1. INTRODUCTION.

Cohen (1960) and Fleiss (1971) developed Kappa statistics of intraclass correlation and measured the agreement between a number of raters when subjects are rated on a nominal scale.

Landis and Koch (1977) introduced the intraclass correlation by random effect model, that is estimated by the components of the analysis of variance. They also measured the agreement or reliability of raters.

These two estimators are asymptotically same when the number of raters for individual subjects are fixed.

However, as mentioned in Fleiss et al (1979), it appears that the variances of the correlation estimators may give inflated values in both papers. In this note, the direct estimation of intraclass correlation and its closed form of variance are presented, and shows some reduced variances as seen in Table 1.

We will briefly describe the intraclass correlation model in Section 2. Above two estimators of correlation are briefly described, following with a direct estimation and its large sample distribution in Section 3. The variance of direct estimators is given in Appendices A, B, and C. A numerical example is included in Section 4.

Brier (1980) and Kleinman(1973) used method of moments; Cohen (1967) obtained the correlation by the maximum likelihood estimation, while Spearman (1904) and Kendall (1955) used nonparametric approaches, using the ranks of variables.

2. CORRELATION MODEL

Suppose we have a population of A clusters (for example, houses or subjects),

\[ U = (U_1, ..., U_A) \]

where the cluster \( i \) included \( B_i \) elementary units (persons in the house or a number of rating on the same subject), expressed as \( U_i = (Y_{i1}, ..., Y_{ij}, ..., Y_{iB_i}) \). The clusters are indexed by \( i = 1, ..., A \) and the units by \( j = 1, ..., B_i \).

Suppose that there are \( r \) categories (or ratings) and each elementary unit can be classified into only one of the categories, indexed by \( h = 1, ..., r \).

Let \( Y_{ijh} = \{ (i,j) \} \) be the response vector of the \((ij)-th\) person with \( \sum_i \sum_j Y_{ijh} = 1 \) and \( \{ 1 \) if the \((ij)-th\) element falls in cell \( h \)

\[ Y_{ijh} = \begin{cases} 1 & \text{if } h = h' \\ 0, \text{otherwise.} \end{cases} \]

Let \( \sum_i \sum_j Y_{ijh} = N_h \) be the population counts in the \( h \)-th cell and \( \sum_h N_h = N \) the entire units in the population. Denote the population proportion of the categories by \( \pi' = (\pi_1, \pi_2, ..., \pi_r) \), where \( \pi_h = N_h / N \) (\( \pi_h > 0 \) and \( \sum \pi_h = 1 \)).

Define the pairwise probability for any two members in the same cluster as

\[
\Pr(Y_{ijh} = 1, Y_{i'j'h'} = 1) = \begin{cases} 
\delta_{hh'} & \text{if } i = i', j = j' \text{ and } h = h' \\
\delta_{hh'} & \text{if } i = i', j = j' \text{ and } h = h' \\
\pi_h & \text{if } i = i', j = j' \text{ and } h = h' \\
0 & \text{if } i = i', j = j' \text{ and } h = h' \\
\pi_h \pi_{h'} & \text{if } i \neq i' \text{ and } h = h' 
\end{cases}
\]

where \( \delta_{hh'} \) is the probability that one member of the pair from the same cluster falls into the \( h \)-th cell and the other into \( h' \)-th cell, \( (h \neq h') \) and \( \theta_{hh} \) arises when both elements fall into the \( h \)-th cell \( (h = h') \).

Note \( \sum_{h,h'} \delta_{hh'} = 1 \) and, if both members of all pairs fall in the same category, the off-diagonal elements are zeros and the sum of the diagonal elements is one.

Denote the intracluster correlation of any two elements in the same cluster by

\[
\rho_{hh'} = \frac{\text{cov}(Y_{ijh}, Y_{ij'h'})}{\sqrt{\text{var}(Y_{ijh})} \sqrt{\text{var}(Y_{ij'h'})}}
\]

where \( \text{cov}(Y_{ijh}, Y_{ij'h'}) = \delta_{hh'} - \pi_h \pi_{h'} \), and \( \text{var}(Y_{ijh}) = \pi_h (1 - \pi_h) \) from the definition (2).

