Republic of Iraq Ministry of Higher Education & Research

University of Anbar

College of Education for Pure Sciences

Department of Mathematics

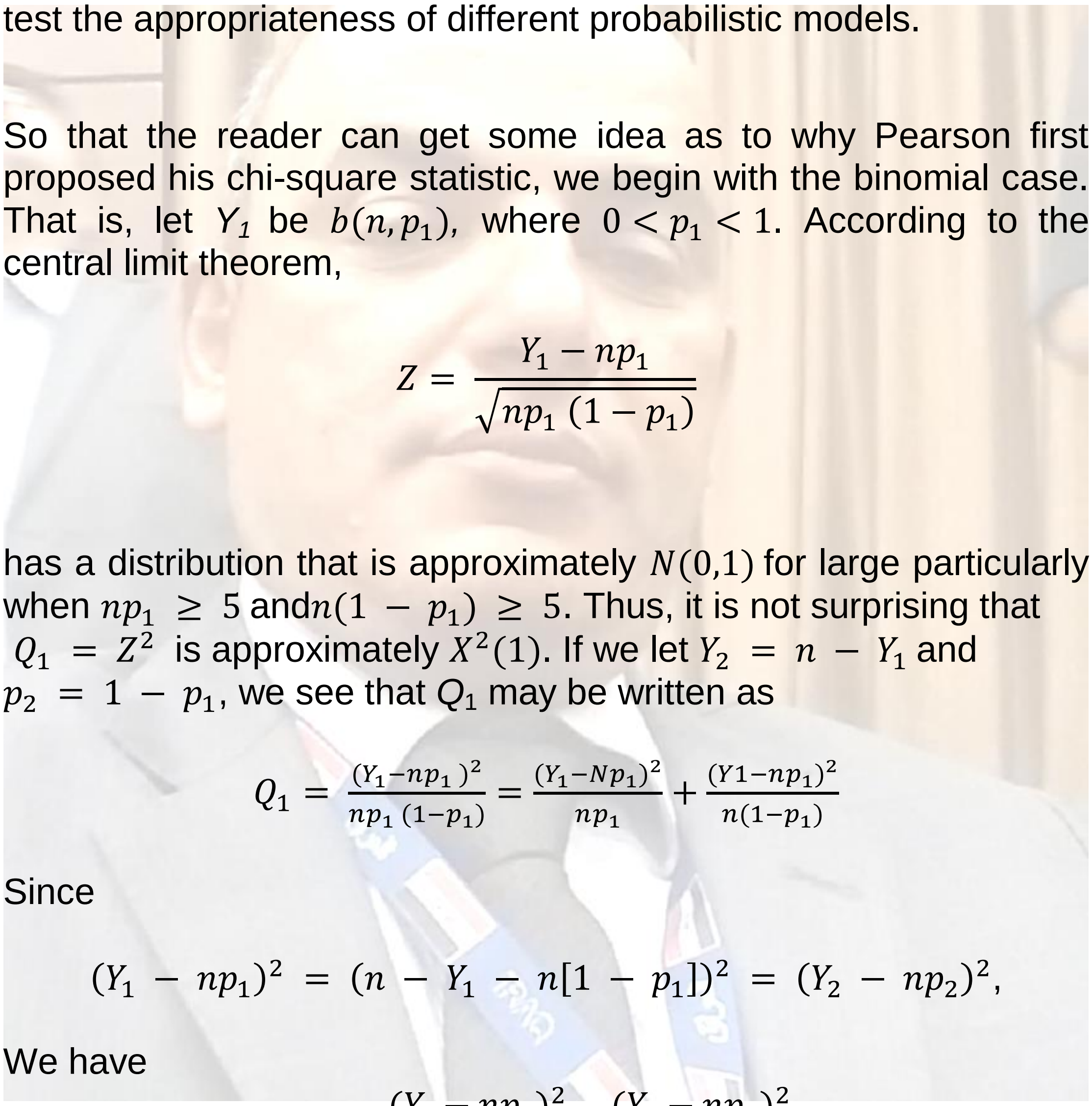
محاضرات الاحصاء ٢

مدرس المادة : الاستاذ المساعد

الدكتور فراس شاكر محمود

CHI-SQUARE GOODNESS-OF-FIT TESTS

We now consider applications of the very important chi-square statistic, first proposed by Karl Pearson in 1900. As the reader will see, it is a very adaptable test statistic and can be used for many different types of tests. In particular, one application allows us to test the appropriateness of different probabilistic models.



So that the reader can get some idea as to why Pearson first proposed his chi-square statistic, we begin with the binomial case. That is, let Y_1 be $b(n,p_1)$, where $0 < p_1 < 1$. According to the central limit theorem,

$$Z = \frac{Y_1 - np_1}{\sqrt{np_1\,(1-p_1)}}$$

has a distribution that is approximately $N(0,1)$ for large particularly when $np_1 \geq 5$ and$n(1-p_1) \geq 5$. Thus, it is not surprising that $Q_1 = Z^2$ is approximately $X^2(1)$. If we let $Y_2 = n - Y_1$ and $p_2 = 1 - p_1$, we see that Q_1 may be written as

$$Q_1 = \frac{(Y_1 - np_1)^2}{np_1\,(1-p_1)} = \frac{(Y_1 - Np_1)^2}{np_1} + \frac{(Y1 - np_1)^2}{n(1-p_1)}$$

Since

$$(Y_1 - np_1)^2 = (n - Y_1 - n[1 - p_1])^2 = (Y_2 - np_2)^2,$$

We have

$$Q_1 = \frac{(Y_1 - np_1)^2}{np_1} + \frac{(Y_2 - np_2)^2}{np_2}$$

Let us now carefully consider each term in this last expression for Q_1. Of course, $Y_1$is the number of “successes,” and $np_1$is the expected number of “successes” that is,$E(Y_1) = np_1$. Likewise, Y_2 and np_2 are, respectively, the number and the expected number of “failures.”

