Biometrika (1985), **72**, 1, *pp.* 91–5
Printed in Great Britain

On heterogeneity tests based on efficient scores

BY ROBERT E. TARONE
National Cancer Institute, Bethesda, Maryland 20205, U.S.A.

SUMMARY

For cases in which a parameter may be estimated from several independent data sets, Mather (1935) proposed a heterogeneity test based on the efficient scores for the individual data sets evaluated at the overall maximum likelihood estimator. Following Fisher (1946), a modification of the heterogeneity score test is presented wherein a consistent but inefficient estimator of the parameter of interest is substituted for the overall maximum likelihood estimator. The modified test is also derived using the $C(\alpha)$ theory of Neyman (1959). Using the modified procedure, we derive a test for a common odds ratio in several 2×2 tables based on the Mantel–Haenszel estimator.

Some key words: $C(\alpha)$ test; Efficient estimator; Efficient score; Heterogeneity test; Maximum likelihood estimator.

1. INTRODUCTION

Data regarding a parameter of interest often may be obtained from several parallel experiments, perhaps with varying designs. For example, information about the recombination fraction for two factors in a given species may be obtained from the offspring of multiple experiments, perhaps representing different types of crosses (Fisher, 1946). Based on a decomposition of the chi-squared goodness-of-fit statistic (Fisher, 1928), Mather (1935) proposed a test for heterogeneity of the recombination fractions from different experiments. This heterogeneity test is based on the efficient scores for the individual parallel experiments evaluated at the maximum likelihood estimator obtained from consideration of all of the experiments: see also Mather (1951, Ch. 7) and Fisher (1958, § 57·3). Rao (1948) demonstrated the asymptotic distribution, in a general setting, of the heterogeneity test based on efficient scores from parallel samples: see also Rao (1952, Ch. 5). Following Fisher (1946), the present paper derives the appropriate modification of the heterogeneity score test when the parameter of interest is estimated by an inefficient, but consistent, estimator. The test is also derived as a $C(\alpha)$ test (Neyman, 1959). The use of the modified test is illustrated by the derivation of a test for heterogeneity of the odds ratios from several 2×2 tables using the Mantel–Haenszel estimator (Mantel & Haenszel, 1959).

2. MODIFICATION OF THE HETEROGENEITY SCORE TEST

Suppose data regarding a parameter θ are obtained from K independent experiments. For experiment k, let x_k represent a random sample of size n_k from an underlying distribution, and let the likelihood function for θ given x_k be denoted by $L_k(x_k;\theta)$ for $k=1,\ldots,K$. Assume that each likelihood function satisfies regularity conditions guaranteeing the asymptotic normality of the corresponding maximum likelihood estimator of

θ, and define

$$S_k(\theta) = \frac{\partial}{\partial\theta}\log L_k(x_k;\theta), \quad V_k(\theta) = -\frac{\partial^2}{\partial\theta^2}\log L_k(x_k;\theta), \quad I_k(\theta) = E\{V_k(\theta)\}.$$

Then the heterogeneity score test suggested by Mather (1935) is based on the statistic

$$H^2 = \sum_{k=1}^{K} \{S_k(\hat{\theta})\}^2/I_k(\hat{\theta}),$$

where $\hat{\theta}$ is the maximum likelihood estimator obtained from consideration of all K experiments, defined by the equation $\Sigma S_k(\hat{\theta}) = 0$. Under the null hypothesis of homogeneity, that is, the hypothesis of a common value of θ for all k, H^2 will have an asymptotic chi-squared distribution with $K-1$ degrees of freedom. Here, as in the remainder of the paper, asymptotic refers to the case in which all n_k tend to ∞ with K fixed. As the development below will demonstrate, if an inefficient estimator is substituted for $\hat{\theta}$ in computing H^2, then H^2 will be stochastically larger under the null hypothesis than a chi-squared random variable with $K-1$ degrees of freedom (Fisher, 1924; Chernoff & Lehmann, 1954).

