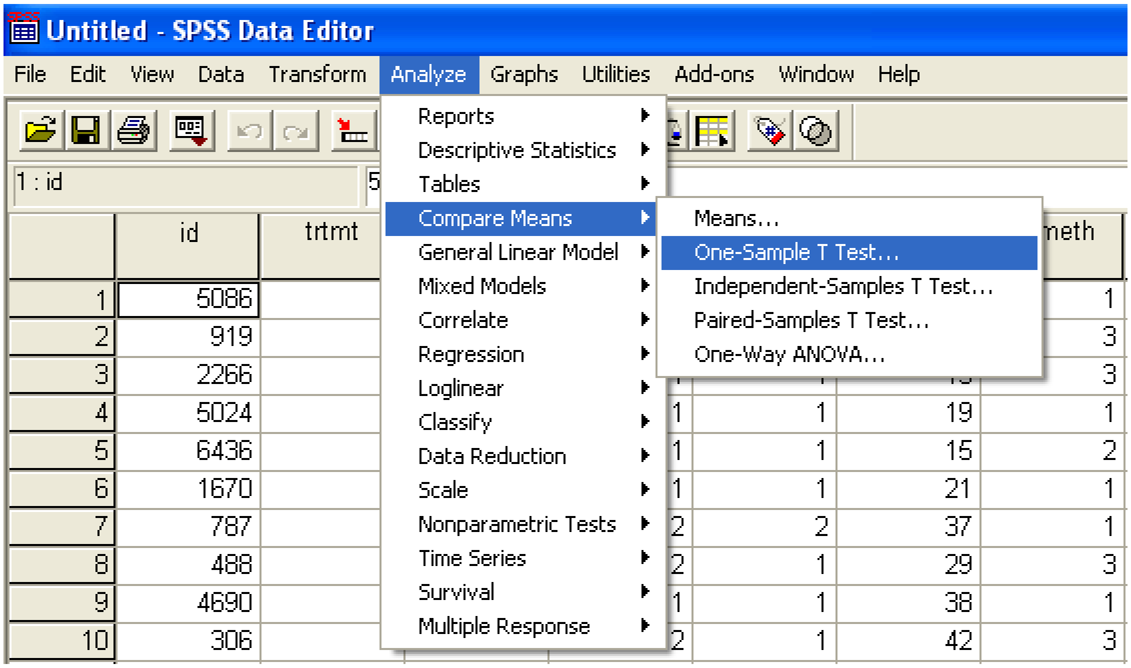
Program Note 8.1 – **Testing a hypothesis about the mean assuming the variance is unknown**

As an example, we use the **DIG200** dataset to test the null hypothesis that the population mean systolic blood pressure is significantly different from 122.3 mmHg. The variable sysbp represents systolic blood pressure in the **DIG200** dataset. The test being conducted is the one-sample t-test because the variance of the population is unknown and is being estimated using the **DIG200** dataset.

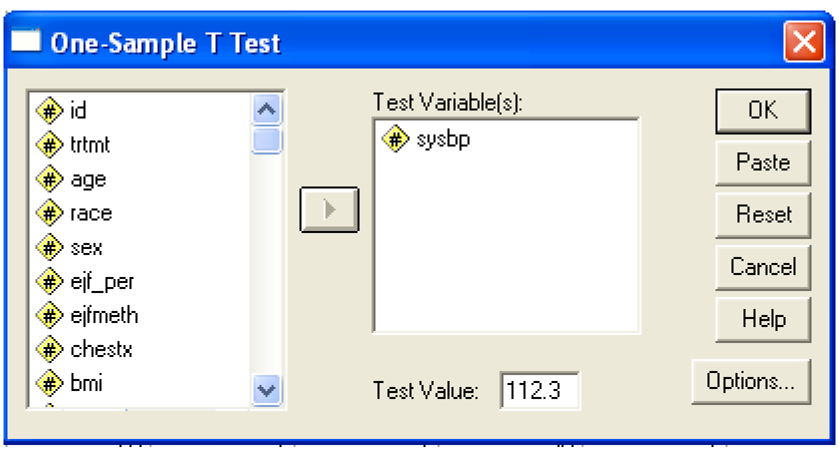


	id	trtmt	age	race	sex	ejf_per
1	5086	0	57	2	1	26
2	919	0	29	1	1	18
3	2266	0	79	1	1	15
4	5024	0	67	1	1	19
5	6436	0	59	1	1	15
6	1670	0	59	1	1	21
7	787	0	58	2	2	37
8	488	0	52	2	1	29
9	4690	0	54	1	1	38
10	306	0	51	2	1	42
11	5735	0	53	1	2	16
12	3690	0	74	1	1	37

To conduct a one-sample t test, we use the SPSS procedure **Analyze** -> **Compare Means** -> **One-Sample T Test…** as shown below.



Because the mean under the null hypothesis is 112.3, we replace **Test Value:** with 112.3 as shown below.



The SPSS output under **One-Sample Statistics** provides the number of observations, the mean, standard deviation, and standard error. Under **One-Sample Test**, SPSS provides the value of the t-statistic, its degrees of freedom, a two-tailed p-value, the mean difference between each observation and the Test Value of 112.3, and a 95% confidence interval.

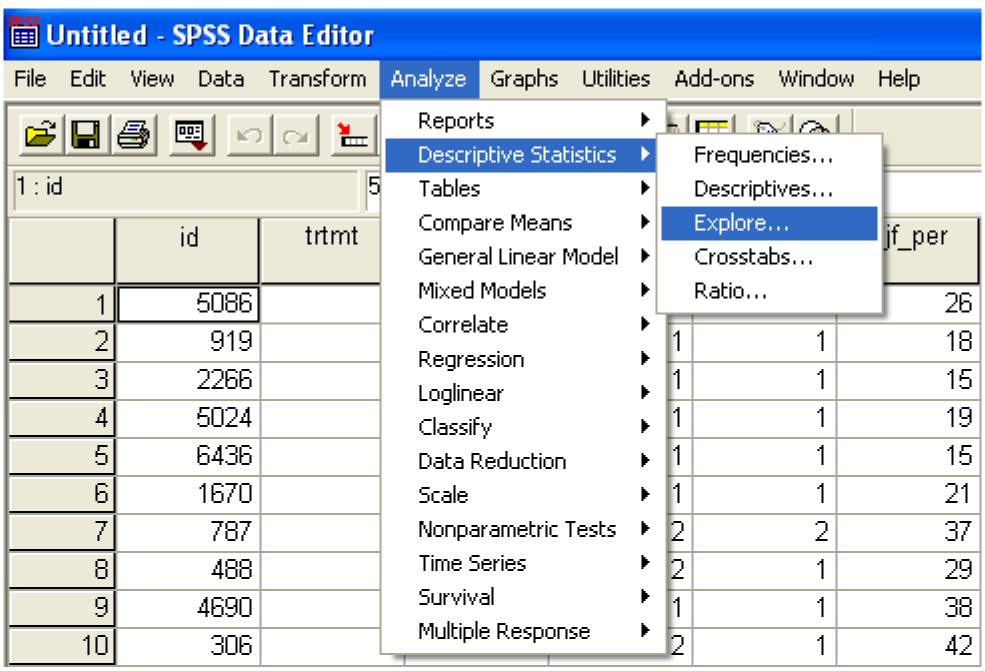
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
sysbp	199	125.82	18.179	1.289

