2013.06.06: 教科書 p.131 (第4章5. 図5) の分布からの標本抽出

データの一覧表を作り、乱数をキーにして並べ替えて上から必要な個数だけ抜き出す(コピーすることでも無作為抽出ができる。「乱数」は乱数関数 "=RAND()" を使う。

図1. 抽出実験に使う母集団分布(教科書 p.131, 図5)

(1) n = 5 を 9回 (k = 9)

図2. n=5, k=9 の抽出実験の結果

(2) n = 9 を 5回 (k = 5)

図3. n=9, k=5 の抽出実験の結果

(3) n = 15 を 3回 (k = 3)

図3. n=9, k=5 の抽出実験の結果

母集団

N	100
母平均 μ	2.78
母標準偏差 σ	1.48
最小値	1
第1四分位数	1.75
中央値	2.5
第3四分位数	4
最大値	6

(1) n = 5 を 9回 (k = 9)		
n =	9	5
	平均値	標準偏差
1回目	2.6	1.8
2回目	2.2	2.2
3回目	2.0	0.7
4回目	3.2	2.2
5回目	2.0	1.2
6回目	2.2	1.6
7回目	2.0	1.7
8回目	3.4	2.1
9回目	3.6	1.9
平均値の平均値	**2.58**	
平均値の標準偏差	**0.65**	
σ / √n	0.660	
標準偏差の平均値	1.72	
	1.54	

(2) n = 9 を 5回 (k = 5)		
n =	9	5
	平均値	標準偏差
1回目	3.0	1.6
2回目	2.4	1.3
3回目	3.6	1.9
4回目	2.7	1.9
5回目	2.7	1.4
	2.87	
	0.43	
	0.48	
	1.62	
	1.52	

(3) n = 15 を 3回 (k = 3)		
n =	15	1
	平均値	標準偏差
1回目	3.1	2.0
2回目	3.1	1.5
3回目	2.5	1.6
	2.87	
	0.35	
	0.388	
	1.70	
	1.64	

標本の個々の平均値は、母平均に比べて大きいものも小さいものもあるが、標本平均の平均値はおおむね母平均に近い値になっている。(標本平均は母平均の**不偏推定量**になっている)

一方、平均値の標準偏差は、母標準偏差よりも小さくなっている。その小さくなり方は、n (標本の個数) が大きいほど小さい。標本抽出の回数 (k) には関係ない。

定理1, 定理2 で示されているように、σ標準偏差を n のルートで割った値を求めてみると、平均値の標準偏差に近くなっていることがわかる。(定理1, 定理2 の確認)

標本の標準偏差は、nにもk にも関係なく、だいたい母標準偏差に近い値で一定となっている(**母標準偏差の不偏推定量**)。

※ただしこの例では、たまたま大きめの値となっている。

標準偏差を n-1 でなく n で割って求めていたらどうなるか

標準偏差を n-1 でなく n で割って求めていたらどうなるか、は、n-1で割って求めた普通の標準偏差に √((n-1)/n) を掛ければよい、そのぶんだけ小さい値になる。

※この例では、n で割って求めた方が母標準偏差に近い値になっているが、普通はそうすると母標準偏差よりやや小さめの値となる。