

標本平均を用いた変動は必ず小さくなるか？

Sampo Suzuki, CC 4.0 BY-NC-SA

2021-06-29

標本平均を用いた変動は必ず小さくなるのか？

『ソフトウェアメトリクス統計分析入門』[小池, 2015] の 3.3 ワンポイント講義「不偏分散を算出する際に自由度を用いる理由」には、標本平均 (\bar{x}) を使って算出した変動 ($\sum_{i=1}^n (x_i - \bar{x})^2$ 、偏差平方和) は母平均 (μ) を使って算出した変動 ($\sum_{i=1}^n (x_i - \mu)^2$) よりも必ず小さな値になるとあります。実際に小さくなるのかを確認します。

母集団データの作成

最初に正規分布を持つ母集団のデータ (x) を作成します。ここでは母平均と母標準偏差は不明であると仮定します。

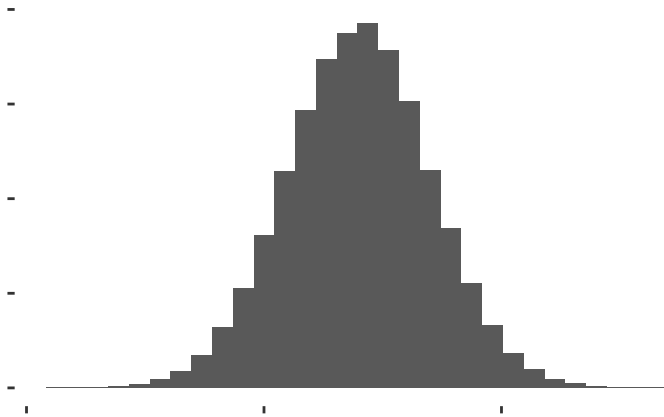


Figure 1: 母集団の分布

簡単なシミュレーション

上記の母集団 (x) から以下の手順で二種類の変動 (偏差平方和) を求めます。

1. 3 つのデータをランダムサンプリングで取り出す (標本 $x_n, n = 1, 2, 3$)
2. 取り出したデータの平均値 (標本平均 \bar{x}) を求める
3. 標本平均 (\bar{x}) を用いて標本の変動 (偏差平方和 $\sum_{i=1}^n (x_i - \bar{x})^2$) を求める
4. 母平均 (μ) を用いて標本の変動 (偏差平方和 $\sum_{i=1}^n (x_i - \mu)^2$) を求める
5. 求めた二つの変動 (偏差平方和) を比較する

この計算を任意の回数繰り返して標本平均 (\bar{x}) を用いた標本の変動 (偏差平方和 $\sum_{i=1}^n (x_i - \bar{x})^2$) の方が小さいことを確認します。

```

1 df <- data.frame()
2 for (i in c(1:30)) {
3   xs <- sample(x, size = 3, replace = FALSE) # 母集団から 3 つのデータを取り出す
4   xb <- mean(xs)                               # 標本平均を求める
5   dssxb <- sum((xs - xb)^2)                     # 標本平均を用いた変動 (偏差平方和)
6   dssmu <- sum((xs - mu)^2)                     # 母平均を用いた変動 (偏差平方和)
7   # 計算結果をデータフレームにまとめる
8   dftmp <- data.frame(no = i,                  # 通し番号
9                       x1 = xs[1], # 標本データ (n = 1)
10                      x2 = xs[2], # 標本データ (n = 2)
11                      x3 = xs[3], # 標本データ (n = 3)
12                      xb,         # 標本平均
13                      mu,         # 母平均
14                      dssxb,      # 標本平均を用いた変動 (偏差平方和)
15                      dssmu,      # 母平均を用いた変動 (偏差平方和)
16                      diff = dssxb - dssmu # 負値なら標本平均による変動が小さい
17                    )
18   df <- dplyr::bind_rows(df, dftmp)
19 }
20
21 df %>%
22   dplyr::rename(`標本平均` = xb, `母平均` = mu,
23                `標本平均での変動` = dssxb, `母平均での変動` = dssmu,
24                `変動差 (標本-母)` = diff) %>%
25   df_print(all = TRUE, scale_down = TRUE, caption = "シミュレーション結果")

