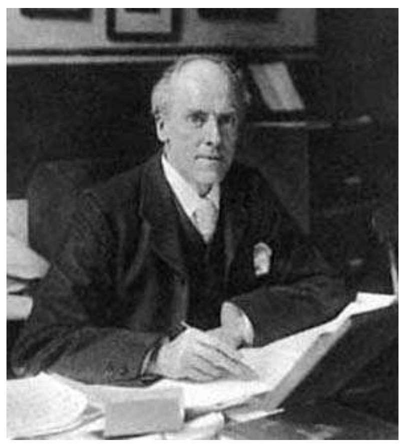
Goodness-of-Fit for STAT112

Charles Fleming

February 4, 2015



Karl Pearson
1857-1936

The goodness-of-fit test is very old. It was presented by Karl Pearson in 1900. One of the principal goals of a statistician is to associate a probability distribution with a histogram of experimental data. Probability distributions lie in the imaginary world of abstract things like events and sample spaces whereas histograms are constructed from actual measurements. The goodness-of-fit test provides an analytical test for determining if a specified distribution may be ascribed to a population. A X^2 quantile will serve the purpose of a measuring stick to judge the fit between the histogram and the probability distribution.

Suppose two dice are tossed twenty times and define a random variable, X, which gives the sum of the faces of the two dice. The observed sum of the faces for each toss is listed here:

2 7 4 9 6 3 8 3 12 4 4 5 7 10 8 3 11 12 4 9

The histogram of the data appears on the left in Figure 1. If the two dice are actually fair, then the distribution of X would assume the Triangle distribution like the one shown on the right in Figure 1. The histogram bears some resemblance to the Triangle distribution, but

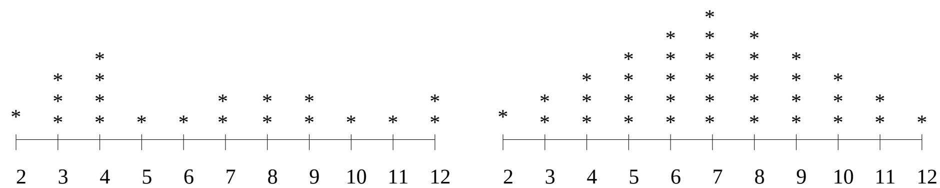


Figure 1

the claim that the histogram and the probability distribution form a good fit is questionable simply based on inspection. Suppose that the dice are indeed fair, then the expected number of 2's which would appear from rolling two dice twenty times will be $np = 20\frac{1}{36}$, and the expected number of 3's will be $np = 20\frac{2}{36}$, and so on. These expectations are listed in Table 1 in which the observed frequency for each value of X appears on the top line, the expected frequency if the dice were fair, appears in the middle row. The bottom row contains the deviations. If the deviations are small, then for practical purposes, the probability distribution agrees with the histogram.

Table 1

Sum of Faces	2	3	4	5	6	7	8	9
Observed	1	3	4	1	1	2	2	2
Expected	$20(\frac{1}{36})$	$20(\frac{2}{36})$	$20(\frac{3}{36})$	$20(\frac{4}{36})$	$20(\frac{5}{36})$	$20(\frac{6}{36})$	$20(\frac{5}{36})$	$20(\frac{4}{36})$
Deviation	.444	1.889	2.333	-1.222	-1.777	-1.333	-.778	-.222

Sum of Faces	10	11	12
Observed	1	1	2
Expected	$20(\frac{3}{36})$	$20(\frac{2}{36})$	$20(\frac{1}{36})$
Deviation	-.667	-.111	1.444

Not surprisingly, the sum of the deviations is equal to zero. In order to eliminate the effect of negative deviations, they are squared and rather unexpectedly, each squared deviation is divided by the expected value.

