Goodness-of-Fit Tests for Describing the Statistical Behavior of Nuclear Counting Equipment

John R. Prince

Research Medical Center, Kansas City, Missouri

A number of goodness-of-fit tests are available to describe the statistical behavior of nuclear counting equipment. We describe the Lexis' divergence coefficient, the reliability factor, the Kolmogrov-Smirnov test, and the chi-square test. Special emphasis is placed on simplified calculations of the chi-square test.

Goodness-of-fit tests are used extensively to evaluate the statistical behavior of nuclear counting equipment *(1–3)*. However, the statistical manipulation required often discourages their routine use. The purpose of this communication is to compare some of the goodness-of-fit tests that have been proposed and to present simplified calculations for the most common test—the chi-square—that are useful to those working with paper and pencil as well as those having access to electronic data processing.

Goodness-of-Fit Tests

Radioactive decay is a random process that is described mathematically by the Poisson distributions. This randomicity is easily demonstrated by taking a series of replicate counts of a radioactive source with a nuclear detector equipped with a scaling unit. It has been observed that replicate counts of the same source do not yield identical count rates. If one assumes that the normal variation of the equipment response is very much smaller than the randomicity of radioactive decay, then a number of statistical tests can be used to test the hypothesis that the observed variation in a series of replicate counts is due solely to variations in the emission rate of the radioactive source. These statistical tests are called goodness-of-fit tests.

The goodness-of-fit tests that can be used include Lexis' divergence coefficient *(4)*, the reliability factor *(2,5)*, the Kolmogrov-Smirnov test *(6)*, and the chi-square test *(1–3)*.

For reprints contact: John R. Prince, Dept. of Nuclear Medicine, Research Medical Center, 2316 East Meyer Blvd., Kansas City, MO 64132.

Lexis' divergence coefficient. The Lexis' divergence coefficient Q^2 has been discussed by Evans *(4)* and is the result of the work by the German economist Lexis. Mathematically, it is defined as

$$Q^2 = \frac{\Sigma(X_i - \bar{X})^2}{n\bar{X}}, \quad (1)$$

where

Σ = the sum of,
X_i = the ith observation,
$\bar{X}$ = the mean value of the observed replicate counts,
n = the number of the observed replicate counts.

Remember that the standard deviation S of a series of replicate counts is given by the following formula:

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}. \quad (2)$$

Combining Eqs. 1 and 2, we have

$$Q^2 = \frac{n-1}{n} * \frac{S^2}{\bar{X}}. \quad (3)$$

Since $S^2 = \bar{X}$ for a Poisson distribution, the above becomes

$$Q^2 = \frac{n-1}{n} \doteq 1 \quad (4)$$

for large n.

We know that as Q^2 becomes closer to 1 we have greater confidence that our equipment is working satisfactorily, and that as Q^2 becomes much larger than 1 we have greater confidence that our equipment is not working satisfactorily. Remember that Q^2 will become larger than 1 when S^2 becomes larger than $\bar{X}$; that is, when the variation in equipment response between replicate counts is greater than can be accounted for by the randomicity of radioactive decay. Unfortunately, quantitative criteria for establishing the dividing line between satisfactory and unsatisfactory equipment performance are not available and this test has never gained acceptance as a quality control tool. However, the value of Q^2 obtained from the sample calculation in Table 1, i.e., 1.19, is sufficiently close to 1 so that we would conclude that our equipment was working satisfactorily.

Reliability factor. The reliability factor was introduced by Bleuler and Goldsmith *(5)* and its utility for quality control has been further discussed by Prince and Schmidt *(2)*. Mathematically, it is defined as

$$R.F. = \frac{S_x}{S_{th}} = \sqrt{\frac{(X_i-\bar{X})^2}{(n-1)\,\bar{X}}}\,, \qquad (5)$$

where

S_x = the observed standard deviation of a group of replicate measurements,

S_{th} = the theoretical standard deviation. (For a Poisson distribution $S_{th}^2 = \bar{X}$.)

Using the same data from Table 1, a sample calculation of the reliability coefficient is given in Table 2. The statistical confidence levels for the reliability coefficient are given in the *Radiological Health Handbook (7)*. Again, since the value of the reliability coefficient is sufficiently close to 1, i.e., $S_x = S_{th}$, we conclude that our instrument is working satisfactorily.

