Article

Analysis of variance: Comfortless questions

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

Received 13 April 2017; *Accepted 20 May 2017*; *Published 1 September 2017*

Abstract
In this paper the simplest variant of analysis of variance is under consideration. Three examples from textbooks by Lakin (1990) and Rokitsky (1973) were re-considered. It was obtained that traditional one-way ANOVA and Kruskal – Wallis criterion can lead to unreal results about factor's influence on value of characteristics. Alternative way to solution of the same problem is under consideration too.

Keywords analysis of variance; tests for normality; goodness measure.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction with Comfortless Questions

Let's consider shortly the simplest case of analysis of variance when we have to determine or to estimate an influence of any factor A on measured characteristics. Final result may be of two types only: factor has confidence influence or not. Factors may have various natures: it can be level of precipitation, increment of tree's diameter, surviving and productivity of insects, concentration of any metal in water and so on (if factor has its own scale of measurements). Sometimes factor may have no scales: for example, it can correspond to Latin names of species if we want to compare different species and take into account some separated characteristics.

First of all, we have to determine so-called gradations of factor A: it must be a system of non-intersected intervals within a range of changing of values of factor (if factor has numerical scale). If factor corresponds to Latin names of species we have determined number of gradations. Let n_A be a number of factor's gradations.

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 $x_{n_A 1}$, $x_{n_A 2}$,…, $x_{n_A m_{n_A}}$. Let N be a sample size, $N = m_1 + m_2 + ... + m_{n_A}$.

Two various questions now arise: does it possible to find a way for correct selection of factor's gradations or not? Honest answer is following: there are no ways for correct selection of factor's gradations, and selection process is based on researcher's experience only. May be, it is impossible to create a common rule for selection of gradations for all possible types of factors. The second question is: is there any influence of selection of factor's gradations onto final results? And again, honest answer is following: it has strong influence on final results.

As it was pointed out in popular textbook on biometrics by Lakin (1990, page 157) "for correct application of analysis of variance we have to have initial sample with distribution which is Normal or close to Normal… And it is also important that variances of all groups (corresponding to all gradations of factor) of numbers are rather close to each other." It is necessary to note that in biology assumption about normality is abnormal (Nedorezov, 2012, 2015). Now we have several questions again: can we say that all researchers provide testing on normality of initial samples before applying of analysis of variance? Can we say that all researchers compare variances for sub-samples, and what's a base for conclusion that observed differences between variances are inappreciable?

Let x_{ij} for all i and j are values of independent stochastic variables with normal distribution. Let's also assume that variances and sample variances are close to each other. In such a situation following sequence of operations is standard (Lakin, 1990; Vasiliev and Melnikova, 2009): first of all, for whole sample x_{11}, x_{12},…, $x_{n_A m_{n_A}}$ 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. 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):

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

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

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

It is important to note now: for calculation of values of all deviates (1)-(3) all elements of initial sample were used. What does it mean? It means that we have no reasons to assume that (1), (2) and (3) are values of independent stochastic variables. For simplest demonstration that (1), (2) and (3) are functionally dependent amounts we can assume that all elements of initial sample are fixed excluding x_{11} (result of first observation or experiment of first gradation). It is obvious that for every concrete value of x_{11} deviates (1)-(3) will have concrete unique values. In such a case formulas (1)-(3) can be considered as parametric presentation of curve in three-dimensional space.

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 (Z.1)-(Z.3) onto respective number of degrees of freedom 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):

$$F_{fact} = \frac{s_x^2}{s_e^2}. \quad (4)$$

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.

And now new question arises: what is a base for comparison of amount (4) with table value for Fisher distribution? It may take a place if and only if relation (4) has direct relation to Fisher distribution. But it is not truth: following a definition of Fisher distribution (Mathematical encyclopedia, Vol. 5, 1985, page 626) we have that it is the relation of two **independent** stochastic variables with chi-squared distributions. As it was pointed out above (4) is the relation of two **functionally dependent** amounts. Thus, for obtaining a final decision about role of factor we compare amount (4) with table value for Fisher distribution but we have no background for this comparison.

Analysis of basic procedure of analysis of variances (when we have one factor only) allowed showing that there are several important questions (presented above) which demonstrate that there are serious problems for application of analysis of variances to real datasets. These questions have direct relation to selection of factor's gradations, to checking of normality of initial sample and sub-samples, and to presenting of background for use of Fisher distribution in procedure.