Definition 1. $X^2 = \sum_{i=1}^{n} \frac{(observed_i - expected_i)^2}{expected_i}$ *is called the* ***chi-squared test statistic****.*

Example 1. *Find the chi-squared test statistic for the previous example of throwing two dice twenty times.*

$$X^2 = \frac{.444^2}{\frac{20}{36}} + \frac{1.889^2}{\frac{40}{36}} + \frac{2.333^2}{\frac{60}{36}} + \frac{(-1.222)^2}{\frac{80}{36}} + \frac{(-1.777)^2}{\frac{100}{36}} + \frac{(-1.333)^2}{\frac{120}{36}} + \frac{(-.778)^2}{\frac{100}{36}} + \frac{(-.222)^2}{\frac{80}{36}} + \frac{(-.667)^2}{\frac{60}{36}} + \frac{(-.111)^2}{\frac{40}{36}} + \frac{1.444^2}{\frac{20}{36}} = 13.44524$$

The X^2 test statistic gives an indication of the discrepancy in the fit between the histogram and the probability distribution. By comparing it to the X^2 quantile, the size of the discrepancy will either be too big to support the claim that the probability distribution adequately fits the data or small enough to say that the fit is not bad. If the X^2 test statistic is too big, then the null hypothesis that the histogram and the probability distribution agree must be rejected. The criterion for rejection is given in the following table.

H_0	Test Statistic	H_1	Reject when
Population has specified distribution	$X^2 = \sum_{i=1}^{n} \frac{(observed_i - expected_i)^2}{expected_i}$	Population does not have specified distribution	$X^2 > X^2_{n-1,\alpha}$

A tabulation of X^2 quantiles is given in Appendix 1.

Example 2. *Test the hypothesis that the empirical distribution shown by the histogram of the frequency of throwing two dice twenty times is the same as the theoretical distribution at a level of significance of* $\alpha = .05$

$X^2_{11-1,.05} = 18.30$

Does $X^2 = 13.445 > 18.307$*? No, cannot reject the null hypothesis that the observed histogram follows the Triangle distribution.*

As imperfect as the shape of the histogram appears in relation to the Triangle distribution, the conclusion of the goodness-of-fit test substantiates the claim that the Triangle distribution may be used to account for the experimental outcomes of the actual tossing of two dice. The implication is that the characteristics of the population which are manifested in the experimental results from tossing of two dice not only twenty times but any number of times may be adequately explained by the Triangle distribution. Furthermore, having successfully made the association between the population which produced the histogram and the sample space consisting of two imaginary fair dice, we may say that the real dice, too, are probably fair. Although we should not say that the goodness-of-fit test proves that the two dice are fair, yet the conclusion of not rejecting the null hypothesis indicates that the dice are probably fair and that the Triangle distribution may adequately describe the population until additional evidence demonstrates otherwise.

1 Contingency Table

The X^2 test can be extended from one to two dimensions, for example:

Problem 1. *A random sample of 200 married men, all of whom are retired, were classified according to education and to the number of children whom they sired.*

	Number of Children			
Education	0-1	2-3	over 3	Row Totals
Elementary	14 (18.675)	37 (39.84)	32 (24.495)	83
Secondary	19 (17.55)	42 (37.44)	17 (23.01)	78
College	12 (8.779)	17 (18.72)	10 (11.505)	39
Column Totals	45	96	59	200

Definition 2. *This table is called a **contingency table**. An element of it is called a **cell**.*

The numbers written within parentheses are the expected number of occurrences if education and number of children are independent.

Question 1. *Are education and number of children independent events?*

Let A be the event of siring 0-1 children. Let B be the event of only getting an elementary school education. If A and B are independent, then $P(A \cap B) = P(A)P(B)$ where $P(A) = \frac{45}{200}$ and $P(B) = \frac{83}{200}$. In the case of independence, what would be the expected number of men who sired 0-1 children but got an elementary school education? $np = 200P(A \cap B) = 200P(A)P(B) = 200\frac{45}{200}\frac{83}{200} = \frac{45 \times 83}{200} = 18.675$. This same number appears in the contingency table within parentheses.

What is the expected number of men who got an elementary education and sired 2-3 children? Let C be the event of siring 2-3 children. Let B be the event of getting an elementary school education. If C and B are independent, then $np = 200P(C \cap B) = 200P(C)P(B) = 200\frac{96}{200}\frac{83}{200} = 39.84$.

It does not take long to see a pattern emerge from calculating the expected frequencies. We will use the pattern to shorten the computations as is done in our final example. The expected number in the cell for siring 3 or more children with a college education is $\frac{59 \times 39}{200} = 11.505$.

