Short communication

What's the distribution of variance ratio?

L.V. Nedorezov

Research Center for Interdisciplinary Environmental Cooperation RAS, Saint-Petersburg, Russia
E-mail l.v.nedorezov@gmail.com

Keywords: Kolmogorov – Smirnov test, Normal distribution, Monte-Carlo methods, Analysis of variance, Distribution of variance ratio

There are two basic requirements for application of ANOVA to analysis of sample (Kruskal, Wallis, 1952; Hudson, 1970; Lakin, 1990; Kobzar, 2006 and others): elements of sample must correspond to Normal distribution, and variances of all gradations of considering factor A must be close to each other. Let's assume that x_{ij} for all i and j are values of independent stochastic variables with Normal distribution and with equal averages (below we'll consider a case when all averages are equal to one). After providing of experiments (or observations) we get a set of numbers x_{11}, x_{12},…, x_{1m_1} which correspond to first gradation of factor A. Number m_1 is a sub-sample size. We have also numbers x_{21}, x_{22},…, x_{2m_2} corresponding to second gradation of factor and so on. Last part of initial sample is x_{n_A1}, x_{n_A2},…, $x_{n_Am_{n_A}}$. Let N be a sample size, $N = m_1 + m_2 + ... + m_{n_A}$.

Let's also assume that sample variances of factor gradations are equal. In such a situation factor must be identified as weak regulator which has no influence on values of considering characteristic.

Following sequence of operations is standard (Lakin, 1990): first of all, for whole sample x_{11}, x_{12},…, $x_{n_Am_{n_A}}$ average $\bar{x}$ is calculated; after that total sum of squared deviations is determined:

$$D_y = \sum_{ij} (x_{ij} - \bar{x})^2 . \qquad (1)$$

In expression (1) summarizing is provided for all possible values of i and j (below we'll consider a particular case when $i, j = 1,2,3,4,5$; in other words, we'll assume that we have five gradations of factor, sample size is equal to 25, $N = 25$, and we have 5 elements for every gradation of factor, $m_1 = ... = m_5 = 5$). D_y is called total deviate.

Next step of process is following: for factor's gradation with number i average $\bar{x}_i$ is calculated, and it continuous for all i. After that squared deviations between averages $\bar{x}_i$ and total average $\bar{x}$ are summarized (under taking into account of sub-sample sizes for all gradations):

$$D_x = \sum_{i=1}^{n_A} \frac{m_i}{N} (\bar{x} - \bar{x}_i)^2 . \tag{2}$$

D_x is called between group deviate (Lakin, 1990). Intragroup deviate D_e is determined by the following expression:

$$D_e = \sum_{i=1}^{n_A} \left[\sum_{j=1}^{m_i} (x_{ij} - \bar{x}_i)^2 \right]. \tag{3}$$

Following relation between three deviates is truthful:

$$D_y = D_x + D_e .$$

For every deviate (1)-(3) there is a certain number of degrees of freedom: $k_y = N - 1$ (for D_y), $k_x = n_A - 1$ (for D_x) and $k_e = k_y - k_x = N - n_A$ (for D_e). We can divide deviates (1)-(3) onto respective number of degrees of freedom, and we get sample variances:

$$s_y^2 = \frac{D_y}{N-1},\ s_x^2 = \frac{D_x}{n_A - 1},\ s_e^2 = \frac{D_e}{N - n_A} .$$

s_y^2 is total sample variance for all initial values; s_x^2 is between group sample variance (factorial variance); and s_e^2 is intragroup sample variance (residual variance).

Final solution (about influence of factor A onto values of characteristics) is based on relation of two sample variances (variance ratio):

$$F_{fact} = \frac{s_x^2}{s_e^2} . \tag{4}$$

Amount (4) we have to compare with table value for Fisher distribution with fixed value of confidence level and numbers of degrees of freedom k_x and k_e. If amount (4) is bigger than table value of Fisher distribution we say that factor A has statistically confident influence on value of characteristics. Note, that we can compare variance ratio with Fisher distribution if and only if (4) has Fisher distribution. Below we'll check a correspondence between variance ration with Fisher distribution using Kolmogorov criterion (Bolshev, Smirnov, 1983) with various confidence levels, and artificial example (with known final results). In our previous publications

(Nedorezov, 2016, 2017 a, b) we tried to discuss problems of analysis of variance, and in particular absence of correspondence between Fisher distribution and distribution of variance ratio. Current publication gives additional arguments for our position.