For solution of the same problem of analysis of variances Kruskal – Wallis test (Kruskal and Wallis, 1952; Likes and Laga, 1985) can also be used. Use of this test doesn't require normality for elements of initial sample, and final result does not depend on monotonic transformation of values of sample. At the same time final verdict about role of factor is based on use of Fisher distribution (see, for example, Kruskal – Wallis approximation and Iman – Davenport approximation; Iman and Davenport, 1976; Kobzar, 2006) in a situation when using statistics doesn't correspond to this distribution.

2 Alternative Way

In both cases – as at use of parametric analysis of variance, as at use of Kruskal – Wallis rank test, – we calculate value of one or other statistic, and this value is used for final determination of role of factor. For example, in first case it is fraction of two sample variances. At the same time it is possible to point out simpler and logic following procedure for estimation of role of factor. First of all, we have to choose significance level α, and after that we can calculate $p-value$ for all pairs of factor's gradations for selected criterion of homogeneity of two samples. For example, it can be Kolmogorov – Smirnov criterion (Hollander and Wolfe, 1973; Bolshev and Smirnov, 1983). In a result of application of Kolmogorov – Smirnov criterion we obtain a symmetric matrix $\|z_{ij}\|$ with units on main diagonal. Every element z_{ij} of this matrix is a probability of error of first kind: we reject (in a result of use of Kolmogorov – Smirnov criterion) hypothesis about equivalence of sub-samples corresponding to i and j gradations of factor A when this hypothesis is truthful. It is obvious that for all i we have $z_{ii} = 1$: with probability one we have an error rejecting hypothesis about equivalence of two distribution functions which are based on one and the same sample.

Taking into account that matrix $\|z_{ij}\|$ is symmetric for analysis of factor's role we can use elements of matrix which are upper of main diagonal. It is obvious if the following inequality is truthful

$$\min_{i,j} z_{ij} > 0.05, \tag{5}$$

where $i = 1,..., n_A - 1$, $i + 1 \le j \le n_A$, then (with 5% significance level) we haven't background for rejecting of Null hypothesis (that factor hasn't a confidence influence on values of characteristic). If the following inequality is truthful

$$\max_{i,j} z_{ij} < 0.05, \tag{6}$$

for the same set of values of i and j and same significance level Null hypothesis must be rejected. If inequality (6) is truthful it allows concluding that factor has strong influence on values of characteristic.

For all other cases when we have values of $p - value$ which are bigger and smaller of critical level 0.05 we have to construct a graph. Every graph node corresponds to one factor's gradation, thus number of nodes is equal to n_A. Let's assume that between nodes i and j ($i = 1,..., n_A - 1$, $i + 1 \le j \le n_A$) we have unoriented edge of a graph if and only if $z_{ij} > 0.05$. And there is no edge if inverse inequality is truthful, $z_{ij} \le 0.05$.

If inequality (5) is truthful we have entire graph with $n_A(n_A - 1)/2$ edges ($i = 1,..., n_A - 1$, $i + 1 \le j \le n_A$). If inequality (6) is truthful we have a set of isolated nodes without edges. For all other situations it is possible to recommend following rules:

- If graph contains two or more isolated groups of nodes (i.e. there are no edges between nodes of these groups) it gives a background for conclusion that factor has confident influence on characteristic.
- If graph doesn't contain any isolated node or isolated group of nodes we have to determine number of existing edges; if this number is less than $n_A(n_A - 1)/4$ (50% of all possible total number of edges) we conclude that factor has confident influence on characteristic; if number of existing edges is bigger than $n_A(n_A - 1)/4$ we can conclude that factor hasn't confident influence on characteristic.

Note that 50% cannot be considered as final suggestion. It can be modified. May be, it is necessary to modify taking into account structure of graph.

3 Example 1

On the field of agrostation influence of methods of organic manuring onto crop of green mass of maize were analyzed (Table 59, page 162, Lakin, 1990). Results of experiments are presented in Table 1 (this is part of Table 59 from Lakin, 1990). As we can see there are four gradations of factor A, $n_A = 4$. For data from table 1 we have $s_e^2 = 11.54$ and $s_x^2 = 6.74$. Taking into account that inequality $s_x^2 < s_e^2$ is truthful (intragroup sample variance is bigger than factorial variance), author (Lakin, 1990) made a conclusion that factor hasn't confident influence on characteristic (crop yield).

