

分散の加法性を視覚的に理解する（その2）

Sampo Suzuki, CC 4.0 BY-NC-SA

2021-06-01

はじめに（小室先生のアドバイスから）

分散の加法性が成り立つには「データが独立」であるという前提条件があります。乱数生成した二つのデータ¹が本当に独立なのかを確認すると共に分散の加法性も確認してみます。

¹ 各々 r r n 個のデータ

関数の定義

最初に以下の処理を行う関数を定義します。

- データを乱数生成する²
- 乱数生成したデータをランダムサンプリングする³
- 作成したデータの統計量を求める
- 無相関検定の結果と統計量をデータフレームにまとめる

² 今回は `rnorm()` 関数による分散が 100 となる正規分布

³ `sampling = TRUE` の場合のみ

```
1 f <- function(i = NA, sampling = FALSE, n = 5000000) {
2   # データを乱数生成する
3   x <- rnorm(n = n, mean = 10, sd = 10)
4   y <- rnorm(n = n, mean = 30, sd = 10)
5   # 乱数生成したデータからサンプリングする場合
6   if (sampling == TRUE) {
7     x <- sample(x, n, replace = TRUE)
8     y <- sample(y, n, replace = TRUE)
9   }
10  # 統計量を求める
11  df <- data.frame(no = i, var.x = var(x), var.y = var(y),
12                  var.xy = var(x + y), var.sum = var(x) + var(y),
13                  cov = cov(x, y), cov2 = cov(x, y) * 2)
14  # 無相関の検定結果と統計量をデータフレームにまとめる
15  df <- cor.test(x, y) %>% broom::tidy() %>% dplyr::bind_cols(df)
16  return(df)
17 }
```

この関数を for ループで一定回数繰り返し、その結果をデータフレームにまとめ、分散がどのようなになっているかを比較します。

Table 1: 変数の意味

変数名	その意味	備考
var.x	データ x の分散	
var.y	データ y の分散	
var.xy	データ x と y を加算したものの分散 ($\text{var}(x + y)$)	加法 1
var.sum	データ x, y の分散を加算したもの ($\text{var}(x) + \text{var}(y)$)	加法 2
var.diff	var.xy から var.sum を減算したもの	加法 1 と加法 2 の差異
cov2	データ x, y の共分散の 2 倍数	
cov	データ x, y の共分散	計算のみで未出力

乱数生成したデータの場合

Table 2: 乱数生成した二つのデータの分散

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
1	-0.0001291	0.7728236	99.94993	100.02913	199.9532	199.9791	-0.0258179	-0.0258179
2	0.0003030	0.4981275	100.03307	100.02382	200.1175	200.0569	0.0606092	0.0606092
3	-0.0004276	0.3389610	99.99416	100.00821	199.9168	200.0024	-0.0855280	-0.0855280
4	0.0004838	0.2793815	100.00953	99.96570	200.0720	199.9752	0.0967391	0.0967391
5	-0.0002187	0.6247927	100.06228	100.07481	200.0933	200.1371	-0.0437737	-0.0437737
6	0.0001338	0.7647540	100.07596	99.99899	200.1017	200.0749	0.0267752	0.0267752
7	0.0005169	0.2477801	99.98430	99.88829	199.9759	199.8726	0.1033083	0.1033083
9	0.0005154	0.2491533	99.98472	99.96028	200.0481	199.9450	0.1030462	0.1030462
10	0.0002321	0.6038112	99.94002	99.94139	199.9278	199.8814	0.0463868	0.0463868
11	0.0005929	0.1848895	99.99015	99.99617	200.1049	199.9863	0.1185795	0.1185795
12	-0.0006464	0.1483490	99.99653	99.94682	199.8141	199.9433	-0.1292428	-0.1292428
13	-0.0002263	0.6127855	99.95193	100.09496	200.0016	200.0469	-0.0452777	-0.0452777
14	0.0001560	0.7272108	100.13666	99.89418	200.0620	200.0308	0.0312058	0.0312058
15	0.0006594	0.1403860	100.03624	99.94274	200.1108	199.9790	0.1318565	0.1318565
16	0.0004778	0.2853648	99.92304	100.07177	200.0904	199.9948	0.0955532	0.0955532
17	-0.0003282	0.4630601	100.01961	100.06666	200.0206	200.0863	-0.0656629	-0.0656629
18	-0.0001478	0.7410526	99.96112	100.17508	200.1066	200.1362	-0.0295773	-0.0295773
19	-0.0006194	0.1660747	99.99906	99.94166	199.8169	199.9407	-0.1238349	-0.1238349
20	-0.0006640	0.1376396	99.97401	100.05617	199.8974	200.0302	-0.1328101	-0.1328101
21	0.0003286	0.4624798	99.90516	99.98661	199.9575	199.8918	0.0656842	0.0656842
22	0.0003411	0.4455869	100.03523	99.96079	200.0642	199.9960	0.0682249	0.0682249
23	-0.0000975	0.8274028	99.93787	99.95452	199.8729	199.8924	-0.0194910	-0.0194910
25	0.0002461	0.5820510	99.98915	100.07462	200.1130	200.0638	0.0492442	0.0492442
26	0.0002765	0.5364458	100.03539	99.96041	200.0511	199.9958	0.0552920	0.0552920
27	0.0000876	0.8447762	99.95505	99.98080	199.9534	199.9358	0.0175062	0.0175062
28	0.0003500	0.4338142	100.03738	99.97071	200.0781	200.0081	0.0700080	0.0700080
29	-0.0002572	0.5652014	99.96146	99.97635	199.8864	199.9378	-0.0514255	-0.0514255
30	0.0000473	0.9157339	99.95647	99.97285	199.9388	199.9293	0.0094605	0.0094605
31	0.0000813	0.8557782	100.03076	99.98796	200.0350	200.0187	0.0162578	0.0162578
32	0.0002302	0.6067843	100.04779	99.98003	200.0739	200.0278	0.0460398	0.0460398
33	0.0004089	0.3604991	100.03151	100.06965	200.1830	200.1012	0.0818291	0.0818291
34	0.0006007	0.1791799	100.13358	100.03684	200.2907	200.1704	0.1202494	0.1202494
35	0.0001666	0.7094596	100.05569	100.06161	200.1507	200.1173	0.0333443	0.0333443
36	-0.0008561	0.0555704	100.08053	100.04790	199.9571	200.1284	-0.1713382	-0.1713382
37	-0.0000164	0.9707969	100.00381	100.01701	200.0175	200.0208	-0.0032747	-0.0032747
38	-0.0000898	0.8408415	100.17939	99.92802	200.0894	200.1074	-0.0179713	-0.0179713
39	-0.0001503	0.7367628	100.00455	100.01916	199.9936	200.0237	-0.0300691	-0.0300691
40	-0.0000712	0.8734625	99.97557	100.07320	200.0345	200.0488	-0.0142483	-0.0142483
41	0.0000246	0.9560569	100.00854	100.02151	200.0350	200.0300	0.0049292	0.0049292
42	0.0004766	0.2865886	100.10682	100.07505	200.2773	200.1819	0.0954000	0.0954000
43	-0.0000857	0.8479615	100.00054	100.02174	200.0051	200.0223	-0.0171499	-0.0171499
44	-0.0004231	0.3440610	100.09685	100.00361	200.0158	200.1005	-0.0846707	-0.0846707
45	-0.0006599	0.1400776	99.96225	100.01485	199.8451	199.9771	-0.1319578	-0.1319578