Procedure described above is well-known, and can be found in various textbooks. But now let’s consider an artificial situation when elements of sample were modeled in the following way:

$$x_{ij} = 1 + \sigma\xi_{ij}, \quad (5)$$

where ξ_{ij} are independent normally distributed stochastic variables with parameters 0,1. Thus, x_{ij} are independent normally distributed stochastic variables with average equal to one and variance equal to σ^2. Amounts of ξ_{ij} were obtained as

$$\xi = \sqrt{\frac{12}{n}}\left(\sum_{k=1}^{n}\alpha_k - \frac{n}{2}\right), \quad (6)$$

where α_k are independent stochastic variables with uniform distribution (Rnd), $n = 12$. For every fixed value of σ a certain number of samples were found with formulas (5) and (6). For every sample variance ratio was calculated. After that dataset of variance ratios (which were modeled independently) was analyzed by Kolmogorov’s criterion onto correspondence to Fisher distribution (with parameters 4 and 20). Results of modeling and analysis are presented in Tables 1-3 (for various values of σ and sample sizes; note, in considering case samples contain values of variance ratios only). Pointed out in tables 1-3 significance levels (5%, 1%, and 0.1%) are critical levels for Kolmogorov’s criterion. For determination of value of every element of table 1 we tested 10^4 independent values of samples of variance ratios.

If we assume that distribution of variance ratios corresponds to Fisher distribution then increasing of sample size must lead asymptotically to decreasing of number of cases when Kolomogorov’s test allows rejecting hypothesis about correspondence of distribution of variance ratios to Fisher distribution. But obtained results (tables 1-3) demonstrate inverse picture: number of negative results increases monotonously. Moreover, number of negative results is much bigger then critical levels of Kolomogorov’s test. It allows concluding that distribution of variance ratios doesn’t correspond to Fisher distribution.

Table 1 (part 1)
Results for 5% significance level

σ	Sample size								
	8	9	10	11	12	13	14	15	16
0.1	0.1169	0.1276	0.1448	0.1575	0.1738	0.1858	0.1952	0.2148	0.2273
0.2	0.115	0.13	0.1455	0.1558	0.1746	0.1867	0.2034	0.2129	0.2282
0.3	0.1086	0.1267	0.1332	0.1491	0.1606	0.174	0.1808	0.1918	0.2121
0.4	0.1187	0.1313	0.1444	0.1642	0.1729	0.1931	0.2085	0.2222	0.2385
0.5	0.1182	0.138	0.1475	0.1601	0.1702	0.1784	0.193	0.2068	0.2189
0.6	0.1031	0.1174	0.1318	0.1425	0.16	0.1716	0.1893	0.1957	0.2104
0.7	0.1234	0.137	0.1568	0.167	0.1799	0.1997	0.2118	0.2275	0.2455
0.8	0.123	0.1374	0.1549	0.1721	0.1899	0.2049	0.22	0.2351	0.2441
0.9	0.111	0.1221	0.143	0.1487	0.1644	0.178	0.1894	0.2051	0.2239
1.0	0.1097	0.1241	0.1363	0.1546	0.1652	0.1805	0.1893	0.2131	0.2216
1.1	0.1256	0.1443	0.1586	0.1722	0.191	0.2038	0.2176	0.2314	0.2399
1.2	0.1128	0.1248	0.137	0.1468	0.1606	0.1769	0.1854	0.203	0.2173
1.3	0.1122	0.1262	0.1394	0.155	0.1658	0.1818	0.1983	0.2129	0.2258
1.4	0.1148	0.1284	0.1426	0.1552	0.1742	0.1931	0.2051	0.2148	0.2337
1.5	0.121	0.1327	0.1443	0.1678	0.1751	0.1914	0.206	0.2192	0.2343
1.6	0.1181	0.1273	0.1396	0.1533	0.1639	0.1796	0.1907	0.2095	0.2158
1.7	0.115	0.1251	0.1461	0.1543	0.171	0.1871	0.2022	0.2167	0.2335
1.8	0.1098	0.1288	0.1371	0.1502	0.1631	0.1761	0.1905	0.2023	0.2155
1.9	0.1264	0.1411	0.1599	0.1785	0.189	0.2018	0.2086	0.2311	0.2352
2.0	0.1001	0.1109	0.1256	0.1382	0.1476	0.1621	0.1737	0.1841	0.1925

