Pearson-Fisher Chi-Square Statistic Revisited

Sorana D. Bolboacă [1], Lorentz Jäntschi [2,*], Adriana F. Sestraş [2,3], Radu E. Sestraş [2] and Doru C. Pamfil [2]

[1] "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, 6 Louis Pasteur, Cluj-Napoca 400349, Romania; E-Mail: sbolboaca@umfcluj.ro
[2] University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca, 3-5 Mănăştur, Cluj-Napoca 400372, Romania; E-Mails: lorentz.jantschi@gmail.com (L.J.); asestras@yahoo.com (A.F.S.); rsestras@yahoo.co.uk (R.E.S.); dpamfil@usamvcluj.ro (D.C.P.);
[3] Fruit Research Station, 3-5 Horticultorilor, Cluj-Napoca 400454, Romania

Supplementary Material

Starting from the well known Gauss-Laplace distribution of errors[1]:

$$\mathrm{PDF}_{GL}(x;\mu,\sigma,p)=\frac{p}{2\sigma}\frac{\Gamma^{1/2}(3/p)}{\Gamma^{3/2}(1/p)}\exp\left(-\frac{\left|\frac{x-\mu}{\sigma}\right|^{p}}{\left(\frac{\Gamma(1/p)}{\Gamma(3/p)}\right)^{p/2}}\right) \quad (1)$$

a further assumption can be made on the observational error: there is no systematic error (mean μ set to 0). The errors for Gauss-Laplace distribution with a mean of zero become:

$$PDF_{GL0}(x;\sigma,p)=PDF_{GL}(x;\sigma,p)=\frac{p}{2\sigma}\frac{\Gamma^{1/2}(3/p)}{\Gamma^{3/2}(1/p)}\exp\left(-\frac{\left|\frac{x}{\sigma}\right|^{p}}{\left(\frac{\Gamma(1/p)}{\Gamma(3/p)}\right)^{p/2}}\right) \quad (2)$$

The (2) relationship can be used to check:

- (Case 1) $O_{i,j}$-$E_{i,j}$ differences (differences between values from Table 2 - $O_{i,j}$ and Table 4 - a_ib_j)
- (Case 2) $O_{i,j}$-a_ib_j differences for $S^2=(O_{i,j}-a_ib_j)^2$ (differences between values from Table 2 - $O_{i,j}$ and Table 3 - $E_{i,j}$)
- (Case 3) $O_{i,j}$-a_ib_j differences for $CV^2=(O_{i,j}-a_ib_j)^2/(a_ib_j)^2$=min. (differences between values from Table 2 - $O_{i,j}$ and Table 4 - a_ib_j)
- (Case 4) $O_{i,j}$-a_ib_j differences for $X^2=(O_{i,j}-a_ib_j)^2/a_ib_j$=min. (differences between values from Table 2 - $O_{i,j}$ and Table 4 - a_ib_j)

The following table presents the above-mentioned differences:

No.	Case 1 E_{ij}	Case 2 S^2 = min.	Case 3 CV^2 = min.	Case 4 X^2 = min.	No.	Case 1 E_{ij}	Case 2 S^2 = min.	Case 3 CV^2 = min.	Case 4 X^2 = min.
1	-2.31	-1.77	-2.27	-2.34	37	1.96	2.36	1.77	1.83
2	1.7	1.58	1.92	1.81	38	0.36	0.25	0.32	0.34
3	0.62	0.66	0.26	0.45	39	0.92	0.94	0.46	0.70
4	-2.51	-1.85	-2.61	-2.6	40	3.05	3.54	2.80	2.89
5	1.19	1.05	1.42	1.31	41	-0.75	-0.86	-0.74	-0.73
6	-0.53	-1.14	-0.81	-0.64	42	-0.45	-0.93	-0.84	-0.62
7	1.44	1.36	1.17	1.32	43	-3.2	-3.26	-3.57	-3.37
8	1.21	1.27	1.21	1.19	44	-2.96	-2.93	-3.13	-3.07
9	0.94	0.37	0.64	0.81	45	-1.43	-1.87	-1.80	-1.59
10	-0.58	-0.78	-0.53	-0.54	46	0.85	0.70	0.77	0.82
11	0.23	0.26	0.01	0.13	47	2.16	2.18	1.88	2.03
12	-1.40	-1.61	-1.56	-1.47	48	-0.49	-0.65	-0.7	-0.58
13	-1.21	-0.66	-1.38	-1.35	49	-2.74	-2.08	-2.97	-2.9
14	1.08	0.98	1.10	1.09	50	-3.23	-3.09	-3.31	-3.28
15	2.05	2.11	1.52	1.79	51	3.36	3.59	2.86	3.11
16	-3.18	-2.52	-3.45	-3.36	52	0.79	1.49	0.49	0.60
17	-0.80	-0.92	-0.74	-0.76	53	0.80	0.89	0.77	0.79
18	3.38	2.80	2.94	3.18	54	-2.01	-2.28	-2.43	-2.20
19	-3.75	-3.82	-4.19	-3.96	55	3.56	3.68	3.14	3.35
20	0.51	0.58	0.35	0.40	56	-2.21	-1.99	-2.42	-2.34
21	3.19	2.64	2.76	3.00	57	-0.35	-0.63	-0.76	-0.54
22	0.54	0.36	0.47	0.52	58	0.24	0.23	0.14	0.20
23	-2.37	-2.33	-2.69	-2.52	59	2.26	2.41	1.96	2.12
24	0.56	0.37	0.32	0.45	60	-0.47	-0.53	-0.70	-0.58
25	0.94	1.59	0.66	0.76	61	3.36	3.21	2.48	2.93
26	-0.55	-0.52	-0.64	-0.60	62	0.65	0.36	-0.14	0.27
27	-6.80	-6.63	-7.39	-7.08	63	-0.15	-0.36	-0.97	-0.53
28	-0.84	-0.11	-1.19	-1.05	64	2.69	2.62	1.94	2.33
29	0.00	-0.01	-0.04	-0.01	65	-0.43	-0.68	-1.07	-0.73
30	2.05	1.61	1.54	1.83	66	-2.45	-2.80	-3.21	-2.80
31	2.39	2.43	1.90	2.15	67	-0.44	-0.66	-1.18	-0.79
32	1.45	1.61	1.20	1.31	68	2.00	1.81	1.33	1.67
33	-0.07	-0.50	-0.56	-0.29	69	-2.26	-2.57	-2.90	-2.56
34	0.15	0.05	0.02	0.10	70	-1.20	-1.40	-1.68	-1.43
35	-1.46	-1.36	-1.82	-1.63	71	-0.82	-0.94	-1.31	-1.05
36	2.74	2.61	2.47	2.62	72	-0.94	-1.1	-1.34	-1.13

