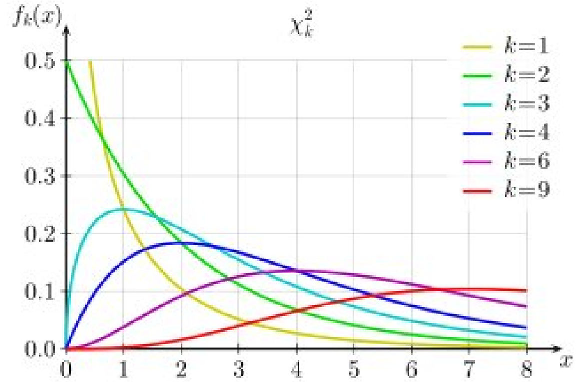
MAT 131 Dayton

χ^2 based tests

The χ^2 (chi-square) distribution is obtained by taking the sum of d independent standard normal distributions squared, that is

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

The parameter k is called the "degrees of freedom", a phrase in statistics which is often used as a parameter for families of distributions. Here k is actually the mean of the χ^2 distribution of d degrees of freedom. This distribution is often useful for modeling discrete statistical phenomena.



χ^2-PDF (Wikipedia)

χ^2 **Goodness of fit (GOF) test** Suppose in a population we expect certain frequencies, $E_i \neq 0$ of a discrete variable with n possibilities when we take a sample of size N. In an actual sample of size N we have observed frequencies O_i and we wish to see if the difference can be attributed to normal sampling error. We calculate

$$\chi^2 = \frac{O_1 - E_1}{E_1} + \frac{O_2 - E_2}{E_2} + \cdots + \frac{O_n - E_n}{E_n}$$

We then test the null hypothesis $H_0 : \chi^2 = 0$ using the χ^2 distribution with the number of *free* variables, often $k = n - 1$. Then $p = \chi^2 CDF(\chi^2, E99, k)$. Here we are more worried about Type II errors so we use large α. If p is large this is a case where we would actually *accept* H_0, even though $\chi^2 = 0$ is impossible in practice, meaning that the difference is completely explained by sampling.

On some TI83/84 calculators the χ^2-GOF is included, otherwise one can calculate χ^2 as above and use the χ^2cdf which is available on all TI83/84 to find p.

Independence of Contingency Tables We can apply the GOF test to a contingency table such as the one on the left which we test against the independent table with same row and column numbers on the left.

Table O	A	B	C	total
D	8	5	3	16
E	12	7	5	24
total	20	12	8	40

Table E	A	B	C	total
D	8	4.8	3.2	16
E	12	7.2	4.8	24
total	20	12	8	40

The 6 cells not including totals in Table O are the observed frequencies while the 6 non-total cells in Table E are the expected frequencies. The χ^2 GOF test is used with $k = 2$ since once two cells are filled in the others can be calculated from the row and column totals. If the null hypothesis $H_0 : \chi^2 = 0$ gives a large p-value then Table O is considered to be *independent*, which is true in this example as $p = 0.98$. When $p < \alpha$ the table would be *dependent*.

All TI83/84 have a χ^2-Test which works by using a `Matrix` to store the non-total cells of Table 0. Table E is calculated by this test and returned to a matrix of your choice. The test calculates k and p.