J Forensic Sci, Apr. 1973, Vol. 18, No. 2

D. D. Lawson,[1] *M.S. and E. P. Framan,*[2] *Ph.D.*

Numerical Correlation and Evaluation in the Comparison of Evidentiary Materials[3]

One of the neglected areas in criminalistic laboratory techniques is the ability of the criminalist to compare and evaluate, quantitatively, sets of numerical data. At the present time, there does not exist in practice any uniformly accepted approach to determining a "figure-of-merit" (or rather a quantitative, reliable expression of the degree of match) for sets of data. The need for such a technique arose in a task designed to apply the phenomena of thermoluminescence to criminalistics [*1*]. This task, funded and supported by NASA, was conducted through the Civil Systems Program Office at the Jet Propulsion Laboratory. In this specific application, each piece of physical evidence provided, after processing, a continuous curve. In comparing the curve of an unknown with that of an exemplar, a need was established for a quantitative expression of the extent to which the two curves matched. A well established statistical procedure was applied and was found to be fully satisfactory in resolving the problem. The same technique, without modifications was also found to be applicable to the analysis of emission spectrography data, neutron activation analysis (NAA), gradient density measurements, and any other criminalistic technique where sets of numerical data are determined. The main attribute of this developed technique is that it allows the criminalist to make judgements on the quality of the evidentiary determinations.

Chi-Square Method

In 1900, Karl Pearson [*2*] devised the chi-square (x^2) test (which represents in this application a goodness of fit index) which occupies a central position in statistical theory, and it is difficult to imagine another test which has the same generality of applications. The use of the chi-square test, which is a hypothesis test, is applicable where data can be represented in the form of a contingency table. An example of this is shown in Table 1A. Since the concern is with the comparison and evaluation of materials on a pair basis, contingency tables thus consist of two rows and *n* columns (the number of attributes measured). Where three or more materials are to be compared, all possible combinations of contingency tables are used.

Presented at the Twenty-fourth Annual Meeting of the American Academy of Forensic Sciences, 2 March 1972, Atlanta, Ga. Received 4 March 1972; accepted for publication 7 Sept. 1972.

[1] Technical Staff, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

[2] Public Safety Systems, Civil Systems Program Office, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, Calif.

[3] This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS7–100, sponsored by the National Aeronautics and Space Administration.

Chi-square for each contingency table may be defined as

$$x^2 = \text{sum of all groups of} \left[\frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}}\right] \quad (1)$$

A problem associated with evidentiary materials is that in some situations it is very difficult to define which is the observed and which is the expected value. Therefore, when comparisons are made, the roles of the expected and observed values are interchanged and two numerical x^2 values are calculated (they are not numerically equal). The pair of x^2 values for the contingency table in Table 1A are calculated and shown in Table 1B. If the situation is such that no problem exists with the role of the expected and observed values, the redundant index can then be discarded.

TABLE 1—*Calculation of a portion of a (chi-square) evaluation array using neutron activation analysis data from hair samples.*

(A) Contingency Table

	Chemical Elements						
	Na[a]	Cl	I	CU	Br	Hg	Zn
Subject #1	6.13	9.26	2.94	2.72	3.89	1.53	5.40
Subject #2	4.90	7.46	3.64	1.94	3.13	1.02	5.25

(B) Calculations

	Subject #1	Subject #2	Difference, Δ	Δ^2	Δ^2 / Subject #1	Δ^2 / Subject #2
Na[a]	6.13	4.90	1.23	1.5129	0.2468	0.3087
Cl	9.26	7.46	1.80	3.2400	0.3498	0.4343
I	2.94	3.64	−0.70	0.4900	0.1666	0.1346
Cu	2.72	1.94	0.79	0.6241	0.2352	0.3216
Br	3.89	3.13	0.76	0.5776	0.1484	0.1845
Hg	1.53	1.02	0.51	0.2601	0.1700	0.2550
Zn	5.40	5.25	0.15	0.0225	0.0041	0.0042
					χ^2 = 1.3209	1.6429

[a] Concentrations are expressed as the natural logarithm of parts per million.