Kolmogrov-Smirnov test. The use of the Kolmogrov-Smirnov statistic as a goodness-of-fit test has been fully described by Conover *(6)*. This test is used as a quality control tool in clinical chemistry *(8–9)*, but has not been widely used in nuclear technology. The test is based on the maximum difference between an assumed theoretical cumulative distribution and an observed experimental distribution.

TABLE 1. Sample Calculation of Lexis' Divergence Coefficient for Representative Counting Data

X_i	$X_i-\bar{X}$	$(X_i-\bar{X})^2$	$\frac{(X_i-\bar{X})^2}{n\bar{X}}$
12036	39.8	1584.04	0.0132
12004	7.8	60.84	0.0005
11850	−146.2	21374.44	0.1782
12152	155.8	24273.64	0.2023
12237	240.8	57984.64	0.4834
11846	−150.2	22560.04	0.1881
11901	−95.2	9063.04	0.0755
11932	−64.2	4121.64	0.0344
12028	31.8	1011.24	0.0084
11976	−20.2	408.04	0.0034
$\Sigma X_i = 119962$	$\Sigma(X_i-\bar{X}) = 0$	$\Sigma(X_i-\bar{X})^2 = 142441.6$	$\frac{\Sigma(X_i-\bar{X})^2}{n\bar{X}} = 1.188$
n = 10	$\bar{X} = 11996.2$	$\therefore Q^2 = 1.19$	

TABLE 2. Sample Calculation of Reliability Coefficient for Representative Counting Data

X_i	$X_i-\bar{X}$	$(X_i-\bar{X})^2$
12036	39.8	1584.04
12004	7.8	60.84
11850*	146.2	21374.44
12152*	155.8	24273.64
12237*	240.8	57984.64
11846*	−150.2	22560.04
11901	−95.2	9063.04
11932	−64.2	4121.64
12028	31.8	1011.24
11976	−20.2	408.04
$\Sigma X_i = 119962$	$\Sigma(X_i-\bar{X}) = 0$	$\Sigma(X_i-\bar{X})^2 = 142441.6$

$$S_t = \sqrt{\bar{X}} = 109.5 \qquad S_x = \sqrt{\frac{\Sigma(X_i-\bar{X})^2}{n-1}} = 125.8$$

$$R.F. = \frac{S_x}{S_{th}} = \frac{125.8}{109.5} = 1.15$$

TABLE 3. Sample Calculation of Kolmogrov-Smirnov Test for Representative Counting Data

X_i	$\frac{(X_i-\bar{X})}{(\bar{X})^{1/2}}$	$\hat{F}$	F	$d = \lvert \hat{F}-F \rvert$
11846	−1.371	0.085	0.100	0.015
11850	−1.335	0.090	0.200	0.110
11901	−0.869	0.194	0.300	0.106
11932	−0.586	0.278	0.400	0.122
11976	−0.184	0.427	0.500	0.073
12004	0.071	0.528	0.600	0.072
12028	0.290	0.614	0.700	0.086
12036	0.363	0.642	0.800	0.158*
12152	−1.422	0.924	0.900	0.024
12237	2.198	0.986	1.000	0.014

*d_{max} = *largest value between $\hat{F}$ and F.*

As a didactic example consider again the data previously used. In Table 3 the counting data are ranked in order from the lowest to the highest value. The column headings have the following meanings.

X_i = the ith observation.

$\frac{(X_i-\bar{X})}{(\bar{X})^{1/2}}$ = the fractional standard deviation of the ith observation.

$\hat{F}$ = the cumulative expected frequency. For this problem we have assumed a Gaussian distribution and F is looked up in a table of Gaussian frequency distributions in any standard statistics book. Remember that the Poisson distribution can be described by a Gaussian distribution when the number of counts in each observation is large.

F = the cumulative relative observed frequency.

For ten observations, each observation increments the cumulative relative observed frequency by 0.10. For 20 observations, each observation increments the cumulative relative observed frequency by 0.05, etc.

$d = |\hat{F} - F|$ = the absolute value of the difference between $\hat{F}$ and F.

The largest value between $\hat{F}$ and F is called d_{max} (Table 3). The value for d_{max} is compared with critical values obtained from standard statistics tables such as is found in Conover *(6)*. If the value in the tables is larger than d_{max} for the statistical confidence level chosen, we conclude that our equipment is operating satisfactorily. If d_{max} is larger than the values found in the statistical tables, we conclude that our equipment is not working satisfactorily.

The data in Table 3 demonstrate that the equipment is working satisfactorily, as the interested reader can confirm as an exercise.