```

Table 1: シミュレーション結果

no	x1	x2	x3	標本平均	母平均	標本平均での変動	母平均での変動	変動差 (標本-母)
1	2.9498361	4.8030369	5.376113	4.3763286	3.984684	3.2165290	3.676686	-0.4601569
2	2.9422607	3.7882078	2.046987	2.9258186	3.984684	1.5163296	4.879916	-3.3635863
3	-1.3890396	6.1825887	2.102006	2.2985182	3.984684	28.7227028	37.252166	-8.5294631
4	5.4850067	2.8509317	2.114541	3.4834930	3.984684	6.2802215	7.033798	-0.7535766
5	4.1080593	4.9947869	1.860016	3.6542875	3.984684	5.2222560	5.549741	-0.3274851
6	9.0280057	-3.1743646	5.726024	3.8598883	3.984684	79.6726125	79.719334	-0.0467218
7	6.5385761	8.0532167	1.003216	5.1983363	3.984684	27.5456179	31.964475	-4.4188573
8	5.9855130	4.0936488	2.889727	4.3229631	3.984684	4.8708215	5.214120	-0.3432987
9	3.9842593	-0.9345524	11.697489	4.9157320	3.984684	81.0857001	83.686252	-2.6005523
10	2.6566400	2.6948555	3.363902	2.9051324	3.984684	0.3164341	3.812728	-3.4962937
11	4.3832517	4.1297346	4.668041	4.3936757	3.984684	0.1450498	0.646873	-0.5018232
12	1.1831673	6.0825864	-5.135020	0.7102445	3.984684	63.2528323	95.418691	-32.1658587
13	4.0711975	1.1562107	3.308821	2.8454097	3.984684	4.5706989	8.464535	-3.8938363
14	4.2007156	4.8714447	1.408004	3.4933882	3.984684	6.7481781	7.472292	-0.7241142
15	4.5775337	4.0539793	8.293052	5.6415218	3.984684	10.6829771	18.918314	-8.2353368
16	4.2242886	7.6331796	0.670143	4.1758704	3.984684	24.2454558	24.355113	-0.1096570
17	6.0319595	1.6244979	0.282768	2.6464085	3.984684	18.0930536	23.465996	-5.3729427
18	3.5105381	1.1650633	1.011195	1.8955989	3.984684	3.9238804	17.016708	-13.0928272
19	5.7761526	7.5624690	1.102853	4.8138250	3.984684	22.2524298	24.314855	-2.0624253
20	4.8925723	3.2339782	5.755922	4.6274908	3.984684	3.2855024	4.525105	-1.2396025
21	4.7781403	3.3230197	9.156763	5.7526409	3.984684	18.4407557	27.817773	-9.3770171
22	5.1021692	5.6443430	5.256119	5.3342104	3.984684	0.1561236	5.619790	-5.4636663
23	5.3166644	2.6901232	3.610101	3.8722960	3.984684	3.5524790	3.590372	-0.0378930
24	-0.0826761	1.2425152	4.480185	1.8800081	3.984684	11.0194465	24.308427	-13.2889801
25	4.9519314	7.0017511	4.258477	5.4040532	3.984684	4.0693974	10.113226	-6.0438284
26	2.4656320	4.7573844	8.669014	5.2973434	3.984684	19.6783063	24.847532	-5.1692257
27	6.1788690	2.7941764	5.130989	4.7013448	3.984684	6.0049633	7.545772	-1.5408090
28	-0.0638104	4.9923238	5.554449	3.4943208	3.984684	19.1484393	19.869807	-0.7213676
29	3.5509714	0.8846013	4.057784	2.8311190	3.984684	5.8118266	9.803962	-3.9921351
30	3.3666071	-2.3742030	4.738442	1.9102821	3.984684	28.4761854	41.385613	-12.9094271