For the situation where there are n sets of data, there will be $[n(n-1)/2]$ contingency tables and, therefore, the same number of pairs of x^2 values. An example is shown in Table 2A where NAA has been applied to 11 subjects. The representation of all possible combinations of pair comparisons is termed an evaluation array. For the example under discussion, the evaluation array is shown in Table 2B in that 55 pairs of x^2 values are shown. The chi-square value for the pair in the array that is the smallest numerically represents the best fit or the closest match of the parameters involved in the particular test. In Table 2B, subjects 1 and 11 are the best match. An exact match will exist when the index goes to zero.

A second factor in the numerical evaluation process is to establish a threshold level for a *first-order criteria* of a match. To provide such a threshold which reflects the reproducibility of the physical analysis method, a series of repetitive determinations on the same sample should be obtained and a chi-square evaluation array then calculated. By using

TABLE 2A—*Neutron activation analysis data given by Parker and Holford* [3].

Subject	Na[a]	Cl	I	Cu	Br	Hg	Zn
1	6.13	9.26	2.94	2.72	3.89	1.53	5.40
2	4.90	7.46	3.64	1.94	3.13	1.02	5.25
3	7.30	8.34	1.63	3.66	3.04	0.74	5.29
4	6.35	7.63	2.33	2.44	3.53	1.70	5.05
5	6.87	7.74	4.48	1.36	2.85	1.29	4.43
6	5.99	5.85	2.58	3.15	5.70	1.24	6.01
7	7.66	6.00	2.35	3.93	2.35	1.52	5.58
8	7.52	8.06	2.94	3.64	3.76	1.80	5.69
9	7.28	7.57	2.17	3.86	2.84	1.41	5.42
10	7.34	8.80	2.10	2.62	3.47	0.51	5.36
11 (Control Mean)	6.63	8.40	2.96	3.04	3.76	1.63	5.24

[a] The natural logarithms of the concentration of different elements in parts per million. This form of expressing concentration is known to be reasonably Gaussian in nature.

TABLE 2B—*Evaluation array for neutron activation analysis of hair using the data of Parker and Holford* [3].

Subject	2	3	4	5	6	7	8	9	10	11
Subject 1	1.32 1.64	1.80 2.64	0.52 0.62	2.14 2.69	2.34 2.80	2.80 3.61	0.85 0.73	1.50 1.56	1.23 2.65	0.17 0.17
2		4.01 4.29	1.54 1.50	1.27 1.07	3.92 2.83	4.80 3.51	3.83 2.39	3.83 2.87	2.64 2.86	1.97 1.44
3			2.24 1.66	6.93 6.04	4.41 3.38	2.05 1.86	2.79 1.39	0.90 0.57	0.58 0.69	2.49 1.41
4				2.76 2.50	2.31 1.89	2.00 1.90	1.09 0.81	1.18 0.91	1.22 3.12	0.43 0.35
5					6.65 4.64	6.48 4.34	4.93 2.68	5.82 4.11	3.21 4.76	3.04 2.06
6						2.74 5.40	2.28 2.22	2.52 3.79	3.34 4.01	2.12 2.09
7							1.78 1.24	0.56 0.45	2.99 3.96	2.33 1.79
8								0.57 0.74	1.56 4.10	0.27 0.31
9									1.31 2.64	0.94 0.83
10										2.99 1.19

The best fit or closest match of parameters are Subjects 1 and 11. This is the same match as Parker and Holford obtained by their procedures.

normal probability paper, the mean ($\bar{X}$) and standard deviation (S) are calculated. The use of normal probability plots allows the assessment of the character of the distribution of the data. In some cases, bimodal distributions or other situations are apparent on the probability plots where otherwise they might be overlooked. When calculating the threshold, five repetitive determinations on the same sample are recommended. This gives 20 chi-square indices which are then arranged in increasing numerical order (or by rank number m, 1 through 20) (refer to Table 5). The plotting position (x axis) on the normal probability paper is the rank number (m) divided by the total number (N) of samples plus one, or

$$\text{plotting position} = \frac{m \cdot 100}{N + 1} \qquad (2)$$

A different plotting position is required for each different sample size. Table 4 depicts the calculated plotting positions for N equalling 1 through 20.