Table 1 (part 2)
Results for 5% significance level

σ	Sample size								
	17	18	19	20	21	22	23	24	25
0.1	0.2436	0.2607	0.2727	0.286	0.3075	0.3136	0.3304	0.3462	0.356
0.2	0.2357	0.2461	0.2666	0.2719	0.2907	0.3056	0.3241	0.3342	0.348
0.3	0.2223	0.2289	0.2439	0.2583	0.2686	0.2924	0.3022	0.3112	0.3351
0.4	0.2502	0.2709	0.2786	0.2889	0.3048	0.3224	0.34	0.3542	0.3736
0.5	0.2291	0.2478	0.2587	0.2694	0.2855	0.2991	0.3078	0.3232	0.3325
0.6	0.222	0.2267	0.2403	0.2495	0.2667	0.2779	0.2941	0.3043	0.3168
0.7	0.26	0.2755	0.2856	0.3001	0.3208	0.3283	0.3485	0.356	0.3709
0.8	0.2621	0.2775	0.2894	0.3085	0.3179	0.3312	0.3517	0.3603	0.3799
0.9	0.2336	0.2544	0.263	0.2775	0.2863	0.2985	0.314	0.3278	0.3432
1.0	0.2352	0.2485	0.2699	0.2806	0.287	0.3099	0.3195	0.3268	0.3488
1.1	0.2592	0.2732	0.2806	0.2912	0.3105	0.3186	0.3385	0.3589	0.3693
1.2	0.2273	0.2445	0.2604	0.269	0.2826	0.3046	0.3185	0.3266	0.3426
1.3	0.2493	0.2617	0.2736	0.2902	0.301	0.3099	0.3286	0.3448	0.3603
1.4	0.2405	0.2575	0.2695	0.2789	0.2976	0.3091	0.326	0.3397	0.3503
1.5	0.2509	0.261	0.2727	0.2961	0.3034	0.3158	0.331	0.335	0.3515
1.6	0.2354	0.2524	0.264	0.2763	0.2871	0.3066	0.321	0.3262	0.3428
1.7	0.2455	0.2649	0.2755	0.2859	0.3071	0.3255	0.3374	0.3561	0.3718
1.8	0.2279	0.2525	0.2563	0.2715	0.2812	0.2919	0.3101	0.3223	0.3244
1.9	0.25	0.2617	0.2675	0.3012	0.3019	0.3156	0.3281	0.3432	0.361
2.0	0.21	0.2197	0.2407	0.2484	0.2609	0.2767	0.2901	0.3042	0.3172

Table 2 (part 1)
Results for 1% significance level

σ	Sample size								
	8	9	10	11	12	13	14	15	16
0.1	0.0284	0.0372	0.0427	0.0503	0.0558	0.0597	0.0721	0.0751	0.0809
0.2	0.0303	0.0335	0.0425	0.0466	0.0541	0.0589	0.0656	0.0741	0.0791
0.3	0.0278	0.0321	0.0386	0.0441	0.05	0.0524	0.0582	0.0607	0.0657
0.4	0.0275	0.0323	0.0424	0.0455	0.0459	0.0585	0.0599	0.0678	0.0746
0.5	0.032	0.0388	0.0423	0.0464	0.0528	0.0564	0.0648	0.0696	0.0786
0.6	0.0275	0.0301	0.0374	0.0388	0.0464	0.0498	0.0543	0.0609	0.0681
0.7	0.032	0.038	0.0438	0.0495	0.0564	0.0604	0.0682	0.074	0.0845
0.8	0.0329	0.0373	0.0448	0.0547	0.0618	0.068	0.0724	0.0824	0.0922
0.9	0.0313	0.0356	0.0413	0.0473	0.0514	0.0599	0.0649	0.072	0.077
1.0	0.0313	0.0331	0.0423	0.047	0.0536	0.0607	0.0605	0.0689	0.074
1.1	0.0341	0.0417	0.05	0.0544	0.0661	0.0727	0.0813	0.0893	0.0951
1.2	0.0305	0.0354	0.0413	0.0453	0.0521	0.056	0.0624	0.0672	0.0705
1.3	0.0307	0.0335	0.0407	0.0444	0.052	0.0562	0.0599	0.07	0.0764
1.4	0.0296	0.034	0.041	0.0455	0.0532	0.0547	0.0645	0.074	0.0764
1.5	0.0314	0.0361	0.0446	0.0472	0.0546	0.0614	0.0665	0.0734	0.0809
1.6	0.0308	0.0377	0.0417	0.0477	0.0514	0.0565	0.0629	0.0688	0.0774
1.7	0.0333	0.0333	0.0412	0.0438	0.0523	0.0582	0.0639	0.0719	0.0812
1.8	0.0309	0.0371	0.0418	0.0476	0.0525	0.0569	0.0647	0.0677	0.0786
1.9	0.0351	0.0401	0.0471	0.0498	0.0607	0.0712	0.0738	0.0891	0.0933
2.0	0.0304	0.0352	0.0381	0.0418	0.0473	0.0539	0.0545	0.0598	0.0639

