2014.06.09: 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。

連番	x	乱数
1	1	0.60492
2	1	0.00345
3	1	0.52607
4	1	0.02140
5	1	0.11756
6	1	0.71975
7	1	0.37273
8	1	0.82808
9	1	0.06391
10	1	0.30649
11	1	0.03334
12	1	0.15997
13	1	0.15956
14	1	0.25623
15	1	0.43564
16	1	0.41480
17	1	0.83426
18	1	0.90207
19	1	0.28958
20	1	0.41757
21	1	0.62584
22	1	0.75452
23	1	0.11551
24	1	0.94218
25	1	0.71669
26	2	0.16838
27	2	0.07579
28	2	0.58832
29	2	0.74095
30	2	0.95900
31	2	0.59955
32	2	0.67585
33	2	0.70354
34	2	0.39365
35	2	0.51967
36	2	0.47411
37	2	0.54230
38	2	0.51499
39	2	0.26733
40	2	0.05837
41	2	0.15285
42	2	0.82766
43	2	0.61091
44	2	0.68960
45	2	0.44421
46	2	0.82397
47	2	0.42403
48	2	0.31043
49	2	0.40146
50	2	0.60634
51	3	0.50578
52	3	0.00028
53	3	0.54171
54	3	0.00476
55	3	0.07943
56	3	0.99155
57	3	0.22777
58	3	0.09207
59	3	0.70009
60	3	0.54903
61	3	0.23237
62	3	0.77106
63	3	0.65249
64	3	0.60085
65	3	0.83561
66	3	0.76854
67	3	0.63919
68	3	0.26449
69	3	0.19690
70	3	0.97938
71	4	0.35324
72	4	0.43144
73	4	0.35488
74	4	0.61144
75	4	0.77435
76	4	0.22848
77	4	0.35966
78	4	0.39350
79	4	0.64959
80	4	0.38109
81	4	0.36075
82	4	0.38148
83	4	0.06108
84	4	0.25081
85	4	0.20074
86	5	0.12310
87	5	0.33765
88	5	0.21125
89	5	0.65648
90	5	0.52845
91	5	0.07823
92	5	0.89879
93	5	0.13514
94	5	0.66409
95	5	0.67842
96	6	0.65725
97	6	0.74230
98	6	0.57063
99	6	0.14362
100	6	0.77617

m	1
x.max	6
x.min	1
R	5
n	100
仮のh	10
仮のc	0.5
c	1

階級下限	上限	階級値	度数
0.5	1.5	1	25
1.5	2.5	2	25
2.5	3.5	3	20
3.5	4.5	4	15
4.5	5.5	5	10
5.5	6.5	6	5
6.5			

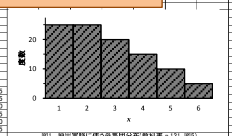

図1. 抽出実験に使う母集団分布(教科書 p.131, 図5)

(1) n = 3 を 12回 (k = 12)

	1回目 連番	1回目 x	2回目 連番	2回目 x	3回目 連番	3回目 x	4回目 連番	4回目 x	5回目 連番	5回目 x	6回目 連番	6回目 x	7回目 連番	7回目 x	8回目 連番	8回目 x	9回目 連番	9回目 x	10回目 連番	10回目 x	11回目 連番	11回目 x	12回目 連番	12回目 x
1	46	2	23	1	36	2	52	3	82	4	34	2	87	5	12	1	78	4	29	2	79	4	53	3
2	97	6	98	6	72	4	74	4	5	1	51	3	61	3	73	4	27	2	100	6	5	1	11	1
3	79	4	46	2	23	1	3	1	80	4	3	1	13	1	63	3	87	5	14	1	35	2	65	3
4																								
5																								
平均値		4.0		3.0		2.3		2.7		3.0		2.0		3.0		2.7		3.7		3.0		2.3		2.3
標準偏差		2.0		2.6		1.5		1.5		1.7		1.0		2.0		1.5		1.5		2.6		1.5		1.2