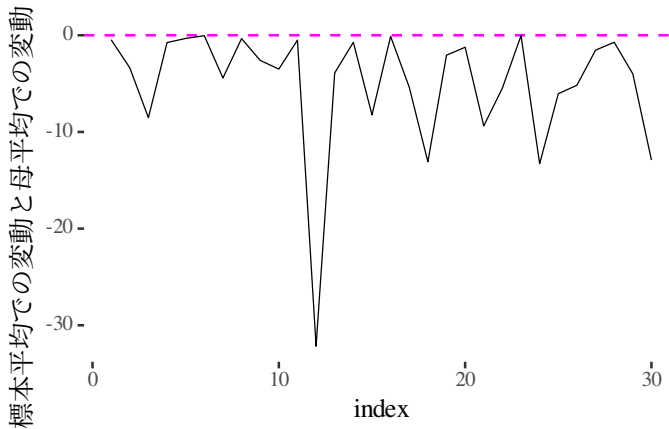


Figure 2: 標本平均を用いた変動と母平均を用いた変動の差

標本標準偏差の補正值を確認する

標本平均 (\bar{x}) を用いた変動 (偏差平方和) は母平均 (μ) を用いた変動 (偏差平方和) よりも小さくなることがわかりました。では、自由度で補正した標準偏差が母平均を用いて求めた標準偏差に本当に近くなるのかを同じ母集団 (x) を使って確認します。

```

1 df <- data.frame()
2 m <- 12
3 for (i in c(1:30)) {
4   xs <- sample(x, size = m, replace = FALSE) # 母集団からデータを取り出す
5   xb <- mean(xs)                             # 標本平均を求める
6   sdxs <- sqrt(sum((xs - xb)^2) / (m - 1))    # 自由度で補正した標本標準偏差
7   sdmu <- sqrt(sum((xs - mu)^2) / m)          # 母平均を用いた標本標準偏差
8   # 計算結果をデータフレームにまとめる
9   dftmp <- data.frame(no = i,                # 通し番号
10                      x1 = xs[1],            # 標本データ (n = 1)
11                      x2 = xs[2],            # 標本データ (n = 2)
12                      x3 = xs[3],            # 標本データ (n = 3)
13                      xb,                    # 標本平均
14                      mu,                    # 母平均
15                      sdxs,                  # 補正した標本標準偏差 (不偏推定値)
16                      sdmu,                  # 母平均を用いた標本標準偏差
17                      diff = sdxs - sdmu     # 負値なら標本平均による変動が小さい
18                      )
19   df <- dplyr::bind_rows(df, dftmp)
20 }
21
22 df %>%
23   dplyr::rename(`標本平均` = xb, `母平均` = mu,
24                 `補正した標準偏差` = sdxs, `母平均による標準偏差` = sdmu,
25                 `差 (標本-母)` = diff) %>%
26   df_print(all = TRUE, scale_down = TRUE, caption = "シミュレーション結果")