Table 3: 乱数生成した二つのデータが独立でない場合

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
8	-0.0009668	0.0306226	100.0167	99.96575	199.7891	199.9825	-0.1933531	-0.1933531
24	-0.0010322	0.0209940	100.1072	100.03828	199.9389	200.1455	-0.2065922	-0.2065922

$$\text{加法 1} = \text{var}(x + y), \text{加法 2} = \text{var}(x) + \text{var}(y)$$

乱数生成したデータをランダムサンプリングした場合

Table 4: ランダムサンプリングしたデータの分散

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
1	-0.0000485	0.9135710	100.04656	99.92857	199.9654	199.9751	-0.0097065	-0.0097065
2	-0.0000652	0.8841659	99.98215	99.99819	199.9673	199.9803	-0.0130297	-0.0130297
3	-0.0002734	0.5409125	100.04045	99.94448	199.9302	199.9849	-0.0546842	-0.0546842
4	0.0002097	0.6390678	100.03122	99.91623	199.9894	199.9474	0.0419379	0.0419379
5	-0.0000656	0.8833763	100.06148	99.98887	200.0372	200.0504	-0.0131237	-0.0131237
6	-0.0000562	0.8999136	100.04036	99.87487	199.9040	199.9152	-0.0112445	-0.0112445
7	0.0000587	0.8955571	99.88474	99.98465	199.8811	199.8694	0.0117340	0.0117340
9	0.0000324	0.9422655	100.18557	100.01365	200.2057	200.1992	0.0064841	0.0064841
10	0.0004657	0.2977174	100.00307	99.91771	200.0139	199.9208	0.0931035	0.0931035
11	-0.0004638	0.2997043	99.94262	99.95997	199.8099	199.9026	-0.0927130	-0.0927130
12	-0.0002869	0.5211490	100.06612	100.07150	200.0802	200.1376	-0.0574237	-0.0574237
13	0.0003968	0.3749637	99.93863	100.00096	200.0189	199.9396	0.0793309	0.0793309
14	0.0008604	0.0543707	99.98128	99.98778	200.1411	199.9691	0.1720497	0.1720497
15	0.0004709	0.2923753	99.98044	100.02723	200.1019	200.0077	0.0941801	0.0941801
16	-0.0000638	0.8866115	99.98576	99.90296	199.8760	199.8887	-0.0127468	-0.0127468
17	-0.0001013	0.8207471	100.08097	100.07135	200.1320	200.1523	-0.0202817	-0.0202817
18	-0.0003408	0.4460659	100.08435	99.94296	199.9591	200.0273	-0.0681638	-0.0681638
19	-0.0005406	0.2267052	99.97086	99.75277	199.6156	199.7236	-0.1079769	-0.1079769
20	-0.0002751	0.5385090	100.00111	99.96321	199.9093	199.9643	-0.0550037	-0.0550037
21	-0.0000966	0.8289283	99.79315	100.03184	199.8057	199.8250	-0.0193095	-0.0193095
22	-0.0004900	0.2732492	100.04193	99.81521	199.7592	199.8571	-0.0979244	-0.0979244
23	-0.0000384	0.9315195	99.98269	99.99324	199.9682	199.9759	-0.0076852	-0.0076852
24	0.0000409	0.9270843	100.09771	100.04714	200.1530	200.1448	0.0081912	0.0081912
25	0.0001548	0.7292396	100.02626	99.98853	200.0458	200.0148	0.0309617	0.0309617
26	0.0000006	0.9989683	99.91310	99.97783	199.8911	199.8909	0.0001156	0.0001156
27	-0.0004062	0.3636902	99.95707	99.97077	199.8466	199.9278	-0.0812165	-0.0812165
28	0.0004554	0.3085175	99.97648	100.03075	200.0983	200.0072	0.0910862	0.0910862
