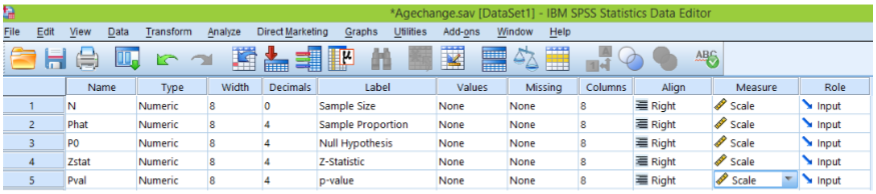
The Process of Statistical Tests in SPSS

This document shows how to perform statistical testing in SPSS. We will show how to find p-values for proportions and means of distributions. There are no built-in functions for proportions in SPSS, and therefore we will use a calculator with appropriate formula to get p-values. For this document we need ‘Skeleton’, ‘Age change’ and ‘Temperature’ 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).

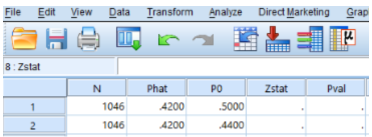
Hypothesis Testing for Proportions

We start this section with the Mayor support example. Remember that a survey of 1046 people was conducted and the sample proportion of support for Rob Ford mayor was 0.42. The goal is test whether the true proportion of the population is 0.5 (null hypothesis) versus that it is less than 0.5 (alternative). To make such a statistical test, open a new data file and in the ’Variable View’ section we introduce some variables (similar to confidence intervals and sample size calculation) :



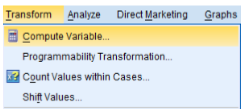
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	N	Numeric	8	0	Sample Size	None	None	8	Right	Scale	Input
2	Phat	Numeric	8	4	Sample Proportion	None	None	8	Right	Scale	Input
3	P0	Numeric	8	4	Null Hypothesis	None	None	8	Right	Scale	Input
4	Zstat	Numeric	8	4	Z-Statistic	None	None	8	Right	Scale	Input
5	Pval	Numeric	8	4	p-value	None	None	8	Right	Scale	Input

Make sure that the ’Measure’ for each variable is ‘Scale’. Now in the ‘Data View’ we enter statistics of this problem:

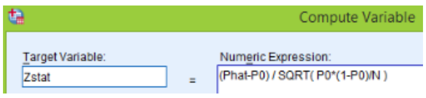


	N	Phat	P0	Zstat	Pval
1	1046	.4200	.5000	.	.
2	1046	.4200	.4400	.	.

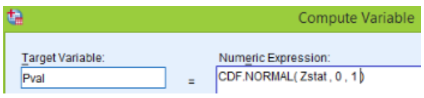
We will do two tests with null hypothesis 0.5 and 0.44 simultaneously to save some time. Next we will use formulas from the video lectures to find z-statistic and the p-value. First go to **Transform > Compute Variable**



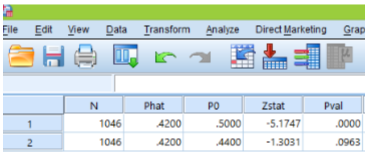
Then enter the following expression and save it into ‘Zstat’ variable:



Click **OK**, next once again we open the calculator and enter this formula:



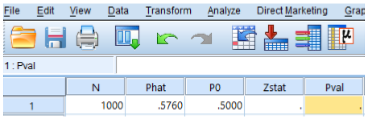
Here ‘CDF.NORMAL’ computes probability below the z-statistic which is what we need since the alternative hypothesis is that the true proportion is less than null. To finish click **OK** and we have the results:



	N	Phat	P0	Zstat	Pval
1	1046	.4200	.5000	-5.1747	.0000
2	1046	.4200	.4400	-1.3031	.0963

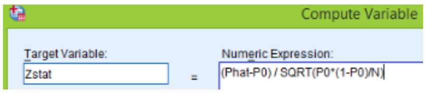
The p-value for $p_0 = 0.5$ null hypotheses is essentially zero and therefore we reject null and conclude that the true proportion is less than 0.5. However the p-value for $p_0 = 0.44$ is 0.096 which is considered as not significant and hence we cannot reject this null hypothesis.

Next consider the 'Flipping the bottle cap' example. Here a bottle cap was flipped 1000 times and proportion of Red was 0.576. We want test whether the true proportion is 0.5 or it is not 0.5 (the alternative here is non-directional). First we fill the table:

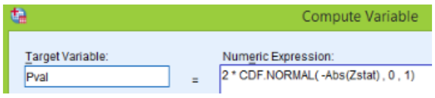


	N	Phat	P0	Zstat	Pval
1	1000	.5760	.5000	.	.

