

IBM SPSS Bootstrapping

- The IBM SPSS Bootstrapping module allows users to derive more robust estimates of a host of statistical values. These values may include means, medians, standard errors and confidence intervals as well as correlations and regression coefficients. Bootstrapping is useful in a number of situations including hypothesis testing as it provides an alternative to classical parametric estimation when the underlying assumptions of such methods are in danger of being violated such as when the error values in a linear regression solution are found to increase linearly. Bootstrapping is also useful when inferential calculations require extremely complex formulas for calculating standard errors. This is especially true when computing standard errors in order to derive confidence intervals for median and percentile values.
- Specifically, Bootstrapping refers to the method of repeatedly resampling subsets of a data file (with replacement) and examining the variation in the resulting calculation of key statistics (such as a mean). By creating an internal sampling distribution of the statistic in question the technique allows for a more realistic estimate of that statistic's parameter value.
- **Example.** A digital retailer adds about 5% to its growing customer base every quarter. The management are interested to know if this figure hold true for customers across three separate age groups. By applying IBM SPSS Bootstrapping, analysts can figure out more accurately the degree to which the customer growth rate varies by age group.

Descriptives				Statistic	Std. Error
Current Salary	Female	Mean		25975.56	520.604
		95% Confidence Interval for Mean	Lower Bound	24949.34	
			Upper Bound	27001.79	
		5% Trimmed Mean		25179.09	
		Median		24150.00	
		Variance		57729155.11	
		Std. Deviation		7597.970	
		Minimum		15750	
		Maximum		58125	
		Range		42375	
	Interquartile Range		7050		
	Skewness		1.880	.167	
	Kurtosis		4.646	.332	
	Male	Mean		41557.01	1230.593
		95% Confidence Interval for Mean	Lower Bound	39133.50	
			Upper Bound	43980.52	
		5% Trimmed Mean		39561.37	
		Median		32850.00	
		Variance		384646969.9	
		Std. Deviation		19612.419	
Minimum			19650		
Maximum			135000		
Range			115350		
Interquartile Range		23250			
Skewness		1.622	.153		
Kurtosis		2.697	.304		

Results from the Explore procedure running with default settings (i.e. Bootstrapping switched off).

Descriptives

			Statistic	Std. Error	Bootstrap ^a				
					Bias	Std. Error	95% Confidence Interval		
							Lower	Upper	
Current Salary	Female	Mean	25975.56	520.604	-28.18	504.43	24979.32	26998.98	
		95% Confidence Interval for Mean	Lower Bound	24949.34					
			Upper Bound	27001.79					
		5% Trimmed Mean	25179.09		6.28	454.13	24359.56	26129.24	
		Median	24150.00		35.85	316.03	23400.00	24750.00	
		Variance	57729155.11		-882475.229	9794490.275	38699831.17	78022224.15	
		Std. Deviation	7597.970		-86.173	648.079	6220.919	8833.019	
		Minimum	15750						
		Maximum	58125						
		Range	42375						
	Interquartile Range	7050		-148	651	5813	8137		
	Skewness	1.880	.167	-.046	.213	1.375	2.208		
	Kurtosis	4.646	.332	-.118	1.079	2.624	6.779		
	Male	Mean	41557.01	1230.593	27.45	1240.12	39370.01	44315.12	
		95% Confidence Interval for Mean	Lower Bound	39133.50					
			Upper Bound	43980.52					
		5% Trimmed Mean	39561.37		55.34	1214.24	37460.60	42211.94	
		Median	32850.00		73.31	1192.14	31200.00	35250.00	
		Variance	384646969.9		-260132.196	51745221.70	291405016.9	496979612.0	
		Std. Deviation	19612.419		-50.778	1315.590	17070.589	22293.038	
Minimum		19650							
Maximum		135000							
Range		115350							
Interquartile Range	23250		-214	3541	16354	30080			
Skewness	1.622	.153	-.032	.231	1.145	2.062			
Kurtosis	2.697	.304	-.167	1.241	.470	5.163			

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Results from the Explore procedure running with Bootstrapping switched on. Note the additional statistical estimates.