SPSS Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery Second Edition
by Ronald N. Forthofer, Eun Sul Lee, Mike Hernandez
Chapter 5: Probability Distributions

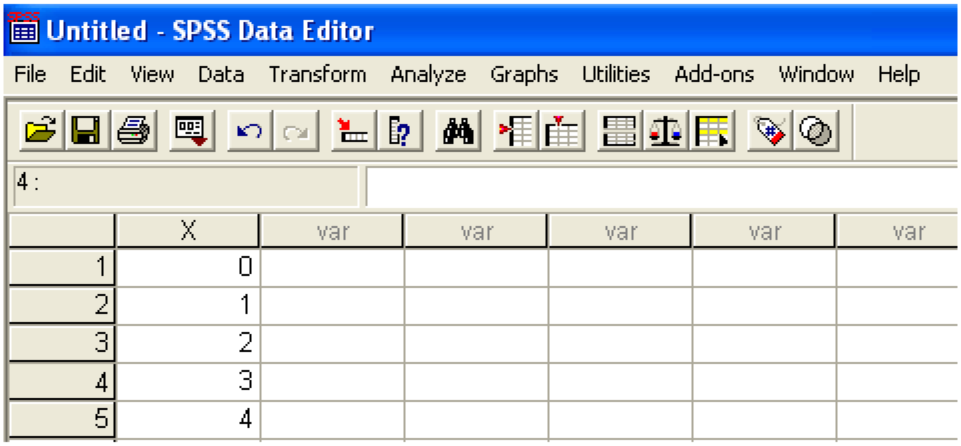
Chapter 5 Formulas

Distribution		Formula	Characteristics
Binomial	Probability Mass Function	$\Pr(X = x) = \binom{n}{x}\pi^{x}(1-\pi)^{1-x}$ where $\mathbf{x} = 0, 1, 2, \ldots$	**π** is the probability of success for any trial and **n** is the number of trials
Poisson	Probability Mass Function	$\Pr(X = x) = \frac{e^{-\mu}\mu^{x}}{x!}$ where $\mathbf{x} = 0, 1, 2, \ldots$	**μ** is the parameter of the Poisson distribution that refers to both the mean and variance
Normal	Probability Density Function	$f(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ where $-\infty < \mathbf{x} < \infty$	Where **μ** is the mean and $\boldsymbol{\sigma^2}$ is the variance

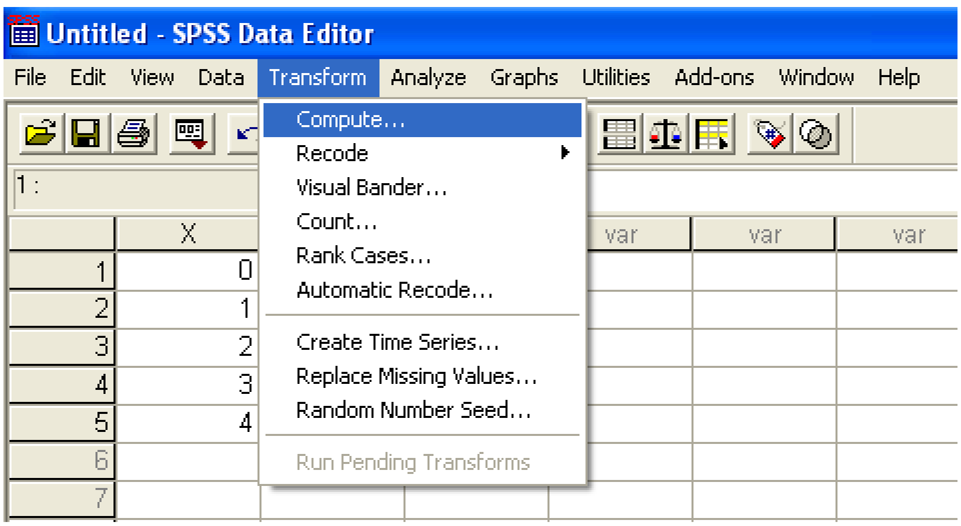
Program Note 5.1 – Finding binomial and Poisson probabilities

1. Finding binomial probabilities:

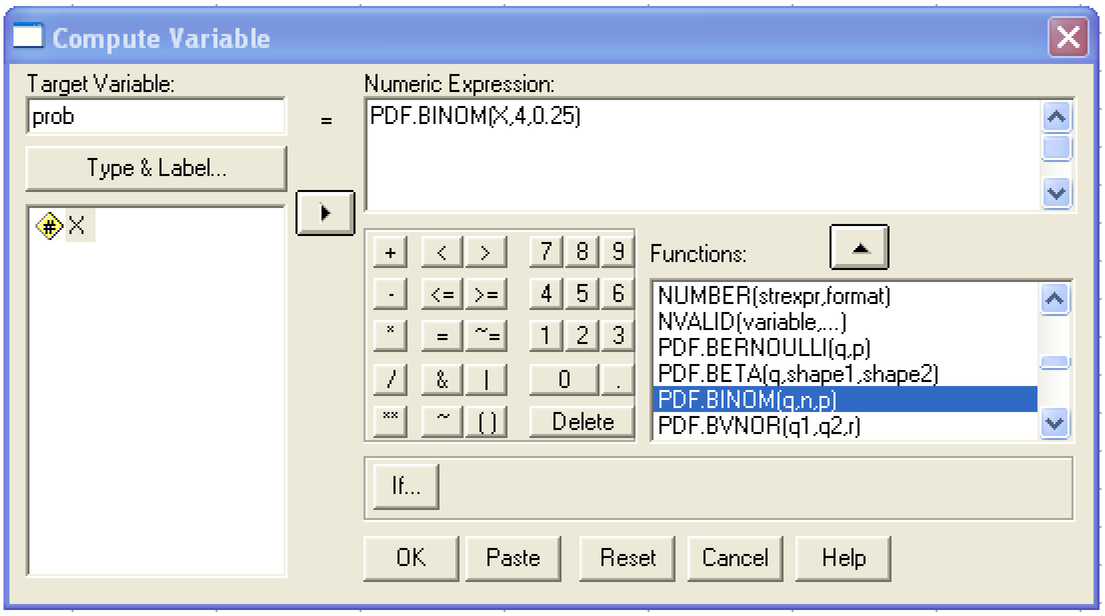
Here we illustrate how to use SPSS to calculate probabilities from a binomial distribution. In the text on page 103, we suppose that four adults have been randomly selected and asked if they currently smoke. Assuming we know the prevalence of smoking in the population to be 25%, the probability of getting a "Yes" response is 25% for every individual. To do this in SPSS, create a variable X with the values 0, 1, 2, 3, and 4 as shown below.

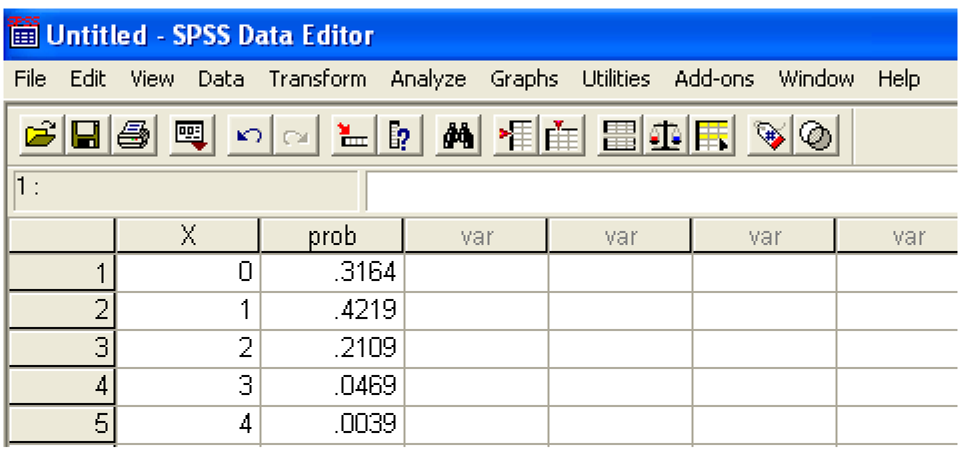


Then use the SPSS procedure **Transform** -> **Compute…** to obtain the **Compute Variable** window below.