The \( \rho_{hh} \) is the intraclass correlation of the \( h \)-th cell when \( h = h' \) and the overall intraclass correlation \( \rho \) is defined as

\[
\rho = \frac{\sum_h (\delta_{hh} - \pi_h^2)}{\sum_h \pi_h (1 - \pi_h)}
\]

3. ESTIMATION

3.1. Sample

Denote a sample design \( \{U, S, P\} \), where a one-stage cluster sample \( S \) is selected from the population \( U \) by the probability \( P \) for the cluster sampling. Denote the sample \( S \) as

\[ S = \{ (i,j): i \in S^*, j \in S_i \} \]

where \( S^* \) is a sample of "a" clusters, and \( S_i \) is a sample of \( B_i \) units from the \( i \)-th cluster. The clusters are indexed by \( i = 1, ..., a \); the units in the \( i \)-th cluster by \( j = 1, ..., B_i \).

We assume the clusters are independent and the units in the clusters are correlated by a common intraclass correlation \( \rho_{hh} \) for the \( h \)-th category and \( \rho \) for overall categories.

The parameters \( \rho_{hh} \) and \( \rho \) defined in the
models (3) and (4), respectively are estimated from the sample S. The probability P used in the sampling is not involved in our model based estimation.

3.2. Multivariate Analysis of Variance (MANOVA) Estimator

Landis and Koch (1977) used a one-way components of variance model for categorical data to estimate the intracluster correlation in one-stage cluster sample involving an unbalanced design. They estimated the parameters \( \hat{p} \) and \( \hat{\delta}_{hh} \) as

\[
\hat{p} = \frac{\sum_{h=1}^{r} (MS_{ch} - MS_{eh})}{\sum_{h=1}^{r} (MS_{ch} + (d-1)MS_{eh})}
\]

where \( d = (n^2 - \sum b_i^2)/n(a-1) \), and \( n = \sum b_i \). \( MS_{ch} \) and \( MS_{eh} \) are the mean squares for the clusters and residual errors with \( a-1 \) and \( n-a \) degrees of freedom, respectively, for the \( h \)-th category in a usual MANOVA table. The intracluster correlation \( \hat{\rho}_{hh} \) for individual cells is estimated from the (5) by dropping the summation signs as

\[
\hat{\rho}_{hh} = \frac{MS_{ch} - MS_{eh}}{MS_{ch} + (d-1)MS_{eh}}
\]

They numerator and denominator of (5) and (6) are unbiased as estimates for those of (3) and (4), respectively.

3.3. Kappa Estimator

Cohen (1960; 1968) used kappa estimator for intracluster correlations. Fleiss (1971) illustrated the use of kappa and weighted kappa with a psychological diagnostic data on 30 patients. Fleiss et al (1979) show the measurement of the extent of agreement beyond chance when for \( b_i = b \) (or the number of units in the clusters are the same):

\[
\kappa_h = 1 - \frac{\sum_{i,j}^{n} y_{i+1,h} (b - y_{i+1,h})}{ab(b-1)\sum_{i,j}^{n} \pi_h (1 - \pi_h)}
\]

where \( \sum_{i,j}^{n} y_{i+1,h} = y_{i+1,h} \) and \( \sum_{i,j}^{n} \pi_h = \pi_h \), and the overall measure of agreement is a weighted average of \( \kappa \) as

\[
\kappa = \frac{\sum_{h=1}^{r} \pi_h (1 - \pi_h) \kappa_h}{\sum_{h=1}^{r} \pi_h (1 - \pi_h)}
\]

The estimators of (7) and (8) are obtained by replacing \( \pi_h \) by \( \hat{\pi}_h \), \( h = 1, \ldots, r \). It is easy to show \( \hat{\kappa}_h = \hat{\rho}_{hh} \) if \( a/a(a-1) \rightarrow 1 \) for a large \( a \). The variances of (7) and (8) are presented in Fleiss et al (1979).