Let $\tilde{\theta}$ be any consistent estimator of θ, and define

$$X^2(\tilde{\theta}) = \sum_{k=1}^{K} \{S_k(\tilde{\theta})\}^2/I_k(\tilde{\theta}) - \left\{\sum_{k=1}^{K} S_k(\tilde{\theta})\right\}^2 \Big/ \sum_{k=1}^{K} I_k(\tilde{\theta}). \tag{1}$$

Note that by the definition of $\hat{\theta}$, $X^2(\hat{\theta}) = H^2$. Let θ_0 denote the unknown true common parameter value under the null hypothesis of homogeneity. It follows from a Taylor series expansion that

$$S_k(\tilde{\theta}) = S_k(\theta_0) - \delta V_k(\theta_k^*),$$

where $\delta = \tilde{\theta} - \theta_0$ and θ_k^* lies between θ_0 and $\tilde{\theta}$ for all k. Since $\tilde{\theta}$ is a consistent estimator of θ_0, it follows that θ_k^* converges in probability to θ_0 for all k.

Following Fisher (1946), let $y_k = S_k(\tilde{\theta})/V_k(\theta_k^*)$ and $z_k = S_k(\theta_0)/V_k(\theta_k^*)$. Since $y_k = z_k - \delta$ for all k, then for any weights $w_k > 0$ and $\pi_k = w_k/w_.$,

$$\sum_{k=1}^{K} w_k(y_k - \bar{y}_w)^2 = \sum_{k=1}^{K} w_k(z_k - \bar{z}_w)^2, \tag{2}$$

where $\bar{y}_w = \Sigma \pi_k y_k$ and $\bar{z}_w = \Sigma \pi_k z_k$. In particular, consider the weights $w_k^0 = I_k(\theta_0)$. Since $V_k(\theta_k^*)/I_k(\theta_0)$ converges in probability to 1 for all k, $\Sigma w_k^0(z_k - \bar{z}_w)^2$ is asymptotically equivalent to $X_0^2 = \Sigma w_k^0(z_k^0 - \bar{z}_w^0)^2$, where $z_k^0 = S_k(\theta_0)/I_k(\theta_0)$. But

$$X_0^2 = \sum_{k=1}^{K} \{S_k(\theta_0)\}^2/I_k(\theta_0) - \left\{\sum_{k=1}^{K} S_k(\theta_0)\right\}^2 \Big/ \sum_{k=1}^{K} I_k(\theta_0) \tag{3}$$

is asymptotically distributed as a chi-squared random variable with $K-1$ degrees of freedom, and thus by (2) it follows that $\Sigma w_k^0(y_k - \bar{y}_w)^2$ is asymptotically chi-squared with $K-1$ degrees of freedom. Now, since $I_k(\theta_0)/I_k(\tilde{\theta})$ and $V_k(\theta_k^*)/I_k(\tilde{\theta})$ converge in probability to 1 for all k, $\Sigma w_k^0(y_k - \bar{y}_w)^2$ is asymptotically equivalent to $\Sigma w_k'(y_k' - \bar{y}_w')^2$, where $w_k' = I_k(\tilde{\theta})$ and $y_k' = S_k(\tilde{\theta})/I_k(\tilde{\theta})$. Since $\Sigma w_k'(y_k' - \bar{y}_w')^2 = X^2(\tilde{\theta})$ given in (1), it follows that $X^2(\tilde{\theta})$ is asymptotically distributed as a chi-squared random variable with $K-1$ degrees of freedom under the null hypothesis of homogeneity. Thus $X^2(\tilde{\theta})$ provides an appropriate modification of H^2 when θ_0 is estimated by an inefficient estimator. The quantities computed in the modification of H^2 also provide an improved estimator of θ_0, since

$$\theta^1 = \tilde{\theta} + \left\{ \sum_{k=1}^{K} S_k(\tilde{\theta}) \right\} \Big/ \sum_{k=1}^{K} I_k(\tilde{\theta}) \tag{4}$$

will be efficient for any consistent estimator, $\tilde{\theta}$, satisfying the condition $\tilde{\theta} - \theta_0 = o_p(n^{-\gamma})$ with $\gamma \geqslant 1/4$ (Fisher, 1925).

