Large Sample Theory
Ferguson

Exercises, Section 24, General Chi-Square Tests.

1. (a) It is suspected that a certain contingency table with r rows and c columns has homogeneous cells except that edge and corner cells differ in some way from the central cells. Based on data from n trials with n_{ij} observations falling in cell (i, j), find the χ^2 test of the hypothesis H_1 that the four corner cells have probability q_1 each, the $2(r + c - 4)$ edge cells not at a corner have probability q_2 each, and the remaining $(r - 2)(c - 2)$ cells have probability q_3 each, where $4q_1 + 2(r + c - 4)q_2 + (r - 2)(c - 2)q_3 = 1$.

(b) Find the χ^2 test of the hypothesis H_0, that all cells are equally likely, against the hypothesis $H_1 - H_0$.

2. A theory predicts that the probabilities for the four cells of a certain multinomial experiment are $p_1 = 3\theta$, $p_2 = (1/2) - \theta$, $p_3 = (1/3) - \theta$ and $p_4 = (1/6) - \theta$, for some unknown parameter $0 < \theta < 1/6$. A sample of size 100 from this multinomial distribution gave the results $n_1 = 30$, $n_2 = 50$, $n_3 = 10$ and $n_4 = 10$ for the four cells respectively.

(a) Find the minimum modified χ^2 estimate of θ.

(b) Compute the χ^2-statistic for testing the correctness of the theory. How many degrees of freedom does this χ^2 have? Do you accept or reject the theory at the 5% level?

3. A large population of individuals is cross-classified into categories (i, j), for $i = 1, 2, 3$, and $j = 1, 2, 3$. A sample of n individuals is taken from this population (with replacement) and the observed frequencies, n_{ij}, are noted. Let p_{ij} denote the true population probabilities.

(a) Give a χ^2 statistic for testing the hypothesis $H_0 : p_{11} = p_{22} = p_{33}$, and $p_{12} = p_{21} = p_{13} = p_{31} = p_{23} = p_{32}$.

(b) Give the approximate large sample distribution of the statistic under H_0.

(c) Give the approximate large sample distribution of the statistic when $p_{11} = p_{22} = p_{33} = .16$, $p_{12} = p_{21} = p_{23} = p_{32} = .10$ and $p_{13} = p_{31} = .06$.

4. Consider a 2 by c contingency table with probability p_{ij} for cell (i, j) for $i = 1, 2$ and $j = 1, \ldots, c$, where $\sum_{i=1}^{2} \sum_{j=1}^{c} p_{ij} = 1$. Data from n trials show that n_{ij} observations fell in cell (i, j).

(a) Find the χ^2 test of the hypothesis H_1 that $p_{1j} = p_{2j}$ for all j. How many degrees of freedom?

(b) Find the χ^2 test of the hypothesis H_0, that all cells are equally likely. How many degrees of freedom?

(c) Find the χ^2 test of H_0 against $H_1 - H_0$. How many degrees of freedom?

5. N balls are distributed at random into $I \times J$ cells, where cell (i, j) has probability $p_{ij} \geq 0$, for $i = 1, \ldots, I$, and $j = 1, \ldots, J$, with $\sum_i \sum_j p_{ij} = 1$. Let n_{ij} represent the number of balls that fall in cell (i, j), $\sum_i \sum_j n_{ij} = N$.
(a) Find the χ^2 test of the hypothesis $H : \sum_j p_{ij} = 1/I$, for $i = 1, \ldots, I$. How many degrees of freedom?
(b) Find the χ^2 test of the hypothesis $H_0 : p_{ij}$ is independent of i (that is, $p_{1j} = p_{2j} =$

$\ldots = p_{Ij}$ for $j = 1, \ldots, J$). How many degrees of freedom?
(c) Find the χ^2 test of H_0 against $H - H_0$. How many degrees of freedom?

6. Some $I \times J$ contingency tables are constructed for K different populations. Let π_{ijk} denote the probability of falling in cell (i, j) for population k, where i goes from 1 to I, j goes from 1 to J, and k goes from 1 to K, and where $\sum_i \sum_j \pi_{ijk} = 1$ for all k. Let n_{ijk} denote the number of observations falling in cell (i, j) for population k, where $n_{..k} = \sum_i \sum_j n_{ijk}$ is the sample size taken from population k.

(a) What is the chi-square test for testing the hypothesis that the π_{ijk} have specific values, and how many degrees of freedom does it have?

(b) Suppose the π_{ijk} are unknown, and we want to test the hypothesis that for each population, k, the factors are independent, and that the populations are homogeneous for the first factor. This is the hypothesis, $H_0 : \pi_{ijk} = p_i q_{jk}$, for some probabilities p_i and q_{jk} such that $\sum_i p_i = 1$ and $\sum_j q_{jk} = 1$ for $k = 1, \ldots, K$. What is the chi-square test of H_0? Give the estimates of the parameters under H_0, the chi-square statistic, and the number of degrees of freedom.

7. N balls are distributed at random into $I \times J$ cells, where cell (i, j) has probability $p_{ij} \geq 0$, for $i = 1, \ldots, I$, and $j = 1, \ldots, J$, and $\sum_i \sum_j p_{ij} = 1$. Let n_{ij} represent the number of balls that fall in cell (i, j), so that $\sum_i \sum_j n_{ij} = N$.

(a) Find the Pearson's χ^2 test statistic for testing the hypothesis $H_0 : p_{ij}$ is independent of i (that is, $p_{1j} = \cdots = p_{Ij}$ for $j = 1, \ldots, J$). How many degrees of freedom does it have?

(b) Suppose the true values of the p_{ij} satisfy

$$p_{ij} = \frac{(1 + \epsilon_i)}{IJ} \quad \text{for all } i \text{ and } j$$

where the ϵ_i are small numbers (say $|\epsilon_i| < 1$ for all i) satisfying $\sum_1^I \epsilon_i = 0$ (so that $\sum_i \sum_j p_{ij} = 1$). For large N, the distribution of the chi-square of part (a) may be approximated by a non-central chi-square with how many degrees of freedom and with what non-centrality parameter?