The criterion for rejecting the null hypothesis of a contingency table is given below.

H_0	Test Statistic	H_1	reject when
Rows and columns are Independent	$X^2 = \sum_{i=1}^{n} \frac{(observed_i - expected_i)^2}{expected_i}$	Not independent	$X^2 > X^2_{\nu,\alpha}$ where $\nu = (r-1)(c-1)$ r=number of rows c=number of columns

Example 3. *In our example, r=3 and c=3; therefore,* $\nu = (3-1)(3-1) = 4$*. Suppose that* $\alpha = .05$*. The appropriate* X^2 *quantile for conducting a goodness-of-fit test is:* $X^2_{4,.05} = 9.48$*.*

$$X^2 = \sum_{i=1}^{9} \frac{(observed_i - expected_i)^2}{expected_i} = \frac{(14-18.675)^2}{18.675} + \frac{(37-39.84)^2}{39.84} + \ldots + \frac{(10-11.505)^2}{11.505} = 7.4626.$$

Is $X^2 = 7.4626 > 9.48$*? No. Therefore, we cannot reject the null hypothesis. Hence, based on the data, a man's education and the number of children whom he sires appear to be independent at a level of significance of* $\alpha = .05$*.*

The choice of α has thus far been arbitrary. For sufficiently large α's, the null hypothesis can be rejected. For sufficiently small α's, the null hypothesis cannot be rejected. That α which lies exactly at the boundary of admitting a rejection or no rejection is called the *p value* of the test. It is often published with the results of an analysis for the benefit of the reader. The p value for the above test is $p = P(X^2_4 > 7.4626) = .113$. We used $\alpha = .05$ in conducting the test but since $\alpha < p$, the null hypothesis could not have been rejected. Only when α exceeds .113 will the null hypothesis be rejected.

Example 4. *An experiment was conducted to investigate the effect of a vaccination on laboratory animals. Some animals when exposed to the disease contracted it, and some did not according to whether the animal was inoculated. The developer hopes that the vaccine and the contraction of the disease are not independent. To prove his belief, the null hypothesis was formulated to assume the worst case in that the vaccination and the likelihood of getting a disease are independent in the hope that it will be rejected at a level of significance of .05. A tabulation of the results appears below.*

	Got the Disease	Did not get the Disease	
Vaccinated	9 (13.84)	42 (37.19)	51
Not Vaccinated	17 (12.19)	28 (32.81)	45
Column Totals	26	70	96

1. $\alpha = .05$; $\nu = (r-1)(c-1) = 1$.

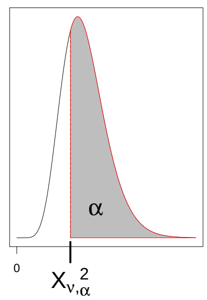
2. *The expected number of cases assuming independence is given in parentheses.*

3. $X^2 = \sum_{i=1}^{4} \frac{(observed_i - expected_i)^2}{expected_i} = \frac{(9-13.84)^2}{13.84} + \frac{(42-37.19)^2}{37.19} + \frac{(17-12.19)^2}{12.19} + \frac{(28-32.81)^2}{32.81} = 4.918$.

4. $X^2_{1,.05} = 3.48146$.

5. *Is* $X^2 = 4.918 > 3.48$? *Yes. Reject the null hypothesis that the vaccination and contracting the disease are independent. In conclusion, based on the data, it appears that the vaccination prevents the contraction of the disease.*

The p value of the test is that α at which the null hypothesis can and cannot be rejected. It is $p = .0266 = P(X^2_1 \geq 4.918)$. For an $\alpha > p$, the null hypothesis will be rejected; for an $\alpha < p$, the null hypothesis cannot be rejected. In other words, suppose α is chosen slightly larger than p, like $\alpha = .0267$, then $X^2_{1,.0267} = 4.910$ and because $4.918 > 4.910$, the null hypothesis is rejected. Suppose, on the other hand, α is chosen slightly smaller that p, like $\alpha = .0265$, then $X^2_{1,.0265} = 4.923$ and because $4.918 \not> 4.923$, the null hypothesis cannot be rejected. The use of the *p-value* offers a reader a way to judge the proximity of the test statistic to the boundary of the rejection region which the p value marks.