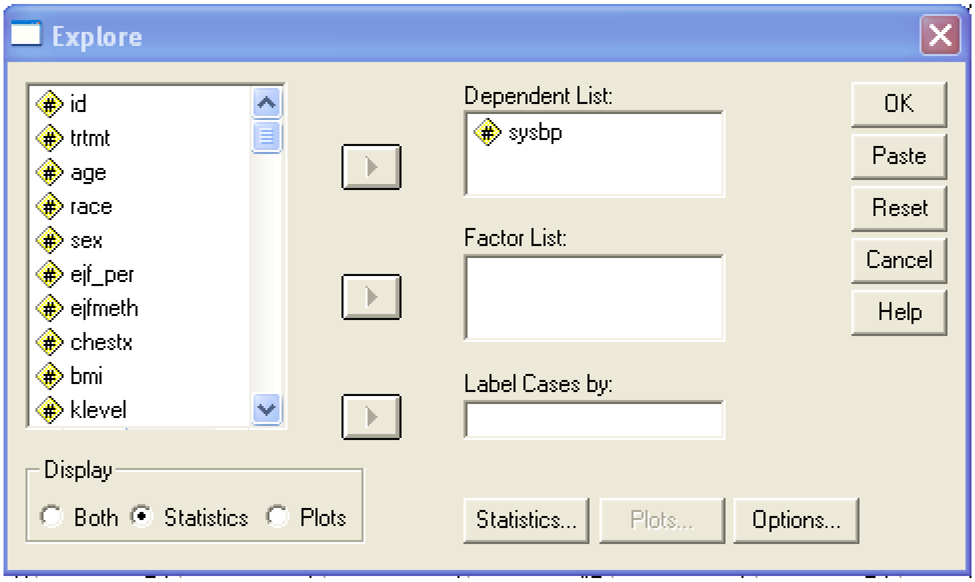
One-Sample Test

	Test Value = 112.3					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
sysbp	10.495	198	.000	13.524	10.98	16.07

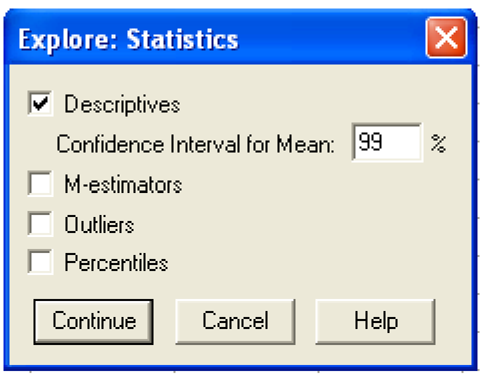
To obtain a 99% confidence interval, we use the SPSS procedure **Analyze** -> **Descriptive Statistics** -> **Explore…** as shown below.



Select sysbp under the **Dependent List:**. Under **Display**, you can choose either statistics, plots, or statistics and plots to be displayed. We chose to display only statistics.



Finally select, **Statistics…** to launch the window below and select the confidence interval for the mean.



The SPSS output provided below under **Descriptives** displays the mean and its corresponding 99% confidence interval.

Descriptives

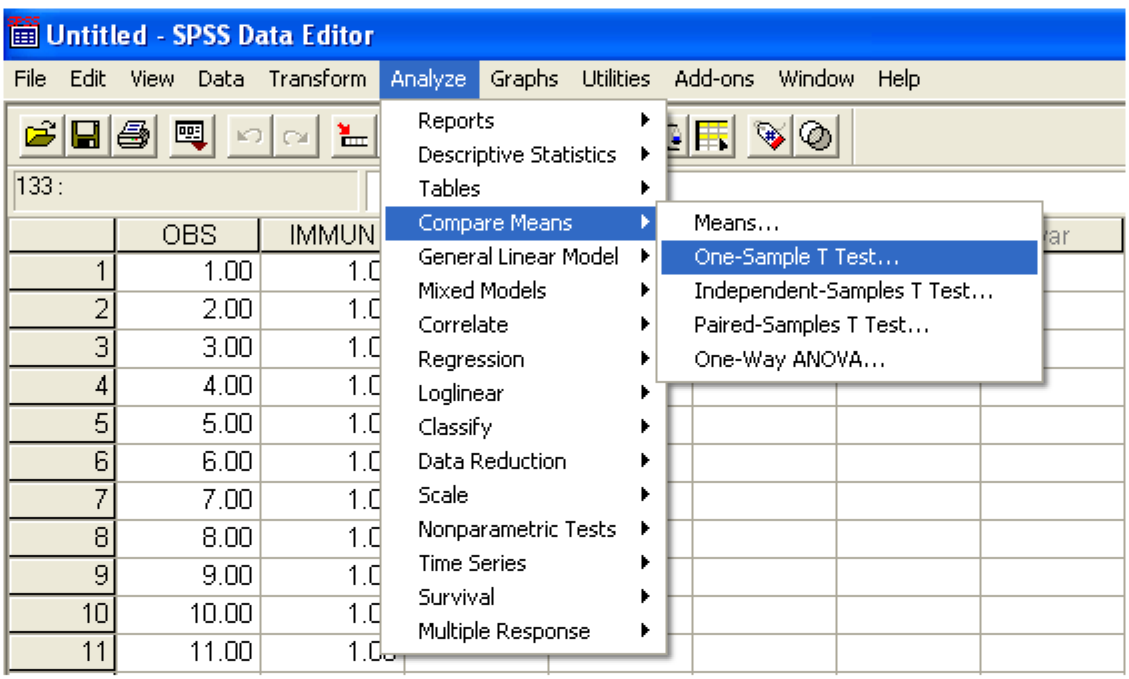
			Statistic	Std. Error
sysbp	Mean		125.82	1.289
	99% Confidence Interval for Mean	Lower Bound	122.47	
		Upper Bound	129.18	
	5% Trimmed Mean		125.48	
	Median		124.00	
	Variance		330.459	
	Std. Deviation		18.179	
	Minimum		85	
	Maximum		170	
	Range		85	
	Interquartile Range		26	
	Skewness		.209	.172
	Kurtosis		-.282	.343

Program Note 8.2 – Testing the hypothesis about a population proportion

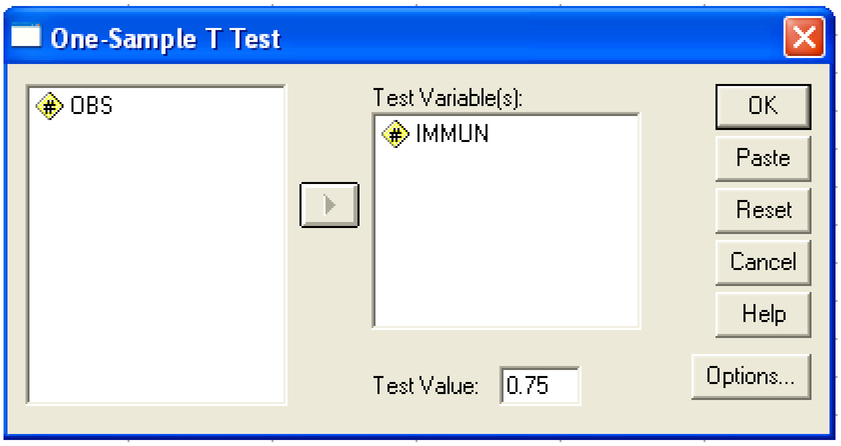
Here we examine the data in Example 8.4 and use a t-test to test the null hypothesis that π equals π_0. To create the data set, we start by creating 140 observations with a value of '1' and then replace 54 of the observations with the value of '0' as shown below using SPSS Syntax.

```
INPUT PROGRAM.
LOOP OBS = 1 TO 140.
   COMPUTE IMMUN = 1.
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

IF OBS >= 87 IMMUN = 0.
EXECUTE.
```

To conduct a One-Sample T Test, we use the SPSS procedure **Analyze** -> **Compare Means** -> **One-Sample T Test…** as shown below.



Because the proportion under the null hypothesis is 0.75, we replace **Test Value:** with 0.75 as shown below.