So each numerator consists of the square of the difference of an observed number and an expected number. Note that Q_1can be written as

$$Q_1 = \sum_{i=1}^{2} \frac{(Y_i - np_i)^2}{np_i},$$

and we have seen intuitively that it has an approximate chi-square distribution with one degree of freedom. In a sense, Q_1 measures the “closeness” of the observed numbers to the corresponding expected numbers. For example, if the observed values of $Y_1$and Y_2equal their expected values, then the computed Q_1 is equal to $q_1 = 0$; but if they differ much from them, then the computed $Q_1 = q_1$is relatively Large.

To generalize, we let an experiment have *k* (instead of only two) mutually exclusive and exhaustive outcomes, say,$A_1, A_2, \ldots, A_k$. Let$p_i = P(A_i)$, and thus$\sum_{i=1}^{k} p_i = 1$. The experiment is repeated *n* independent times, and we let *Y_i* represent the number of times the experiment results in $A_i, i = 1, 2, \ldots, k$. This joint distribution of $Y_1, Y_2, \ldots, Y_{k-1}$ is a straightforward generalization of the binomial distribution, as follows. In considering the joint pmf, we see that

$$f(y_1, y_2, \ldots, y_{k-1}) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_{k-1} = y_{k-1}),$$

where $y_1, y_2, \ldots, y_{k-1}$ are nonnegative integers such that $y_1 + y_2 + \cdots + y_{k-1} \leq n$. Note that we do not need to consider Y_k, since, once the other *k*–1 random variables are observed to equal $y_1, y_2, \ldots, y_{k-1}$, respectively, we know that

$$Y_k = n - y_1 - y_2 - \cdots - y_{k-1} = y_k$$

From the independence of the trials, the probability of each particular arrangement of $y_1 A_1 s, y_2 A_2 s, \ldots, y_k A_k s$ is

$$p_1^{y_1} p_2^{y_2} \cdot \cdot \cdot \ p_k^{y_k}$$

The number of such arrangements is the multinomial coefficient

$$\binom{n}{y1, y2, \ldots, yk} = \frac{n!}{y1!\, y2! \ldots yk!}$$

Hence, the product of these two expressions gives the joint pmf of $Y_1, Y_2, \dots, Y_{k-1}$

$$f(y_1, y_2, \dots, y_{k-1}) = \frac{n!}{y1!\, y2! \dots yk!} p_1^{y_1} p_2^{y_2} .\ .\ .\ p_k^{y_k}$$

(Recall that $y_k = n - y_1 - y_2 - \cdots - y_{k-1}$)

Pearson then constructed an expression similar to Q_1, which involves Y_1 and $Y_2 = n - Y_1$, that we denote by Q_{k-1}, which involves $Y_1, Y_2, \dots, Y_{k-1}$, and $Y_k = n - Y_1 - Y_2 - \cdots - Y_{k-1}$, namely,

$$Q_1 = \sum_{i=1}^{k} \frac{(Y_i - np_i)^2}{np_i}$$

He argued that Q_{k-1} has an approximate chi-square distribution with $k-1$degrees of freedom in much the same way we argued that Q_1 is approximately $X^2(1)$. We accept Pearson’s conclusion, as the proof is beyond the level of this text. Some writers suggest that *n* should be large enough so that $np_i \geq 5, i = 1,2, \dots, k,$ to be certain that the approximating distribution is adequate. This is probably good advice for the beginner to follow, although we have seen the approximation work very well when $np_i \geq 1, i = 1,2, \dots, k$. The important thing to guard against is allowing some particular np_i to become so small that the corresponding term in Q_{k-1}, namely, $(Y_i - np_i)2/np_i$, tends to dominate the others because of its small denominator. In any case, it is important to realize that Q_{k-1} has only an approximate chi-square distribution. We shall now show how we can use the fact that Q_{k-1} is approximately $X^2(k-1)$ to test hypotheses about probabilities of various outcomes. Let an experiment have *k* mutually exclusive and exhaustive outcomes, $A_1, A_2, \dots, A_k$. We would like to test whether $p_i = P(A_i)$ is equal to a known number $p_{i0}, i = 1, 2, \dots, k$. That is, we shall test the hypothesis

$$H_0 : p_i = p_{i0}, \qquad i = 1, 2, \dots, k.$$

In order to test such a hypothesis, we shall take a sample of size *n*; that is, we repeat the experiment *n* independent times. We tend to favor H_0 if the observed number of times that A_i occurred, say,

y_i, and the number of times $A_i$was expected to occur if H_0were true, namely, np_{i0}, are approximately equal. That is, if

$$q_{k-1} = \sum_{i=1}^{k} \frac{(yi - np_{i0})^2}{np_{i0}}$$

is “small,” we tend to favor H_0. Since the distribution of $Q_{k-1}$is approximately $X^2(k-1)$ we shall reject H_0if $q_{k-1} \geq X^2_{\propto}(k-1)$, where $\propto$ is the desired significance level of the test.

