NOTES ON CORRELATED BINARY DATA

The first two pages of these notes describe estimation of the correlation parameter κ.

The third page describes a modified estimator of the probability *p*, based on a weighted combination of the results in the different units, with weights given by the inverse variance. In theory, this weighting can lead to a reduction in the variance of the estimate. In practice, it often makes little difference, but it can make a difference in certain cases where the number of subunits m_i varies widely across the units.

The notation in these notes is somewhat different from the notation I used in class, but you should be able to follow what is going on.

Estimation of κ (Fleiss, Sec. 13.2)

ANOVA sums of squares type formulation

$$\hat{p} = \bar{X}_{..} = \sum_{jk} X_{jk} \Big/ \sum_{j} m_j$$

$$\hat{p}_j = \bar{X}_{j.} = \sum_{k} X_{jk} \,/\, m_j$$

Between-subject mean square

$$BMS = \frac{1}{N-1} \sum_j m_j (\bar{X}_{j.} - \bar{X}_{..})^2$$

$$= \frac{1}{N-1} \sum_j m_j (\hat{p}_j - \hat{p})^2 .$$

Within-subject mean square

$$WMS = \frac{1}{\sum_j m_j - N} \sum_{jk} (X_{jk} - \bar{X}_{j.})^2$$

$$= \frac{1}{N(\bar{m}-1)} \sum_j m_j \hat{p}_j \hat{q}_j .$$

For the case of X continuous, we have (proof = exercise):

$$E[BMS] = \frac{1}{N-1} \sum_{j=1}^{N} \left(1 - \frac{m_j}{M}\right) [1 + \kappa(m_j - 1)]\, \sigma_X^2$$

$$= \sigma_X^2 \, [1 + (m^* - 1)\kappa] = \sigma_X^2 \, [m^* \kappa + (1-\kappa)]$$

$$\text{with } m^* = \bar{m} - \frac{1}{\bar{m}N(N-1)} \sum_{j=1}^{N} (m_j - \bar{m})^2$$

$$E[WMS] = \sigma_X^2 \, (1-\kappa) . \qquad [\leftrightarrow pq(1-\kappa)]$$

This leads to

$$\hat{\kappa} = (BMS - WMS) / [BMS + (m^* - 1) WMS] .$$

For the binary case, above formulas for $E[BMS]$ + $E[WMS]$ hold with $\sigma_X^2 = pq$, and $\hat{\kappa}$ is as above.

For N large, $m^* \doteq \bar{m}$ and $BMS = BMS'$, where

$$BMS' = \frac{1}{N} \sum_{j=1}^{N} m_j (\hat{p}_j - \hat{p})^2 = \frac{N-1}{N} BMS .$$

This leads to

$$\hat{\kappa} = (BMS' - WMS) / [BMS' + (\bar{m} - 1) WMS]$$

$$= 1 - WMS / \hat{p}\hat{q} \qquad \left[\begin{array}{l}\text{also suggested by} \\ E[WMS] = pq(1-\kappa)\end{array}\right]$$

$$[\text{because } BMS' + (\bar{m} - 1) WMS = \bar{m}\hat{p}\hat{q}] .$$

b) <u>Weighted combination of individual proportions</u>

$$\hat{p}_j = Y_j / m_j$$

$$\boxed{\text{Cf.} \quad \hat{p} = \frac{\Sigma m_j \hat{p}_j}{\Sigma m_j}}$$

$$E[\hat{p}_j] = p$$

$$Var(\hat{p}_j) = \frac{1}{m_j} pq \, [1 + (m_j - 1)\kappa] .$$

$$\hat{p}_w = \sum_{j=1}^{N} m_j [1 + (m_j - 1)\kappa]^{-1} \hat{p}_j \Big/ \sum_{j=1}^{N} m_j [1 + (m_j - 1)\kappa]^{-1}$$

$$= \sum_{j=1}^{N} [1 + (m_j - 1)\kappa]^{-1} Y_j \Big/ \sum_{j=1}^{N} [1 + (m_j - 1)\kappa]^{-1} m_j .$$

(need preliminary estimate of κ) .

$$Var(\hat{p}_w) = \left[\sum_{j=1}^{N} m_j [1 + (m_j - 1)\kappa]^{-1} \right]^{-1} pq .$$

$$m_j' = m_j [1 + (m_j - 1)\kappa]^{-1}$$

= "effective" number of subunits
(number of independent bits of info)

$$\frac{\Sigma w_i Y_i}{\Sigma w_i m_i} = \frac{1}{\Sigma w_i m_i} \Sigma w_i (Y_i - m_i p)$$

$$N \left(\frac{1}{\Sigma w_i m_i} \right)^2 \left(\frac{1}{N-1} \Sigma \, w_i^2 \, (Y_i - m_i p)^2 \right)$$