The SPSS output is shown below:

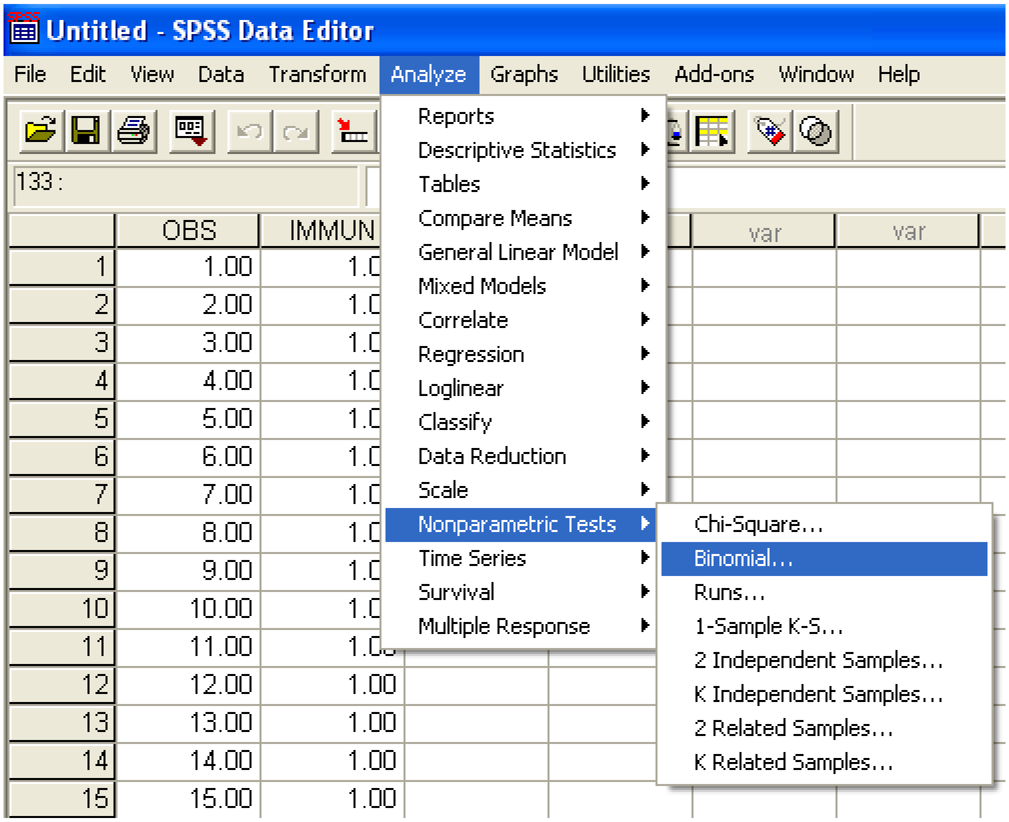
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
IMMUN	140	.6143	.48851	.04129

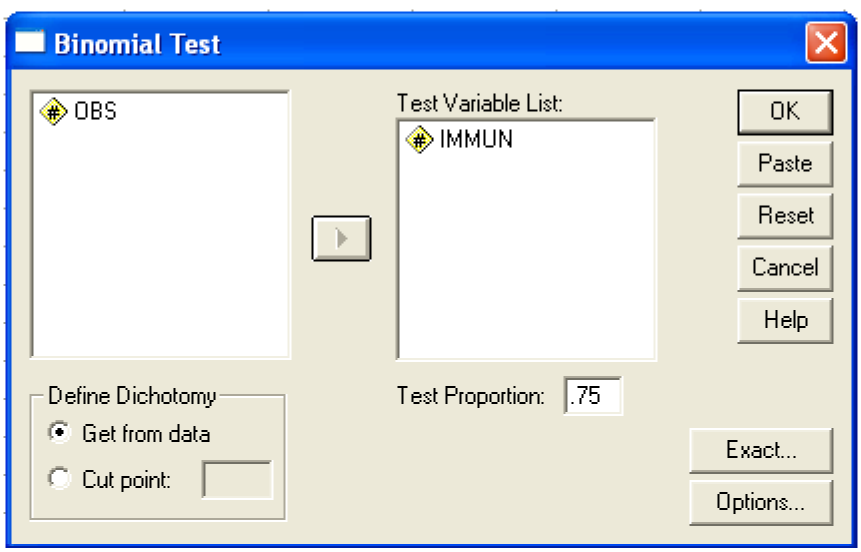
One-Sample Test

	Test Value = 0.75					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
IMMUN	-3.287	139	.001	-.13571	-.2173	-.0541

To conduct a binomial test use the SPSS procedure **Analyze** -> **Nonparametric Tests** -> **Binomial...** as shown below.



After selecting the variable IMMUN into the **Test Variable List:**, change the **Test Proportion:** to 0.75 which is the value of the proportion under the null hypothesis.



The SPSS output is shown below:

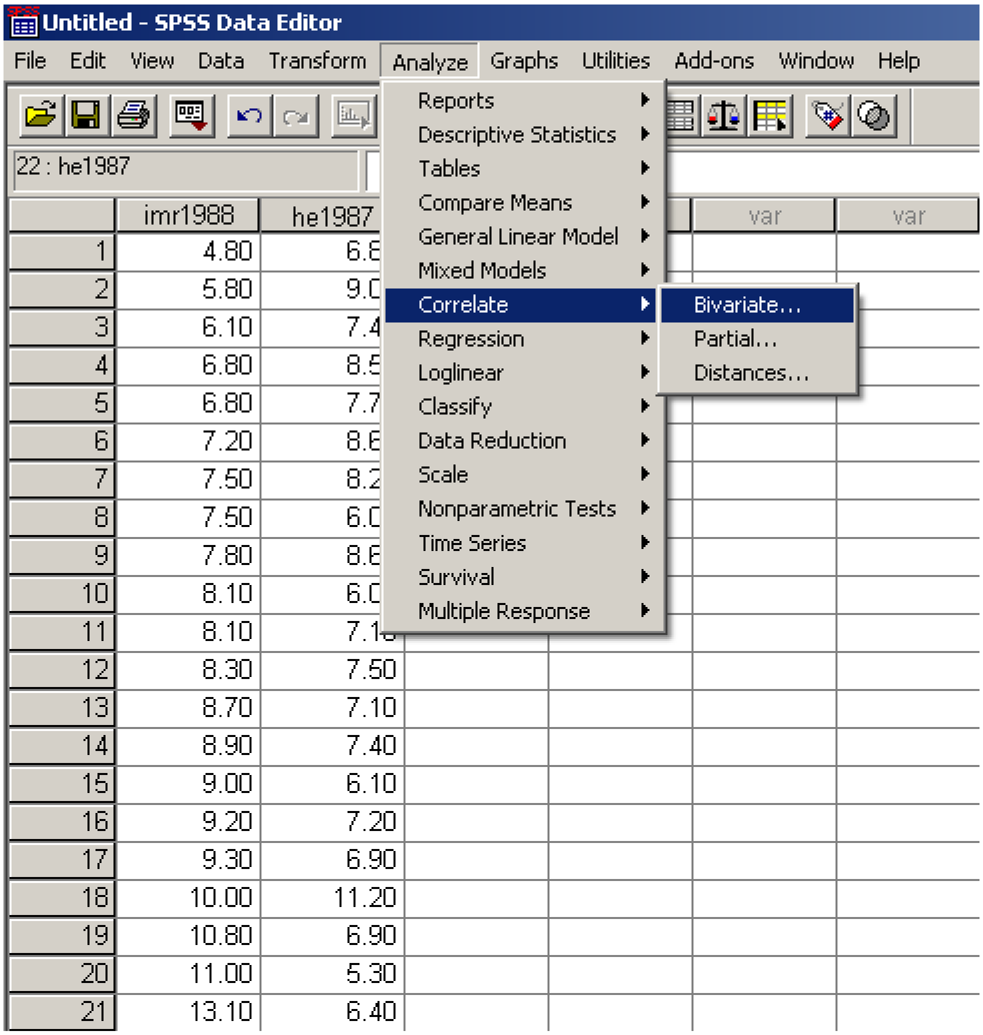
Binomial Test

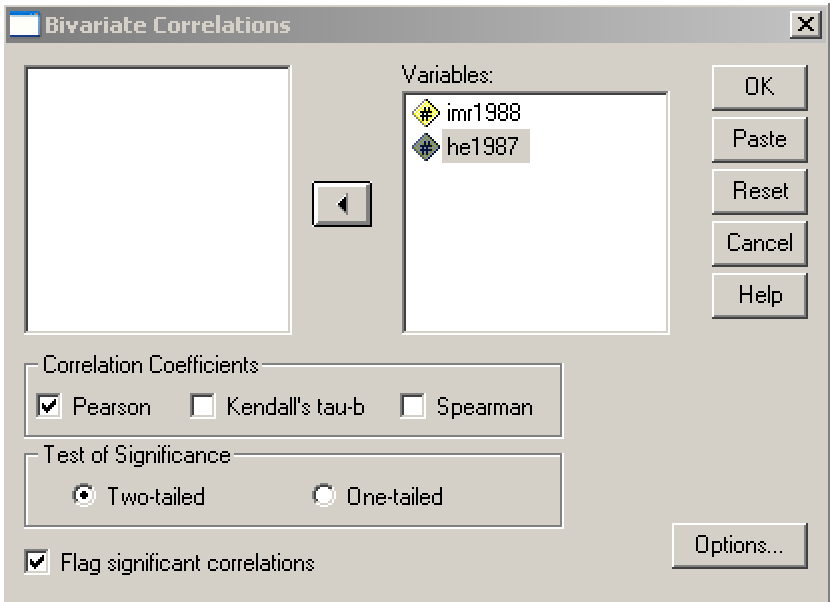
		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (1-tailed)
IMMUN	Group 1	1.00	86	.61	.75	.000(a,b)
	Group 2	.00	54	.39		
	Total		140	1.00		

a Alternative hypothesis states that the proportion of cases in the first group < .75.
b Based on Z Approximation.