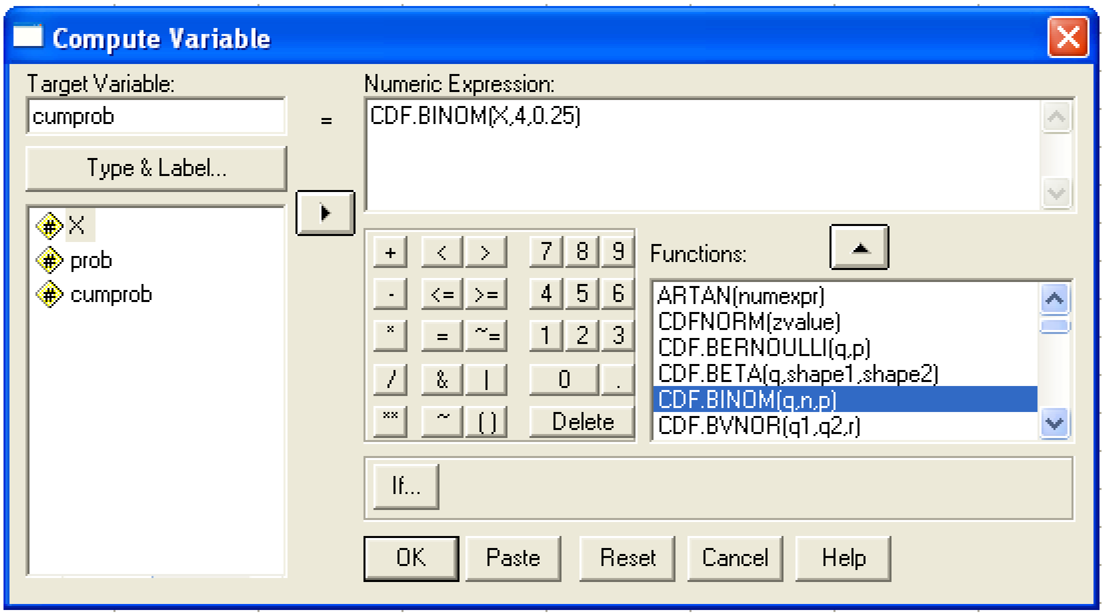


The SPSS function **PDF.BINOM(q,n,p)** refers to the probability mass function for a binomial distribution. The value, **q**, is the number of successes, **n** is the number of trials, and **p** is the probability of success in each trial. Notice that we use the values of the variable X to represent to the number of successes. Because there are four individuals, the number of trials, **n**, is equal to 4. Finally the probability of success, **p**, is 0.25 or 25%.

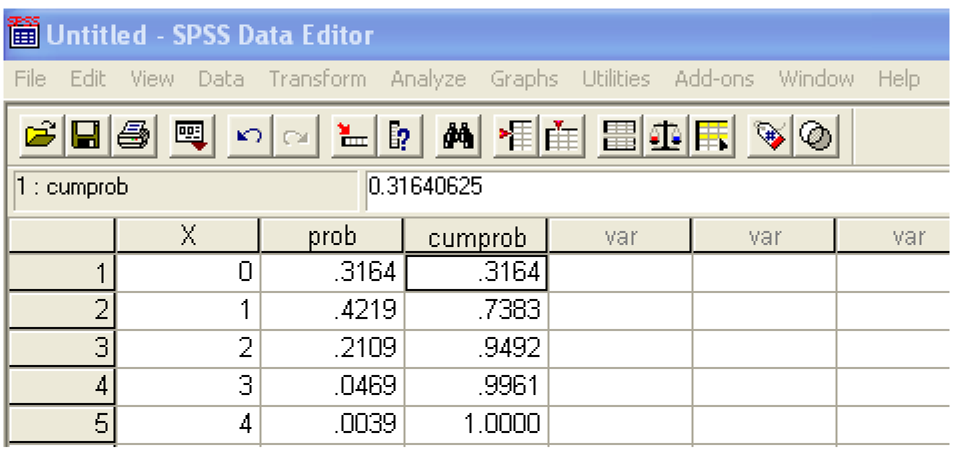




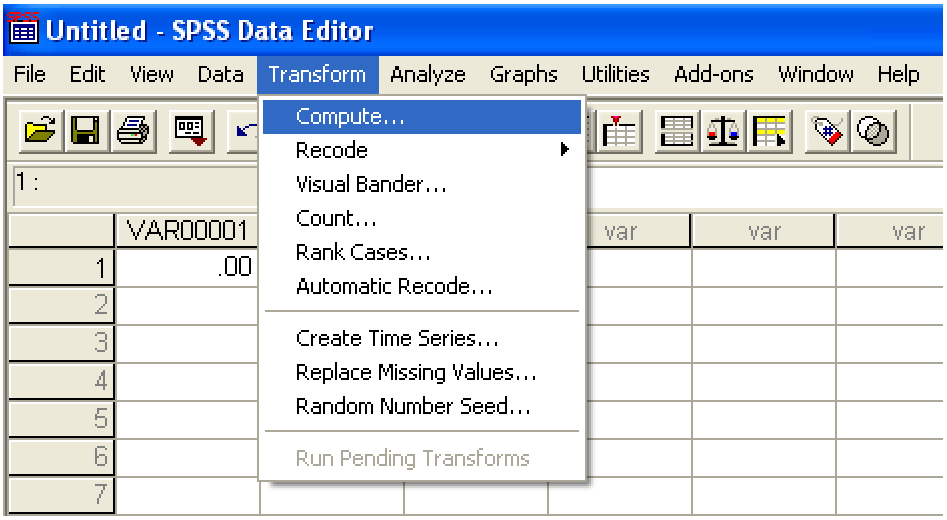
To find values for the cumulative distribution function for the binomial distribution with n = 4 and = 0.25, use the SPSS procedure **Transform** -> **Compute...** to obtain the **Compute Variable** window as was discuss before. However, the SPSS function **CDF.BINOM(q,n,p)** is used to obtain values from the cumulative distribution function. Like before, the value, **q**, is the number of successes, **n** is the number of trials, and **p** is the probability of success in each trial.