Quantiles for a X^2 Distribution

ν	$X^2_{\nu,.20}$	$X^2_{\nu,.15}$	$X^2_{\nu,.10}$	$X^2_{\nu,.05}$	$X^2_{\nu,.025}$	$X^2_{\nu,.01}$	$X^2_{\nu,005}$
1	1.64237	2.07225	2.70554	3.84146	5.02389	6.63490	7.87944
2	3.21888	3.79424	4.60517	5.99146	7.37776	9.21034	10.59663
3	4.64163	5.31705	6.25139	7.81473	9.34840	11.34487	12.83816
4	5.98862	6.74488	7.77944	9.48773	11.14329	13.27670	14.86026
5	7.28928	8.11520	9.23636	11.07050	12.83250	15.08627	16.74960
6	8.55806	9.44610	10.64464	12.59159	14.44938	16.81189	18.54758
7	9.80325	10.74790	12.01704	14.06714	16.01276	18.47531	20.27774
8	11.03009	12.02707	13.36157	15.50731	17.53455	20.09024	21.95495
9	12.24215	13.28804	14.68366	16.91898	19.02277	21.66599	23.58935
10	13.44196	14.53394	15.98718	18.30704	20.48318	23.20925	25.18818
11	14.63142	15.76710	17.27501	19.67514	21.92005	24.72497	26.75685
12	15.81199	16.98931	18.54935	21.02607	23.33666	26.21697	28.29952
13	16.98480	18.20198	19.81193	22.36203	24.73560	27.68825	29.81947
14	18.15077	19.40624	21.06414	23.68479	26.11895	29.14124	31.31935
15	19.31066	20.60301	22.30713	24.99579	27.48839	30.57791	32.80132
16	20.46508	21.79306	23.54183	26.29623	28.84535	31.99993	34.26719
17	21.61456	22.97703	24.76904	27.58711	30.19101	33.40866	35.71847
18	22.75955	24.15547	25.98942	28.86930	31.52638	34.80531	37.15645
19	23.90042	25.32885	27.20357	30.14353	32.85233	36.19087	38.58226
20	25.03751	26.49758	28.41198	31.41043	34.16961	37.56623	39.99685
21	26.17110	27.66201	29.61509	32.67057	35.47888	38.93217	41.40106
22	27.30145	28.82245	30.81328	33.92444	36.78071	40.28936	42.79565
23	28.42879	29.97919	32.00690	35.17246	38.07563	41.63840	44.18128
24	29.55332	31.13246	33.19624	36.41503	39.36408	42.97982	45.55851
25	30.67520	32.28249	34.38159	37.65248	40.64647	44.31410	46.92789
26	31.79461	33.42947	35.56317	38.88514	41.92317	45.64168	48.28988
27	32.91169	34.57358	36.74122	40.11327	43.19451	46.96294	49.64492
28	34.02657	35.71499	37.91592	41.33714	44.46079	48.27824	50.99338
29	35.13936	36.85383	39.08747	42.55697	45.72229	49.58788	52.33562
30	36.25019	37.99025	40.25602	43.77297	46.97924	50.89218	53.67196
40	47.26854	49.24385	51.80506	55.75848	59.34171	63.69074	66.76596
50	58.16380	60.34599	63.16712	67.50481	71.42020	76.15389	79.48998
60	68.97207	71.34110	74.39701	79.08194	83.29767	88.37942	91.9517
70	79.71465	82.25535	85.52704	90.53123	95.02318	100.42518	104.2149
80	90.40535	93.10575	96.57820	101.87947	106.62857	112.32879	116.32106
90	101.05372	103.90406	107.56501	113.14527	118.13589	124.11632	128.29894
100	111.66671	114.65882	118.49800	124.34211	129.56120	135.80672	140.16949
150	164.34919	167.96177	172.58121	179.58063	185.80045	193.20769	198.36021
200	216.60878	220.74413	226.02105	233.99427	241.05790	249.44512	255.26416