Although the above derivation was carried out for a scalar-valued parameter, the derivation extends easily to the case of an $r \times 1$ vector-valued parameter $\theta' = (\theta_1, \ldots, \theta_r)$. If S_k denotes the $r \times 1$ vector of scores from experiment k evaluated at a consistent estimator $\tilde{\theta}$ and I_k denotes the corresponding $r \times r$ information matrix evaluated at $\tilde{\theta}$, then $\Sigma S'_k I_k^{-1} S_k - (\Sigma S_k)'(\Sigma I_k)^{-1}(\Sigma S_k)$ will be asymptotically distributed as a chi-squared random variable with $r(K-1)$ degrees of freedom under the null hypothesis of homogeneity.

3. Derivation of $X^2(\tilde{\theta})$ as a $C(\alpha)$ test statistic

Suppose that the likelihood for the kth data set is represented as $L_k(x_k; \theta_0 + \phi_k)$ for $k = 1, \ldots, K$, with $\phi_K = 0$. Then testing the null hypothesis of homogeneity reduces to testing H_0: $\phi_k = 0$ for all k, with θ_0 treated as a nuisance parameter. In this setting, the theory of $C(\alpha)$ tests may be used to derive a heterogeneity test using any root-n consistent estimator of θ_0, that is, any $\tilde{\theta}$ satisfying $(\tilde{\theta} - \theta_0) = O_p(n^{-\frac{1}{2}})$ (Neyman, 1959; Moran, 1970). Let $\tilde{\theta}$ be a root-n consistent, but possibly inefficient estimator of θ_0, and let $L = \Pi L_k$. Derivation of the $C(\alpha)$ test statistic requires the following scores evaluated at $\tilde{\theta}$ under H_0: $\partial \log L / \partial \phi_k = S_k(\tilde{\theta})$ for $k = 1, \ldots, K-1$, and $\partial \log L / \partial \theta = S_.(\tilde{\theta})$, where the dot notation denotes the sum from 1 to K. The corresponding information matrix evaluated under H_0 is

$$\begin{bmatrix} D & A \\ A' & B \end{bmatrix},$$

where D is the $(K-1) \times (K-1)$ diagonal matrix with kth diagonal entry $I_k(\tilde{\theta})$, A is the $(K-1) \times 1$ vector with kth entry $I_k(\tilde{\theta})$ and $B = I_.(\tilde{\theta})$. The $C(\alpha)$ test is based on $T_k(\tilde{\theta}) = S_k(\tilde{\theta}) - \beta_k S_.(\tilde{\theta})$, where β_k is the regression coefficient of S_k on $S_.$ for $k = 1, \ldots, K-1$. Thus, $\beta_k = I_k(\tilde{\theta})/I_.(\tilde{\theta})$, and letting T denote the $(K-1) \times 1$ vector with kth entry $S_k(\tilde{\theta}) - I_k(\tilde{\theta}) S_.(\tilde{\theta})/I_.(\tilde{\theta})$, the $C(\alpha)$ test statistic is

$$C^2 = T'(D - AA'/B)^{-1} T = T'(D^{-1} - JJ'/I_K) T,$$

where J is the $(K-1) \times 1$ vector of ones. Hence

$$\begin{aligned} C^2 &= \sum_{k=1}^{K-1} T_k^2(\tilde{\theta})/I_k(\tilde{\theta}) - \left\{ \sum_{k=1}^{K-1} T_k(\tilde{\theta}) \right\}^2 \Big/ I_K(\tilde{\theta}) \\ &= \sum_{k=1}^{K} \{ S_k(\tilde{\theta}) - I_k(\tilde{\theta}) S_.(\tilde{\theta})/I_.(\tilde{\theta}) \}^2 / I_k(\tilde{\theta}) = X^2(\tilde{\theta}). \end{aligned}$$

Although in general the validity of the $C(\alpha)$ test would require that $\tilde{\theta}$ be root-n consistent, the derivation in §2 demonstrates that consistency of $\tilde{\theta}$ is sufficient for the validity of the heterogeneity test based on $X^2(\tilde{\theta})$.