3.4. Direct Estimator

The estimator of \( \hat{p} \) for overall categories can be obtained by direct substitution of \( \hat{\pi}_h \) and \( \hat{\delta}_{hh} \) in the definition (3) and (4) as

\[
\hat{\rho} = \frac{\sum_{h=1}^{r} (\hat{\delta}_{hh} - \hat{\pi}_h^2)}{1 - \sum_{h=1}^{r} \hat{\pi}_h^2}
\]

where \( \hat{\pi}_h = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b_i} y_{i,j+h} \), \( H = \sum_{i=1}^{a} b_i(b_i-1) \);

and \( \hat{\delta}_{hh} = \frac{1}{n} [\sum_{i=1}^{a} \sum_{j=1}^{b_i} y_{i,j+h}^2] - \sum_{i=1}^{a} \sum_{j=1}^{b_i} y_{i,j+h}^2 \].

* denotes the sumation over the corresponding subscript. The \( \text{var}(\hat{\rho}) \) is given in the Appendix B, (B14).

The estimator of \( \hat{\rho}_{hh} \) for the intracluster correlation of the \( hh \)-th category is obtained by (9) without summation signs as

\[
\hat{\rho}_{hh} = \frac{(\hat{\delta}_{hh} - \hat{\pi}_h^2)}{(\hat{\pi}_h^2 - (1 - \hat{\pi}_h)}
\]

The \( \text{var}(\hat{\rho}_{hh}) \) is given in the Appendix A, (A13).

The numerator and denominator of ANOVA estimators \( \hat{\rho} \) and \( \hat{\rho}_{hh} \) can be expressed as the estimators \( \tilde{\rho} \) and \( \tilde{\rho}_{hh} \) if \( \hat{\delta}_{hh} = \hat{\delta}_{i,h} \) and \( \hat{\pi}_h = \tilde{\pi}_i \) for all \( i \)-s. But this is not true in general.

If the numerator and denominator of \( \hat{\rho} \) and \( \hat{\rho}_{hh} \) are consistent, then the estimators \( \tilde{\rho} \) and \( \tilde{\rho}_{hh} \) are also consistent since the estimates of a parameter function is consistent as the same function of consistent estimators of the parameters.

However, \( \hat{\rho} \) and \( \hat{\rho}_{hh} \) include biased estimators of numerator and denominator, and we may use the unbiased numerator and denominator, and improve \( \tilde{\rho}_{hh} \) and \( \tilde{\rho} \) as

\[
\tilde{\rho}_{hh} = \frac{\hat{\rho}_{hh} (1 - \frac{1}{n}) + \frac{1}{n}}{\hat{\rho}_{hh} H + (1 - \frac{H}{n})} \frac{1}{n}
\]

with \( \text{var}(\hat{\rho}_{hh}) = \text{var}(\hat{\rho}_{hh}) (1 - \frac{1}{n} - \frac{H}{n^2}) \), (13)

where \( \text{var}(\hat{\rho}_{hh}) \) is given in (A13) and \( H \) in (12).

Note both \( \hat{\rho}_{hh} \rightarrow \tilde{\rho}_{hh} \) and \( \text{var}(\hat{\rho}_{hh}) \rightarrow \text{var}(\tilde{\rho}_{hh}) \) as \( n \rightarrow \infty \). Following the same steps of the cell intracluster correlations, an unbiased estimator over all categories is obtained as
\[ \hat{\rho} = \frac{1}{n \sum_{h} (\hat{\delta}_{hh} - \hat{\nu}_{h})^2 + \frac{1}{n} \left( 1 - \sum_{h} \hat{\nu}_{h} \right)} \cdot (14) \]

\[ \text{var}(\hat{\rho}) \text{ is the same as var}(\hat{\rho}) \text{ except for the partial derivatives shown in (B2) and (B3), and (C2) and (C3), respectively. With the new partial derivatives (C2) and (C3), we obtain the asymptotic variance var}(\hat{\rho}) \text{ as shown in (C4). Note} \hat{\rho} \rightarrow \rho \text{ and var}(\hat{\rho}) \rightarrow \text{var}(\rho) \text{ for a large n.} \]

Large sample distribution

The theorems on large sample distribution (for example, Cheung, 1969; Bishop, Fienberg, and Holland, 1979) are applied to the above estimators under the usual regularity conditions. A null hypothesis that specifies the cell correlation \( \rho_{hh} \) may be tested by the statistic,

\[ Z_h = \frac{\hat{\rho}_{hh} (1 - 1/n) + 1/n}{\left[ 1 - \frac{H}{n^2} \right] \left[ 1 - \frac{1}{n} \right] \sqrt{\text{var}(\hat{\rho}_{hh})}} \]

where \( Z_h \approx N(0, 1) \) for a large "a". \( \text{var}(\hat{\rho}_{hh}) \) is given in (A13) and \( H \) is a constant multiplier (A12).