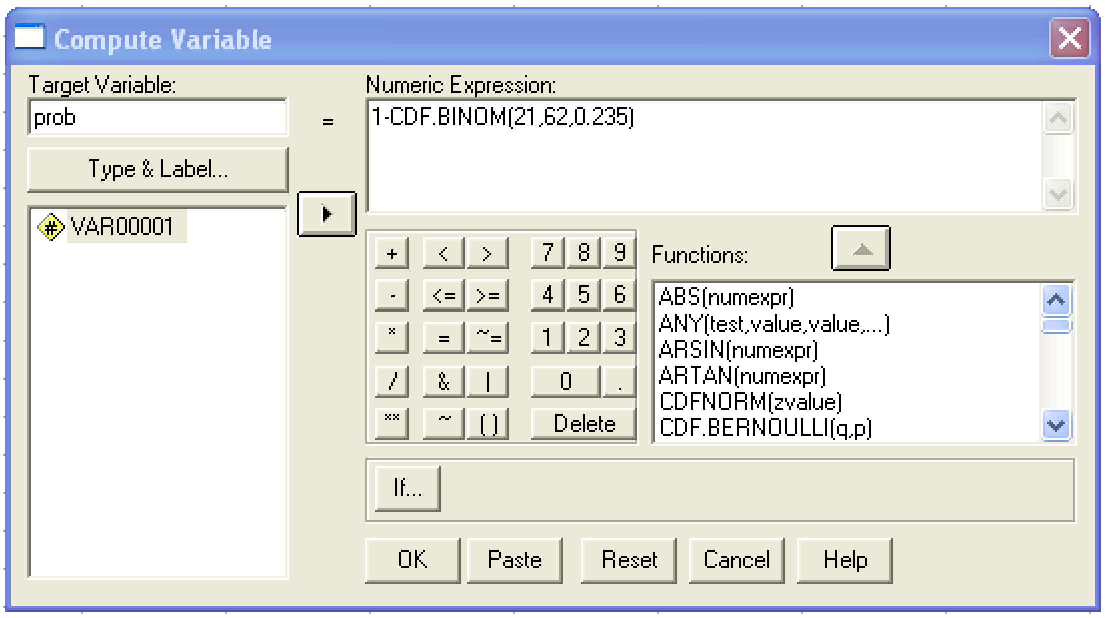


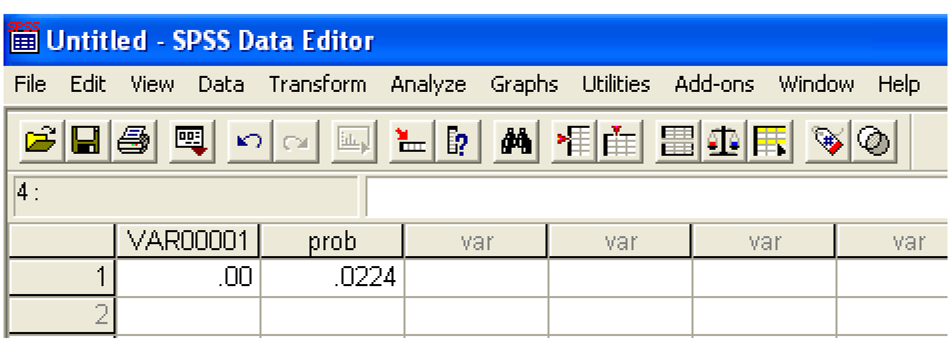
The values below correspond to Table 5.2 on page 107 in the textbook.



In Example 5.1, we calculate the probability of observing 22 or more cesarean sections. Recall that of all deliveries in the United States, c-sections represented 23.5%. The probability of 22 or more c-sections can be expressed as $\Pr\{X \geq 22\}$ or as $1 - \Pr\{X \leq 21\}$.