4. EXAMPLE

Consider K pairs of independent binomial variates x_k and y_k with corresponding parameters p_{1k} and p_{2k} and sample sizes n_{1k} and n_{2k}, where $k = 1, \ldots, K$. The heterogeneity score test may be applied to test the assumption of a constant odds ratio, that is, $(p_{1k} q_{2k})/(p_{2k} q_{1k}) = \psi$ for all k, where $q_{jk} = 1 - p_{jk}$ for $j = 1, 2$ and $k = 1, \ldots, K$. Using the conditional distribution of x_k given $m_k = x_k + y_k$ in order to eliminate nuisance parameters, we find the likelihood function for ψ, given $x = (x_1, \ldots, x_K)'$, is $L(x;\psi) = \Pi f(x_k; m_k; \psi)$, where $f(x_k; m_k; \psi)$ is the density function of the noncentral hypergeometric distribution (Gart, 1970). Letting $e_k(\psi) = E(x_k \mid m_k; \psi)$ and $v_k(\psi) = \operatorname{var}(x_k \mid m_k; \psi)$, we have that $S_k(\psi) = \psi^{-1}\{x_k - e_k(\psi)\}$ and $I_k(\psi) = \psi^{-2} v_k(\psi)$. Thus, for a consistent estimator $\tilde{\psi}$, the modified heterogeneity score test statistic is

$$X^2(\tilde{\psi}) = \sum_{k=1}^{K} \{x_k - e_k(\tilde{\psi})\}^2 / v_k(\tilde{\psi}) - \{x_. - e_.(\tilde{\psi})\}^2 / v_.(\tilde{\psi}).$$

Although the exact moments of the x_k can be expressed as finite sums, their evaluation can be time consuming for $\psi \neq 1$ and large n_{jk}. For $\psi = 1$ the moments have closed form expressions, and it is of interest that $X^2(1)$ is the heterogeneity test proposed by Zelen (1971). As noted by Halperin et al. (1977), Zelen's test is valid only for a common odds ratio value of 1, the test statistic being of the type X_0^2 given in (3). Because computation of exact moments of the x_k can be time consuming, it will, in general, be convenient to replace the exact moments in the computation of $X^2(\tilde{\psi})$ by the asymptotic moments (Gart, 1970). Thus, let $\hat{e}_k(\psi)$ be the appropriate solution to the quadratic equation

$$\frac{\hat{e}_k(n_{2k} - m_k + \hat{e}_k)}{(m_k - \hat{e}_k)(n_{1k} - \hat{e}_k)} = \psi,$$

and let

$$\hat{v}_k(\psi) = \{\hat{e}_k^{-1} + (m_k - \hat{e}_k)^{-1} + (n_{1k} - \hat{e}_k)^{-1} + (n_{2k} - m_k + \hat{e}_k)^{-1}\}^{-1}$$

for $k = 1, \ldots, K$. For the consistent estimator, $\tilde{\psi}$, a test statistic which is asymptotically equivalent to $X^2(\tilde{\psi})$ is

$$Y^2(\tilde{\psi}) = \sum_{k=1}^{K} \{x_k - \hat{e}_k(\tilde{\psi})\}^2 / \hat{v}_k(\tilde{\psi}) - \{x_. - \hat{e}_.(\tilde{\psi})\}^2 / \hat{v}_.(\tilde{\psi}).$$

Both $X^2(\tilde{\psi})$ and $Y^2(\tilde{\psi})$ are asymptotically distributed as chi-squared random variables with $K-1$ degrees of freedom under the null hypothesis of a constant odds ratio. A heuristic derivation of $Y^2(\tilde{\psi})$ can be obtained by substituting an inefficient, but consistent, estimator of ψ for the unconditional maximum likelihood estimator in the extension by Halperin et al. (1977) of the original development by Zelen (1971).

To test for a common odds ratio, Breslow & Day (1980, p. 142) propose the statistic

$$Z^2(\hat{\psi}) = \sum_{k=1}^{K} \{x_k - \hat{e}_k(\hat{\psi})\}^2 / \hat{v}_k(\hat{\psi}),$$

where $\hat{\psi}$ is the unconditional maximum likelihood estimator of ψ. Thus, $Z^2(\hat{\psi}) = Y^2(\hat{\psi})$, the same statistic suggested by Bartlett (1935) for $K = 2$ and by Norton (1945) for $K > 2$. Breslow & Day also suggest that a valid test can be based on $Z^2(\hat{\psi}_{MH})$, where $\hat{\psi}_{MH}$ is the Mantel–Haenszel estimator (Mantel & Haenszel, 1959). However, since $\hat{\psi}_{MH}$ is not efficient (Tarone, Gart & Hauck, 1983), the heterogeneity test using $\hat{\psi}_{MH}$ should be based on $Y^2(\hat{\psi}_{MH})$.