<u>Example</u>:-

If persons are asked to record a string of random digits, such as

$$3 \quad 7 \quad 2 \quad 4 \quad 1 \quad 9 \quad 7 \quad 2 \quad 1 \quad 5 \quad 0 \quad 8 \ldots,$$

we usually find that they are reluctant to record the same or even the two closest numbers in adjacent positions. And yet, in true random-digit generation, the probability of the next digit being the same as the preceding one is $p_{10} = 1/10$, the probability of the next being only one away from the preceding (assuming that 0 is one away from 9) is $p_{20} = 2/10$, and the probability of all other possibilities is $p_{30} = 7/10$. We shall test one person’s concept of a random sequence by asking her to record a string of 51 digits that seems to represent a random-digit generation. Thus, we shall test

$$H_0 : p_1 = p_{10} = \frac{1}{10}\, , p_2 = p_{20} = \frac{2}{10}\, , p_3 = p_{30} = \frac{7}{10}$$

The critical region for an $\propto = 0.05$ significance level is $q_2 \geq X^2_{0.005}(2) = 5.991$. The sequence of digits was as follows:

$$5 \quad 8 \quad 3 \quad 1 \quad 9 \quad 4 \quad 6 \quad 7 \quad 9 \quad 2 \quad 6 \quad 3 \quad 0$$

$$8 \quad 7 \quad 5 \quad 1 \quad 3 \quad 6 \quad 2 \quad 1 \quad 9 \quad 5 \quad 4 \quad 8 \quad 0$$

$$3 \quad 7 \quad 1 \quad 4 \quad 6 \quad 0 \quad 4 \quad 3 \quad 8 \quad 2 \quad 7 \quad 3 \quad 9$$

$$8 \quad 5 \quad 6 \quad 1 \quad 8 \quad 7 \quad 0 \quad 3 \quad 5 \quad 2 \quad 5 \quad 2$$

We went through this listing and observed how many times the next digit was the same as or was one away from the preceding one:

	Frequency	Expected Number
Same	0	50(1/10) = 5
One way	8	50(2/10) = 10
Other	42	50(7/10) = 35
Total	50	50

The computed chi-square statistic is

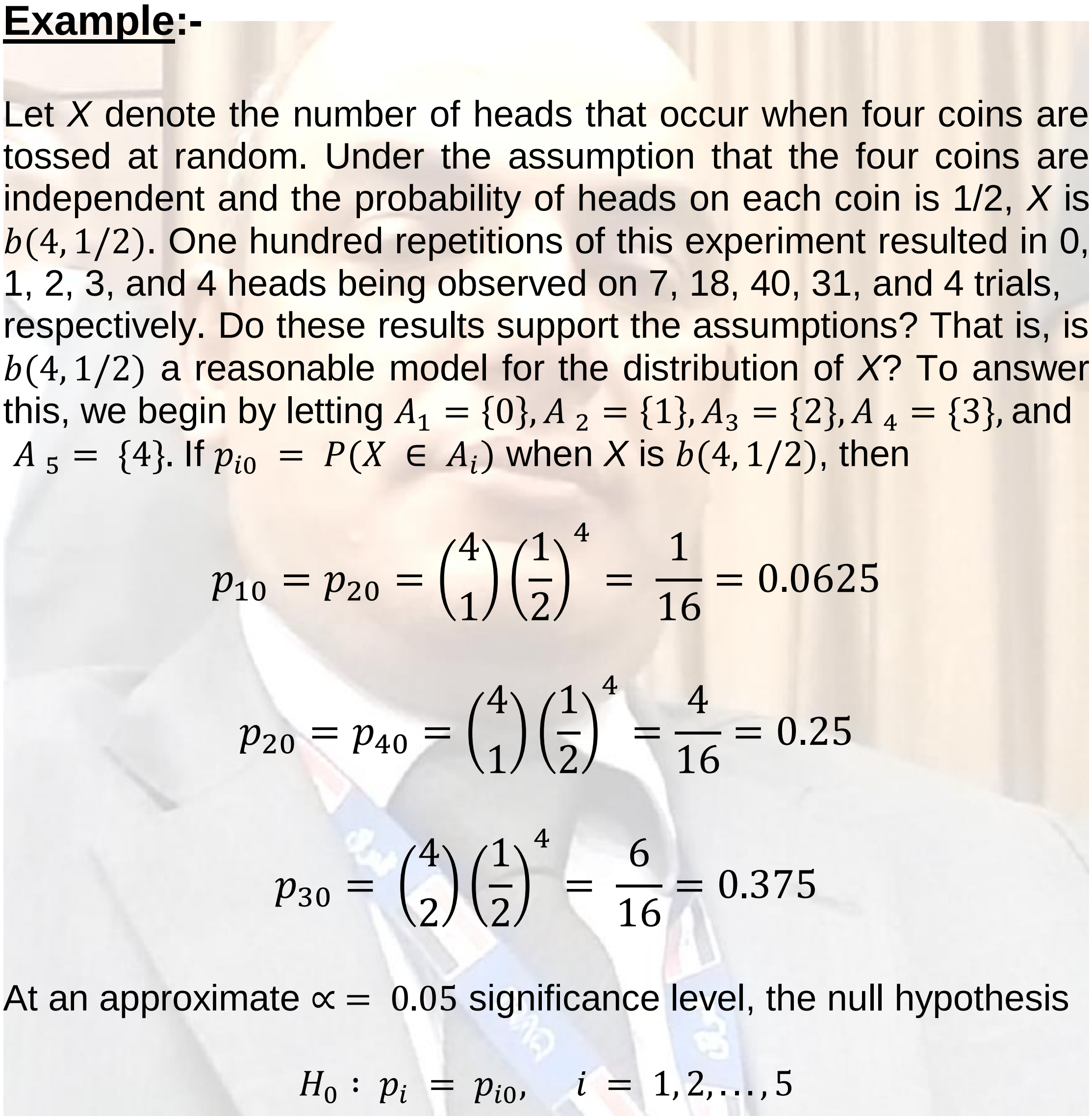
$$\frac{(0-5)^2}{5}+\frac{(8-10)^2}{10}+\frac{(42-35)^2}{35}=6.8\ >5.991=x^2_{0.005}(2)$$

Thus, we would say that this string of 51 digits does not seem to be random.