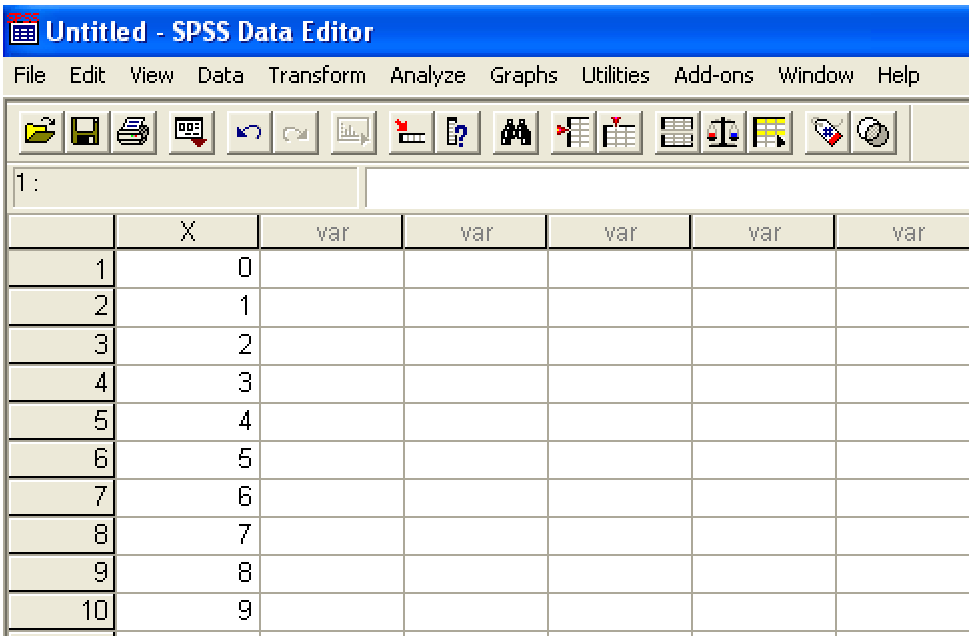




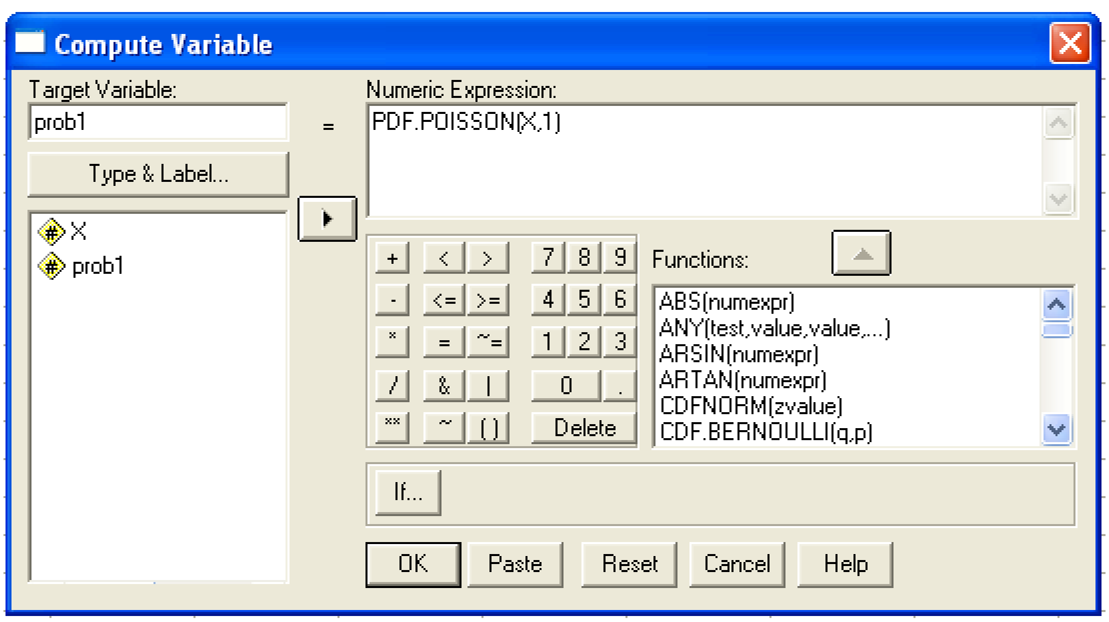


2. Finding Poisson probabilities:

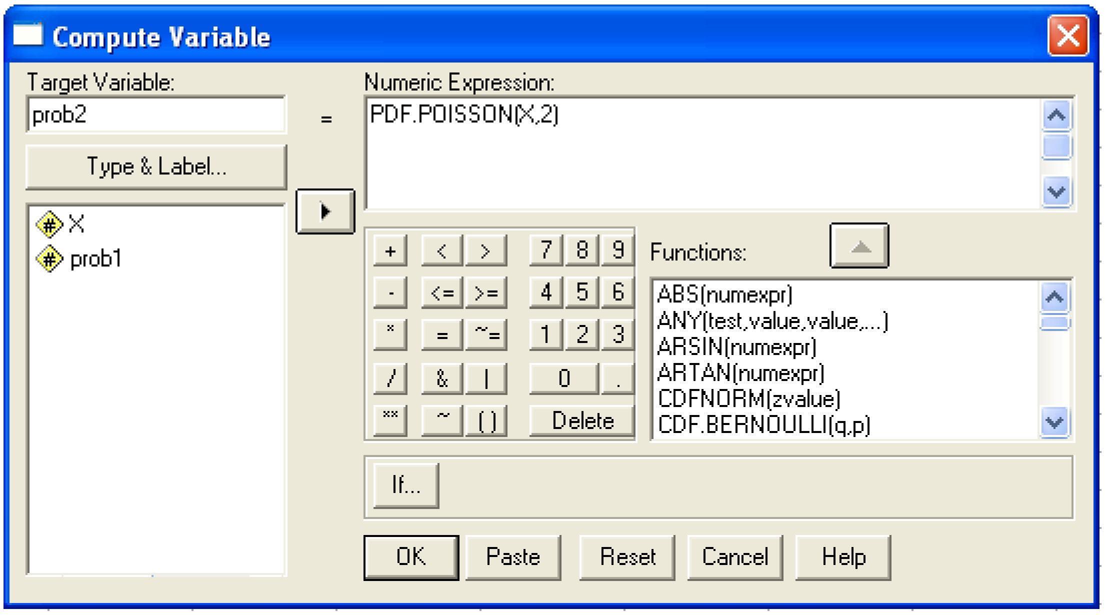
Here we show how SPSS can be used to calculate probabilities from a Poisson distribution. We use the values of the variable X along with the SPSS function **PDF.POISSON(q,mean)** which refers to the probability mass function for a Poisson distribution. The value, **q**, represents the number of occurrences, and the **mean** is just the mean of the Poisson distribution. Here we will reproduce the values in Table 5.3 on page 112 of the textbook.



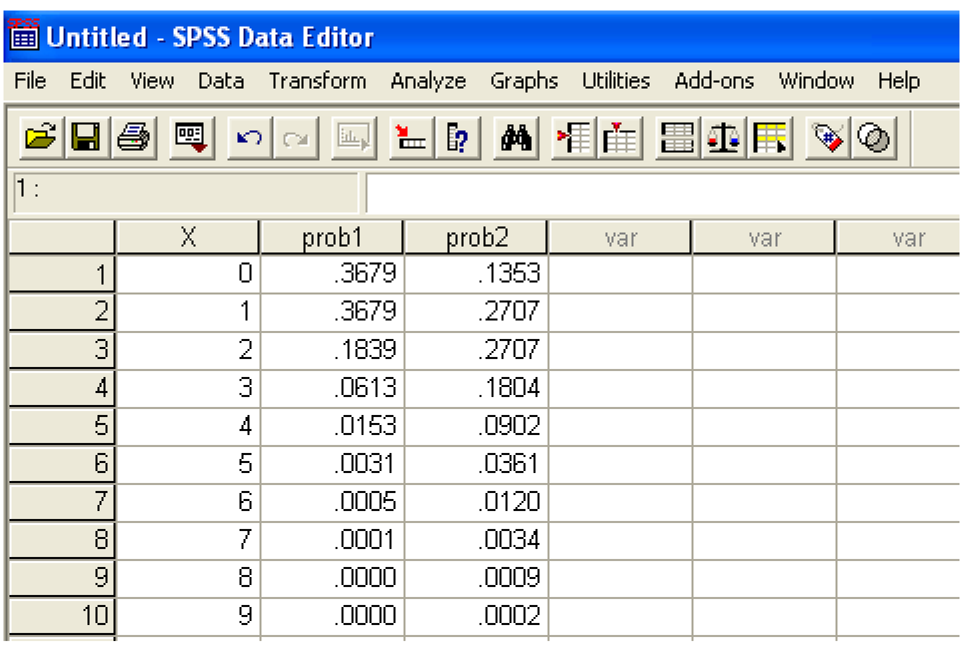
Using the SPSS procedure **Transform** -> **Compute…** to obtain the **Compute Variable** window below. We are using the variable prob1 to store Poisson probabilities with a mean equal to 1.



We will use the same strategy to store Poisson probabilities with a mean equal to 2 under the variable prob2.



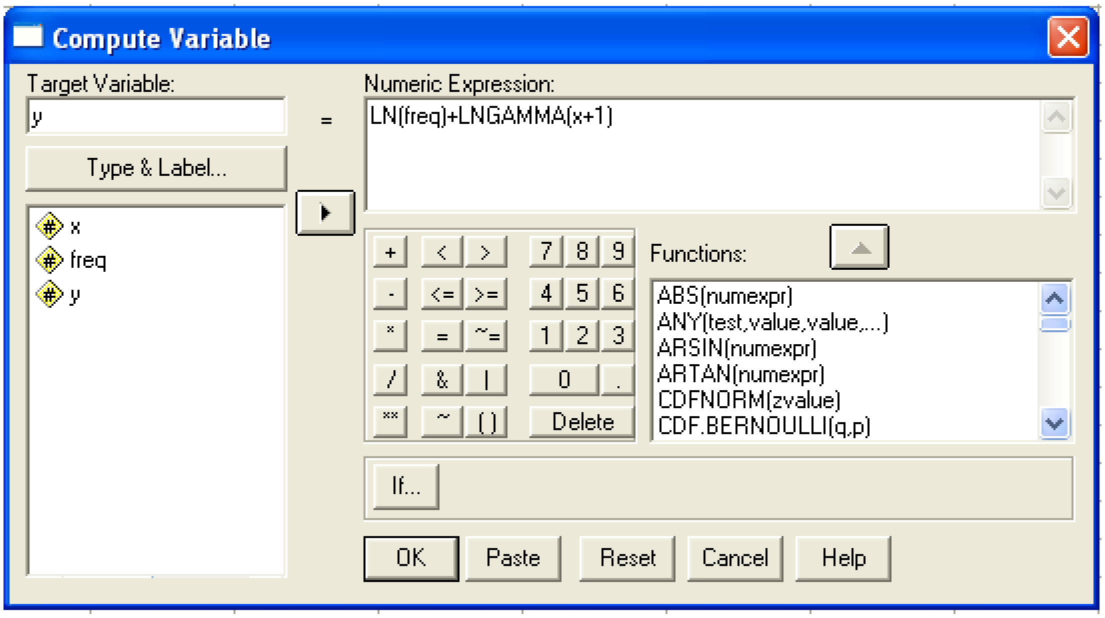
After clicking **OK**, the probabilities are displayed in the **SPSS Data Editor** as shown below.