A null hypothesis that specifies overall correlation \( \rho \) may also be tested by the statistic,

\[ Z = \frac{\sum_{h} (\hat{\delta}_{hh} - \hat{\nu}_{h})^2 + \frac{1}{n} \left( 1 - \sum_{h} \hat{\nu}_{h} \right)}{\left[ 1 - \frac{1}{n^2} \right] \left[ 1 - \frac{1}{n} \right] \sqrt{\text{var}(\hat{\rho})}} \]

where \( Z \approx N(0, 1) \) for a large "a" or \( n \)

4 EXAMPLE

Fleiss (1971) and Landis and Koch (1977) used a psychiatric diagnostic data of 30 patients each patient classified separately by six psychiatrists into one of the five response categories: 1) depression, 2) personality disorder, 3) schizophrenia, 4) neurosis, and 5) others. Here patients are considered as clusters and the six diagnosis as the elements in the cluster. They measured the intraclass correlation and its variances of overall and individual categories in order to test the reliability of the psychiatrists. Table 1 compares the estimates of intraclass correlations and their variances of Landis and Koch (1977) and those of the Fleiss et al (1979) to the direct results.

The three sets of estimates for the intraclass correlations are quite close as expected, and for the direct estimation, the tests of significance for the hypothesis that there is no correlation is rejected at \( \alpha = 0.01 \) level of significance.

We may test the correlation with a hypothesis of nonzero correlation.

The estimator of MANOVA and direct methods are expected to be the same, and all three results are approximately equal. However, the wide difference of these standard errors requires careful evaluations of these variance estimators.

The variance of MANOVA estimators may be obtained under multinomial assumptions of cell distribution, and compared with the asymptotic variance of the direct estimator.

The further study on the variance of these estimator are required.

REFERENCES


Appendix A: Variance of \( \hat{\rho}_{hh} \)

We only consider \( \hat{\rho}_{hh} \) as the function of \( (\hat{\nu}_{h}, \hat{\delta}_{hh}) \) and is not in the from of a direct function of \( \gamma_{ijh} \)'s, we will use linear approximation twice to obtain \( \text{var}(\hat{\rho}_{hh}) \) as shown below.

Denote the partial derivative of \( \hat{\rho}_{hh} \) with respect to \( \hat{\nu}_{h} \) by \( f_{1} \), and the partial derivative of \( \hat{\rho}_{hh} \) with respect to \( \hat{\delta}_{hh} \) by \( f_{2} \), both evaluated at the parameters \( \hat{\nu}_{h} \) and \( \hat{\delta}_{hh} \), assuming that the first order partial derivatives exist.
The linear approximation for \( \text{var}(\hat{\rho}_{hh}) \) is expressed as

\[
\text{var}(\hat{\rho}_{hh}) = \begin{bmatrix} f_1 & f_2 \\ f_1 & f_2 \end{bmatrix} \begin{bmatrix} \text{var}(\hat{\pi}_h) & \text{cov}(\hat{\pi}_h, \hat{\delta}_{hh}) \\ \text{cov}(\hat{\pi}_h, \hat{\delta}_{hh}) & \text{var}(\hat{\delta}_{hh}) \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}
\]

The partial derivatives are given by

\[
f_1 = \frac{\partial \hat{\rho}_{hh}}{\partial \pi_h} = \frac{[(2\pi_h - 1)\delta_{hh} - \pi_h^2]}{(\pi_h(1-\pi_h))^2}
\]

\[
f_2 = \frac{\partial \hat{\rho}_{hh}}{\partial \delta_{hh}} = \frac{1}{\pi_h(1-\pi_h)}
\]

