HG

Nov. 06

Rice Ex. 9.45 (9.35 in ed. 2)

In the following table we have 13 cells and 6115 trials, and it is reasonable to choose the multinomial model for the cell frequencies, $Y_0, Y_1, \ldots, Y_{12}$, with cell probabilities, $p_0, p_1, \ldots, p_{12}$. Without any structure on the p_j 's, the mle estimates are $\hat{p}_j = Y_j / 6115$.

Let X_i be the number of boys in an arbitrary family with 12 children ($i = 1, 2, \ldots, n$ where n = 6115). It is reasonable to assume that, for family *i*, $X_i \sim \text{Bin}(12, q_i)$ where q_i is the probability of a boy in any birth for family *i*. The hypothesis we want to test is that the probability of a boy does not vary between families (i.e. $q_1 = q_2 = \cdots = q_{6115}$). Call the common value, *q*. This hypothesis implies a structure on the cell-probabilities (p_j), i.e.,

$$p_j = P(X_i = j) = \binom{12}{j} q^j (1-q)^{12-j}, \quad j = 0, 1, 2, \ldots, 12 .$$

The mle estimator for q is found by maximizing the likelihood

$$L(q) = \prod_{i=1}^{6115} P(X_i = x_i) = \left(\prod\nolimits_i \binom{12}{x_i} \right) \cdot q^{\sum_i x_i} (1-q)^{\sum_i (12 - x_i)}$$

which gives

$$\hat{q} = \frac{\sum_i x_i}{(6115) \cdot (12)} = \frac{35280}{73380} = 0,48078$$

where $\sum x_i = 7 \cdot 0 + 45 \cdot 1 + 181 \cdot 2 + \cdots + 3 \cdot 12 = 35280$

> [**Note:** This result indicates that there is probably an error in the data as given by Rice, i.e., that X_i actually denotes the number of girls, not boys, in family *i*. It is generally accepted that the frequency of girl-births is about 0,485 for large groups of births in Europe.]

The expected frequency in cell *j* under H_0 then becomes

$$E_j = 6115 \cdot \hat{P}(X_i = j) = 6115 \cdot \binom{12}{j} \hat{q}^j (1-\hat{q})^{12-j}, \quad j = 0, 1, 2, \ldots, 12$$

Data: Based on a classical genetics study from Germany (1889). The first two columns of the table shows, according to Rice, the number of male children in families with 12 children, for 6115 families (based on hospital records in Saxony).

Results:

x	Observed frequency (O)	$X \sim Bin(12, \hat{q})$ $P(X = x)$	Expected freq. under H_0 (*E*)	*O-E*	$\frac{(O-E)^2}{E}$
0	7	0.000384	2.347267	4.652733	9.222611
1	45	0.004265	26.08236	18.91764	13.72104
2	181	0.021723	132.836	48.16402	17.46343
3	478	0.06705	410.0125	67.98755	11.27358
4	829	0.139697	854.2465	-25.2465	0.746139
5	1112	0.206972	1265.63	-153.631	18.64867
6	1343	0.223594	1367.279	-24.2792	0.431132
7	1033	0.177467	1085.211	-52.2107	2.511914
8	670	0.102707	628.0551	41.94495	2.801313
9	286	0.042269	258.4751	27.52487	2.931108
10	104	0.011742	71.80317	32.19683	14.43719
11	24	0.001977	12.08884	11.91116	11.7361
12	3	0.000153	0.932839	2.067161	4.580802
Sum	6115	1	6115	0	$\chi^2 = 110.5$

The degrees of freedom (df) is

df = the number of estimated parameters under the alternative (13-1 cell probabilities) minus the number of estimated parameters under H_0 is: df = 13-1-1 = 11

Pearson's $\chi^2 = 110{,}5$ is highly significant.

> [**Note.** To calculate binomial (bin(12,q)-probabilities in stata, use the function Binomial(12,x,q) that calculates $P(X \geq x)$. Hence, if the numbers, 0,1,2,….,12 are collected in the column z, the probabilities $P(X = x)$ can be generated by the command: gen r = Binomial(12,z,q) − Binomial(12,z+1,q), where q is the success-probability.]