Program Note 8.3 – **Correlation coefficients and their p-values**

Here we use the data presented in Example 8.7 to calculate the Pearson correlation coefficient. Using the SPSS procedure **Analyze** -> **Correlate** -> **Bivariate…**, we have the option to choose the Pearson or Spearman correlation coefficient along with either a one or two-tailed test.





The SPSS output is shown below:

Correlations

		imr1988	he1987
imr1988	Pearson Correlation	1	-.243
	Sig. (2-tailed)	.	.289
	N	21	21
he1987	Pearson Correlation	-.243	1
	Sig. (2-tailed)	.289	.
	N	21	21

Program Note 8.4 – Testing the hypothesis of no difference in two population means assuming equal variances

Here we present the data referred to in Example 8.9 which comes from **Chapter 7 Table 7.7**. The Independent Samples T-Test is used to test the hypothesis of no difference between two population means.

Untitled.sav - SPSS Data Editor

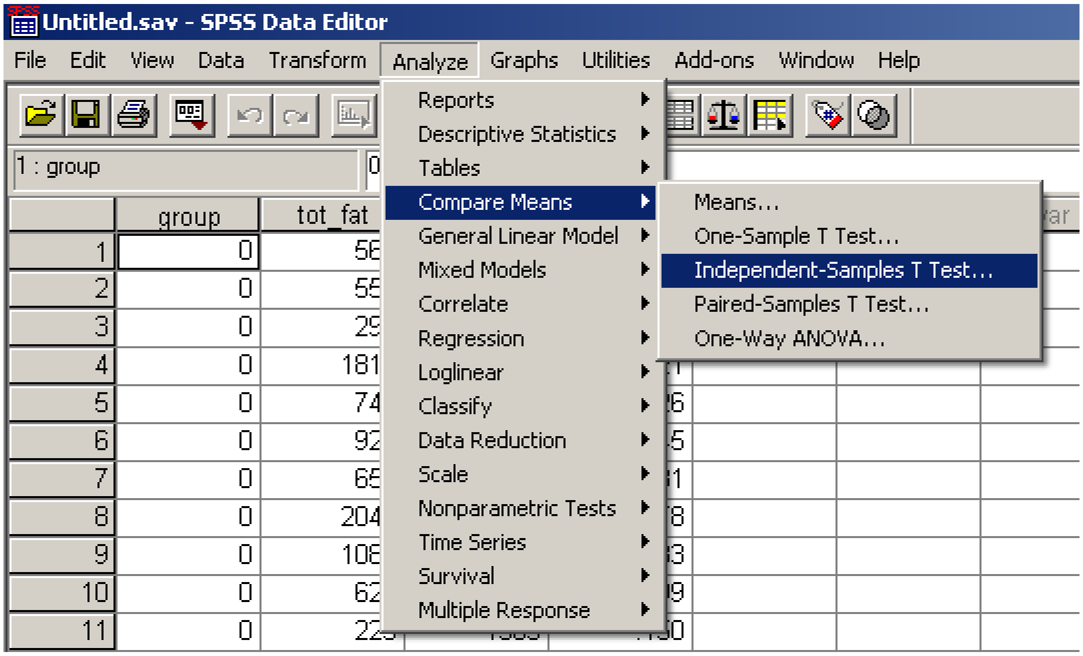
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

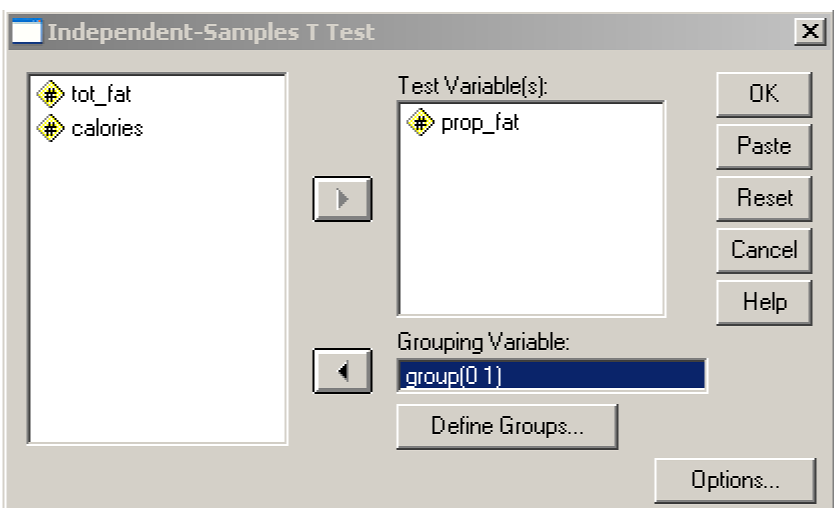
1 : group | 0

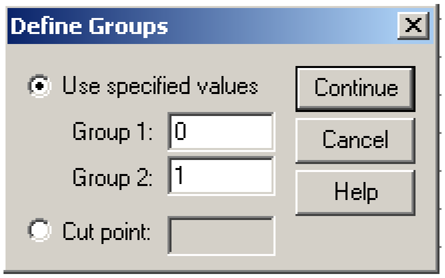
	group	tot_fat	calories	prop_fat	var	var
1	0	567	1823	.311		
2	0	558	2007	.278		
3	0	297	1053	.282		
4	0	1818	4322	.421		
5	0	747	1753	.426		
6	0	927	2685	.345		
7	0	657	2340	.281		
8	0	2043	3532	.578		
9	0	1089	2842	.383		
10	0	621	2074	.299		
11	0	225	1505	.150		
12	0	783	2330	.336		
13	0	1035	2436	.425		
14	0	1089	3076	.354		
15	0	621	1843	.337		
16	0	666	2301	.289		
17	0	1116	2546	.438		
18	0	531	1292	.411		
19	0	1089	3049	.357		
20	1	1197	3277	.365		
21	1	891	2039	.437		
22	1	495	2000	.248		
23	1	756	1781	.424		
24	1	1107	2748	.403		
25	1	792	2348	.337		
26	1	819	2773	.295		
27	1	738	2310	.319		
28	1	738	2594	.285		
29	1	882	1898	.465		
30	1	612	2400	.255		
31	1	252	2011	.125		
32	1	702	1645	.427		
33	1	387	1723	.225		

Data View / Variable View

Using the SPSS procedure **Analyze** -> **Compare Means -> Independent-Samples T Test...**, we can compare the mean proportions of total calories from fat between 5th/6th grade and 7th/8th grade boys.