	平均値	標準偏差
1回目	4.0	2.0
2回目	3.0	2.6
3回目	2.3	1.5
4回目	2.7	1.5
5回目	3.0	1.7
6回目	2.0	1.0
7回目	3.0	2.0
8回目	2.7	1.5
9回目	3.7	1.5
10回目	3.0	2.6
11回目	2.3	1.5
12回目	2.3	1.2
平均値の平均値	2.8	
平均値の標準偏差	0.6	

標準偏差の平均値 1.7

抽出結果の集計

階級値	1回目	2回目	3回目	4回目	5回目	6回目	7回目	8回目	9回目	10回目	11回目	12回目	
1	0	1	1	1	1	1	1	1	0	1	1	1	10
2	1	1	1	0	0	1	0	0	1	1	1	0	7
3	0	0	0	1	0	1	1	1	0	0	0	2	6
4	1	0	1	1	2	0	0	1	1	0	1	0	8
5	0	0	0	0	0	0	1	0	1	0	0	0	2
6	1	1	0	0	0	0	0	0	0	1	0	0	3
													36

抽出結果の集計(相対度数)

階級値	1回目	2回目	3回目	4回目	5回目	6回目	7回目	8回目	9回目	10回目	11回目	12回目	
1	0%	3%	3%	3%	3%	3%	3%	3%	0%	3%	3%	3%	28%
2	3%	3%	3%	0%	0%	3%	0%	0%	3%	3%	3%	0%	19%
3	0%	0%	0%	3%	0%	3%	3%	3%	0%	0%	0%	6%	17%
4	3%	0%	3%	3%	6%	0%	0%	3%	3%	0%	3%	0%	22%
5	0%	0%	0%	0%	0%	0%	3%	0%	3%	0%	0%	0%	6%
6	3%	3%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	8%
													100%

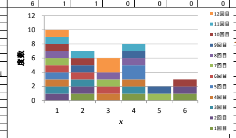

図2. n=3, k=12 の抽出実験の結果

個々の標本はバラバラに抽出されるが、積み上げ式の棒グラフでまとめてみると、全体としては母集団の分布を再現していることがわかる。

(2) n = 6 を 6回 (k = 6)

	1回目 連番	1回目 x	2回目 連番	2回目 x	3回目 連番	3回目 x	4回目 連番	4回目 x	5回目 連番	5回目 x	6回目 連番	6回目 x
1	78	4	29	2	60	3	75	4	75	4	33	2
2	21	1	73	4	11	1	63	3	48	2	77	4
3	16	1	28	2	4	1	53	3	76	4	36	2
4	43	2	13	1	15	1	84	4	31	2	89	5
5	73	4	26	2	72	4	34	2	81	4	26	2
6	63	3	80	4	48	2	44	2	83	4	20	1
7												
8												
9												
平均値		2.5		2.5		2.0		3.0		3.3		2.7
標準偏差		1.4		1.2		1.3		0.9		1.0		1.5

	平均値	標準偏差
1回目	2.5	1.4
2回目	2.5	1.2
3回目	2.0	1.3
4回目	3.0	0.9
5回目	3.3	1.0
6回目	2.7	1.5
平均値の平均値	2.7	
平均値の標準偏差	0.5	

標準偏差の平均値 1.2

抽出結果の集計

階級値	1回目	2回目	3回目	4回目	5回目	6回目	
1	2	1	3	0	0	1	7
2	1	3	1	2	2	3	12
3	1	0	1	2	0	0	4
4	2	2	1	2	4	1	12
5	0	0	0	0	0	1	1
6	0	0	0	0	0	0	0
							36

抽出結果の集計(相対度数)

階級値	1回目	2回目	3回目	4回目	5回目	6回目	
1	6%	3%	8%	0%	0%	3%	19%
2	3%	8%	3%	6%	6%	8%	33%
3	3%	0%	3%	6%	0%	0%	11%
4	6%	6%	3%	6%	11%	3%	33%
5	0%	0%	0%	0%	0%	3%	3%
6	0%	0%	0%	0%	0%	0%	0%
							100%