The threshold level can be set at one to three or more standard deviations ($\overline{X} + 1S$ to $\overline{X} + 3S$). Table 3 depicts the percentiles on the probability paper for a given standard deviation, its associative percentage of error, and the chance that the determination will be wrong for some given threshold level. The errors mentioned here, for example, 15.87 percent at one standard deviation, are the percentages of truly positive matches that will appear incorrectly as mismatches. As the numerical threshold level increases the more false matches will be obtained. False matches here are those coincidental matches of data for samples that are really different. The selection of a standard deviation level should be made by the criminalist and not adjusted to fit the circumstances.

Applications

In order to illustrate the procedure and application, two examples using the discrete (number) form may be discussed along with a description of the continuous curve evaluation method.

1. The example shown earlier in Table 2 is clearly one type of application involving eleven hair samples with two from the same head. This is a standard representation of NAA data whereby the concentrations of each element are given in natural logarithms in

TABLE 3—*Standard deviations and related percentile of error.*

Standard Deviations (normal)	Percentile on Normal Probability Paper	Percentile of Error	Chances that the Determination will be Wrong
1.0	84.13	15.87	6.3[a]
1.5	93.32	6.68	15.2
2.0	97.73	2.27	45
2.5	99.38	0.62	166
3.0	99.86	0.14	714

[a] In terms n to one.

TABLE 4—*Plotting position for use with normal probability paper.*

RANK	N=1	N=2	N=3	N=4	N=5	N=6	N=7	N=8	N=9	N=10	RANK
1	50.0	33.0	25.0	20.0	16.6	14.4	12.5	11.1	10.0	9.1	1
2		66.6	50.0	40.0	33.3	28.6	25.0	22.2	20.0	18.2	2
3			75.0	60.0	50.0	42.8	37.5	33.3	30.0	27.3	3
4				80.0	66.6	57.1	50.0	44.4	40.0	36.4	4
5					83.3	71.4	62.5	55.5	50.0	45.4	5
6						85.7	75.0	66.6	60.0	54.5	6
7							87.5	77.7	70.0	63.6	7
8								88.8	80.0	72.7	8
9									90.0	81.8	9
10										90.9	10

	N=11	N=12	N=13	N=14	N=15	N=16	N=17	N=18	N=19	N=20	
1	8.3	7.7	7.1	6.6	6.2	5.9	5.5	5.3	5.0	4.8	1
2	16.6	15.4	14.3	13.3	12.5	11.8	11.1	10.5	10.0	9.5	2
3	25.0	23.1	21.4	20.0	18.7	17.6	16.6	15.9	15.0	14.3	3
4	33.3	30.8	28.5	26.6	25.0	23.5	22.2	21.0	20.0	19.0	4
5	41.6	38.4	35.7	33.3	31.2	29.4	27.7	26.3	25.0	23.8	5
6	50.0	46.1	42.8	40.0	37.5	35.3	33.3	31.6	30.0	28.6	6
7	58.3	53.8	50.0	46.6	43.7	41.2	38.9	36.8	35.0	33.3	7
8	66.6	61.5	57.2	53.4	50.0	47.0	44.4	42.1	40.0	38.1	8
9	75.0	69.2	64.2	60.0	56.2	52.9	50.0	47.4	45.0	42.8	9
10	83.3	76.9	71.4	66.6	62.5	58.8	55.5	52.6	50.0	47.6	10
11	91.7	84.6	78.5	73.3	68.7	64.7	61.1	57.9	55.0	52.4	11
12		92.3	85.7	80.0	75.0	70.6	66.6	63.1	60.0	57.1	12
13			92.8	86.6	81.2	76.5	72.2	68.4	65.0	61.9	13
14				93.3	87.5	82.3	77.7	73.7	70.0	66.6	14
15					93.7	88.2	83.3	78.9	75.0	71.4	15
16						94.1	88.8	84.2	80.0	76.2	16
17							94.4	89.5	85.0	80.9	17
18								94.7	90.0	85.7	18
19									95.0	90.5	19
20										95.2	20

