

統計的推定とブートストラップ法についての考察

Some remarks on statistical estimation using the Bootstrap methods

石村 貞夫・石村 友二郎

Sadao ISHIMURA and Yujiro ISHIMURA

1. 序文

統計学といえば、統計的推定や統計的検定が中心的话题となる。研究対象としての母集団のパラメータを推定したり検定したりする。このとき、母集団は正規分布に従っているという正規性を仮定するのが一般的なのだが、研究対象によっては、正規性を仮定できない場合も多い。

統計的検定の場合には、ノンパラメトリックな方法が開発され、ウィルコクソンの順和検定やクラスカルウォリスの検定など、多くのノンパラメトリック検定が知られている。

それに対し、統計的推定の場合には、B.Efron によって開発されたブートストラップ法という新しい統計処理がある。

この方法は、標本から母集団を再生するという方法で、乱数を利用している点に特徴がある。

統計解析用ソフトSPSSでは、新しいバージョンから、このブートストラップ法が追加されている。そこで、この掌篇では、正規性を仮定した母平均の推定方法と、ブートストラップ法による母平均の推定を比較検討してみよう。

データは、次のように入力される。

Case	x
1	800
2	1000
3	1100
4	900
5	1200
6	900
7	1000
8	800

図 1

2. 正規性を仮定した母平均の区間推定

SPSSを使って母平均の区間推定を行う場合には、次のようにデータの変数を従属変数の枠に移動し、OK ボタンをクリックする。

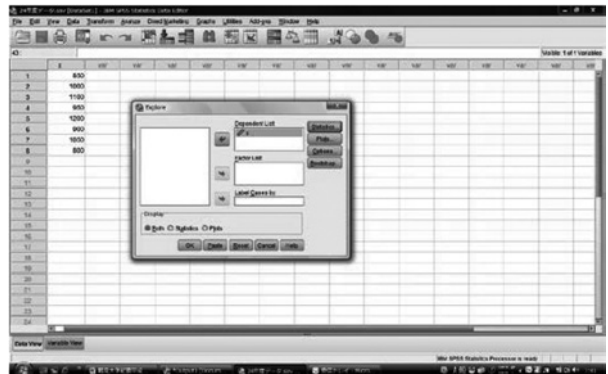


図 2

この出力は、次の表1、表2のようになる。したがって、表2を見ると、信頼係数95%における母平均の信頼区間は、869.67 ~ 1092.83 であることがわかる。

Explore

表 1 Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
x	8	100.0%	0	.0%	8	100.0%

表 2 Descriptives

	Statistic	Std. Error
x Mean	981.25	47.186
95% Confidence Interval for Mean	Lower Bound: 869.67 Upper Bound: 1092.83	
5% Trimmed Mean	979.17	
Median	975.00	
Variance	17812.500	
Std. Deviation	133.463	
Minimum	800	
Maximum	1200	
Range	400	
Interquartile Range	225	
Skewness	.296	.752
Kurtosis	-.652	1.481

3. ブートストラップ法による母平均の区間推定

SPSSにおけるブートストラップ法の手順は、次の図3のようなになる。



図 3

この出力は、表3、表4のようなになる。したがって、このブートストラップ法では、取り出すサンプル数が1000であることがわかる。表2を見ると、信頼係数95%における母平均の信頼区間は、900.00 ~ 1075.00 になっている。

表 3 Bootstrap Specifications

Sampling Method	Simple
Number of Samples	1000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

表 4 Descriptives

	Bootstrap ^a			
	Bias	Std. Error	95% Confidence Interval	
			Lower	Upper
x Mean	1.01	43.27	900.00	1075.00
95% Confidence Interval for Mean				
5% Trimmed Mean	1.95	45.31	890.35	1073.61
Median	2.80	61.15	850.00	1100.00
Variance	-2660.31	6238.98	4242.20	28526.78
Std. Deviation	2	2	6	6
Minimum	-13.175	26.144	65.132	168.899
Maximum				
Range				
Interquartile Range	-19	68	88	350
Skewness	-.092	.649	-1.045	1.577
Kurtosis	.175	1.319	-2.187	3.113

