ANNALS OF FAMILY MEDICINE

Online Supplementary Material

Reed JF III. Adjusted chi-square statistics: application to clustered binary data in primary care. *Ann Fam Med.* 2004;2:201-203.

http://www.annfammed.org/cgi/content/full/2/3/201/DC1

Appendix 1. Computing Adjustment to Intraclass Correlation in Analytic Methods Specific to Randomized Cluster Trials

- Suppose there are m_i, i = 1, 2, clusters randomized to either treatment or control group. Let x_{ij} denote the number of successes among the n_{ij} observations in the jth cluster (j = 1, 2) with $p'_i = x_{ij}/n_{ij}$ the cluster-specific success rate. We let $p'_i = x_i/n_i$, where $x_i = \Sigma_j x_{ij}$ and $n_i = \Sigma_j n_{ij}$ and $p' = \Sigma_i x_i / \Sigma_i n_i$. Then an ANOVA estimator of the intracluster correlation coefficient (ICC) p'(5-6) is: p' = (MSC - MSE)/(MSC + (n_o - 1)MSE). where: MSC = $\Sigma_i \Sigma_j n_{ij} (p'_{ij} - p'_i)^2 / (m - 2)$, MSE = $\Sigma_i \Sigma_j n_{ij} p'_{ij} (1 – p'_{ij})^2/(n - m)$, $n_o = (n - \Sigma_i n'_{ai}) / (m - 2)$, $n'_{ai} = \Sigma n^2_{ij}/n_i$, $m = \Sigma_i m_i$, and $n = \Sigma_i n_i$.
- **Brier's statistic (χ^2_b).** This adjustment is based on the clustering effect within the treatment (C_1) and control (C_2) groups.[1] An underlying assumption is that C_1 and C_2 are homogeneous across treatment groups. $\chi^2_b = \chi^2/C$, where $C' = \Sigma_i (n_iC_i)/N$ and $C_i = 1 + (n_{ai} - 1) p'$. χ^2 is the usual Pearson chi-square.
- **Rosner and Milton's chi-square (χ^2_{rm}).** This statistic estimates the presence of dependency both between and within the comparison groups.[2] $\chi^2_{rm} = \chi^2 /(\Sigma_i n_i (1 + (n_{ai} - 1) p' p'_e/n$. χ^2 is Pearson's chi-square, and p'_e is the estimated ICC for treatment status. Rosner and Milton, p' = [Pr (++) - Pr^2(+)]/[Pr(+)Pr(-)]. When the competitive treatments are independent, $p'_e = 1$.
- **Donner and Donald's chi-square (χ^2_{dd}).** χ^2_{dd} is dependent on clustering "correction" factors estimated separately in each of the treatment groups.[3] $\chi^2_{dd} = \Sigma_i ((x_i - n_ip')2/ (n_i C_i p'(1 - p'))$, with $C_i = 1 + (n_{ai} - 1) p'$. χ^2_{dd} assumes that C_1 and C_2 are homogeneous.
- **Rao and Scott's chi-square (χ^2_{rs}).** $\chi^2_{rs} = \Sigma_i (x'_i - n'_i p')^2/[n'_i p'(1 - p')]$ with $p' = \Sigma x'_i/ \Sigma n'_i$. Where, p_i is the overall sample proportion ($p'_i = x_i/n_i$, $x_i = \Sigma_j x_{ij}$, and $n_i = \Sigma_j n_{ij}$) and the variance of p'_i for large m_i is: $v_i = m_i(m_i-1)^{-1} n^{-2} \Sigma_j r_{ij}^2$, where $r_{ij} = x_{ij}-n_{ij} p'$. The variance inflation due to clustering d_i is the ratio of v_i to the estimated binomial $p'_i (1 - p'_i)/n_i$. This inflation factor is called the design effect and $n'_i = n_i/d_i$ is termed the effective sample size. The data (x_i, n_i) is transformed to (x'_i, n'_i), I = 1, 2, . . ., k, where $x'_i = x_i/d_i$. χ^2_{rs} statistic makes no assumption regarding the nature of the clustering or homogeneity of design effects.[4]
- **Rao and Scott's chi-square (χ^2_{prs}).** $\chi^2_{prs} = \chi^2_{rs} / d$, where $d = \Sigma_i [(1 - n_i/n) p'_i (1 - p'_i)d_i]/[p'(1 - p')]$. χ^2_{prs} assumes a pooled estimate of a common design effect.[4]

References

1. Brier SS. Analysis of contingency tables under cluster sampling. *Biometrika*. 1980;67:591-596.
2. Rosner B, Milton RC. Significance testing for correlated binary outcome data. *Biometrics*. 1988;44:505-512.
3. Donner A, Donald A. The statistical analysis of multiple binary measurements. *J Clin Epidem.* 1988;41:899-906.
4. Rao JNK, Scott AJ. A simple method for the analysis of clustered binary data. *Biometrics*. 1992;48:577-585.