Table 1 Results of experiments for four gradations of factor.

Gradations of factor A			
A_1	A_2	A_3	A_4
1.2	3.6	4.0	9.2
8.0	2.6	10.0	8.0
11.2	8.0	9.2	7.0

At the beginning we have to check initial sample onto normality. But for every gradation of factor experiments were repeated three times only. It means that there are no possibilities to check distributions of sub-samples on normality and check hypotheses about equivalence of variances. We can check normality for whole sample only. But as it was pointed out above, assumption about normality is abnormal (Nedorezov, 2012, 2015): this assumption means that with positive probability we may have negative values for crop yield.

For combined sample (12 values) we have: for Shapiro – Wilk test $p-value = 0.2738$; for Cramer – von-Mises test $p-value = 0.1177$; for Anderson – Darling test $p-value = 0.1737$; for Lilliefors test $p-value = 0.09252$; for chi-squared test $p-value = 0.1718$; for Shapiro – Francia test $p-value = 0.309$ (Shapiro and Wilk, 1965; Anderson and Darling, 1952, 1954; Lilliefors, 1967, 1969; Thode, 2002). Thus, all used criterions showed that $p-value$ isn't big; moreover, Lilliefors test shows that with 10% significance level we must reject Null hypothesis.

Application of Kruskal – Wallis test to data from the table 1 shows that $p-value = 0.5258$; for median test $p-value = 0.3916$. It can be considered as additional background for conclusion that there are no reasons for rejecting Null hypothesis.

Application of Munn – Whitney criterion gives following results: $z_{12} = 0.8248$, $z_{13} = 1$, $z_{14} = 1$, $z_{23} = 0.2$, $z_{24} = 0.2683$, $z_{34} = 0.8248$. Thus, inequality (5) is truthful, and we have no reasons for rejecting Null hypotheses about equivalence of averages. Similar results were obtained after application of Kolmogorov – Smirnov criterion: $z_{12} = 0.9963$, $z_{13} = 1$, $z_{14} = 0.9963$, $z_{23} = 0.6$, $z_{24} = 0.5176$, $z_{34} = 0.9963$. Finally, obtained results show that conclusion made in book by Lakin (1990) is correct.

4 Example 2

Considering below data are presented in book by Lakin (1990), Table 60, page 164. On fields of biostation six local breeds of wheat were tested. Results of experiments (in units "centners per hectare") are presented in Table 2. Thus, in considering case factor has six gradations, $n_A = 6$.

Table 2 Results of experiments.

Gradations of factor (breeds of wheat)					
1	2	3	4	5	6
26.1	25.0	27.2	23.6	30.0	23.0
29.2	24.3	26.4	27.2	33.0	26.0
30.0	28.5	31.0	25.2	36.0	26.0
27.3	29.0	26.4	24.8	29.8	24.8

Analysis of dataset from Table 2 showed that $s_e^2 = 6.29$, $s_x^2 = 27.66$, $p-value = 0.0086$. It allowed concluding that with big probability (which is more than 0.99) observed differences in productivity of breeds of wheat are not stochastic.

Like in a previous case it is impossible to check normality of sub-samples corresponding different gradations of factor, and it is impossible to check differences between sample variances. For combined sample we have following results: for Shapiro – Wilk test $p-value = 0.1309$; for Cramer – von-Mises test $p-value = 0.2172$; for Anderson – Darling test $p-value = 0.2027$; for Lilliefors test $p-value = 0.1772$; for chi-squared test $p-value = 0.2466$; for Shapiro – Francia test $p-value = 0.09939$. Like in a previous example all used criterions showed that $p-value$ isn't big; moreover, Shapiro – Francia test shows that with 10% significance level we must reject Null hypothesis.

Application of Kruskal – Wallis test to data from the Table 2 shows that $p-value = 0.1167$; for median test $p-value = 0.0752$. Consequently, with 5% significance level Null hypothesis cannot be rejected.