TABLE 5—*Headlamp glass (AC4201) goodness of fit index on five repetitive determinations by emission spectrography.*

	2	3	4	5
1	2.92	4.31	1.86	6.85
	5.91	0.79	3.84	5.25
2		0.40	1.30	1.94
		0.79	5.08	9.11
3			7.49	2.88
			1.84	12.0
4				6.65
				0.88

(A) Evaluation array for reproducibility of glass AC 4201.

N	RANKED CHI SQ. M	M*100/N+1
1	0.40	4.8
2	0.79	9.5
3	0.79	14.3
4	0.88	19.0
5	1.30	23.8
6	1.84	28.6
7	1.86	33.3
8	1.94	38.1
9	2.88	42.8
10	2.92	47.6
11	3.84	52.4
12	4.31	57.1
13	5.08	61.9
14	5.21	66.6
15	5.91	71.4
16	6.65	76.2
17	6.85	80.9
18	7.49	85.7
19	9.11	90.5
20	12.03	95.2

(B) Ranked indexes and the appropriate plotting position ready to be plotted in Fig. 2.

parts per million. From the evaluation array of Table 2B, the two hair samples from the same head are indicated by the lowest x^2 values of 0.17 and 0.17 for the match between subjects 1 and 11. This agrees with the results of Parker and Holford [*3*].

2. Ten samples of auto head lamp glass were examined by emission spectroscopy with 20 elements measured in each sample.

In the previous NAA example, no threshold level (reproducibility factor) was determined. In this example, one of the auto head lamp samples was selected as a reference material (AC4201), and five repetitive determinations were obtained. An evaluation array was calculated (20 indices) from this reference glass sample and then arranged in numerical order (or by rank number, *m*), ranging from the smallest to the largest value (Table 5). The threshold level that was used was one standard deviation ($\overline{X} + 1S$) or 84.2 percentile on the normal probability paper. For other possible levels see Table 3.

Figure 1 shows the data from Table 3 plotted on normal probability paper and shows a threshold of 4.20. Out of 45 pairs compared in the evaluation array (Table 6), 23 (51.1 percent) are below the reproducibility threshold. Therefore, the use of this data for establishing commonality of materials would be open to serious doubt. Threshold levels are generally about the same for each category of material and technique at the same S level, but should, if possible, be determined in each specific case. However, if available sample sizes are small, the use of past threshold values for that technique and type of material may be warranted.

3. In the criminalistic laboratory, continuous curves are obtained with infrared, gas chromatography, and other techniques which employ strip-chart recorders. During the development of the criminalistic aspects of thermoluminescence, it was found that a need for the analysis of glow curves was required. The major problem is that there is too much data in a curve, and one needs a way of choosing an appropriate and convenient number of points on the curve so that these can be used in the chi-square test.

Mann and Wald [*4*] developed a method of calculating the number of class intervals in the application of the chi-square test. The number of classes to be used can be computed by means of the following formula:

$$K = \left[4 \sqrt[5]{\frac{2(N-1)^2}{C^2}} \right] \tag{3}$$

TABLE 6—*Evaluation array for auto headlamp glass calculated from emission spectra data. Out of 45 pairs, 23 are below the 7.20 threshold level.*