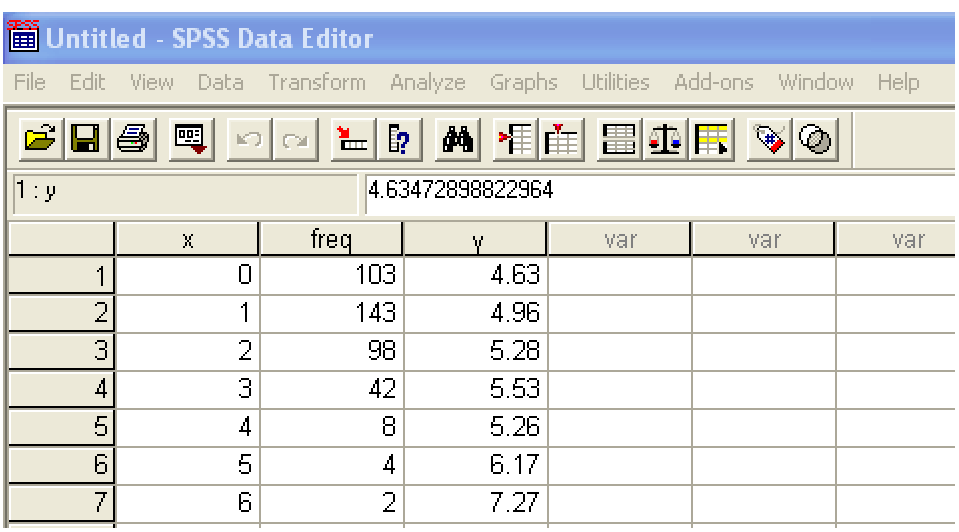


	X	prob1	prob2	var	var	var
1	0	.3679	.1353			
2	1	.3679	.2707			
3	2	.1839	.2707			
4	3	.0613	.1804			
5	4	.0153	.0902			
6	5	.0031	.0361			
7	6	.0005	.0120			
8	7	.0001	.0034			
9	8	.0000	.0009			
10	9	.0000	.0002			

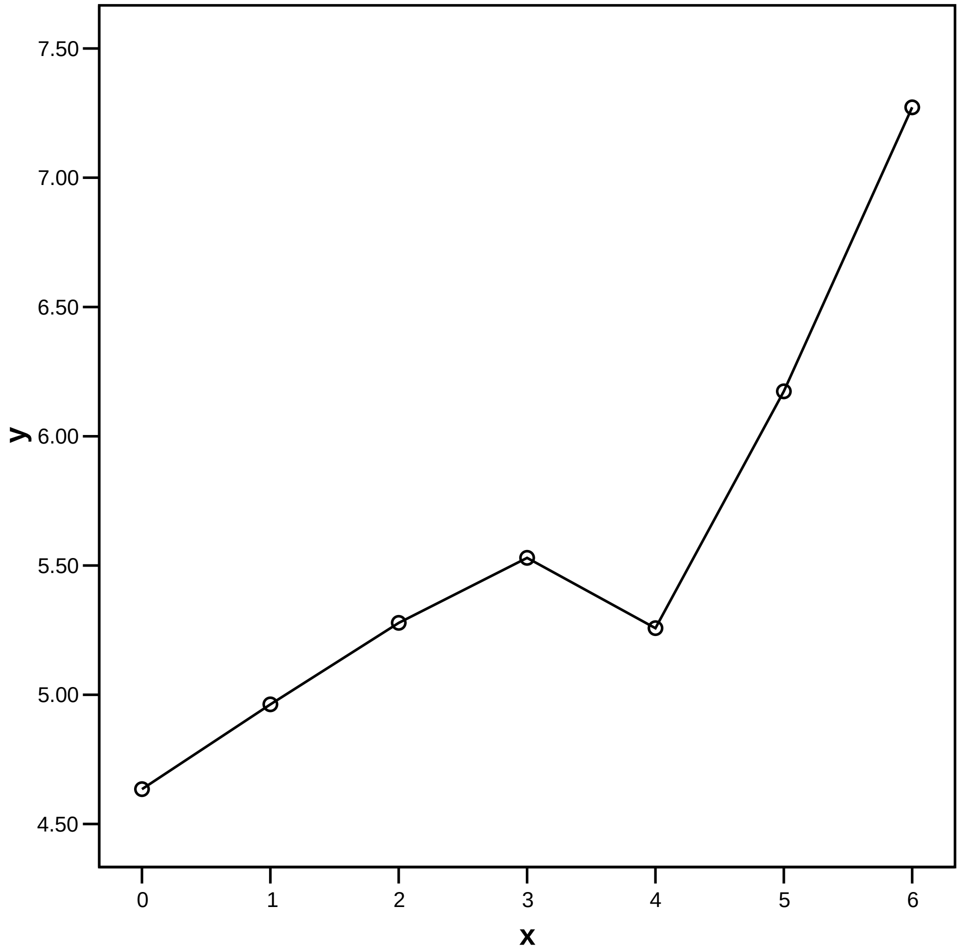
Program Note 5.2 – Creating a Poissonness plot

We can create a Poissonness plot using the frequency values provided in Table 5.4. Basically we are plotting: $\{\ln(\text{freq}(x)) + \ln(x!)\}$ versus x. The SPSS function **LN(numexpr)** and **LNGAMMA(numexpr)** are required to conduct the necessary calculations. Note that **GAMMA(x+1) = x!**, therefore **LNGAMMA(x+1) = LN(x!)**.





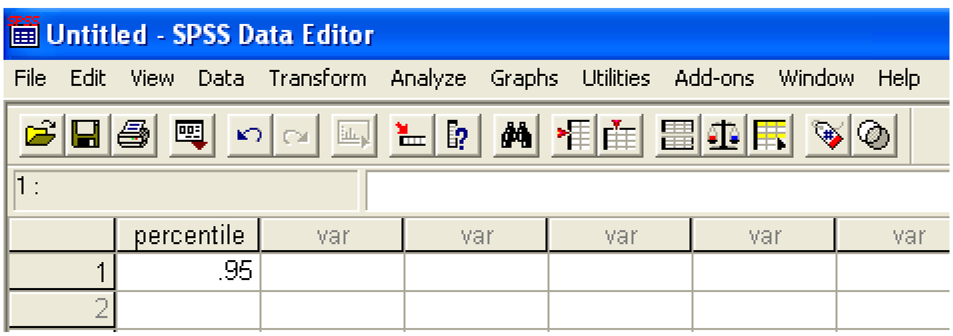
Now we make a scatter plot of y versus x.