Application of Munn – Whitney criterion gives following results: $z_{12} = 0.3429$, $z_{13} = 0.8845$, $z_{14} = 0.05714$, $z_{15} = 0.08143$, $z_{16} = 0.0294$, $z_{23} = 0.6631$, $z_{24} = 0.4857$, $z_{25} = 0.02857$, $z_{26} = 0.4678$, $z_{34} = 0.1441$, $z_{35} = 0.1102$, $z_{36} = 0.02843$, $z_{45} = 0.02857$, $z_{46} = 1$, $z_{56} = 0.0294$. These results show that situation isn't so obvious like we have using standard analysis of variance. If we have 1% significance level there are no reasons for rejecting Null hypothesis: minimum value of $p-value$ is bigger than 0.01. For 5% significance level we have unilaterally connected graph (with six nodes, Fig. 1). For 10% significance level we have also unilaterally connected graph with smaller number of edges (we haven't edges between nodes 1 and 4, and 1 and 5).

In first case (Fig. 1) graph hasn't 5 edges of 15 (we haven't edge if and only if for selected level of significance sub-samples of corresponding gradations are confidently different); in second case seven edges of 15 are absent. In both cases number of absent edges is less than 50% of total number of edges, and graphs are unilaterally connected graphs. The question is: can it be a background for conclusion that factor has a confident influence on values of characteristic? The honest answer is following: no, it can't be a background for such a conclusion at all.

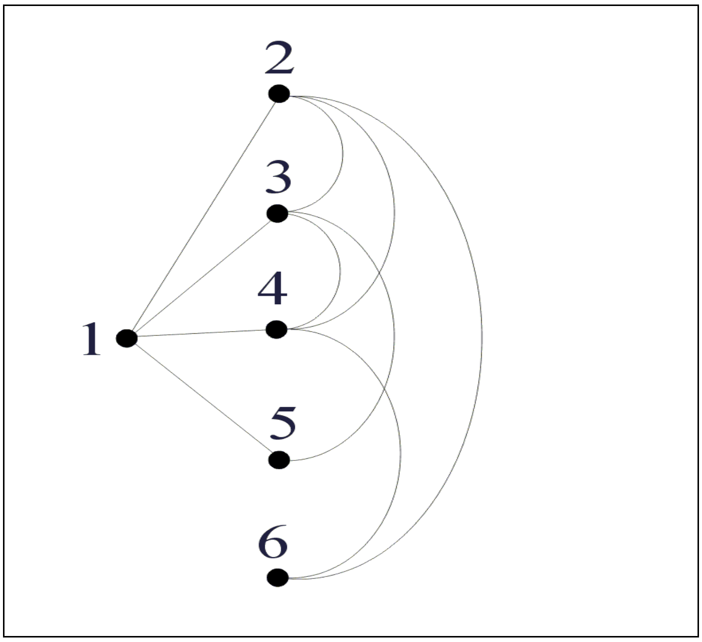

Fig. 1 Unilaterally connected graph, corresponding to results obtained by application of Munn – Whitney test.

Similar results were obtained after application of Kolmogorov – Smirnov criterion: $z_{12} = 0.7714$, $z_{13} = 0.6994$, $z_{14} = 0.2286$, $z_{15} = 0.2106$, $z_{16} = 0.03663$, $z_{23} = 0.6994$, $z_{24} = 0.7714$, $z_{25} = 0.02857$, $z_{26} = 0.6994$, $z_{34} = 0.2106$, $z_{35} = 0.2106$, $z_{36} = 0.03663$, $z_{45} = 0.02857$, $z_{46} = 0.9996$, $z_{56} = 0.03663$. Like in previous case for 1% significance level we have no reasons for rejecting Null hypothesis: all amounts of $p-value$ is bigger than 0.01. This conclusion is in contradiction with conclusion from book by Lakin (1990). In this book it was pointed out that even with this significance level we can conclude that factor has confident influence onto values of characteristic. For 5% and 10% significance levels we have one and the same graph (Fig. 1). The question is the same: can we say that presented results of calculations allow concluding that factor has confident influence on values of characteristic? One of possible answers is following: presented sample is rather small, it doesn't conclude that factor has confident influence on values of characteristic, and it is necessary to continue experiments.

5 Example 3

Lengths of wings for three different species for toms of starlings were analyzed. Results are presented in Table 3 (1 – *Sturnus contra*, 2 – *S. ginginiamus*, 3 – *S. fuscus*; Table 45, page 198, Rokitsky, 1973).

Table 3 Data of bird's measurements.

Species	Results for separated birds											
1	120	120	121	122	122	126	122	123	125	125	126	
2	123	124	125	125	126	127	127	127	128	128	129	129
3	122	122	125	127	127	127	128	129				

For datasets from Table 3 it was obtained (Rokitsky, 1973) that $s_e^2 = 5.028$, $s_x^2 = 40.54$, and $p-value = 0.001715$. It allowed concluding that with probability $p > 0.99$ (with significance level $\alpha < 0.01$) there are interspecies differences (with respect to lengths of wings) between toms of starlings.