One major disadvantage in the use of the chi-square test is that it is a many sided test. That is, the alternative hypothesis is very general, and it would be difficult to restrict alternatives to situations such as $H_1\colon p_1 > p_{10}, p_2 > p_{20}, p_3 < p_{30}$ (with $k = 3$). As a matter of fact, some statisticians would probably test H_0 against this particular alternative H_1 by using a linear function of Y_1, Y_2, and Y_3. However, that sort of discussion is beyond the scope of the book because it involves knowing more about the distributions of linear functions of the dependent random variables Y_1, Y_2, and Y_3. In any case, the student who truly recognizes that this chi-square statistic tests $H_0\colon p_i = p_{i0}, i = 1, 2, \ldots, k$, against all alternatives can usually appreciate the fact that it is more difficult to reject H_0 at a given significance level $\propto$ when the chi-square statistic is used than it would be if some appropriate “one-sided” test statistic were available.

Many experiments yield a set of data, say, $x_1, x_2, \ldots, x_n$, and the experimenter is often interested in determining whether these data can be treated as the observed values of a random sample $X_1, X_2, \ldots, X_n$ from a given distribution. That is, would this proposed distribution be a reasonable probabilistic model for these sample items? To see how the chi-square test can help us answer questions of this sort, consider a very simple example.

Example:-



Let *X* denote the number of heads that occur when four coins are tossed at random. Under the assumption that the four coins are independent and the probability of heads on each coin is 1/2, *X* is $b(4, 1/2)$. One hundred repetitions of this experiment resulted in 0, 1, 2, 3, and 4 heads being observed on 7, 18, 40, 31, and 4 trials, respectively. Do these results support the assumptions? That is, is $b(4, 1/2)$ a reasonable model for the distribution of *X*? To answer this, we begin by letting $A_1 = \{0\}, A_2 = \{1\}, A_3 = \{2\}, A_4 = \{3\}$, and $A_5 = \{4\}$. If $p_{i0} = P(X \in A_i)$ when *X* is $b(4, 1/2)$, then

$$p_{10} = p_{20} = \binom{4}{1}\left(\frac{1}{2}\right)^4 = \frac{1}{16} = 0.0625$$

$$p_{20} = p_{40} = \binom{4}{1}\left(\frac{1}{2}\right)^4 = \frac{4}{16} = 0.25$$

$$p_{30} = \binom{4}{2}\left(\frac{1}{2}\right)^4 = \frac{6}{16} = 0.375$$

At an approximate $\propto = 0.05$ significance level, the null hypothesis

$$H_0 : p_i = p_{i0}, \quad i = 1, 2, \ldots, 5$$

is rejected if the observed value of Q_4 is greater than $X^2_{0.005}(4) = 9.488$ If we use the 100 repetitions of this experiment that resulted in the observed values $y_1 = 7, y_1 = 18, y_3 = 40, y_4 = 31, and \; y_5 = 4$, of $Y_1, Y_2, \ldots, Y_5$, respectively, then the computed value of Q_4 is

$$q_4 = \frac{(7\ -\ 6.25)^2}{6.25} + \frac{(18\ -\ 25)^2}{25} + \frac{(40\ -\ 37.5)^2}{37.5} + \frac{(4\ -\ 6.25)^2}{6.25} + \frac{(31\ -\ 25)^2}{25} = 4.47$$

Since $4.47 < 9.488$, the hypothesis is not rejected. That is, the data support the hypothesis that $b(4, 1/2)$ is a reasonable probabilistic model for *X*. Recall that the mean of a chi-square random variable is its number of degrees of freedom. In this example, the mean is 4 and the observed value of Q_4 is 4.47, just a little greater than the mean. Thus far, all the hypotheses H_0 tested with the chi-square statistic Q_{k-1}have been simple ones (i.e., completely specified_ namely, in $H_0 : p_i = p_{i0},\ i = 1, 2, \ldots, k$, each p_{i0} has been known). This is not always the case, and it frequently happens that $p_{10}, p_{20}, \ldots, p_{k0}$ are functions of one or more unknown parameters. For example, suppose that the hypothesized model for X in (Example 2) was $H_0 \colon X$ is $b(4, p)$, $0 < p < 1$. Then

$$p_{i0} = P(X \in A_i) = \frac{4!}{(i-1)!(5-i)!} p^{i-1}(1-p)^{5-i}, \quad i = 1, 2, \ldots, 5,$$

which is a function of the unknown parameter *p*. Of course, if $H_0 : p_i = p_{i0}, \quad i = 1, 2, \ldots, 5$, is true, then, for large *n*,

$$Q_4 = \sum_{i=1}^{5} \frac{(Yi -\ np_{i0})^2}{np_{i0}}$$

still has an approximate chi-square distribution with four degrees of freedom. The difficulty is that when $Y_1, Y_2, \ldots, Y_5$ are observed to be equal to $y_1, y_2, \ldots, y_5$, Q_4 cannot be computed, since $p_{10}, p_{20}, \ldots, p_{50}$ (and hence Q_4) are functions of the unknown parameter *p*.

One way out of the difficulty would be to estimate *p* from the data and then carry out the computations with the use of this estimate. It is interesting to note the following: Say the estimation of *p* is carried out by minimizing Q_4 with respect to *p*, yielding *p*˜. This *p*˜ is sometimes called a minimum chi-square estimator of *p*. If, then, this *p*˜ is used in Q_4, the statistic Q_4 still has an approximate chi-square distribution, but with only 4 − 1 = 3 degrees of freedom. That is, the number of degrees of freedom of the approximating chi-square distribution is reduced by one for each parameter estimated by the minimum chi-square technique. We accept this result without proof (as it is a rather difficult one). Although we have considered it when $p_{i0}, i = 1, 2, \ldots, k$, is a function of only one parameter, it holds when there is more than one unknown parameter, say, *d*. Hence, in a more general situation, the test would be completed by computing Q_{k-1}, using Y_i and the estimated $p_{i0}, i = 1, 2, \ldots, k$, to obtain q_{k-1}(i.e., $q_{k-1}$is the minimized chi-square). This value q_{k-1} would then be compared with a critical value $X_{\propto}^{2}(k-1-d)$. In our special case, the computed (minimized) chi-square q_4 would be compared with $X_{\propto}^{2}(3)$.