Chi-square test. The chi-square test is perhaps the most commonly used goodness-of-fit test used for assessing the reliability of nuclear counting equipment. However, as previously mentioned in this paper, the statistical manipulation required often precludes its routine use. One of the primary purposes of this communication is to show how this statistic can easily be calculated so that its use can be more widely encouraged in quality control programs.

The definition of the chi-square statistic is defined mathematically as

$$\text{chi-square} = \frac{\Sigma(X_i-\bar{X})^2}{\bar{X}}, \qquad (6)$$

where the symbols have all been previously defined. The above formulation assumes a validity of the Poisson probability distribution wherein the mean value of a series of replicate counts is the best estimate of the expected value. If the variation of the equipment response is substantial, then the value of the chi-square calculated from Eq. 6 will exceed the normal bounds predicted by the Poisson distribution. Further, if the equipment response is too uniform then the value of the chi-square calculated from Eq. 6 will be smaller than the normal bounds predicted by the Poisson distribution. The condition of too uniform a counting rate could theoretically occur, for example, when a constant noise is fed into the system and the counting instrument does not respond to the radioactive source. In the experience of this author, the condition of too uniform a counting rate is easily recognized without resorting to any statistical tests. However, both the upper and lower bounds for the chi-square test have been adapted from Chase and Rabinowitz *(1)* (Table 4).

The samle calculations of the chi-square shown in Table 5 are simple with the use of a programmable calculator, but tedious if programs are not available. However, the recent widespread availability and use of pocket calculators with hardwired functions for the calculation of the mean and standard deviation make an alternate formulation of the chi-square useful. Consider again the definition of the standard deviation S (Eq.2). A simple manipulation of Eq. 2 yields the chi-square value in terms of the standard deviation and the mean of a series of replicate counts of a radioactive source:

$$\text{chi-square} = \frac{S^2 * (n-1)}{\bar{X}}. \qquad (7)$$

Using the data given in Table 5, the interested reader can demonstrate, as an exercise, that Eqs. 6 and 7 yield identical results. From the data in Tables 4 and 5 we conclude, again, that the equipment is working satisfactorily.

In the event that the analyst has neither a program-

TABLE 4. Critical Values for Chi-Square Test

Number of observations	P = 0.98*	P = 0.90†
3	0.02– 9.21	0.10– 5.99
4	0.12–11.34	0.35– 7.82
5	0.30–13.28	0.71– 9.49
6	0.55–15.09	1.14–11.07
7	0.87–16.81	1.64–12.59
8	1.24–18.48	2.17–14.07
9	1.65–20.09	2.74–15.51
10	2.09–21.67	3.33–16.92

**If chi-square value is between tabulated values, there is a 98% confidence level that instrument is working satisfactorily.*

†*If chi-square value is between tabulated values, there is a 90% confidence level that instrument is working satisfactorily.*

TABLE 5. Sample Calculation of Chi-Square Test for Representative Counting Data

X_i	$X_i-\bar{X}$	$(X_i-\bar{X})^2$	$\frac{(X_i-\bar{X})^2}{\bar{X}}$
12036	39.8	1584.04	0.132
12004	7.8	60.84	0.005
11850	−146.2	21374.44	1.782
12153	155.8	24273.64	2.023
12237	240.8	57984.64	4.834
11846	−150.2	22560.04	1.881
11901	−95.2	9063.04	0.755
11932	−64.2	4121.64	0.344
12028	31.8	1011.24	0.084
11976	−20.2	408.04	0.034
$\Sigma X_i = 119962$	$\Sigma(X_i-\bar{X}) = 0$	$\Sigma(X_i-\bar{X})^2 = 142441.6$	$\frac{\Sigma(X_i-\bar{X})^2}{\bar{X}} = 11.87$
n = 10	$\bar{X} = 11996.2$	Chi-square = 11.87	

TABLE 6. Multipliers for Converting Range to Standard Deviation

Number of observations (N)	Multiplier (K)
2	0.886
3	0.591
4	0.486
5	0.430
6	0.395
7	0.379
8	0.351
9	0.337
10	0.325

TABLE 7. Sample Calculation of Chi-Square Test Using Range to Approximate Standard Deviation

X_i	X_i
12237	11996
12152	11976
12036	11901
12028	11850
12004	11846
R = 391	N−1 = 9
$\bar{X}$ = 11996.2	Chi-square = 12.11

mable calculator nor a pocket calculator with the requisite functions, a simple approximation using small sample statistics can be used that is adequate for most circumstances, and the approximation can easily be calculated with a simple adding machine or paper and pencil.