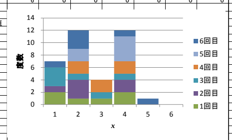

図3. n=6, k=6 の抽出実験の結果

(3) n = 12 を 3回 (k = 3)

	1回目 連番	1回目 x	2回目 連番	2回目 x	3回目 連番	3回目 x
1	100	6	91	5	59	3
2	13	1	9	1	98	6
3	32	2	37	2	13	1
4	49	2	43	2	8	1
5	43	2	74	4	82	4
6	15	1	77	4	54	3
7	48	2	16	1	79	4
8	91	5	93	5	91	5
9	68	3	6	1	94	5
10	35	2	80	4	60	3
11	59	3	71	4	75	4
12	26	2	78	4	2	1
13						
14						
15						
平均値		2.6		3.1		3.3
標準偏差		1.5		1.6		1.7

	平均値	標準偏差
1回目	2.6	1.5
2回目	3.1	1.6
3回目	3.3	1.7
平均値の平均値	3.0	
平均値の標準偏差	0.4	

標準偏差の平均値 1.6

抽出結果の集計

階級値	1回目	2回目	3回目	
1	2	3	3	8
2	6	2	0	8
3	2	0	3	5
4	0	5	3	8
5	1	2	2	5
6	1	0	1	2
				36

抽出結果の集計(相対度数)

階級値	1回目	2回目	3回目	
1	6%	8%	8%	22%
2	17%	6%	0%	22%
3	6%	0%	8%	14%
4	0%	14%	8%	22%
5	3%	6%	6%	14%
6	3%	0%	3%	6%
				100%

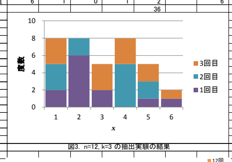

図3. n=12, k=3 の抽出実験の結果

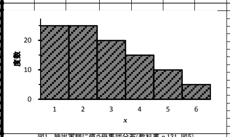

図1. 抽出実験に使う母集団分布(教科書 p.131, 図5)

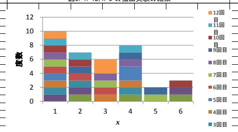

図2. n=3, k=12 の抽出実験の結果

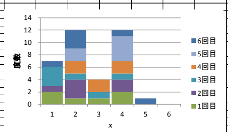

図3. n=6, k=6 の抽出実験の結果

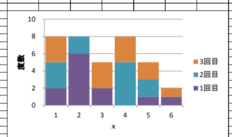

図3. n=12, k=3 の抽出実験の結果

母集団	
N	100
母平均 μ	**2.75**
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 = 3 を 12回 (k = 12)

n =	3	
	平均値	標準偏差
1回目	4.0	2.0
2回目	3.0	2.6
3回目	2.3	1.5
4回目	2.7	1.5
5回目	3.0	1.7
6回目	2.0	1.0
7回目	3.0	2.0
8回目	2.7	1.5
9回目	3.7	1.5
10回目	3.0	2.6
11回目	2.3	1.5
12回目	2.3	1.2

(2) n = 6 を 6回 (k = 6)

n =	6	
	平均値	標準偏差
1回目	2.5	1.4
2回目	2.5	1.2
3回目	2.0	1.3
4回目	3.0	0.9
5回目	3.3	1.0
6回目	2.7	1.5

(3) n = 12 を 3回 (k = 3)

n =	12	
	平均値	標準偏差
1回目	2.6	1.5
2回目	3.1	1.6
3回目	3.3	1.7

	(1) n = 3	(2) n = 6	(3) n = 12
平均値の平均値	**2.83**	**2.67**	**3.00**
平均値の標準偏差	**0.58**	**0.46**	**0.38**
σ / √n	**0.85**	**0.60**	**0.43**
標準偏差の平均値	1.73	1.22	1.58
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	1.42	1.11	1.51

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**不偏推定値**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、ぼ標準偏差を 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値**)。

※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、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。