Table 2 (part 2)
Results for 1% significance level

σ	Sample size								
	17	18	19	20	21	22	23	24	25
0.1	0.0882	0.097	0.1047	0.1131	0.1206	0.1284	0.1427	0.1416	0.1521
0.2	0.0854	0.0929	0.1001	0.1056	0.1116	0.1241	0.1229	0.1351	0.1407
0.3	0.0721	0.077	0.0821	0.09	0.0951	0.1018	0.1075	0.1202	0.1236
0.4	0.0786	0.0903	0.0973	0.1091	0.1129	0.1256	0.1286	0.1413	0.1548
0.5	0.0844	0.0899	0.1014	0.1051	0.1178	0.1226	0.1305	0.1405	0.1438
0.6	0.0738	0.0759	0.086	0.0921	0.0949	0.1025	0.1067	0.1071	0.1236
0.7	0.0893	0.1013	0.1082	0.1151	0.1193	0.1311	0.1428	0.1458	0.1553
0.8	0.1003	0.1065	0.1161	0.1265	0.1374	0.1463	0.1552	0.1627	0.1751
0.9	0.08	0.0896	0.0966	0.1045	0.1087	0.1212	0.124	0.1346	0.138
1.0	0.0792	0.0859	0.0882	0.0957	0.1028	0.1104	0.1125	0.1224	0.1312
1.1	0.1052	0.1108	0.1211	0.1259	0.1347	0.1402	0.1516	0.1573	0.1673
1.2	0.0784	0.0834	0.0905	0.095	0.1023	0.1108	0.1153	0.1273	0.1298
1.3	0.08	0.0903	0.099	0.1103	0.118	0.1211	0.1263	0.1387	0.1433
1.4	0.0847	0.0932	0.0987	0.1049	0.1104	0.1235	0.1231	0.1349	0.1403
1.5	0.0889	0.1	0.1066	0.1156	0.1225	0.1254	0.1375	0.1425	0.148
1.6	0.0833	0.0914	0.0945	0.1041	0.1138	0.1198	0.1289	0.138	0.1496
1.7	0.0875	0.0972	0.1014	0.1145	0.1187	0.1293	0.1387	0.149	0.1537
1.8	0.0805	0.0931	0.0959	0.1038	0.1104	0.1243	0.1253	0.1383	0.1408
1.9	0.0987	0.1117	0.1134	0.1229	0.1275	0.1336	0.1391	0.1476	0.1557
2.0	0.0675	0.0716	0.0795	0.0837	0.0879	0.0982	0.101	0.103	0.1133

Table 3 (part 1)
Results for 0.1% significance level

σ	Sample size								
	8	9	10	11	12	13	14	15	16
0.1	0.0029	0.005	0.0064	0.0088	0.0104	0.0108	0.0134	0.0146	0.0175
0.2	0.0033	0.0047	0.0047	0.0063	0.0079	0.0087	0.0116	0.0119	0.0131
0.3	0.0025	0.0028	0.0051	0.0048	0.0064	0.0084	0.0092	0.011	0.0138
0.4	0.0044	0.0045	0.0067	0.008	0.0081	0.0099	0.0091	0.0114	0.0119
0.5	0.0031	0.005	0.0054	0.0068	0.0078	0.0085	0.0099	0.0111	0.0133
0.6	0.0025	0.0037	0.0038	0.0047	0.0065	0.0093	0.0083	0.0104	0.011
0.7	0.0021	0.0046	0.0059	0.008	0.0081	0.0085	0.0101	0.0114	0.0125
0.8	0.0041	0.0057	0.0071	0.0094	0.0094	0.0115	0.012	0.0129	0.0179
0.9	0.003	0.0048	0.0053	0.0076	0.0071	0.0104	0.0118	0.0141	0.0142
1.0	0.0045	0.005	0.0065	0.0069	0.008	0.0086	0.0084	0.0118	0.0116
1.1	0.0036	0.0059	0.0071	0.0091	0.01	0.0107	0.0156	0.0162	0.0205
1.2	0.0014	0.0033	0.0032	0.007	0.0083	0.0099	0.0108	0.0129	0.0152
1.3	0.0039	0.0046	0.005	0.0055	0.0068	0.0086	0.0091	0.0107	0.0114
1.4	0.0025	0.0039	0.0035	0.0058	0.007	0.0072	0.0105	0.0101	0.0116
1.5	0.004	0.0055	0.0062	0.0073	0.0091	0.0111	0.013	0.015	0.0174
1.6	0.004	0.0063	0.0062	0.0071	0.0085	0.0086	0.011	0.0122	0.0135
1.7	0.0036	0.004	0.0048	0.0056	0.0061	0.0075	0.0092	0.0116	0.0139
1.8	0.0032	0.0042	0.0049	0.0073	0.0077	0.0103	0.0109	0.0136	0.0131
1.9	0.0039	0.0068	0.0073	0.0074	0.0106	0.0105	0.0136	0.0154	0.0178
2.0	0.0028	0.0069	0.0073	0.0084	0.0103	0.0126	0.0118	0.0147	0.0156