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	
1 Tungsol 4001	29.74	8.28	1.44[a]	2.86[a]	8.05	12.04	11.26	5.74[a]	14.59	(1)
New	12.34	7.43	1.38[a]	7.25	8.85	24.04	2.45[a]	2.60[a]	160.7	
2 GE 4012		15.08	5.28[a]	2.23[a]	148.6	58.46	33.02	56.42	38.00	(2)
New		107.8	84.98	13.41	18.34	67.04	31.77	6.00[a]	269.9	
3 AC 4201 (center section)			3.85[a]	3.87[a]	4.89[a]	37.03	9.21	4.33[a]	18.21	(3)
New			15.86	9.04	5.60[a]	8.85	22.17	14.58	7.96	
4 AC 4201				2.15[a]	7.34	9.60	20.72	4.63[a]	12.90	(4)
3 years old, removed from auto				2.91[a]	16.45	50.23	3.00[a]	1.88[a]	332.6	
5 Tungsol 4413					4.76[a]	15.08	27.45	5.98[a]	10.09	(5)
New					11.88	55.23	5.35[a]	3.45[a]	182.3	
6 Chevrolet 1948 Headlamp						3.21[a]	6.19[a]	2.23[a]	22.29	(6)
auto salvage yard–El Monte, Calif.						5.89[a]	7.24	10.10	10.02	
7 Ford 1956 Headlamp							5.09[a]	5.03[a]	27.68	(7)
auto salvage yard–El Monte, Calif.							4.81[a]	28.05	53.95	
8 Italian Headlamp								12.53	34.15	(8)
auto salvage yard–El Monte, Calif.								3.09[a]	148.4	
9 English Lucas Headlamp									241.1	(9)
from MG–TD									8.46	
10 Japan Toshiba Headlamp										
auto salvage yard–Monrovia, Calif.										

[a] Threshold below 7.20.

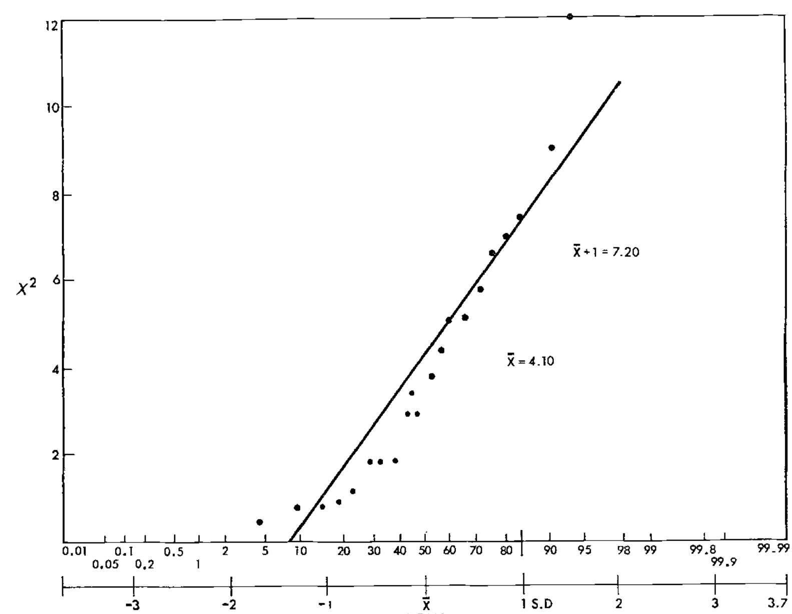

FIG. 1—*Threshold level of glass samples (AC4201) from emission spectrography data.*

TABLE 7—*Percentiles of the standard distribution for a normal curve (for use with the Mann-Wald equation).*

Percentiles	C
0.001	−3.09
0.005	−2.58
0.01	−2.33
0.02	−2.05
0.03	−1.88
0.04	−1.75
0.05	−1.64
0.10	−1.28
0.15	−1.04
0.20	−0.84
0.30	−0.52
0.40	−0.25
0.50	0
0.60	0.25
0.70	0.52
0.80	0.84
0.85	1.04
0.90	1.28
0.95	1.645
0.96	1.75
0.97	1.88
0.98	2.05
0.99	2.33
0.995	2.58
0.999	3.09

K is the number of classes, N is the number of items in the sample, and C is obtained from a table of areas under the normal curve (Table 7). It was found in the case of thermoluminescence that the useful information occurs over a range of 330°C in temperature. By use of Mann-Wald calculations at 99 percent level of significance (C), each class should be 9.88°C or 10°C to yield 33 temperature classes. Figure 2 shows a glow curve of a soil sample with the light level measured and ready to be compared to another similar glow curve.