SPSS output is shown below:

Group Statistics

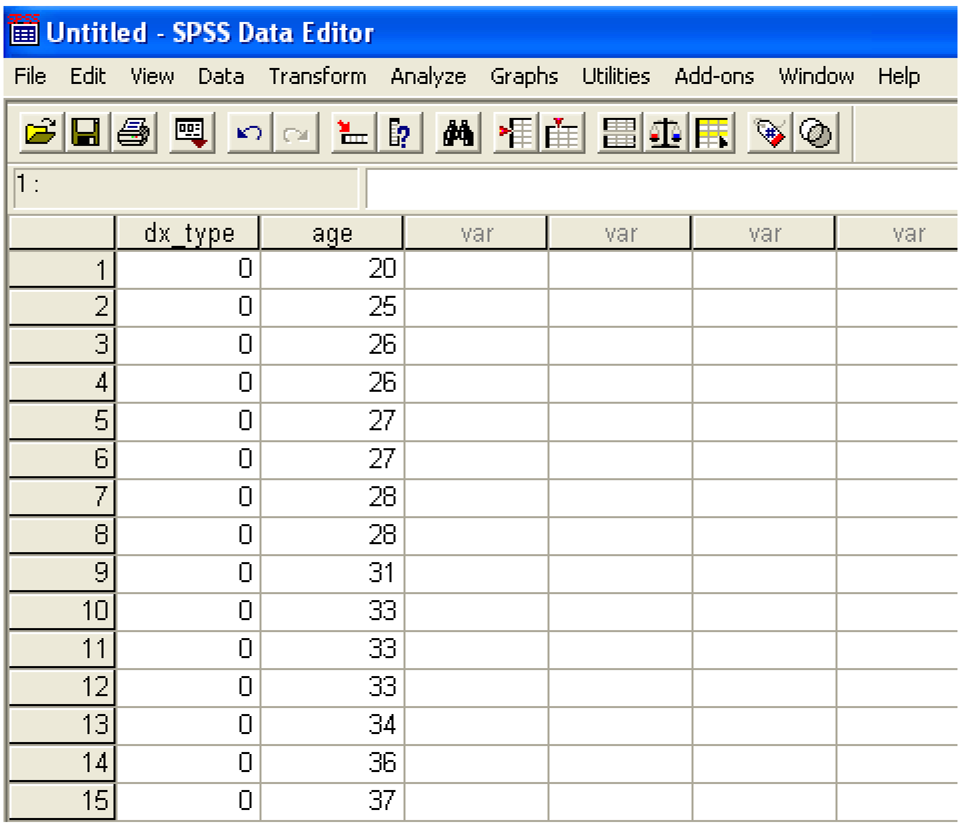
	group	N	Mean	Std. Deviation	Std. Error Mean
prop_fat	0	19	.35268	.089353	.020499
	1	14	.32929	.097383	.026027

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					
									95% Confidenc of the Differ
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower
prop_fat	Equal variances assumed	.485	.491	.716	31	.479	.023398	.032688	-.043269
	Equal variances not assumed			.706	26.708	.486	.023398	.033130	-.044613

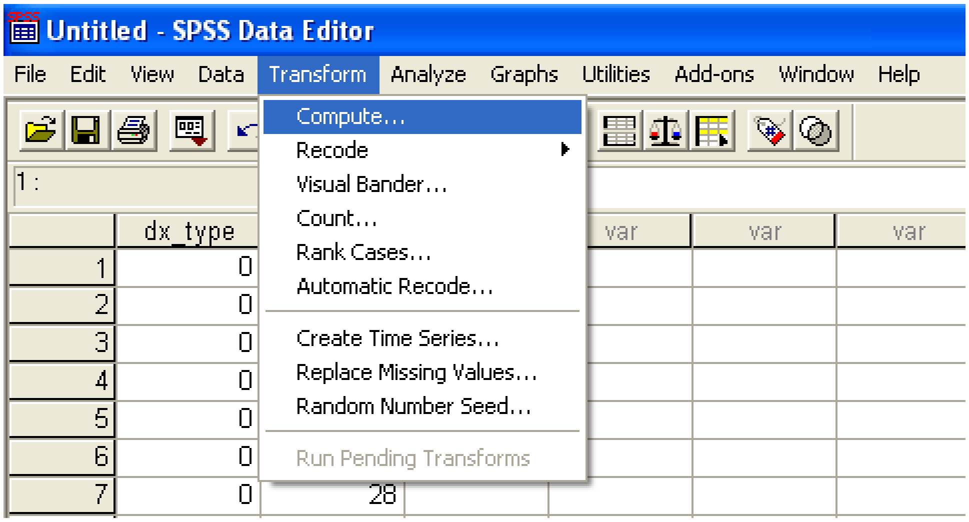
Program Note 8.5 – Testing the hypothesis of no difference in two population means assuming unequal variances

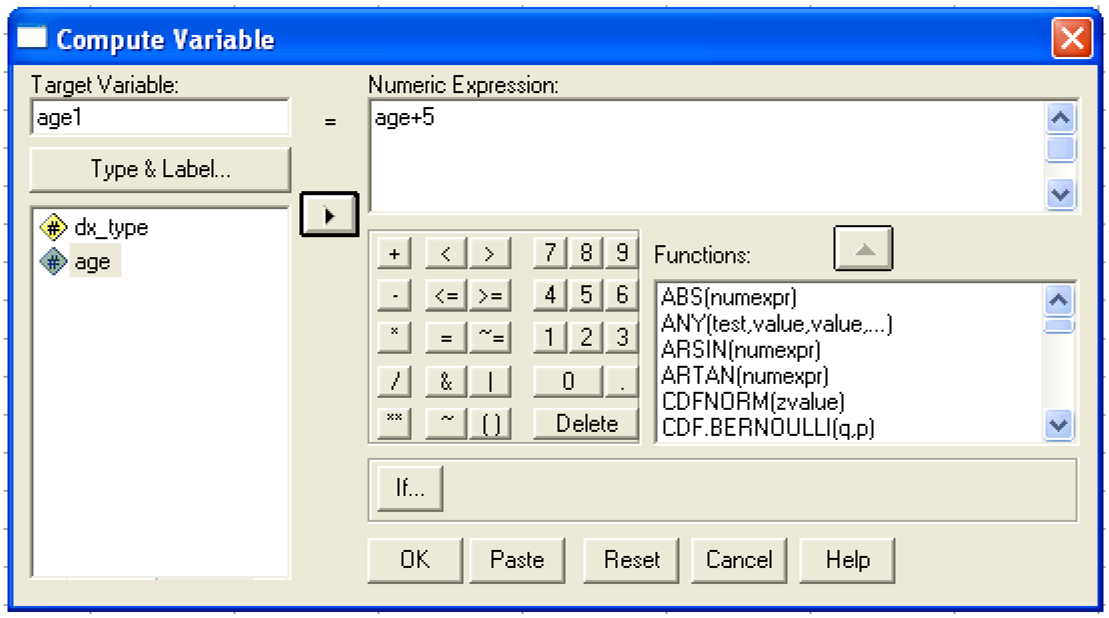
As was stated in Program Note 8.4, the Independent Samples T-Test is used to test the hypothesis of no difference between two population means. We now use the test statistic associated with the assumption that the variances are **unequal**. In Example 8.10, we refer to the data on ages of AML and ALL patients from **Chapter 7 Table 7.8** on **page 193**. We wish to test that the mean AML age is less than or equal to 5 years greater than the mean ALL age versus the alternative that the difference is greater than 5 years. The first 51 observations are from the AML patients and the last 20 are from the ALL patients.



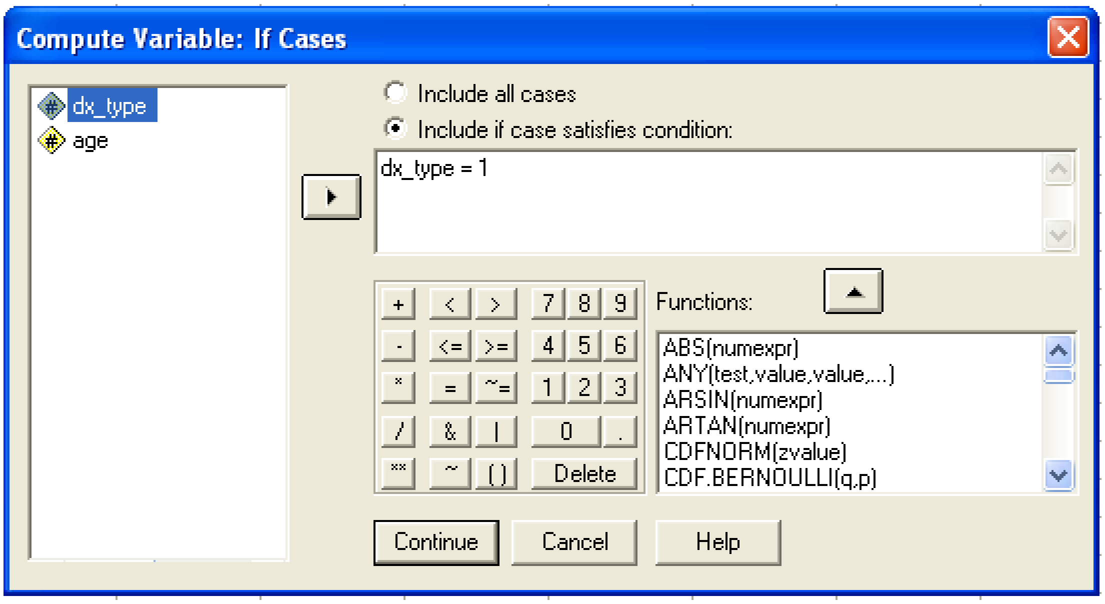
	dx_type	age	var	var	var	var
1	0	20				
2	0	25				
3	0	26				
4	0	26				
5	0	27				
6	0	27				
7	0	28				
8	0	28				
9	0	31				
10	0	33				
11	0	33				
12	0	33				
13	0	34				
14	0	36				
15	0	37				

In order to test the null hypothesis that the difference between the mean ages is less than or equal to 5 years, we need to do some data management and add 5 years to the ages of the ALL patients. We begin by using the SPSS procedure **Transform** -> **Compute…** to create a new variable age1.