In order to obtain the maximum likelihood estimations of the parameters for the distribution given in equation (2) the following system of equations must be solved:

$$\begin{cases} \frac{\partial}{\partial \sigma} MLE\_PDF_{GL0}(\sigma, p) = 0 \\ \frac{\partial}{\partial p} MLE\_PDF_{GL0}(\sigma, p) = 0 \end{cases} \quad (3)$$

where:

$$\mathrm{MLE\_PDF}_{GL0}(\sigma,p)=\sum_{i=1}^{72}\frac{p}{2\sigma}\frac{\Gamma^{1/2}(3/p)}{\Gamma^{3/2}(1/p)}\exp\left(-\frac{\left|\frac{z_i}{\sigma}\right|^p}{\left(\frac{\Gamma(1/p)}{\Gamma(3/p)}\right)^{p/2}}\right) \quad (4)$$

and z_i are the values listed in the columns of the table above.

MathCad was used to solve this system, and EasyFit to obtain the statistics of the agreement.

The following table contains the solutions for Eq.(3) together with the Kolmogorov-Smirnov (K-S), Anderson-Darling (A-D), and Chi-Square (C-S) probabilities from which Fisher's Chi-Square [2] statistic (F-C-S=-ln(p_{K-S})-ln(p_{A-D})-ln(p_{C-S})) giving global agreement between observational error and Gauss-Laplace model and their associated probability (from χ^2 distribution with 3 degrees of freedom) were calculated:

Error	**Distribution**	p_{K-S}	p_{A-D}	p_{C-S}	**F-C-S**	p_{F-C-S}	**Skewness, Kurtosis**
E_{ij}	PDF_{GL0}(2.005,1.859)	0.972	0.760	0.348	1.358	0.715	Sk=0.0; eKu=0.156
S^2	PDF_{GL0}(1.980,1.983)	0.955	0.831	0.682	0.614	**0.893**	Sk=0.0; eKu=0.172
CV^2	PDF_{GL0}(2.027,1.796)	0.510	0.418	0.285	2.801	0.423	Sk=0.0; eKu=0.238
X^2	PDF_{GL0}(2.009,1.834)	0.709	0.687	0.076	3.296	0.348	Sk=0.0; eKu=0.188

K-S = Kolmogorov-Smirnov statistic, A-D = Anderson-Darling statistic,
C-S = Chi-Square statistic; F-C-S = Fisher's Chi-Square statistics

The above-presented results clearly indicate that the best agreement between observational errors and the Gauss-Laplace distribution are obtained when $S^2 = (O_{i,j}-a_ib_j)^2$ = min is used to obtain the a_i and b_j values. Moreover, the errors in Fisher's experiment are more likely absolute errors (which are subject to minimization in order to obtain the best agreement between the estimate from the treatment and variety factors and the observed production).

Thus, this supplementary study on Fisher's reported results on treatment and variety factors influencing production clearly indicates that combining the proposed pool of objective functions ($E_{i,j}$, S^2=min., CV^2=min., X^2=min.) with the modified form of the Gauss-Laplace distribution according to the assumption that there is no systematic error (μ=0) and according to the Chi-Square measure of global agreement developed by Fischer is able to provide an excellent estimate for a certain type of experimental error.

[1] Jäntschi, L.; Bolboacă, S.D. Observation *vs.* Observable: Maximum Likelihood Estimations according to the Assumption of Generalized Gauss and Laplace Distributions. *Leonardo El. J. Pract. Technol.* **2009**, *8*, 81-104.

[2] Fisher, R.A. Combining independent tests of significance. *Am. Stat.* **1948**, *2*, 30.