Like in two previous cases let's consider a question about normality of combined sample and sub-samples of datasets presented in table 3. For combined sample we have following results: for Shapiro – Wilk test $p-value = 0.05668$; for Cramer – von-Mises test $p-value = 0.06115$; for Anderson – Darling test $p-value = 0.05665$; for Lilliefors test $p-value = 0.07843$; for chi-squared test $p-value = 0.006228$; for Shapiro – Francia test $p-value = 0.1196$. Like in a previous example all used criterions showed that $p-value$ isn't big. Moreover, application of chi-squared test allows rejecting Null hypothesis with 1% significance level. Additionally, all criterions (excluding Shapiro – Francia test) allow rejecting Null hypothesis with with 10% significance level. The question is: can we say that results presented in book by Rokitsky (1973), are confident if all values of $p-value$ are rather small and close to critical values (5%)?

For first sub-sample of factor gradation (11 elements) we have: for Shapiro – Wilk test $p-value = 0.1638$; for Cramer – von-Mises test $p-value = 0.2221$; for Anderson – Darling test $p-value = 0.206$; for Lilliefors test $p-value = 0.2391$; for chi-squared test $p-value = 0.6718$; for Shapiro – Francia test $p-value = 0.2791$. Situation is better than for the whole sample: Null hypothesis cannot be rejected even with 10% significance level.

For second sub-sample of factor gradation (12 elements) we have: for Shapiro – Wilk test $p-value = 0.5468$; for Cramer – von-Mises test $p-value = 0.5117$; for Anderson – Darling test $p-value = 0.5469$; for Lilliefors test $p-value = 0.3043$; for chi-squared test $p-value = 0.8013$; for Shapiro – Francia test $p-value = 0.6389$. Situation is much better than for dataset of first gradation.

For third sub-sample of factor gradation (8 elements) we have: for Shapiro – Wilk test $p-value = 0.1202$; for Cramer – von-Mises test $p-value = 0.1141$; for Anderson – Darling test $p-value = 0.1026$; for Lilliefors test $p-value = 0.04652$; for chi-squared test $p-value = 0.05642$; for Shapiro – Francia test $p-value = 0.1453$. Application of Lilliefors test allows rejecting Null hypothesis with 5% significance level; application of chi-squared test allows rejecting with 10% significance level. Other amounts of $p-value$ are close to critical level (10%). Thus, the question about confidence to results presenting in book by Rokitsky (1973), is open.

Application of Kruskal – Wallis test to data from the Table 3 shows that $p-value=0.004298$; for median test – $p-value=0.0437$. Consequently, with 5% significance level Null hypothesis must be rejected. It corresponds to results from book by Rokitsky (1973).

Comparison of sample variances for factor gradations produced following results: for Ansari – Bradley test $w_{12}=0.9422$, $w_{13}=0.2887$, $w_{23}=0.6425$, where w_{ij} is a $p-value$ for gradations i and j. For Mood two-sample test we have following results: $w_{12}=0.912$, $w_{13}=0.3058$, $w_{23}=0.5521$. Thus, we have no reasons for rejecting Null hypothesis about equivalence of sample variances even with 28% significance level.

Application of Munn – Whitney criterion delivers following results: $z_{12}=0.001953$, $z_{13}=0.02138$, $z_{23}=0.6944$. Thus, we can conclude that with 3% significance level first species differs from two other species. Graph contains two components: node 1 has no edges to other nodes. Taking into account recommendations described above, we can say that factor has confident influence onto values of characteristic. At the same time we have to note that species 2 and 3 cannot be divided with respect to considering characteristic.

Other results were obtained after application of Kolmogorov – Smirnov criterion: $z_{12}=0.04025$, $z_{13}=0.05365$, $z_{23}=0.9251$. In this situation for 5% significance level graph contains two edges, and respectively we have no reasons for conclusion that factor has a confidence influence on characteristic.

6 Conclusion

Analysis of variance plays very important role in ecology (Lakin, 1990; Rokitsky, 1973; Plokhinsky, 1970; Lyubishev, 1986, and many others). Analyses of some particular cases show that results which can be obtained with traditional analysis of variance and Kruskal – Wallis test, can lead to incomprehension and arising of wondering.