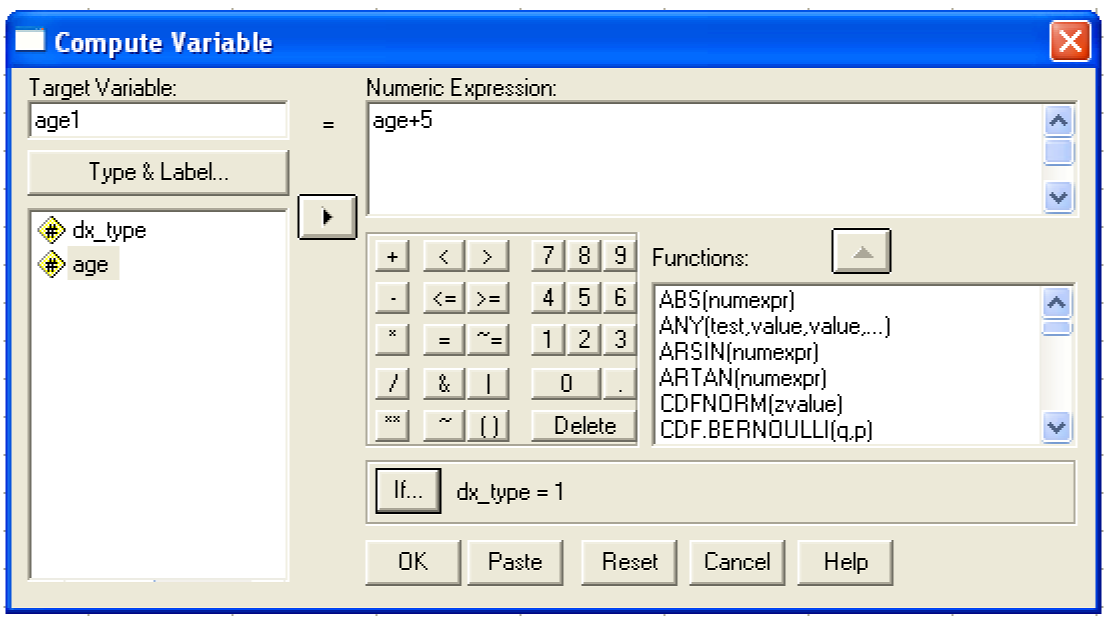




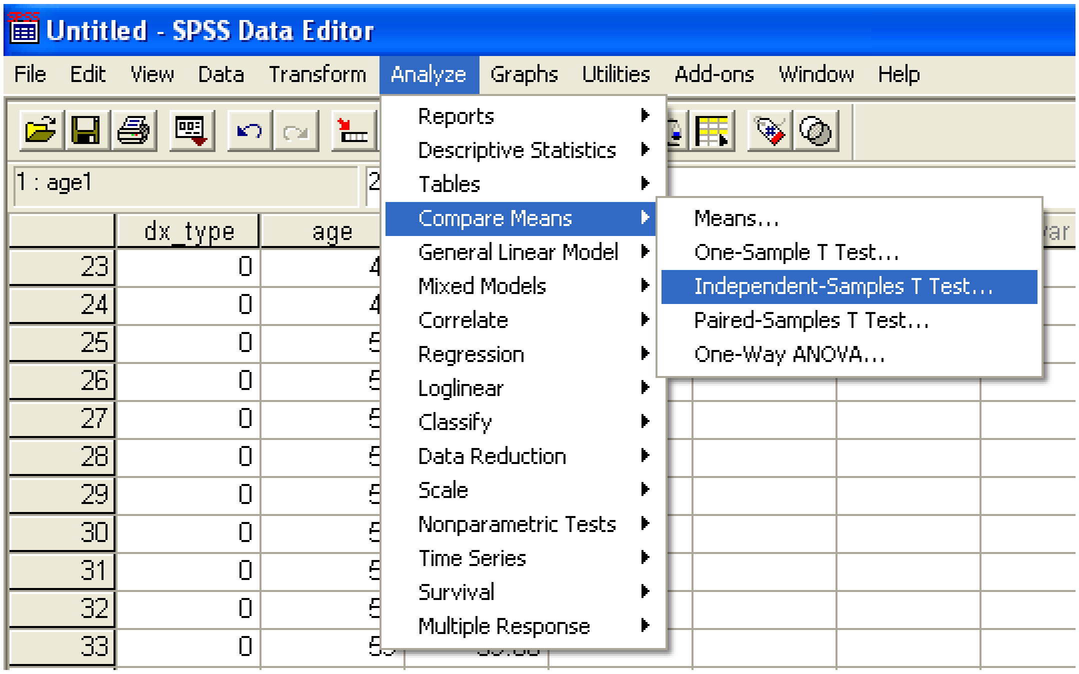
Click on the **If...** button to add five years only to the ALL patients.



After selecting the cases where dx_type = 1, click on **Continue**.



After clicking on **OK**, you should notice that the values under age1 are blank where dx_type = 0, but have increased by 5 where dx_type = 1. Just copy and past the values under age where dx_type = 0 onto the age1 column rather than using the **Compute Variable** window. Then use the SPSS procedure **Analyze** -> **Compare Means** -> **Independent-Samples T Test**.



SPSS output is shown below:

Group Statistics

	dx_type	N	Mean	Std. Deviation	Std. Error Mean
age1	0	51	49.8627	16.51063	2.31195
	1	20	41.6500	17.84812	3.99096

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence I of the Differen	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	U
age1	Equal variances assumed	.019	.890	1.843	69	.070	8.21275	4.45600	-.67674	17
	Equal variances not assumed			1.781	32.501	.084	8.21275	4.61225	-1.17643	17

Program Note 8.6 – Paired t-test

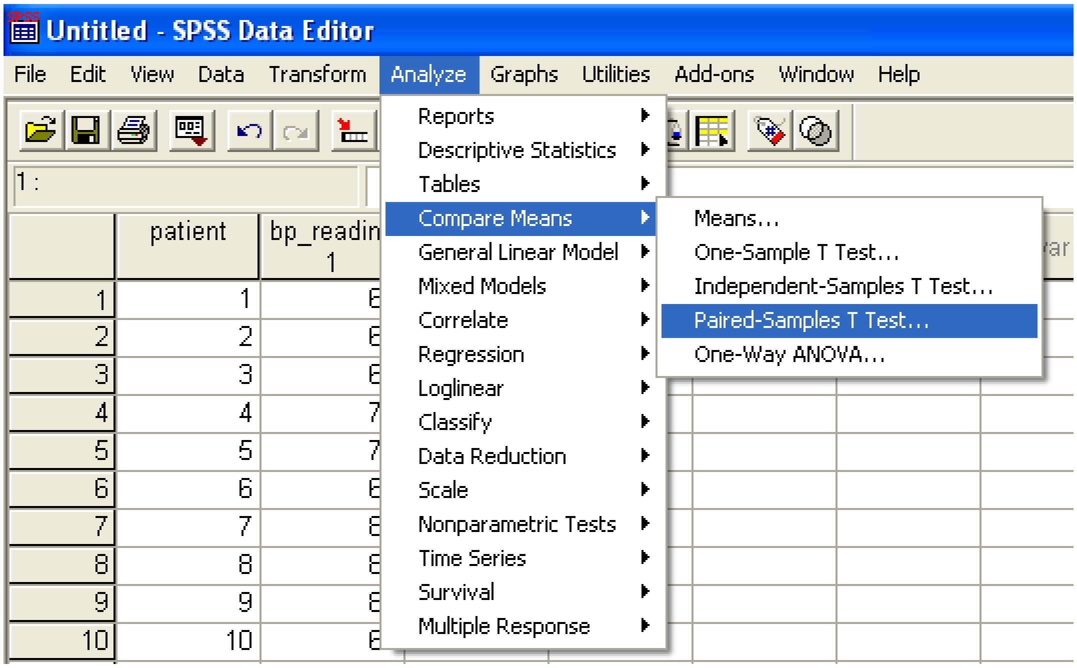
In Example 8.11, the mean difference between diastolic blood pressures before and after Ramipril is given. Because the raw data was not provided, we present two diastolic blood pressure readings on the same patient as an example of when to use a paired t-test. Assume that the variable bp_reading1 represents the first diastolic blood pressure reading recorded and bp_reading2 represents the second one.