Table 3 (part 2)
Results for 0.1% significance level

σ	Sample size								
	17	18	19	20	21	22	23	24	25
0.1	0.0186	0.0206	0.0237	0.0252	0.0302	0.0299	0.0328	0.0353	0.0393
0.2	0.0156	0.0189	0.0213	0.023	0.0259	0.0287	0.0308	0.0307	0.036
0.3	0.0139	0.0156	0.0171	0.018	0.0212	0.0208	0.0234	0.0245	0.0268
0.4	0.0139	0.0158	0.0179	0.0192	0.0212	0.0229	0.026	0.0274	0.0323
0.5	0.0154	0.0172	0.0185	0.0218	0.0241	0.029	0.0302	0.0339	0.0388
0.6	0.0129	0.0143	0.0159	0.0182	0.0185	0.0231	0.0236	0.0284	0.0274
0.7	0.0143	0.0154	0.0183	0.0196	0.0219	0.0262	0.0289	0.0311	0.035
0.8	0.0191	0.0211	0.0222	0.0261	0.0302	0.0298	0.036	0.0395	0.0426
0.9	0.0172	0.0197	0.0195	0.0227	0.025	0.0251	0.0254	0.0307	0.0319
1.0	0.0137	0.0137	0.0148	0.0167	0.0189	0.0201	0.0187	0.0239	0.0228
1.1	0.0225	0.0251	0.0256	0.0295	0.0307	0.0332	0.0386	0.0401	0.0407
1.2	0.0178	0.0187	0.019	0.02	0.0235	0.0227	0.0244	0.0265	0.0261
1.3	0.0144	0.0143	0.017	0.0213	0.0209	0.0254	0.0277	0.0298	0.033
1.4	0.0149	0.0178	0.0195	0.0211	0.0249	0.0273	0.029	0.0295	0.0342
1.5	0.0164	0.0191	0.0222	0.02	0.0225	0.0239	0.0279	0.0294	0.0294
1.6	0.0168	0.0164	0.0178	0.0212	0.023	0.0249	0.0269	0.0305	0.0311
1.7	0.0148	0.0164	0.0211	0.0214	0.0231	0.0251	0.0269	0.0286	0.0331
1.8	0.0167	0.0161	0.0174	0.0195	0.0181	0.0231	0.026	0.0249	0.03
1.9	0.0192	0.0233	0.0237	0.0289	0.03	0.0325	0.0346	0.0356	0.0383
2.0	0.0183	0.0187	0.0183	0.0224	0.0222	0.0232	0.0254	0.0271	0.0282

References

Bolshev L.N., Smirnov N.V. 1983. Tables of Mathematical Statistics. Moscow: Nauka, 416 p.

Hudson D.J. 1970. Statistics for physicists. Moscow: Mir, 296 p.

Kobzar A.I. 2006. Applied Mathematical Statistics. For engineers and scientists. Moscow: Fizmatlit, 816 p.

Kruskal W.H., Wallis W.A. 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260): 583-621

Lakin G.F. 1990. Biometrics. Moscow: Visshaya Shkola, 352 p.

Nedorezov L.V. 2016. Shine and poverty of Ordinary Least Squares// Samarskaya Luka: Problems of Regional and Global Ecology 25(1): 13-17.

Nedorezov L.V. 2017 a. Linear Regression as Great Joke of Great Scientists// Proc. of the International Academy of Ecology and Environmental Sciences 7(3): 67-77.

Nedorezov L.V. 2017 b. Analysis of Variance: Comfortless Questions// Computational Ecology and Software, 7(3): 91-99.