7.50
7.00
6.50
6.00
5.50
5.00
4.50
y
0
1
2
3
4
5
6
x

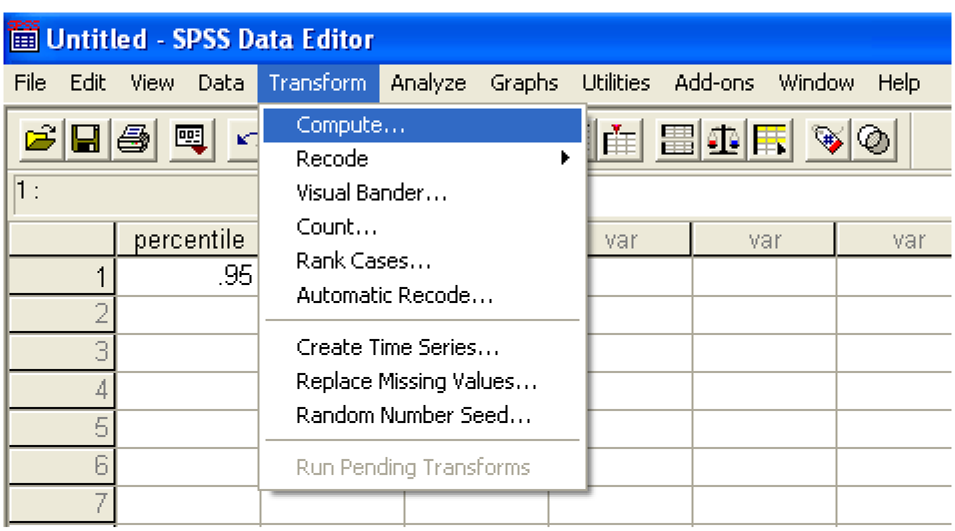
Program Note 5.3 – Finding normal probabilities

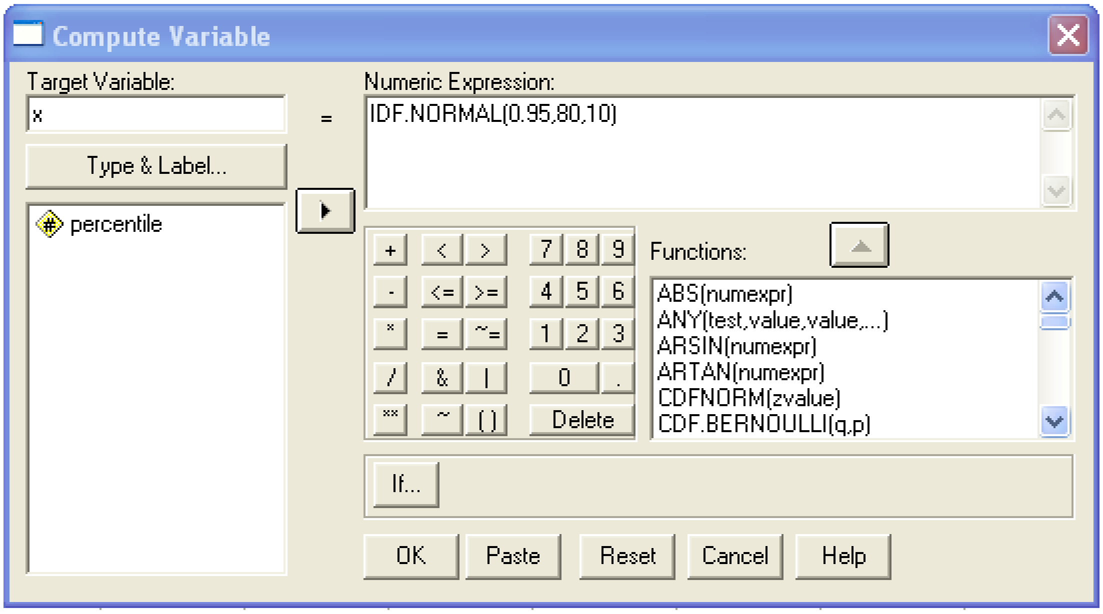
The SPSS function **PDF.NORMAL(q,mean,stddev)** evaluates the normal distribution's probability density function at specified values of **q**. For this function, **q** is the value of the random variable, **mean** is the normal distribution's mean, and **stddev** is the normal distribution's standard deviation. We can also use SPSS to calculate lower tale probabilities using the cumulative distribution function with the SPSS function **CDF.NORMAL(q,mean,stddev)** where **q** is the value of the random variable, **mean** is the normal distribution's mean, and **stddev** is the normal distribution's standard deviation. The SPSS function **IDF.NORMAL(p,mean,stddev)** returns the **p** percentile from a normal distribution with mean equal to **mean** and standard deviation equal to **stddev**.

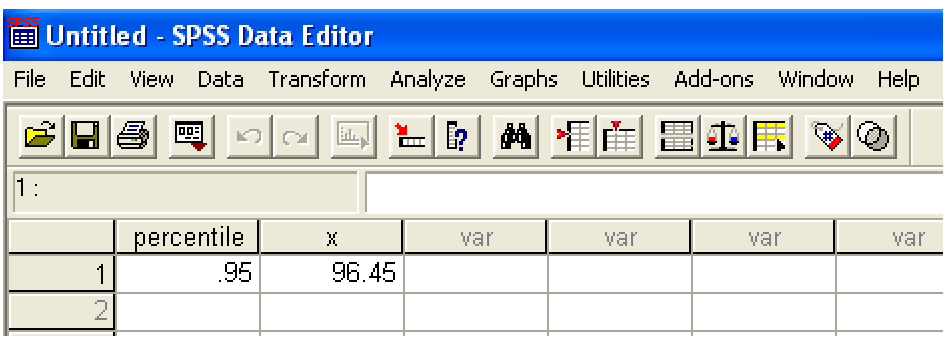
In Example 5.4, we compute the 95th percentile of diastolic blood pressures for adult women. The diastolic blood pressures follow a normal distribution with mean equal to 80mmHg and standard deviation equal to 10mmHg. We begin by creating the variable percentile with the value 0.95.



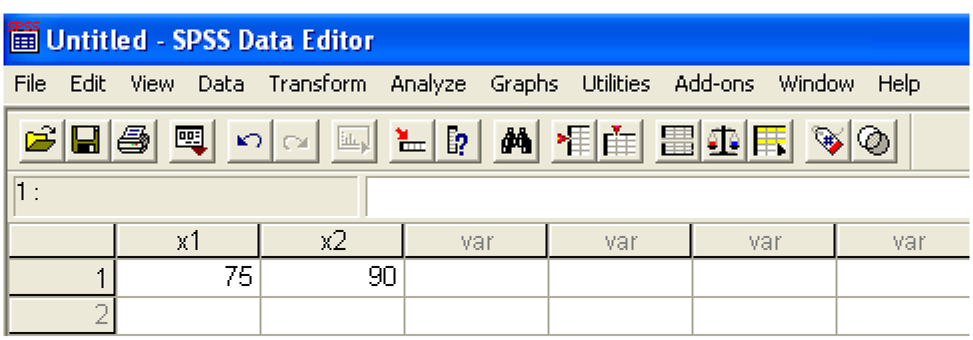
We use the SPSS procedure **Transform** -> **Compute...** to calculate the value of the 95th percentile.

