



## Confidence Intervals in SPSS

This document shows how to construct confidence intervals in SPSS. The confidence intervals for mean of distribution can be found very easily in this software. However there are no built-in functions to find confidence intervals for proportions and hence we show how to use SPSS as a calculator that uses appropriate formula. For this document we need ‘Skeleton’ and ‘Age change’ data sets. It is assumed that you have managed to upload all these data into SPSS (please refer to ‘Data sets import in SPSS’ document for detailed explanation).

### Introduction to confidence intervals

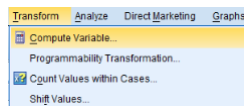
In this section we introduce a very important discrete distribution which is Binomial. Suppose we are flipping a coin 10 times and that probability of a head is 50% (fair coin) then binomial distribution gives the probabilities of getting  $x$  heads. Clearly in this case the possible values for  $x$  is 0,1,2 up to 10. Let’s make a plot of probabilities. First open a new data file in SPSS, then go to ‘Variable View’ section to introduce new variables. We construct two variables, one is the number of successes (‘X’) and second is the probability that we need to find (‘Prob’):

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	X	Numeric	8	2	Number of Heads	None	None	8	Right	Scale	Input
2	Prob	Numeric	8	2	Probability	None	None	8	Right	Scale	Input
3											

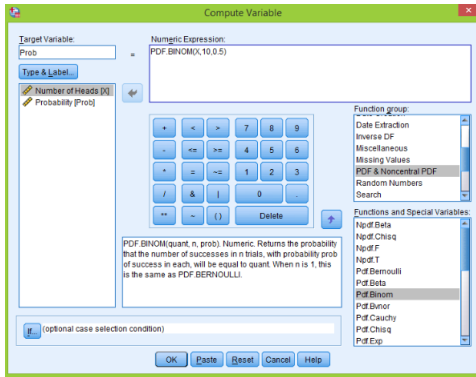
Now go to the ‘Data View’ section to enter values into our variables. We enter 0,1,2 up to 10 into the first variable.

	X	Prob	var
1	0	.	.
2	1.00	.	.
3	2.00	.	.
4	3.00	.	.
5	4.00	.	.
6	5.00	.	.
7	6.00	.	.
8	7.00	.	.
9	8.00	.	.
10	9.00	.	.
11	10.00	.	.
12			

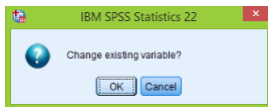
The goal is of course to find probability of the binomial distribution with number of trials equals to 10 and probability of success 0.5. We clearly need a calculator here and therefore go to **Transform > Compute Variable:**



The target variable here is the ‘Prob’ variable and ‘PDF.BINOM’ computes the probabilities that we need:



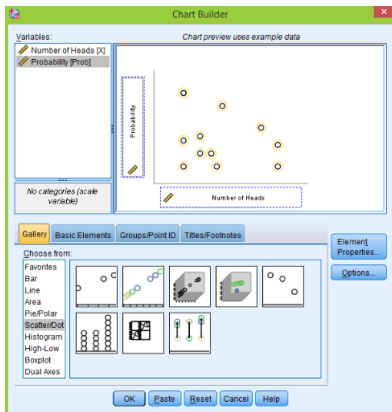
Click **OK** and the following window appears:



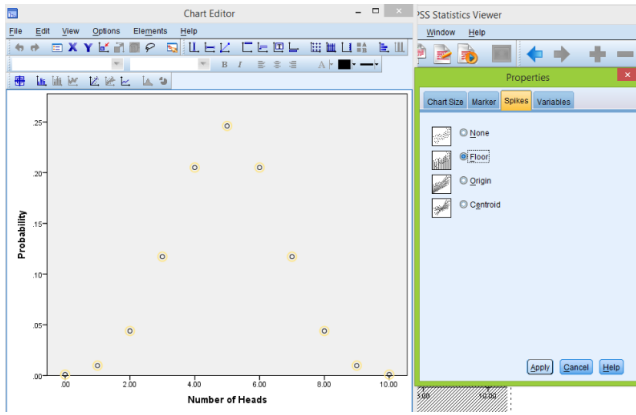
Click **OK** again and a column of binomial probabilities is created:

X	Prob
.00	.00
1.00	.01
2.00	.04
3.00	.12
4.00	.21
5.00	.25
6.00	.21
7.00	.12
8.00	.04
9.00	.01
10.00	.00