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

次に、取り出すサンプルの数を5000個にしてみよう。SPSSの場合、サンプル数の枠の中を5000とするだけでよい。

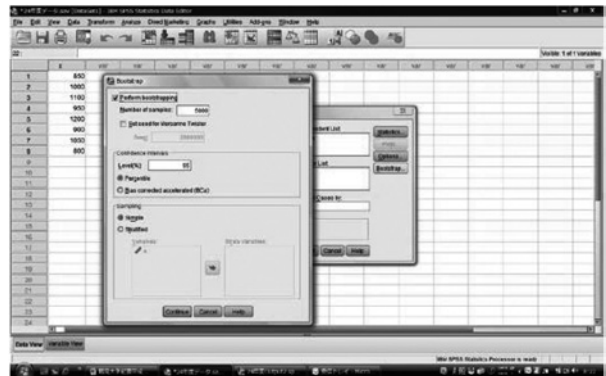


図 4

サンプルの数を5000個にした場合の出力は、次のようになる。表6を見ると、

信頼係数95%における母平均の信頼区間は、900.00 ~ 1068.75 になっている。

表 5 Bootstrap Specifications

Sampling Method	Simple
Number of Samples	5000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

表 6 Descriptives

	Bootstrap ^a			
	Bias	Std. Error	95% Confidence Interval	
			Lower	Upper
x Mean	.44	44.38	900.00	1068.75
95% Lower Confidence Interval for Mean				
5% Trimmed Mean	1.38	46.56	893.06	1075.00
Median	.33	61.97	850.00	1100.00
Variance	-2088.87	6310.31	4824.78	29107.14
Std. Deviation	-10.810	26.076	69.461	170.608
Minimum				
Maximum				
Range				
Interquartile Range	-13	70	88	350
Skewness	-.090	.654	-1.081	1.539
Kurtosis	.161	1.342	-2.185	3.193

a. Unless otherwise noted, bootstrap results are based on 5000 bootstrap samples

サンプルの数を10000個にした場合の出力は、次のようになる。

信頼係数95%における母平均の信頼区間は、893.75 ~ 1068.75 になっている。

表 7 Bootstrap Specifications

Sampling Method	Simple
Number of Samples	10000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

表 8 Descriptives

	Bootstrap ^a			
	Bias	Std. Error	95% Confidence Interval	
			Lower	Upper
x Mean	-.50	44.45	893.75	1068.75
95% Lower Confidence Interval for Mean				
5% Trimmed Mean	.41	46.61	890.28	1073.61
Median	-.76	61.87	850.00	1100.00
Variance	-2226.81	6326.59	4821.42	28571.42
Std. Deviation	-11.425	26.312	69.437	169.031
Minimum				
Maximum				
Range				
Interquartile Range	-15	69	88	350
Skewness	-.088	.646	-1.092	1.538
Kurtosis	.150	1.307	-2.185	2.947

a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

サンプルの数を100000個にした場合の出力は、次のようになる。

このように、再生させるサンプル数を多くすると、時間の関係で、もはやコンピュータでは計算ができなくなる。

表 9 Bootstrap Specifications

Sampling Method	Simple
Number of Samples	100000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

表 10 Warnings

The total number of pivot table cells across split files exceeds 733923. This limit can be altered by using the SET MXCELLS command. If the working file was being created or modified during the current procedure, any changes were probably lost. Execution of this command stops.

ところで、サンプルの数をもう一度、5000個にしてみよう。すると、その出力は次のようになる。信頼係数95%における母平均の信頼区間は、893.75 ~ 1068.75 になっている。

したがって、ブートストラップ法の場合、取り出すサンプルの数を同じにしても、信頼区間の値が異なっていることがわかる。

このことは、乱数を使って、サンプルを再生していることを示していることに他ならない。

表 11 Bootstrap Specifications

Sampling Method	Simple
Number of Samples	5000
Confidence Interval Level	95.0%
Confidence Interval Type	Percentile

表 12 Descriptives

	Bootstrap ^a			
	Bias	Std. Error	95% Confidence Interval	
			Lower	Upper
x Mean	.24	43.70	893.75	1068.75
95% Lower Confidence Interval for Mean				
5% Trimmed Mean	1.23	45.85	890.28	1070.83
Median	.51	61.24	850.00	1100.00
Variance	-2227.27	6361.68	4598.21	29107.14
Std. Deviation	7	5	4	3
Minimum	-11.463	26.479	67.810	170.608
Maximum				
Range				
Interquartile Range	-15	71	88	350
Skewness	-.099	.658	-1.081	1.587
Kurtosis	.191	1.348	-2.157	3.150

a. Unless otherwise noted, bootstrap results are based on 5000 bootstrap samples

参考文献

- 1) M.G.Kendall Kendall's Advanced Theory of Statistics, Volume 1,2,3 CHARLS&GRIFIN
- 2) 石村貞夫 入門はじめての統計解析 東京図書
- 3) 石村貞夫 他 SPSSによる統計処理の手順 第6版 東京図書

統計的推定とブートストラップ法についての考察

Some remarks on statistical estimation using the Bootstrap methods

歯学部 准教授 石村貞夫
早稲田大学大学院 基幹理工学研究科 応用数学科
石村友二郎