29	-0.0002536	0.5707257	99.97185	99.96773	199.8889	199.9396	-0.0506972	-0.0506972
30	0.0005116	0.2525925	100.07632	100.03861	200.2173	200.1149	0.1023880	0.1023880
31	-0.0000701	0.8755172	100.09967	99.96810	200.0537	200.0678	-0.0140163	-0.0140163
32	0.0002323	0.6033995	99.82793	100.03516	199.9095	199.8631	0.0464353	0.0464353
33	0.0007507	0.0932231	100.09395	99.96132	200.2055	200.0553	0.1501832	0.1501832
34	-0.0002231	0.6178539	99.77223	100.08013	199.8078	199.8524	-0.0445895	-0.0445895
35	0.0003078	0.4913172	100.05562	99.93641	200.0536	199.9920	0.0615534	0.0615534
36	0.0000576	0.8974954	99.94412	99.96231	199.9180	199.9064	0.0115172	0.0115172
37	-0.0006663	0.1362344	99.98475	99.93338	199.7849	199.9181	-0.1332117	-0.1332117
38	-0.0000595	0.8942027	100.13083	100.05302	200.1719	200.1838	-0.0119058	-0.0119058
39	0.0005318	0.2343975	100.00880	99.95418	200.0693	199.9630	0.1063373	0.1063373
40	-0.0001554	0.7281648	100.10111	99.86696	199.9370	199.9681	-0.0310824	-0.0310824
41	0.0002603	0.5604638	99.93540	100.00392	199.9914	199.9393	0.0520535	0.0520535
42	0.0002006	0.6537639	100.09478	99.95661	200.0915	200.0514	0.0401289	0.0401289
43	0.0005878	0.1887052	100.10992	99.98966	200.2172	200.0996	0.1176238	0.1176238
44	0.0000720	0.8720832	100.00396	99.89731	199.9157	199.9013	0.0143943	0.0143943
45	-0.0002899	0.5168766	100.04341	99.96053	199.9460	200.0039	-0.0579750	-0.0579750

Table 5: ランダムサンプリングしたデータが独立でない場合

No	相関係数	p 値	標本 x	標本 y	加法 1	加法 2	差異	cov2
8	0.0011228	0.0120532	100.0812	100.2103	200.5164	200.2915	0.2248807	0.2248807

$$\text{加法 1} = \text{var}(x + y), \text{ 加法 2} = \text{var}(x) + \text{var}(y)$$

まとめ

データが独立であれば分散の加法性が成り立っていることがわかります。データが独立とは言い難い無相関の検定が成功するケース (95% 信頼区間に 0 が入らない) では、分散の差 (共分散の 2 倍数) が一桁大きいので加法性が成り立っているとは言い難いように思えますがこのケースでは数値だけを見ている限り差はよくわかりません。

`cor.test()` 関数について

`cor.test()` 関数は無相関の検定を行う関数です。対立仮説 (H_1) は下記の出力の通り「true correlation is **not** equal to 0 (相関係数はゼロではない)」ですので、帰無仮説 (H_0) は「相関係数はゼロである (相関はない)」となります。有意水準 α で検定が失敗すれば (帰無仮説が棄却されない、 $p \geq \alpha$ である) 帰無仮説が採択されますので相関係数はゼロ (データ間には相関がない) と考えられます。

```
##
## Pearson's product-moment correlation
##
## data:  rnorm(n) and rnorm(n)
## t = -0.71677, df = 4999998, p-value = 0.4735
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.0011970704  0.0005559746
## sample estimates:
##          cor
## -0.0003205482
```

Appendix

About handout style

The Tufte handout style is a style that Edward Tufte uses in his books and handouts. Tufte's style is known for its extensive use of sidenotes, tight integration of graphics with text, and well-set typography. This style has been implemented in LaTeX and HTML/CSS⁴, respectively.

⁴ See Github repositories `tufte-latex` and `tufte-css`