There is still one trouble with all of this: It is usually very difficult to find minimum chi-square estimators. Hence, most statisticians usually use some reasonable method of estimating the parameters. (Maximum likelihood is satisfactory.) They then compute q_{k-1}, recognizing that it is somewhat larger than the minimized chi-square, and compare it with $X_{\propto}^{2}\,(k-1-d)$. Note that this approach provides a slightly larger probability of rejecting H_0 than would the scheme in which the minimized chi-square were used because the computed q_{k-1} is larger than the minimum q_{k-1}.

CONTINGENCY TABLES

We demonstrate the flexibility of the chi-square test. We first look at a method for testing whether two or more multinomial distributions are equal, sometimes called a ***test for homogeneity***. Then we consider a ***test for independence of attributes of classification***. Both of these lead to a similar test statistic. Suppose that each of two independent experiments can end in one of the *k*

mutually exclusive and exhaustive event $A_1, A_2, . . . A_k$. Let

$$p_{ij} = P(A_i) \qquad i = 1,2, . . . K \qquad j = 1,2$$

That is, $p_{11}, p_{21}, . . . p_{k1}$are the probabilities of the events in the first experiment, and $p_{12}, p_{22}\,, . . . p_{k2}$are those associated with the second experiment. Let the experiments be repeated $n_1$and n_2 independent times, respectively. Also, let $Y_{11}, Y_{12}, . . ., Y_{k1}$ be the frequencies of $A_1, A_2, . . . , A_k$ associated with the n_1 independent trials of the first experiment. Similarly, let $Y_{12}, Y_{22}, . . . , Y_{k2}$be the respective frequencies associated with the n_2 trials of the second experiment. Of course, $\sum_{i=1}^{k} Y_{ij} = n_j$, $\quad j = 1,2$ From the sampling distribution theory corresponding to the basic chi-square test, we know that each of

$$\sum_{i=1}^{k} \frac{(Y_{ij} - n_j p_{ij})^2}{n_i p_{ij}}, \qquad j = 1,2$$

has an approximate chi-square distribution with $k - 1$ degrees of freedom. Since the two experiments are independent (and thus the two chi-square statistics are independent), the sum

$$\sum_{j=1}^{2} \sum_{i=1}^{k} \frac{(Y_{ij} - n_j p_{ij})^2}{n_j p_{ij}}$$

is approximately chi-square with $k - 1 + k - 1 = 2k - 2$ degrees of freedom. Usually, the $p_{ij},\ i = 1,2 ,. . . ,k ,j = 1,2$ are unknown, and frequently we wish to test the hypothesis

$$H_0: P_{11} = P_{12}, P_{21} = P_{22}, \quad . . . \quad , P_{K1} = P_{K2}$$

that is, H_0 is the hypothesis that the corresponding probabilities associated with the two independent experiments are equal. Under H_0, we can estimate the unknown $p_{i1} = p_{i2} \qquad i = 1,2, . \ . \ . \ ,k$ by using the relative frequency (Y_{i1}+Y_{i2}) / (n_1+ n_2), *i* = 1, 2, . . . , *k*. That is, if H_0

is true, we can say that the two experiments are actually parts of a larger one in which Y_{i1}+Y_{i2} is the frequency of the event $A_i\ ,i = 1,2,. . .,k$Note that we have to estimate only the $k - 1$ probabilities

$p_{i1} = p_{i2}$, using

$$\frac{Y_{i1} + Y_{i2}}{n_1 + n_2}, \qquad i = 1,2,. . .k - 1$$

since the sum of the k probabilities must equal 1. That is, the estimator of $p_{k1} = p_{k2}$ is

$$1 - \frac{Y_{11} + Y_{12}}{n_1 + n_2} -. . .- \frac{Y_{k-1,1} + Y_{k-1,2}}{n_1 + n_2} = \frac{Y_{k1} + Y_{k2}}{n_1 + n_2}$$

Substituting these estimators, we find that

$$Q = \sum_{j=1}^{2} \sum_{i=1}^{k} \frac{\left[Y_{ij} - \frac{n_j (Y_{i1} + Y_{i2})}{n_1 + n_2}\right]^2}{\frac{n_j (Y_{i1} + Y_{i2})}{n_1 + n_2}}$$

has an approximate chi-square distribution with $2k - 2 - (k - 1) = k - 1$ degrees of freedom. Here $k - 1$ is subtracted from $2k - 2,$ because that is the number of estimated parameters. The critical region for testing H_0 is of the form

$$q \geq X^2_{\propto}(K-1)$$

Example:

To test two methods of instruction, 50 students are selected at random from each of two groups. At the end of the instruction period, each student is assigned a grade (A, B, C, D, or F) by an evaluating team. The data are recorded as follows:

	Grade					
	A	B	C	D	F	Totals
Group I	8	13	16	10	3	50
Group II	14	9	14	16	7	50

Accordingly, if the hypothesis H_0 that the corresponding probabilities are equal is true, then the respective estimates of the probabilities are