Summary

At the present time, this statistical procedure has only been used for sorting the results to find the best fit of the data tested with the parameters involved. It is possible to calculate the probability of change that other parameter combinations would give a better fit [5]. We have refrained from this until more experience is gained. Investigations have been limited to

1. Sorting the results or getting the smallest numerical index so that the closest match of the parameters involved can be easily evaluated.
2. Use of the technique to establish a threshold or level of reproducibility of the particular technique and material involved.

To make the calculations convenient, computer programs have been written for an IBM 1620. These programs are developed primarily for thermoluminescence glow curve analysis, but can be used for general applications.

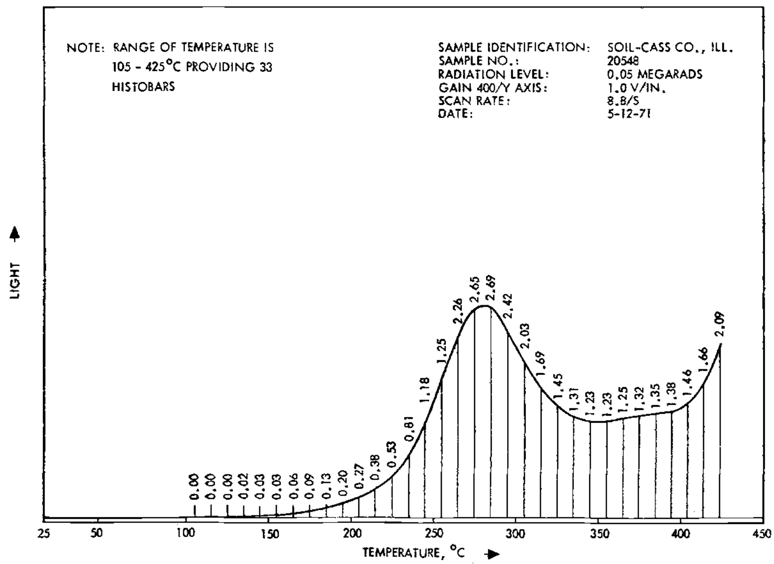

FIG. 2—*Thermoluminescence glow curve depicting light intensities at specified temperature intervals as calculated by the Mann-Wald technique.*

GLOSSARY

Column—A vertical array of terms.

Contingency Table—If a set of items can be classified jointly on the basis of two factors, one of which has q subclasses and another set of subclasses (p), the resulting table of classification is a q by p table.

Evaluation Array—Arrangement of the indices (x^2) so that all possible combinaions are orderly presented. It must have $n(n - 1)/2$ pair terms, where n equals the total number of items to be compared.

Plotting Position—Found by calculating the cumulative frequencies. In this paper it is $m \cdot 100/N + 1$, where m equals the rank number of the sample and N equals the total sample number.

Probability Paper—Graph paper which has one axis scaled so that the graph of the cumulative frequencies of the normal disbtribution function form a straight line.

Row—An arrangement of terms in a horizontal line.

References

[1] Lawson, D. D. and Ingham, J. D., *Thermoluminescence—A New Criminalistic Technique*, to be published.

[2] Pearson, K, "On the Criterion that a Given System of Deviation from the Probable in the Case of a Correlated System of Variables is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling," *Philosophical Magazine*, Vol. 50, Series 50, 1900, pp. 157–174.

[3] Parker, J. B. and Holford, A., "Optimum Test Statistics with Particular Reference to a Forensic Science Problem," *Applied Statistics*, Vol. 17, 1968, pp. 237–251.

[4] Mann, H. B. and Wald, A., "On the Choice of the Number of Class Intervals in the Application of the Chi-Square Test," *Annals of Mathematical Statistics*, Vol. 13, 1942, pp. 306–317.

[5] Cochran, W. G., "The Chi-Square Test of Goodness of Fit," *Annals of Mathematical Statistics*, Vol. 23, 1952, pp. 315–345.