Table 1. *Example of a* $2 \times 2 \times 2$ *table from Halperin et al.* (1977)

190	810	750	250
10	990	250	750

An example with $K = 2$ from the paper of Halperin et al. (1977) is presented in Table 1. For this example

$$\hat{\psi}_{\mathrm{MH}} = 10{\cdot}63, \quad x_{.} - \hat{e}_{.}(\hat{\psi}_{\mathrm{MH}}) = -5{\cdot}0735, \quad \hat{v}_{.}(\hat{\psi}_{\mathrm{MH}}) = 107{\cdot}2674.$$

Thus $Z^2(\hat{\psi}_{\mathrm{MH}}) = 8{\cdot}57$, while $Y^2(\hat{\psi}_{\mathrm{MH}}) = 8{\cdot}33$. Since $Y^2(\hat{\psi}_{\mathrm{MH}})$ is asymptotically chi-squared with one degree of freedom under the null hypothesis of a common odds ratio, $Y^2(\hat{\psi}_{\mathrm{MH}})$ provides strong evidence of heterogeneity. By way of comparison, $Z^2(\hat{\psi}) = 8{\cdot}08$. The efficient estimate obtained from (4) is

$$\psi^1 = \hat{\psi}_{\mathrm{MH}} + \hat{\psi}_{\mathrm{MH}}\{x_{.} - \hat{e}_{.}(\hat{\psi}_{\mathrm{MH}})\}/\hat{v}_{.}(\hat{\psi}_{\mathrm{MH}}) = 10{\cdot}63 - 0{\cdot}50 = 10{\cdot}13,$$

compared to the unconditional maximum likelihood estimate of 10·14 and the conditional maximum likelihood estimate of 10·13.

References

Bartlett, M. S. (1935). Contingency table interactions. *J. R. Statist. Soc. Sup.* **2**, 248–52.

Breslow, N. E. & Day, N. E. (1980). *Statistical Methods in Cancer Research*, **1**, *The Analysis of Case-Control Studies*. Lyon: International Agency for Research on Cancer.

Chernoff, H. & Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Ann. Math. Statist.* **25**, 579–86.

Fisher, R. A. (1924). The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *J. R. Statist. Soc.* **87**, 442–9.

Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.* **22**, 700–25.

Fisher, R. A. (1928). On a property connecting the χ^2 measure of discrepancy with the method of maximum likelihood. *Atti del Congresso Internazionale dei Matematici* **6**, 94–100.

Fisher, R. A. (1946). A system of scoring linkage data, with special reference to the pied factors in mice. *Am. Nat.* **80**, 568–78.

Fisher, R. A. (1958). *Statistical Methods for Research Workers*, 13th ed. Edinburgh: Oliver & Boyd.

Gart, J. J. (1970). Point and interval estimation of the common odds ratio in the combination of 2×2 tables with fixed marginals. *Biometrika* **57**, 471–5.

Halperin, M., Ware, J. H., Byar, D. P., Mantel, N., Brown, C. C., Koziol, J., Gail, M. & Green, S. B. (1977). Testing for interaction in an $I \times J \times K$ contingency table. *Biometrika* **64**, 271–5.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Nat. Cancer Inst.* **22**, 719–48.

Mather, K. (1935). The combination of data. *Ann. Eugen.* **6**, 399–410.

Mather, K. (1951). *The Measurement of Linkage in Heredity*. London: Methuen.

Moran, P. A. P. (1970). On asymptotically optimal tests of composite hypotheses. *Biometrika* **57**, 47–55.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics*, Ed. U. Grenander, pp. 213–34. New York: Wiley.

Norton, H. W. (1945). Calculation of chi-square from complex contingency tables. *J. Am. Statist. Assoc.* **40**, 251–8.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proc. Camb. Phil. Soc.* **44**, 50–7.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: Wiley.

Tarone, R. E., Gart, J. J. & Hauck, W. W. (1983). On the asymptotic inefficiency of certain noniterative estimators of a common relative risk or odds ratio. *Biometrika* **70**, 519–22.

Zelen, M. (1971). The analysis of several 2×2 contingency tables. *Biometrika* **58**, 129–37.

[*Received June* 1984. *Revised September* 1984]