$$\frac{8+4}{100} = 0.12, 0.22, 0.30, 0.26, \frac{3+7}{100} = 0.10$$

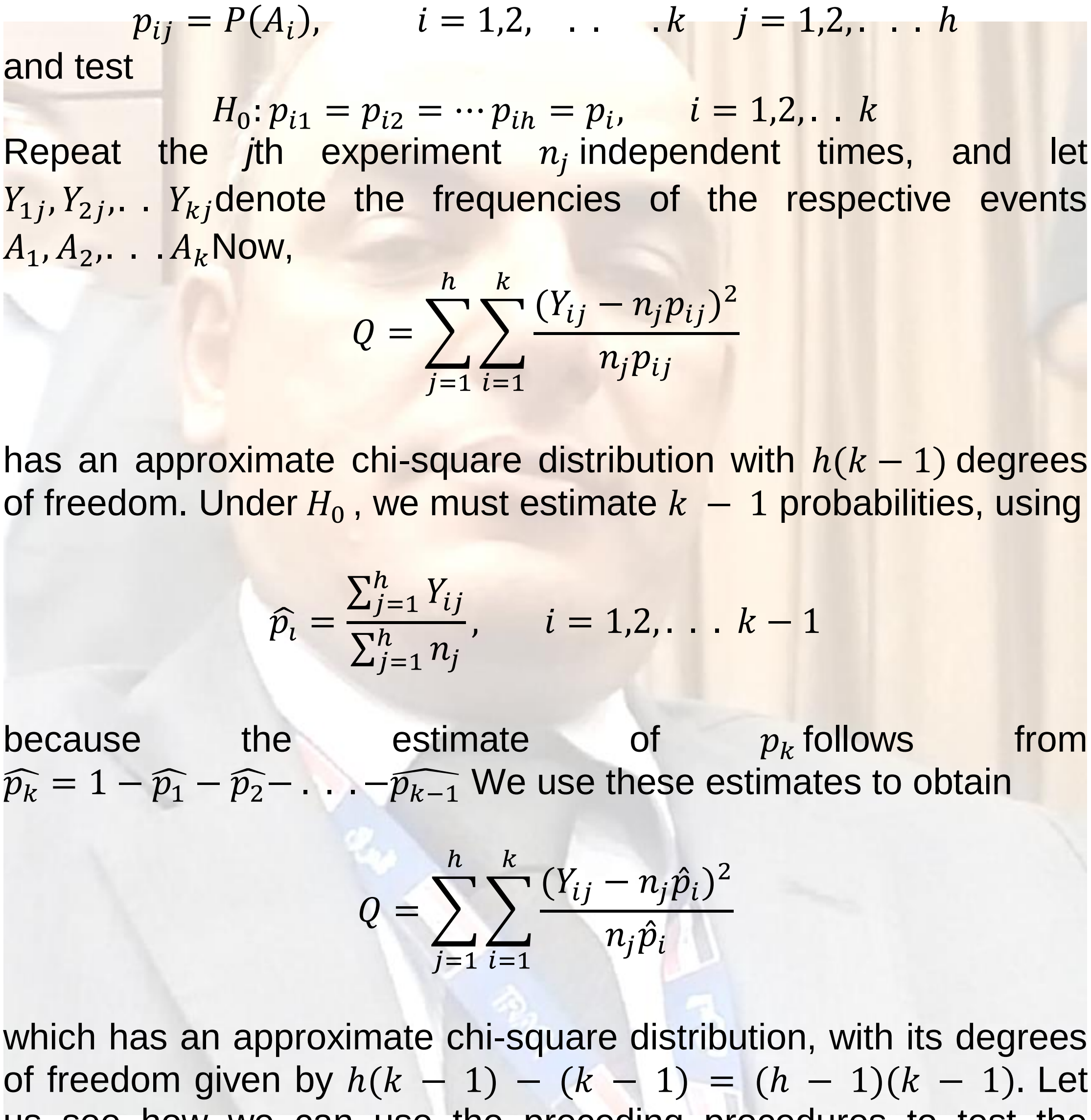
Thus, the estimates of $n_1 p_{i1} = n_2 p_{i2}$ are 6, 11, 15, 13, and 5, respectively. Hence, the computed value of Q is

$$\begin{aligned} q = \frac{(8-6)^2}{6} + \frac{(13\ -\ 11)^2}{11} + \frac{(10\ -\ 13)^2}{13} + \frac{(3\ -\ 5)^2}{5} + \frac{(4\ -\ 6)^2}{6} \\ + \frac{(9\ -\ 11)^2}{11} + \frac{(14\ -\ 15)^2}{15} + \frac{(16\ -\ 13)^2}{13} + \frac{(7\ -\ 5)^2}{5} \\ = \frac{4}{6} + \frac{4}{11} + \frac{1}{15} + \frac{9}{13} + \frac{4}{5} + \frac{4}{6} + \frac{4}{11} + \frac{1}{15} + \frac{9}{13} + \frac{4}{5} = 5.18 \end{aligned}$$

Now, under H_0, Q has an approximate chi-square distribution with $k - 1 = 4$ degrees of freedom, so the $\propto = 0.05$ critical region is

$q \geq 9.488 = X^2_{0.005}(4)$ Here $q = 5.18 < 9.488$, and hence H_0 is not rejected at the 5% significance level. Furthermore, the *p*-value for $q = 5.18$ is 0.269, which is greater than most significance levels. Thus, with these data, we cannot say that there is a difference between the two methods of instruction.

It is fairly obvious how this procedure can be extended to testing the equality of *h* independent multinomial distributions. That is, let



$$p_{ij} = P(A_i), \qquad i = 1,2, \ . \ . \quad . k \quad j = 1,2, . \ . \ . \ h$$

and test

$$H_0: p_{i1} = p_{i2} = \cdots p_{ih} = p_i, \qquad i = 1,2, . \ . \ k$$

Repeat the *j*th experiment n_j independent times, and let $Y_{1j}, Y_{2j}, . \ . \ Y_{kj}$denote the frequencies of the respective events $A_1, A_2, . \ . \ . A_k$Now,

$$Q = \sum_{j=1}^{h} \sum_{i=1}^{k} \frac{(Y_{ij} - n_j p_{ij})^2}{n_j p_{ij}}$$

has an approximate chi-square distribution with $h(k-1)$ degrees of freedom. Under H_0 , we must estimate $k - 1$ probabilities, using

$$\widehat{p_\iota} = \frac{\sum_{j=1}^{h} Y_{ij}}{\sum_{j=1}^{h} n_j}, \qquad i = 1,2, . \ . \ . \ k-1$$