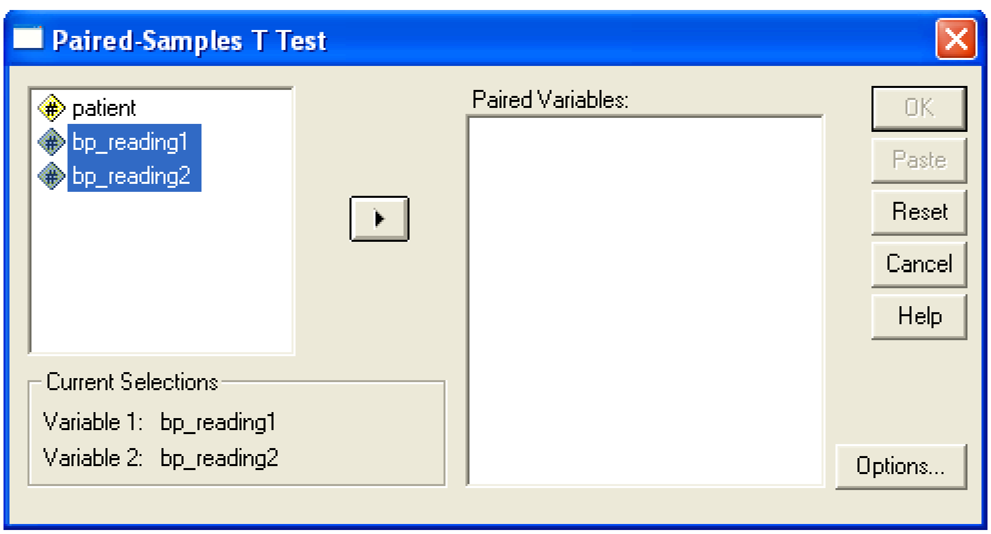
Untitled - SPSS Data Editor

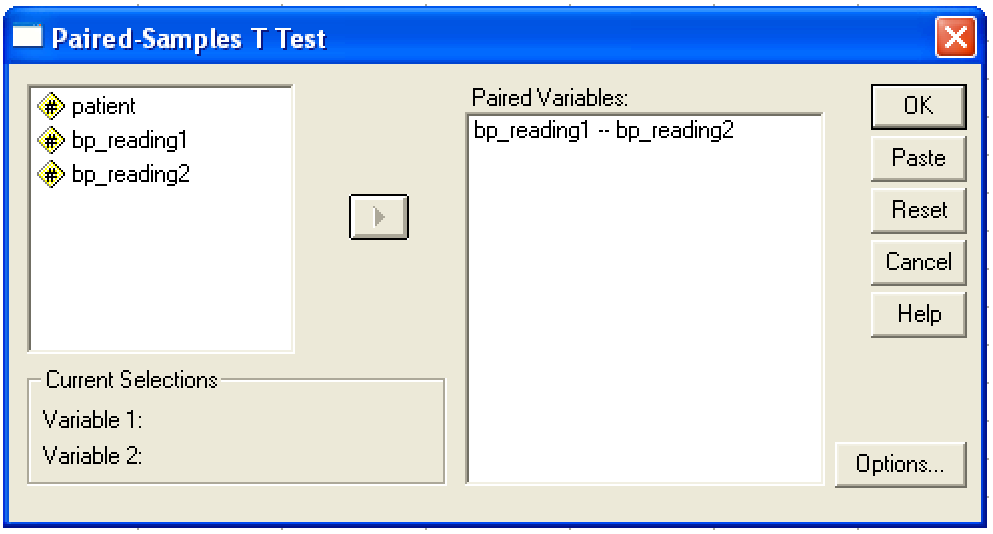
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	patient	bp_reading1	bp_reading2
1	1	60	82
2	2	60	66
3	3	60	80
4	4	73	68
5	5	70	70
6	6	68	74
7	7	87	98
8	8	80	70
9	9	80	96
10	10	60	70
11	11	60	90
12	12	82	70
13	13	66	78
14	14	80	70
15	15	68	80
16	16	70	88
17	17	74	90
18	18	98	70
19	19	70	78
20	20	96	80
21	21	70	68
22	22	90	70
23	23	70	74
24	24	78	98
25	25	70	70

Using the SPSS procedure **Analyze** -> **Compare Means -> Paired-Samples T Test…**, we can conduct a paired t-test.


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 :
patient
bp_readin 1
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
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...


Paired-Samples T Test
patient
bp_reading1
bp_reading2
Paired Variables:
OK
Paste
Reset
Cancel
Help
Current Selections
Variable 1: bp_reading1
Variable 2: bp_reading2
Options...



SPSS output is shown below:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	bp_reading1	73.60	25	10.958	2.192
	bp_reading2	77.92	25	10.058	2.012

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	bp_reading1 & bp_reading2	25	.053	.800

Paired Samples Test

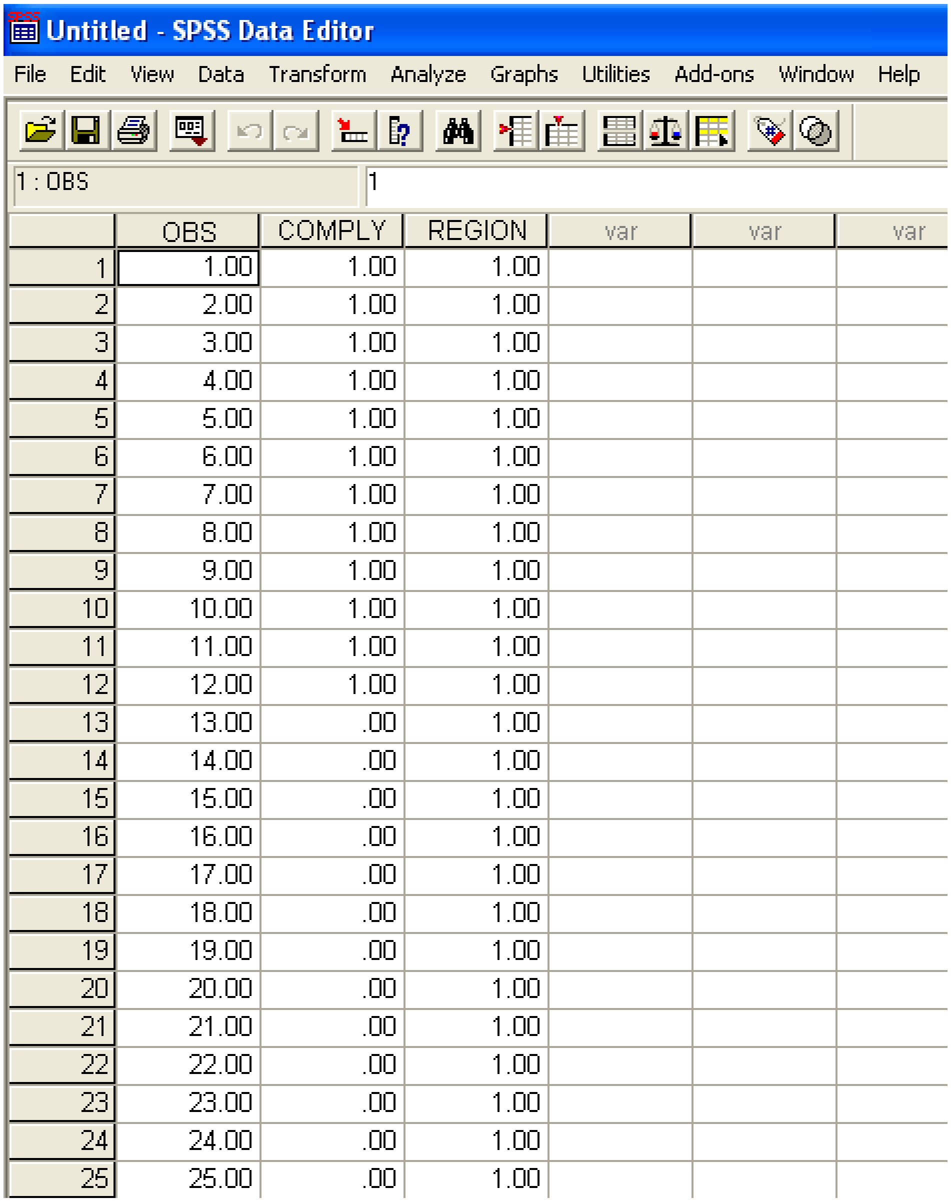
		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	bp_reading1 - bp_reading2	-4.320	14.473	2.895	-10.294	1.654	-1.492	24	.149