```

Table 2: シミュレーション結果

no	x1	x2	x3	標本平均	母平均	補正した標準偏差	母平均による標準偏差	差 (標本-母)
1	3.3696071	-0.4807766	4.3063727	3.941167	3.984684	2.849916	2.728934	0.1209822
2	4.7973404	-0.5017052	0.8790420	2.233852	3.984684	2.330122	2.835916	-0.5057949
3	3.9409217	4.3090122	4.6869091	4.970903	3.984684	2.437623	2.533666	-0.0960437
4	0.6384337	4.9499169	7.6676053	3.689281	3.984684	3.708505	3.562890	0.1456146
5	8.8716897	3.8549278	1.8699868	4.428316	3.984684	2.668223	2.592863	0.0753599
6	0.9779053	2.1566581	5.1314726	4.143060	3.984684	2.075048	1.993010	0.0820381
7	4.6233985	7.5856011	0.3109877	3.757260	3.984684	3.420127	3.282410	0.1377167
8	4.2261305	0.8805910	6.6261815	3.097315	3.984684	2.446461	2.504761	-0.0583006
9	-2.0145137	6.1512267	-0.0291733	3.307421	3.984684	2.876382	2.835982	0.0404001
10	7.6280570	5.7775197	5.8629411	4.476783	3.984684	3.290492	3.188608	0.1018840
11	1.7516972	3.8582515	5.4316160	4.191824	3.984684	2.051548	1.975100	0.0764484
12	2.0572194	3.8017517	-0.4898263	4.094754	3.984684	2.654731	2.544093	0.1106374
13	5.4974431	4.7694066	5.0203960	4.067797	3.984684	2.769818	2.653201	0.1166171
14	4.8845877	5.6430005	5.9589743	4.518539	3.984684	2.357336	2.319256	0.0380800
15	1.2769413	0.7777181	1.3530467	3.070003	3.984684	3.606769	3.572304	0.0344647
16	3.7381707	4.0422341	1.5614187	3.890184	3.984684	3.273092	3.135171	0.1379205
17	2.1633681	2.0131472	5.2402933	2.985413	3.984684	2.712794	2.782898	-0.0701041
18	4.0875833	3.3836800	0.9202103	2.635534	3.984684	2.565132	2.802103	-0.2369710
19	1.4154756	5.2815788	6.8912748	4.331742	3.984684	1.990321	1.936934	0.0533872
20	9.9226998	-1.0064291	2.8427723	4.371727	3.984684	3.512037	3.384722	0.1273155
21	5.6160721	4.1564319	10.5469797	4.786439	3.984684	3.441183	3.390832	0.0503512
22	3.7435084	4.1756659	5.5536661	4.823473	3.984684	2.453382	2.494205	-0.0408233
23	3.8786155	7.1202550	7.2847092	6.491526	3.984684	2.100989	3.214120	-1.1131312
24	3.4363045	3.8960033	8.0357531	4.202687	3.984684	2.767739	2.658860	0.1088784
25	-0.9928191	3.6903414	1.8810476	3.664850	3.984684	2.794451	2.694533	0.0999188
26	1.7804977	1.9495721	7.1510235	3.333250	3.984684	3.208572	3.140285	0.0682871
27	5.4760104	3.8982844	6.7556193	4.290551	3.984684	1.940303	1.882711	0.0575924
28	-0.0896337	7.8124540	7.1423155	4.391903	3.984684	2.826535	2.736668	0.0898669
29	8.5725004	3.8306904	1.6058561	3.434627	3.984684	2.885195	2.816597	0.0685984
30	3.3304177	5.0800070	5.3682500	3.656751	3.984684	2.060631	1.999972	0.0606585

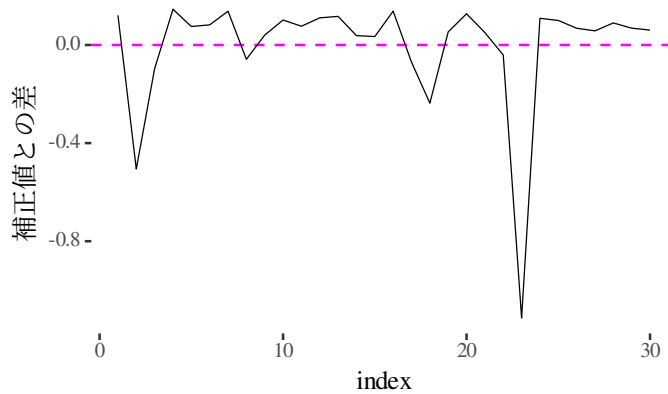


Figure 3: 標本標準偏差の補正值（不偏推定値）と母平均による標準偏差の差

おわりに

詳細で理論的な説明が必要な場合は『なぜ不偏分散は $N-1$ で割るのか』[est] を参照してください。

ちなみに母集団 (x) の平均値 (μ) と標準偏差 (s) は以下の通りでした。

```
1 mean(x)                # 平均値
## [1] 3.984684
1 (n / (n - 1)) * sd(x)  # 標準偏差
## [1] 3.0058
```

References

なぜ不偏分散は $n-1$ で割るのか. URL <http://kosugitti.sakura.ne.jp/wp/wp-content/uploads/2013/08/est.pdf>.

利和小池. 『ソフトウェアメトリクス統計分析入門』. 日科技連出版, first edition, 2015. URL <https://www.juse-p.co.jp/products/view/545>. ISBN 978-4-8171-9558-6.