because the estimate of p_k follows from $\widehat{p_k} = 1 - \widehat{p_1} - \widehat{p_2} - . \ . \ . - \widehat{p_{k-1}}$ We use these estimates to obtain

$$Q = \sum_{j=1}^{h} \sum_{i=1}^{k} \frac{(Y_{ij} - n_j \hat{p}_i)^2}{n_j \hat{p}_i}$$

which has an approximate chi-square distribution, with its degrees of freedom given by $h(k - 1) - (k - 1) = (h - 1)(k - 1)$. Let us see how we can use the preceding procedures to test the equality
of two or more independent distributions that are not necessarily multinomial. Suppose first that we are given random variables U and V with distribution functions $F(u)$and $G(v)$ respectively. It is sometimes of interest to test the hypothesis $H_0: F(x) = G(x)$ for all x. Previously, we considered tests of $\mu_U = \mu_{V,}$ $\sigma_U^2 = \sigma_V^2$ we will look at the two-sample Wilcoxon test. Now we shall assume only

that the distributions are independent and of the continuous type. We are interested in testing the hypothesis $H_0\colon F(x) = G(x)$ for all x . This hypothesis will be replaced by another one. Partition the real line into *k* mutually disjoint sets $A_1, A_2, . . . , A_k$Let

$$p_{i1} = P(U \in A_i), \quad i = 1,2, . . . k.$$

and

$$p_{i2} = P(V \in A_i), \quad i = 1,2, . . . k.$$

We observe that if $F(x) = G(x)$ for all x , then $p_{i1} = p_{i2}, i = 1,2, . . . k$We replace hypothesis $H_0\colon F(x) = G(x)$ with the less restrictive hypothesis $H'_0\colon p_{i1} = p_{i,}, \ i = 1,2, . . ., k$That is,

we are now essentially interested in testing the equality of two multinomial distributions.

Let n_1 and n_2 denote the number of independent observations of $U$and V, respectively. For$I = 1,2, . . . k$, let Y_{ij} denote the number of these observations of U and V, $j = 1, 2$, respectively, that fall into a set Ai. At this point, we proceed to make the test of H'_0 as described earlier. Of course, if H'_0 is rejected at the (approximate) significance level $\propto$ then H_0 is rejected with the same probability. However,

if H'_0 is true, H_0 is not necessarily true. Thus, if H'_0 is not rejected, then we do not reject H_0.

In applications, the question of how to select $A_1, A_2, \ldots, A_k$ is frequently raised. Obviously, there is no single choice for *k* or for the dividing marks of the partition. But it is interesting to observe that the combined sample can be used in this selection

without upsetting the approximate distribution of Q. For example, suppose that $n_1 = n_2 = 20$. Then we could easily select the dividing marks of the partition so that $k = 4$, and one fourth of the combined sample falls into each of the four sets.

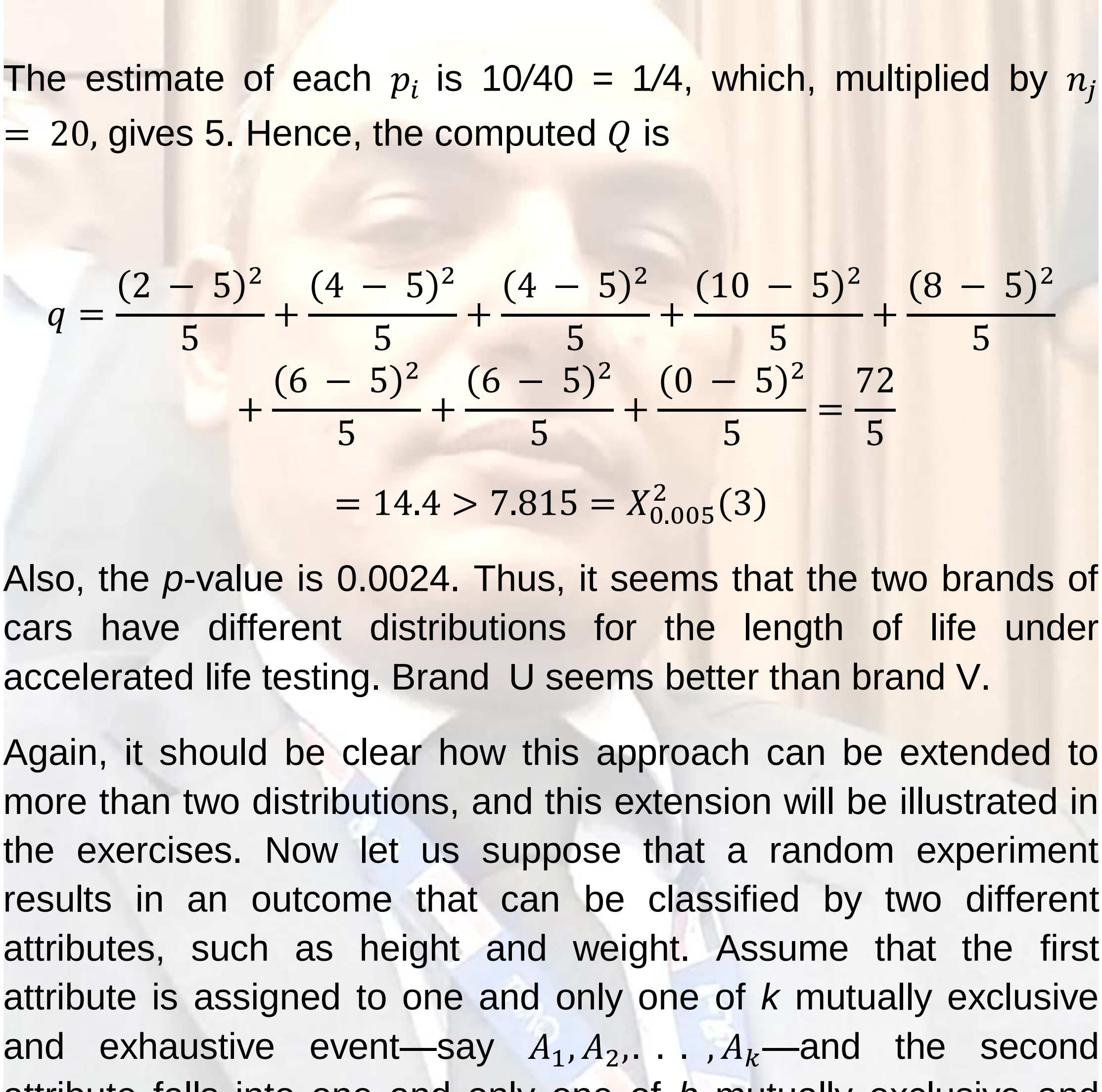
Example:

Select, at random, 20 cars of each of two comparable major-brand models. All 40 cars are submitted to accelerated life testing; that is, they are driven many miles over very poor roads in a short time, and their failure times (in weeks) are recorded as follows:

Brand U:	25	31	20	42	39	19	35	36	44	26
	38	31	29	41	43	36	28	31	25	38
Brand V:	28	17	33	25	31	21	16	19	31	27
	23	19	25	22	29	32	24	20	34	26

If we use 23.5, 28.5, and 34.5 as dividing marks, we note that exactly one fourth of the 40 cars fall into each of the resulting four sets. Thus, the data can be summarized as follows:

	A_1	A_2	A_3	A_4	Totals
Brand U	2	4	4	10	20
Brand V	8	6	6	0	20



The estimate of each p_i is 10/40 = 1/4, which, multiplied by $n_j = 20$, gives 5. Hence, the computed Q is

$$q = \frac{(2-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(10-5)^2}{5} + \frac{(8-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(6-5)^2}{5} + \frac{(0-5)^2}{5} = \frac{72}{5}$$

$$= 14.4 > 7.815 = X^2_{0.005}(3)$$

Also, the *p*-value is 0.0024. Thus, it seems that the two brands of cars have different distributions for the length of life under accelerated life testing. Brand U seems better than brand V.

Again, it should be clear how this approach can be extended to more than two distributions, and this extension will be illustrated in the exercises. Now let us suppose that a random experiment results in an outcome that can be classified by two different attributes, such as height and weight. Assume that the first attribute is assigned to one and only one of *k* mutually exclusive and exhaustive event—say $A_1, A_2, . . . , A_k$—and the second attribute falls into one and only one of *h* mutually exclusive and exhaustive events—say, $B_1, B_2, . . . B_h$. Let the probability of $A_i \cap B_j$be defined by

$$p_{ij} = P(A_i \cap B_j), \quad i = 1, 2, \dots, k, \qquad j = 1, 2, \dots, h.$$

The random experiment is to be repeated n independent times, and Y_{ij} will denote the frequency of the event $A_i \cap B_j$. Since there are kh such events as $A_i \cap B_j$, the random variable

$$Q_{kh-1} = \sum_{j=1}^{h}\sum_{i=1}^{k}\frac{(Y_{ij} - np_{ji})^2}{np_{ji}}$$

has an approximate chi-square distribution with $kh - 1$ degrees of freedom, provided that n is large.

Suppose that we wish to test the hypothesis of the independence of the A and B attributes, namely,

$$H_0{:}\, P\big(A_i \cap B_j\big) = P(A_i)P\big(B_j\big) \quad i \;=\; 1,2,\ldots,k, \qquad j \;=\; 1,2,\ldots,h.$$

Let us denote $P(A_i)$ by $p_{i.}$and $P(B_j)$ by $p_{.j}$; that is,

$$p_{i.} = \sum_{j=1}^{h} p_{ij} = p(A_i) \qquad and \qquad p_{.j} = \sum_{i=1}^{k} p_{ij} = p\big(B_j\big)$$

Of course,

$$1 = \sum_{j=1}^{h}\sum_{i=1}^{k} p_{ij} = \sum_{j=1}^{h} p_{.j} = \sum_{i=1}^{k} p_{i.}$$

Then the hypothesis can be formulated as

$$H_0{:}\, p_{ij} = p_{i.}.p_{.j} \qquad i \;=\; 1,2,\ldots,k, \qquad j \;=\; 1,2,\ldots,h.$$

To test H_0, we can use $Q_{kh-1}$with p_{ij} replaced by $p_{.j}p_{i.}$. But if $p_{i.}$ $i \;=\; 1,2,\ldots,k,$ and $p_{.j}$ $j \;=\; 1,2,\ldots,h$, are unknown, as they usually are in applications, we cannot compute Q_{kh-1} once the frequencies are observed. In such a case, we estimate these unknown parameters by

$$\hat{p}_{.j} = \frac{y_{.j}}{n} \qquad where \quad y_{.j} = \sum_{i=1}^{k} y_{ij}$$

is the observed frequency of B_j, $j \;=\; 1,2,\ldots,h$. Since $\sum_{i=1}^{k} p_i = \sum_{j=1}^{h} p_{.j} = 1$ we actually estimate only $k - 1 + h - 1 \;=\; k + h - 2$

parameters. So if these estimates are used in Q_{kh-1}, with $p_{ij} = p_{i.}p_{.j}$ then, according to the rule stated earlier, the random variable

$$Q = \sum_{j=1}^{h}\sum_{i=1}^{k}\frac{(Y_{ij} - n(\frac{Y_{i.}}{n})(\frac{Y_{.j}}{n}))^2}{n(\frac{Y_{i.}}{n})(\frac{Y_{.j}}{n})}$$

has an approximate chi-square distribution with $kh - 1 - (k + h - 2) = (k - 1)(h - 1)$ degrees of freedom, provided that H_0 is true. The hypothesis H_0 is rejected if the computed value of this statistic exceeds $X^2_{\propto}[(k - 1)(h - 1)]$.