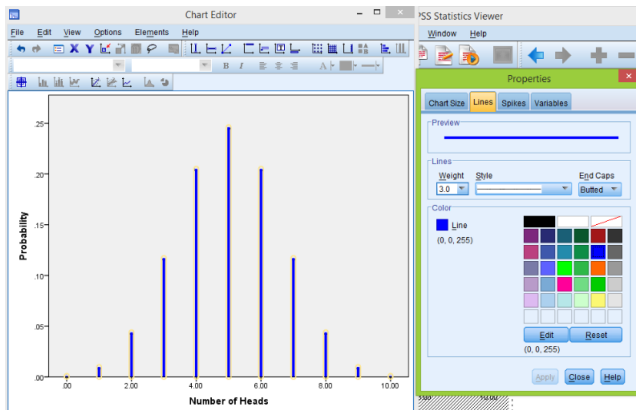
To make a visual representation of these probabilities go to **Graphs > Chart Builder > Scatter/Dot > double click on Simple Scatter** then drag 'X' variable to horizontal and 'Prob' variable to vertical axis:



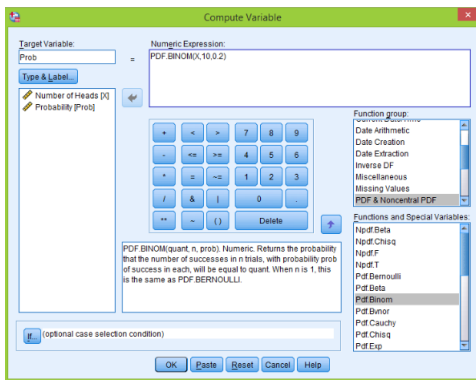
Click **OK** to finish. To improve the plot double click on it to open 'Chart Editor'. Then double click on any dot in this chart to open 'Properties' window. Now click on 'Spikes' button and select 'Floor':



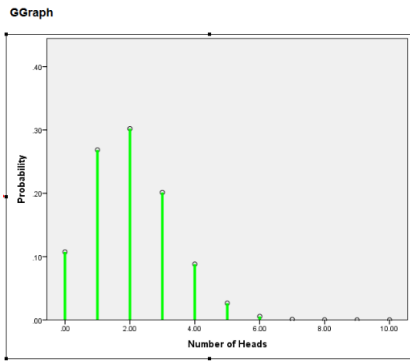
Click **Apply** and vertical bars are added to this plot. To change characteristic of these bars, double click on any bar and then select 'Lines' in the 'Properties' window. Here we can change color and style of these vertical bars.



Note, the distribution of this binomial is symmetric and unimodal. Now let's make a similar plot of probabilities but with probability of success is equal to 0.2: As before go to **Transform > Compute Variable** and enter the following expression (the only change here is that instead of 0.5 we enter 0.2):



Finally exactly as above we construct a plot for this binomial distribution:



This distribution is no longer symmetric but shifted to the right.

## Confidence intervals for proportions

In this section we show how to find confidence intervals for one proportion. As said in the introduction there is no convenient way to find these intervals in SPSS. Hence we will use a calculator to get the confidence intervals.

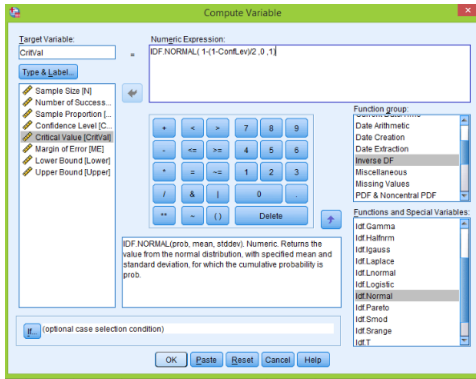
First we consider the 'Bottle cap flipping' example. Here 1000 flips were done and we get Red 576 times. Based on this information we want to get 95% confidence interval for the true proportion. First open a new data file and click on the 'Variable View' section. Then introduce the next variables:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	N	Numeric	8	0	Sample Size	None	None	8	Right	Scale	Input
2	NumSucc	Numeric	8	0	Number of Successes	None	None	8	Right	Scale	Input
3	Phat	Numeric	8	4	Sample Proportion	None	None	8	Right	Scale	Input
4	ConfLev	Numeric	8	4	Confidence Level	None	None	8	Right	Scale	Input
5	CritVal	Numeric	8	4	Critical Value	None	None	8	Right	Scale	Input
6	ME	Numeric	8	4	Margin of Error	None	None	8	Right	Scale	Input
7	Lower	Numeric	8	4	Lower Bound	None	None	8	Right	Scale	Input
8	Upper	Numeric	8	4	Upper Bound	None	None	8	Right	Scale	Input
9											