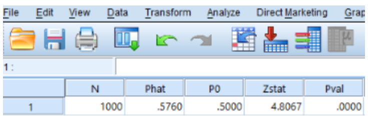
Next we find 'Zstat' as before using the calculator:



To get the p-value for non-directional alternative we must find probability above this z-statistic and multiply it by 2. The most convenient way to find it, is by following expression:



'Abs' here is the absolute value. Finally we get the results:

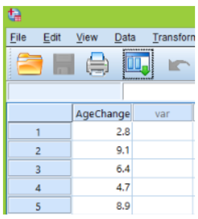


	N	Phat	P0	Zstat	Pval
1	1000	.5760	.5000	4.8067	.0000

The p-value is almost zero and therefore we reject the null hypothesis that the true proportion of red is 0.5 and conclude that it is not 0.5.

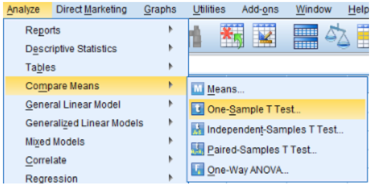
Hypothesis Testing for Means

We start this section with the 'Age change' data set.

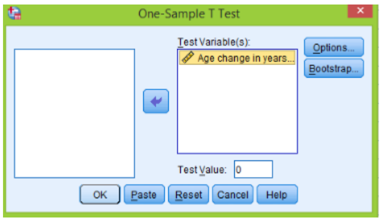


	AgeChange	var
1	2.8	
2	9.1	
3	6.4	
4	4.7	
5	8.9	

These data record by how many years a subject looks younger after a plastic surgery. We want to test whether the true mean is 0 (plastic surgery does not make any difference) versus that it is greater than 0 (people look younger after surgery). It is very easy to do in SPSS. Go to **Analyze > Compare Means > One-Sample T Test**



Send the variable across using the arrow:



Note that the 'Test Value' is zero by default; click **OK** and the next results are produced:

One-Sample Statistics

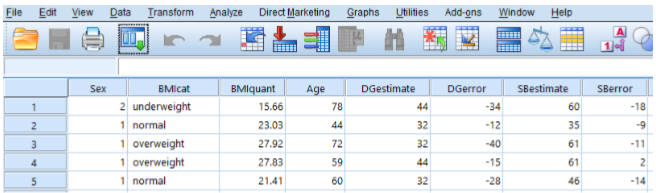
	N	Mean	Std. Deviation	Std. Error Mean
Age change in years	60	7.177	2.9481	.3806

One-Sample Test

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

We observe that the t-statistic is larger than 18 which is very large and degrees of freedom is 59. The most important here, is that two sided p-value is almost zero (SPSS finds only two sided p-values). Since we want one sided p-value (true mean is greater than 0) than it would be just half of the two sided one. In either case the p-value is tiny and hence we reject null hypothesis and conclude that plastic surgery does make people look younger.

Next let's analyze 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

In this example the goal is to test whether the true mean of difference between actual and estimated age is 0 or not. We concentrate on 'DGerror' variable. Imitating the above procedure we get the next table:

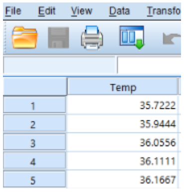
One-Sample Statistics

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

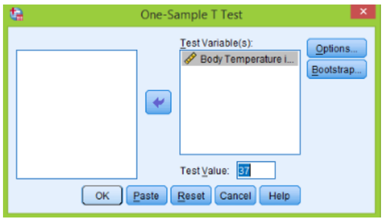
One-Sample Test

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

The t-statistic is approximately -20 which is very small and consequently the p-value is almost zero. Based on the p-value we reject null hypothesis and conclude that the true mean of 'DGerror' is not 0.
We finish this section with the 'Temperature' data set which consists of body temperatures in Celsius of 130 subjects.



We want test whether the true average body temperature is 37 or not. As usual go to **Analyze** > **Compare Means** > **One-Sample T Test**, move the variable to the right window and enter '37' in the 'Test Value' option:



Click **OK** and get:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Body Temperature in Celsius	130	36.805128	.4073240	.0357247

One-Sample Test

	Test Value = 37					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Body Temperature in Celsius	-5.455	129	.000	-.1948718	-.265554	-.124190

Note that the two sided p-value is very small and therefore we reject the null hypothesis and conclude that the actual average body temperature is not 37 degrees Celsius.