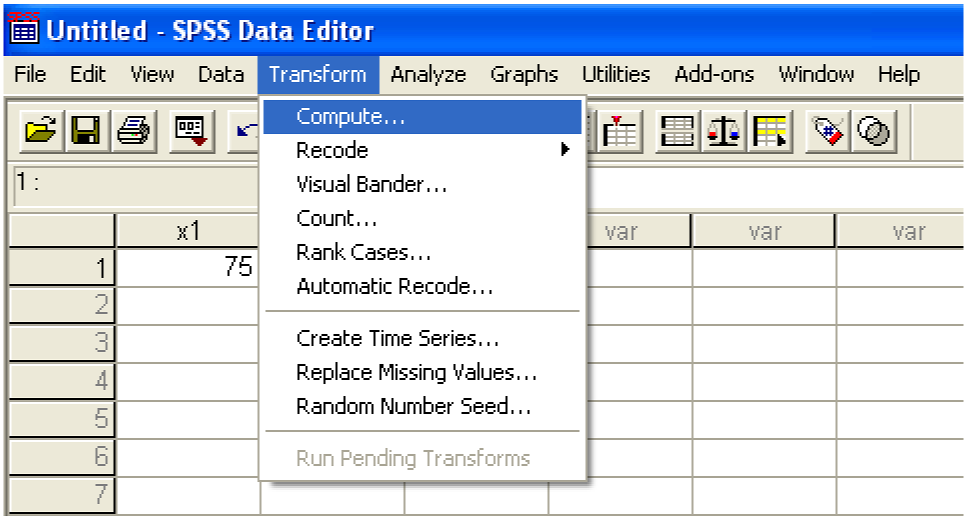


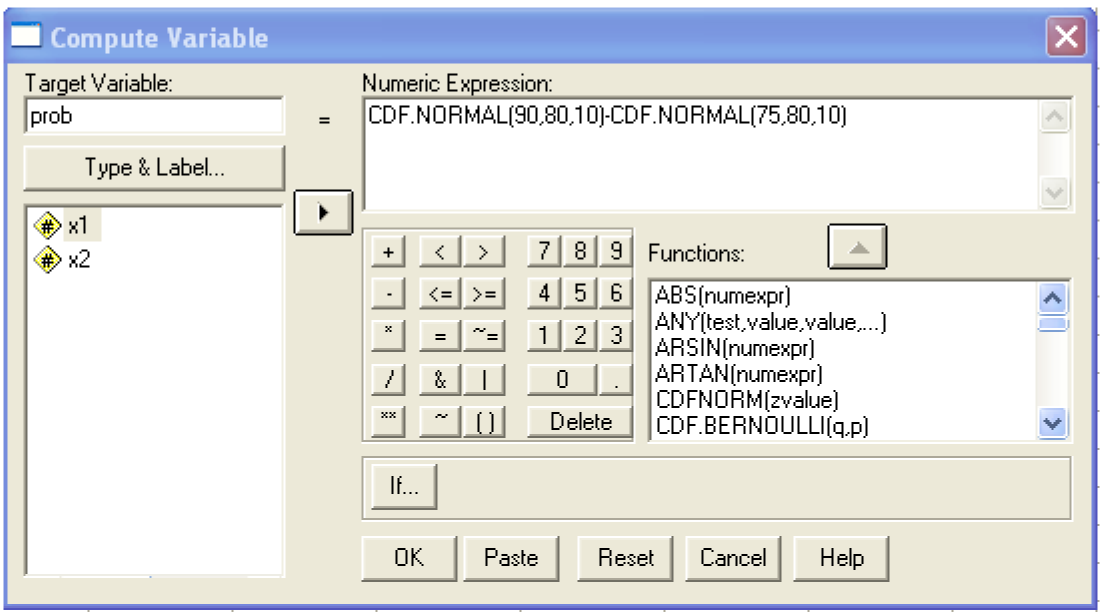


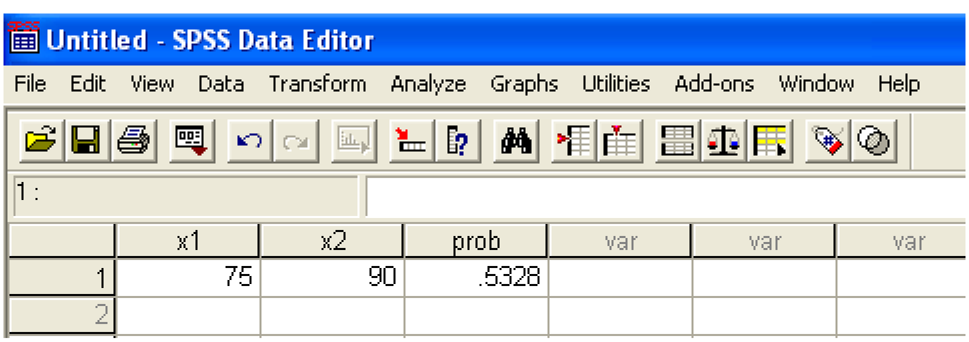


In Example 5.5, we calculate the probability of having diastolic blood pressures between 75 and 90mmHg using the SPSS procedure below.



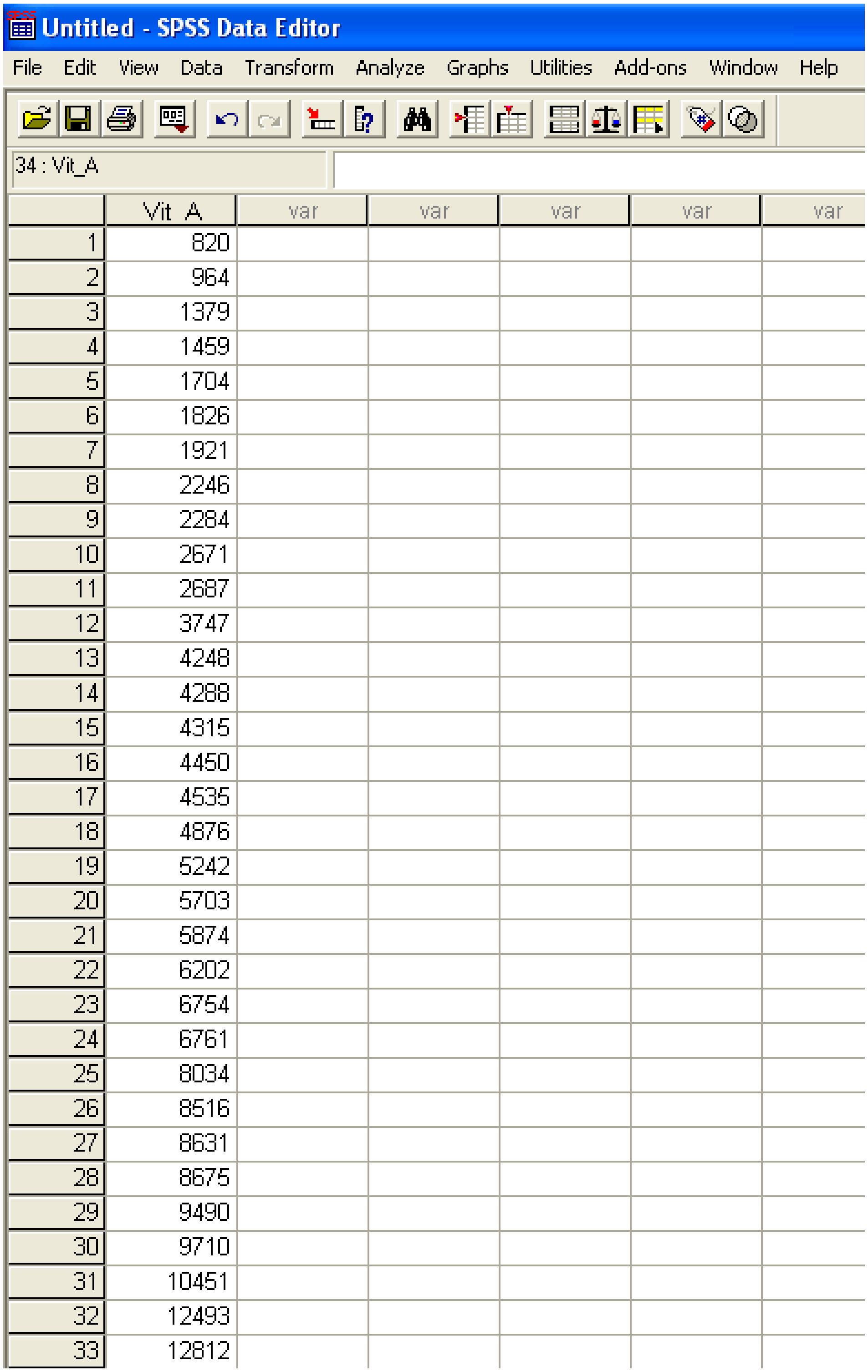

Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
Compute...
Recode
Visual Bander...
Count...
Rank Cases...
Automatic Recode...
Create Time Series...
Replace Missing Values...
Random Number Seed...
Run Pending Transforms
x1
75
var var var