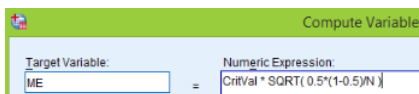
Most of the table is filled automatically once you enter name of a variable. We only change 'Decimals', in the 'Measure' column select 'Scale' and type labels for clarity. Next go to 'Data View' section and enter the information from the above example:

1: N	N	NumSucc	Phat	ConfLev	CritVal	ME	Lower	Upper
1	1000	576	.5760	.9500				

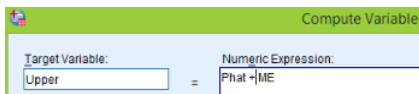
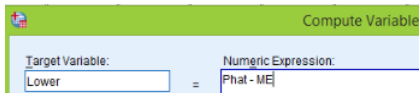
Note that we use SPSS in an unusual way, since usually each row corresponds to one independent observation while now we have only one row of statistics. Next we will use standard formulas to compute the confidence interval. First we find the critical value which is used in the construction of confidence intervals. Go to **Transform > Compute Variable**, and enter the following expression:



'IDF.NORMAL(p,0,1)' is defined as the number, such that  $p \cdot 100\%$  of  $\mathcal{N}(0,1)$  is below it. Also note that we use confidence level in this calculation. Once we find critical value we can calculate the margin of error:



At this stage we have sample proportion and the marginal of error and therefore we can easily get lower and upper bounds for our confidence interval:



Now we look at our variables and see the result:

	N	NumSucc	Phat	ConfLev	CritVal	ME	Lower	Upper
1	1000	576	.5760	.9500	1.9600	.0310	.5450	.6070
2								

So we can say that we are 95% confident that the true proportion lies between 0.5450 and 0.6070. Next we want to calculate confidence intervals for the same problem but with 90% and 99% confidence levels. We enter these data into our variables:

	N	NumSucc	Phat	ConfLev	CritVal	ME	Lower	Upper
1	1000	576	.5760	.9500	1.9600	.0310	.5450	.6070
2	1000	576	.5760	.9000				
3	1000	576	.5760	.9900				
4								

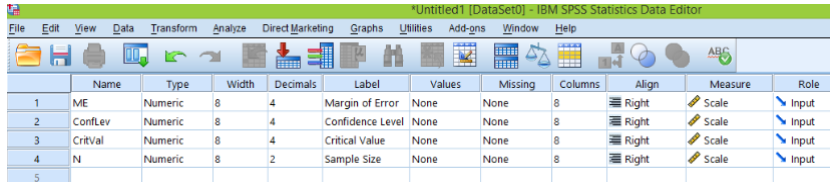
Next we repeat the above procedure and get:

	N	NumSucc	Phat	ConfLev	CritVal	ME	Lower	Upper
1	1000	576	.5760	.9500	1.9600	.0310	.5450	.6070
2	1000	576	.5760	.9000	1.6449	.0260	.5500	.6020
3	1000	576	.5760	.9900	2.5758	.0407	.5353	.6167

Note that 95% confidence interval (CI) is shorter than 99% CI but wider than 90% CI. Hence the more confident we want to be, the less precise the confidence interval is.

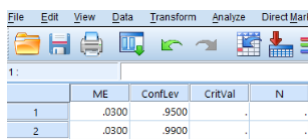
## Sample size for estimating a proportion

In this section we investigate how to estimate the sample size based on the margin of error and confidence level. We should estimate the sample size before an experiment or survey. Suppose we are seeking 95% confidence interval and we want the margin of error to be 3%. Based on these conditions, the goal is to estimate the sample size. We use the same strategy as in the previous section. First we construct new variables that will be used in this example:



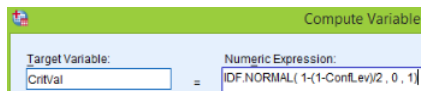
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ME	Numeric	8	4	Margin of Error	None	None	8	Right	Scale	Input
2	ConfLev	Numeric	8	4	Confidence Level	None	None	8	Right	Scale	Input
3	CritVal	Numeric	8	4	Critical Value	None	None	8	Right	Scale	Input
4	N	Numeric	8	2	Sample Size	None	None	8	Right	Scale	Input
5											

Next in the 'Data View' section we fill part of the table with 95% and 99% confidence level:



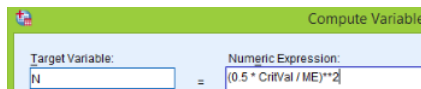
	ME	ConfLev	CritVal	N
1	.0300	.9500		
2	.0300	.9900		

We need to compute the critical value and finally the sample size using formulas from video lectures. Go to **Transform > Compute Variable** and enter the next expression and click **OK**:



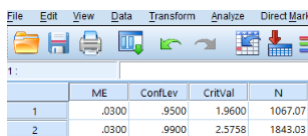
Target Variable: CritVal = Numeric Expression: IDF.NORMAL(1-(1-ConfLev)/2, 0, 1)

As you see the formula is exactly the same that we have used in the last section of this document. Finally we calculate the sample size with appropriate formula:



Target Variable: N = Numeric Expression: (0.5 \* CritVal / ME)\*\*2

Here \*\*2 means power of two. Click **OK** and we get the results:



	ME	ConfLev	CritVal	N
1	.0300	.9500	1.9600	1067.07
2	.0300	.9900	2.5758	1843.03

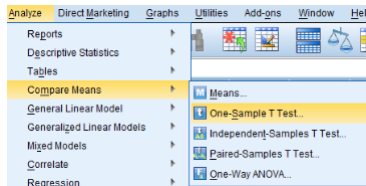
Hence the sample size that we will need for 95% confidence level is around 1067 and for 99% is approximately 1843. So we see that the more confident we want to be, the larger the sample size is (with the same margin of error).

## Confidence Intervals for Means

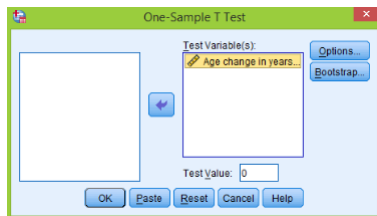
We start this section with the 'Age change' data set. This data set includes 60 observations of people who made a plastic surgery. For each subject it is estimated by how many years a person looks younger than before the surgery. The first six observations are displayed below:

	AgeChange	var
1	2.8	
2	9.1	
3	6.4	
4	4.7	
5	8.9	
6	11.5	

The goal here is to produce a 95% confidence interval for the average age change. Since we do not know the true variance of the population we have to use student-t distribution. Go to **Analyze > Compare Means > One-Sample T Test**



Then move the 'Age change' variable across using the arrow:



Click **OK** and the next table appears in the Output window:

**One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
Age change in years	60	7.177	2.9481	.3806

**One-Sample Test**

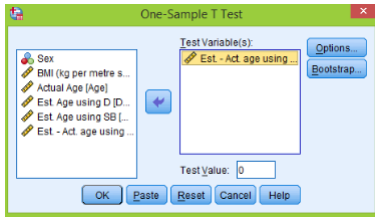
	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Age change in years	18.856	59	.000	7.1767	6.415	7.938

This table shows that we are 95% confident that on average a person looks from 6.4 to 7.9 years younger after a plastic surgery.

Next let's move to the 'Skeleton' data set:

	Sex	BMIcat	BMIquant	Age	DGestimate	DGerror	SBestimate	SBerror
1	2	underweight	15.66	78	44	-34	60	-18
2	1	normal	23.03	44	32	-12	35	-9
3	1	overweight	27.92	72	32	-40	61	-11
4	1	overweight	27.83	59	44	-15	61	2
5	1	normal	21.41	60	32	-28	46	-14

Here we need 95% CI for the DGerror variable. As before go to **Analyze > Compare Means > One-Sample T Test** and send 'DGerror' variable to the right window:



Click **OK**

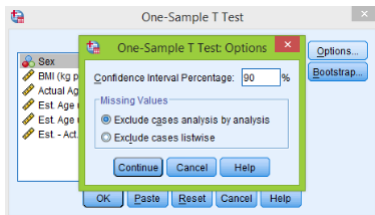
**One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
Est. - Act. age using D (years)	400	-14.15	14.126	.706

**One-Sample Test**

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Est. - Act. age using D (years)	-20.034	399	.000	-14.150	-15.54	-12.76

Hence we are 95% sure that the true mean of DGerror is between  $-15.53$  and  $-12.76$ . If we need for example 90% confidence interval then in the 'One-Sample T Test' window click on the 'Options' button and change 95 to 90%:



Click **Continue** > **OK** and 90% confidence interval is produced:

**One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
Est. - Act. age using D (years)	400	-14.15	14.126	.706

**One-Sample Test**

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
Est. - Act. age using D (years)	-20.034	399	.000	-14.150	-15.31	-12.99