When only a few replicate counts are obtained, which is the usual practice for routine quality control evaluations, the standard deviation can be approximated conveniently from the range, i.e., the difference between the highest and the lowest value. The multipliers required for converting the range into a standard deviation are given in Table 6. These multipliers have been abstracted from Dixon and Massey *(10)*. Thus, Eq. 7 can be transformed as follows:

$$\text{chi-square} = \frac{(K*R)*(n-1)}{\bar{X}}, \qquad (8)$$

where

R = the range,

K = the multiplier (from Table 6) which converts the range into a standard deviation.

Using the same data as previously used, a sample caculation using Eq. 8 is given in Table 7. It is apparent that the variation between the two chi-square values from Tables 4 and 7 does not make any practical difference.

In order to compare the results obtained from either Eq. 7 or 8, a portion of our operating experience with a single instrument is summarized in Table 8.

For routine paper and pencil work in calculating the chi-square value, it is convenient to standardize on ten replicate counts and to round off the multiplier in Table 6 to 1/3. If a question arises as to the interpretation of a particular set of data using this simplified procedure, then the more precise calculations can be carried out. In the experience of this author, the use of the simplified chi-square calculation based on the range has proven entirely adequate.

Practical Example

Our routine quality control program calls for a daily count of a standard ^{137}Cs source to monitor possible changes in instrument sensitivity. The ^{137}Cs source is counted using an energy range of 1 MeV. In spite of an acceptable counting rate with the ^{137}Cs source, the technologists noted that spurious counts were obtained when routine blood samples for a blood volume were counted for ^{125}I using an energy range of 0.25 MeV. In order to differentiate between equipment malfunction and errors in sample preparation, a chi-square test was performed on both the 1- and 0.25-MeV energy ranges (Table 9). Apparently, the equipment was performing satisfactorily on an energy range of 1 MeV but not satisfactorily on an energy range of 0.25 MeV.

The service representative was called in for equipment repair and it was discovered that the contacts on the energy range selector switch were corroded on the 0.25-

TABLE 8. Comparison of Chi-Square Values Obtained Using Calculated Standard Deviation with Standard Deviation Approximated from Range

X^2 using Eq. 7	X^2 using Eq. 8	X^2 using Eq. 7	X^2 using Eq. 8
26.5	27.7	17.8	14.0
7.1	8.5	4.1	3.6
9.5	6.4	20.7	17.4
3.6	5.1	8.8	7.1
19.5	25.4	8.6	4.4

TABLE 9. Comparison of Chi-Square Values Obtained on Different Energy Ranges Before and After Equipment Maintenance

Energy range (MeV)	X^2 value before maintenance	X^2 value after maintenance
0.25	1107	8.64
1.0	4.58	9.98

MeV setting. After the switch was repaired the chi-square test was repeated with satisfactory values being obtained on both the 1- and 0.25-MeV energy ranges (Table 9). Note that cleaning the switch contacts did restore the equipment to a satisfactory operating status.

References

1. Chase GC, Rabinowitz JL: *Principles of Radioisotope Methodology.* Minneapolis, Burgess, 1967, 3rd ed, pp 102-104

2. Prince JR, Schmidt LD: *Statistical and Mathematical Methods in Nuclear Medicine.* Chicago, American Society of Clinical Pathology, 1976, pp 31-32

3. Sodee DB, Early PJ: *Technology and Interpretation of Nuclear Medicine Procedures.* St. Louis, C. V. Mosby, 1975, pp 46-50

4. Evans RD: *The Atomic Nucleus.* New York, McGraw-Hill, 1955, p 774

5. Bleuler E, Goldsmith GD: *Experimental Nucleonics.* New York, Hold, Rinehart and Winston, 1952, p 77

6. Conover WJ: *Practical Nonparametric Statistics.* New York, John Wiley, 1971, p 295

7. Bureau of Radiological Health, eds: *Radiological Health Handbook.* Washington, DC, Government Printing Office, 1970

8. Gindler EM: Some Nonparametric Statistical Tests for Quick Evaluation of Clinical Data. *Clin Chem* 21: 309-314, 1975

9. Wu GT, Twomey SL, Thiers RE: Statistical Evaluation of Method-Comparison Data. *Clin Chem* 21: 315-320, 1975

10. Dixon WD, Massey FJ: *Introduction to Statistical Analysis.* New York, McGraw-Hill, 1969, p 462