Program Note 8.7 – Testing a hypothesis about the difference of two proportions

Here we create the data used in Example 8.12. The data are the compliance status of 42 milk producers in the East and 50 milk producers in the Southwest. We will use the variable comply to indicate compliance status and the variable region to distinguish between the East and the Southwest. The SPSS Syntax used to create the data is shown below.

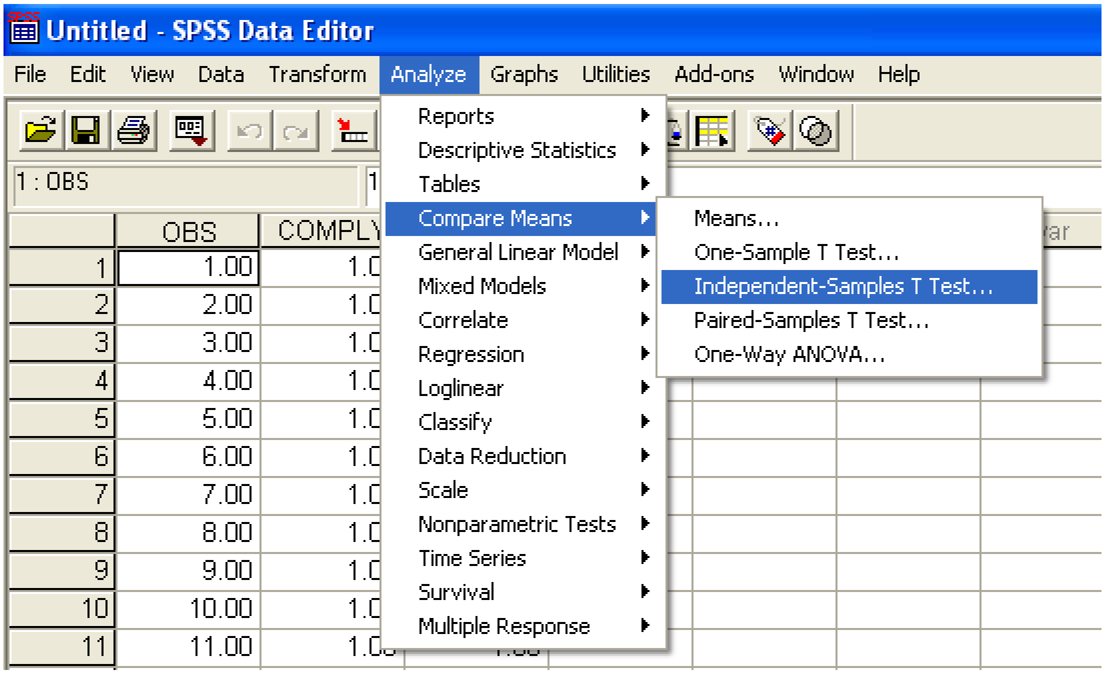
```
INPUT PROGRAM.
LOOP OBS = 1 TO 92.
   COMPUTE COMPLY = 0.
   COMPUTE REGION = 1.
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

IF OBS <= 12 COMPLY = 1.
IF OBS >= 43 AND OBS <= 63 COMPLY = 1.
IF OBS >= 43 REGION = 2.
EXECUTE.
```

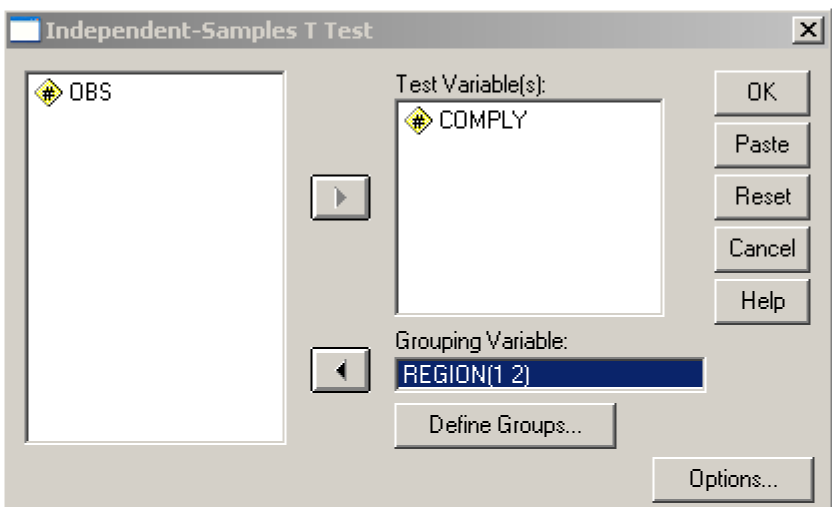
The variable COMPLY equals “0” to indicate compliance and “1” to indicate non-compliance. The variable REGION equals “1” when referring to the East and equals “2” when referring to the Southwest. A t-test at the 0.01 significance level is conducted using the SPSS procedure below:

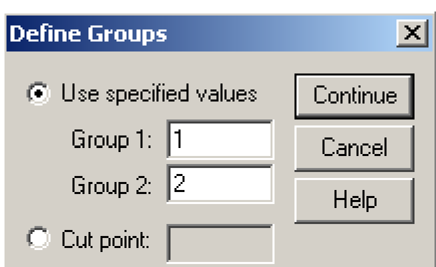


	OBS	COMPLY	REGION	var	var	var
1	1.00	1.00	1.00			
2	2.00	1.00	1.00			
3	3.00	1.00	1.00			
4	4.00	1.00	1.00			
5	5.00	1.00	1.00			
6	6.00	1.00	1.00			
7	7.00	1.00	1.00			
8	8.00	1.00	1.00			
9	9.00	1.00	1.00			
10	10.00	1.00	1.00			
11	11.00	1.00	1.00			
12	12.00	1.00	1.00			
13	13.00	.00	1.00			
14	14.00	.00	1.00			
15	15.00	.00	1.00			
16	16.00	.00	1.00			
17	17.00	.00	1.00			
18	18.00	.00	1.00			
19	19.00	.00	1.00			
20	20.00	.00	1.00			
21	21.00	.00	1.00			
22	22.00	.00	1.00			
23	23.00	.00	1.00			
24	24.00	.00	1.00			
25	25.00	.00	1.00			

Here, we will use SPSS to conduct a t-test, therefore the test statistic obtained from using SPSS will vary slightly from the test statistic presented in the book.


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
1 : OBS
OBS
COMPLY
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
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...


Independent-Samples T Test
OBS
Test Variable(s):
COMPLY
Grouping Variable:
REGION(1 2)
Define Groups...
OK
Paste
Reset
Cancel
Help
Options...


Define Groups
Use specified values
Group 1: 1
Group 2: 2
Cut point:
Continue
Cancel
Help

The SPSS output is given below:

Group Statistics

	REGION	N	Mean	Std. Deviation	Std. Error Mean
COMPLY	1.00	42	.2857	.45723	.07055
	2.00	50	.4200	.49857	.07051

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
COMPLY	Equal variances assumed	6.804	.011	-1.336	90	.185	-.13429	.10050	-.33396	.06538
	Equal variances not assumed			-1.346	89.279	.182	-.13429	.09974	-.33247	.06390