Problem which we have in considering situation is following: this is a problem of rendering of expression "confidence influence on values of characteristic". It is obvious, if factor has no scale, and we can use two sample variances only – we have to use these variances. But this is self-limitation and nothing more.

Idea of analysis of variance is rather simple. In ideal case when factor has extra strong influence on values of characteristic, every sub-sample will contain equal numbers. If so, every sub-sample variation will be equal to zero, and, respectively, $s_e^2=0$. In other ideal case when factor hasn't influence on values of characteristic, for every gradation average will be equal to average of whole sample, and, finally, $D_e=D_y$, $D_x=0$, and $s_x^2=0$. Consequently, analyzing relation s_x^2/s_e^2 we can say something about influence of factor onto values of characteristic: factor becomes stronger at decreasing of s_e^2, and factor becomes weaker at decreasing of s_x^2. Final conclusion about power of factor we obtain after comparison of relation s_x^2/s_e^2 with table value for Fisher distribution. But as it was pointed above fraction s_x^2/s_e^2 has no relation to Fisher distribution because for calculation of both variations s_e^2 and s_x^2 we use all elements of initial sample. In other words, these variations s_e^2 and s_x^2 cannot be considered as independent stochastic variables.

One of serious problems of modern statistics (and, in particular, for analysis of variances) is following: for solution of one or other problem researches try to construct any statistic which gives (after respective calculations) an answer of the type "yes" or "no" (this is ideal variant) or "yes" or "no" or "unobvious" (for example, like Durbin – Watson criterion: there exists a zone of ambiguity; Draper and Smith, 1981). But within the framework of this approach we have regular problems. For example, we have it when for solution we use several different statistical criterions (like for testing of normality of samples). Application of different statistical criterions can lead (and we had it in examples considered above) to contradiction when one criterion gives positive answer and another criterion gives negative answer. When we have a contradiction we have to follow a principle “if one of used statistical criterions deliver a negative answer we have to reject Null hypothesis”.

References

Anderson TW, Darling DA. 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics, 23: 193-212

Anderson TW, Darling DA. 1954. A test of goodness-of-fit. Journal of the American Statistical Association, 49: 765-769

Bolshev LN, Smirnov NV. 1983. Tables of Mathematical Statistics. Nauka, Moscow, Russia

Draper NR, Smith H. 1981. Applied Regression Analysis. Wiley and Sons Inc, New York, USA

Hollander MDA, Wolfe DA. 1973. Nonparametric Statistical Methods. John Wiley and Sons Inc, USA

Iman RL, Davenport JM. 1976. New approximations to the exact distribution at the Kruskal – Wallis test statistic. Communications in Statistics - Theory and Methods, 5: 1335-1348

Kobzar AI. 2006. Applied Mathematical Statistics. For engineers and scientists. Fizmatlit, Moscow, Russia

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

Lakin GF, 1990. Biometrics. Visshaya Shkola, Moscow, Russia

Likes J, Laga J. 1985. Basic Tables of Mathematical Statistics. Finance and Statistics, Moscow, Russia

Lilliefors H. 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, 62(318): 399-402

Lilliefors H. 1969. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. Journal of the American Statistical Association, 64(325): 387-389

Lyubishev AA. 1986. Analysis of Variance in Biology. Moscow State University, Moscow, Russia

Mathematical encyclopedia. 1985. Vol. 5. Soviet encyclopedia, Moscow, Russia

Nedorezov LV. 2012. Chaos and Order in Population Dynamics: Modeling, Analysis, Forecast. LAP Lambert Academic Publishing, Saarbrucken, USA

Nedorezov LV. 2015. Application of generalized discrete logistic model for fitting of pine looper moth time series: Feasible sets and estimations of model parameters. Proceedings of the International Academy of Ecology and Environmental Sciences, 5(1): 38-48

Plokhinsky NA. 1970. Biometrics. Moscow State University, Moscow, Russia

Rokitsky PF. 1973. Biological Statistics. Visshaya Shkola, Minsk, Russia

Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality. Biometrika, 52(3): 591-611

Thode JrHC. 2002. Testing for Normality. Marcel Dekker, New York, USA

Vasiliev AV, Melnikova IN. 2009. Methods of applied analysis of field measurements in environment. Baltic State Technical University, Saint Petersburg, Russia