Compute Variable
Target Variable:
prob
Numeric Expression:
CDF.NORMAL(90,80,10)-CDF.NORMAL(75,80,10)
Type & Label...
x1
x2
Functions:
ABS(numexpr)
ANY(test,value,value,...)
ARSIN(numexpr)
ARTAN(numexpr)
CDFNORM(zvalue)
CDF.BERNOULLI(q,p)
Delete
If...
OK Paste Reset Cancel Help


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
x1 x2 prob var var var
1 75 90 .5328

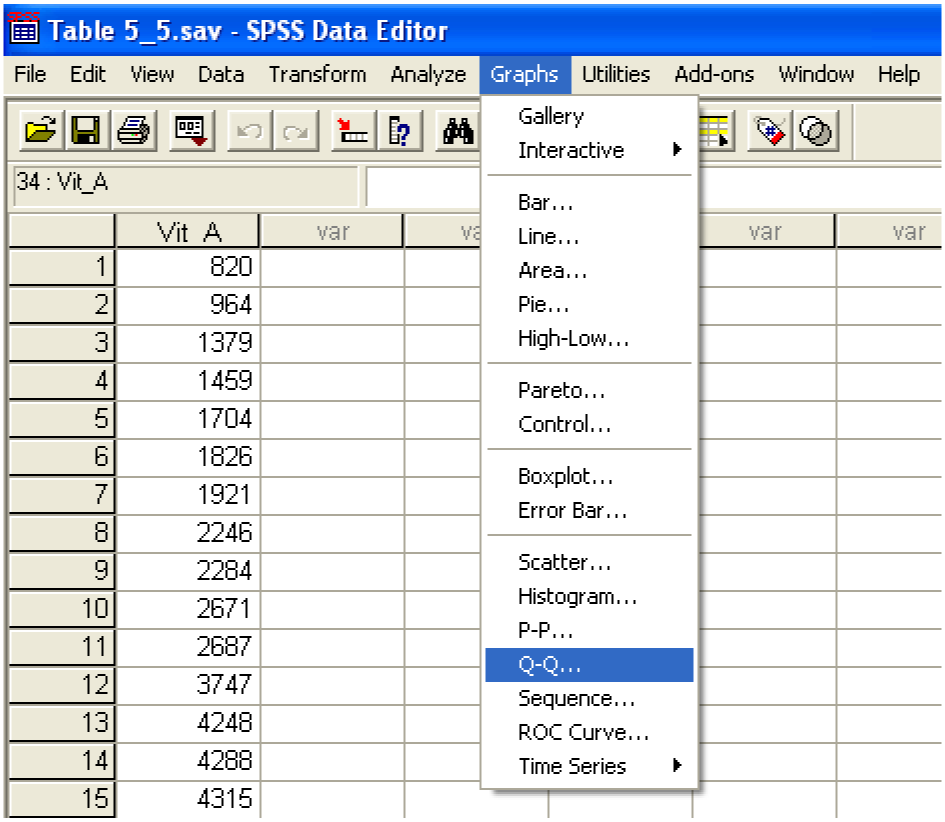
Program Note 5.4 – **Creating a normal probability plot**

Below is the data from Table 5.5 on vitamin A values for 33 boys. Before going on, use the SPSS procedure **File** -> **Save As…** to save the SPSS dataset as **Table 5_5**.

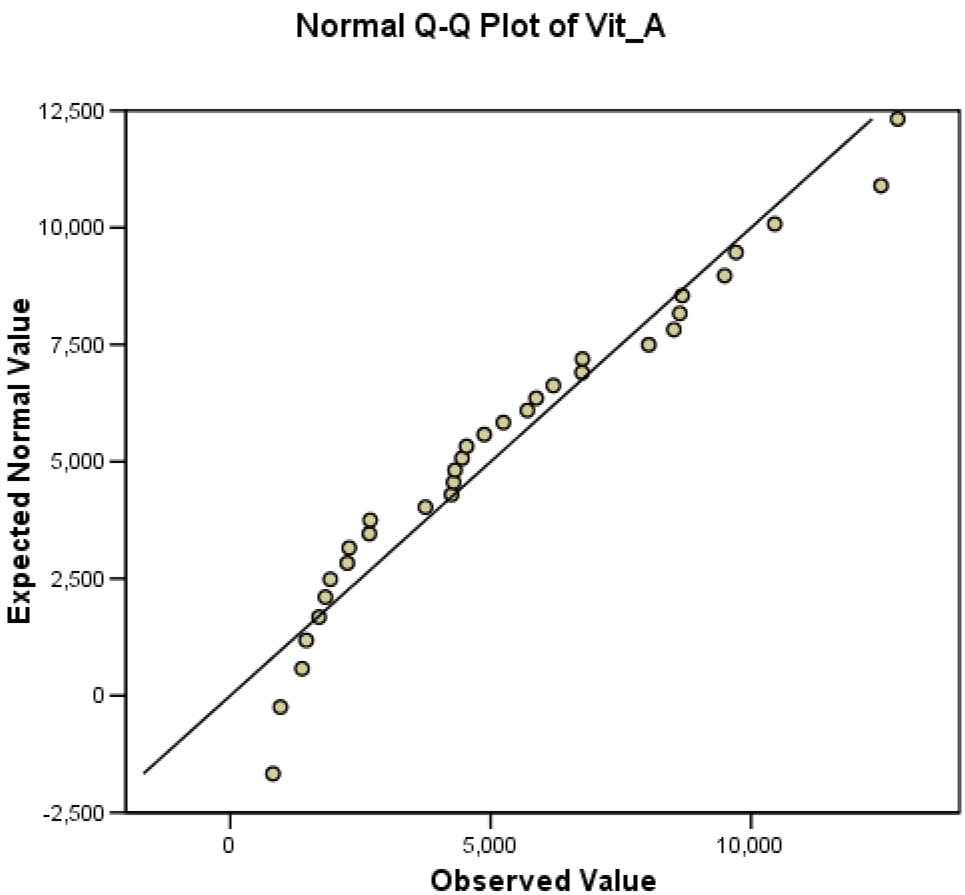


	Vit_A	var	var	var	var	var
1	820					
2	964					
3	1379					
4	1459					
5	1704					
6	1826					
7	1921					
8	2246					
9	2284					
10	2671					
11	2687					
12	3747					
13	4248					
14	4288					
15	4315					
16	4450					
17	4535					
18	4876					
19	5242					
20	5703					
21	5874					
22	6202					
23	6754					
24	6761					
25	8034					
26	8516					
27	8631					
28	8675					
29	9490					
30	9710					
31	10451					
32	12493					
33	12812					

To create a normal probability plot similar to the one in Figure 5.11 use the SPSS procedure **Graphs** -> **Q-Q…** as shown below:



The graph produced by SPSS is shown below:


Normal Q-Q Plot of Vit_A
Expected Normal Value
Observed Value
12,500
10,000
7,500
5,000
2,500
0
-2,500
0
5,000
10,000