We obtain the \( \text{var}(\hat{\delta}_{hh}) \), \( \text{var}(\hat{\pi}_h) \), and \( \text{cov}(\hat{\delta}_{hh}, \hat{\pi}_h) \) by linear approximation for the second time. Since both \( \hat{\delta}_{hh} \) and \( \hat{\pi}_h \) are now the function of \( Y_h = (Y_{h1}, Y_{h2}, \ldots, Y_{hn}) \), \( (n = \sum_i b_i) \) for the \( h \)-th category, and these asymptotic variances and covariance can be obtained by the second linear approximation as

\[
\begin{vmatrix} \text{var}(\hat{\pi}_h) & \text{cov}(\hat{\pi}_h, \hat{\delta}_{hh}) \\ \text{cov}(\hat{\pi}_h, \hat{\delta}_{hh}) & \text{var}(\hat{\delta}_{hh}) \end{vmatrix} = \begin{bmatrix} J'_{\pi h} & J_{\pi h} \\ J_{\pi h} & J'_{\delta_{hh}} \end{bmatrix} W_{hh} \begin{bmatrix} J_{\pi h} & J_{\delta_{hh}} \end{bmatrix}
\]

where \( W_{hh} \) is now the \( n \times n \) covariance matrix of \( y_h \), obtained from the definition (2), and the \( n \times 1 \) vector \( J_{\pi h} \) is the partial derivatives of \( \hat{\pi}_h \) with respect to \( y_h \), and \( J_{\delta_{hh}} \) is the \( n \times 1 \) partial derivative vector of \( \hat{\delta}_{hh} \) with respect to \( y_h \), both evaluated at the parameters \( \pi_h \) and \( \delta_{hh} \). These are

\[
J'_{\pi h} = \frac{\partial \hat{\pi}_h}{\partial \pi_h} = \frac{1}{\pi_h} \begin{bmatrix} 1, 1, \ldots, 1 \end{bmatrix}
\]

\[
J_{\pi h} = \frac{\partial \hat{\pi}_h}{\partial y_h} = \frac{2\pi_h[(b_1-1)(b_1-1),(2-1)(2-1),(b_1-1),\ldots]}{\pi_h}
\]

Substituting (A4) into (A1), we obtain

\[
\text{var}(\hat{\rho}_{hh}) = \begin{bmatrix} f_1 & f_2 \\ f_1 & f_2 \end{bmatrix} \begin{bmatrix} J'_{\pi h} & W_{hh} [J_{\pi h} J_{\delta_{hh}}] \\ J_{\pi h} & J'_{\delta_{hh}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} f_1 & f_2 \end{bmatrix} \begin{bmatrix} J'_{\pi h} & W_{hh} [J_{\pi h} J_{\delta_{hh}}] & f_1 \\ J_{\pi h} & J'_{\delta_{hh}} & f_2 \end{bmatrix}
\]

\[
= f_1^2 (J_{\pi h} W_{hh} J_{\pi h}) + 2 f_1 f_2 (J_{\pi h} W_{hh} J_{\delta_{hh}}) + f_2^2 (J_{\delta_{hh}} W_{hh} J_{\delta_{hh}})
\]

where

\[
(J_{\pi h} W_{hh} J_{\pi h}) = \text{var}(\hat{\pi}_h),
\]

\[
(J_{\pi h} W_{hh} J_{\delta_{hh}}) = \text{cov}(\hat{\delta}_{hh}, \hat{\pi}_h),
\]

\[
(J_{\delta_{hh}} W_{hh} J_{\delta_{hh}}) = \text{var}(\hat{\delta}_{hh})
\]

and these are:

\[
\text{var}(\hat{\pi}_h) = \frac{\pi_h(1-\pi_h)}{n} + \frac{H(\delta_{hh} - \pi_h^2)}{n^2}
\]

\[
\text{var}(\hat{\delta}_{hh}) = \frac{4\pi_h^2 \left[ \pi_h(1-\pi_h)D + (L-D)(\delta_{hh} - \pi_h^2) \right]}{nH}
\]

\[
\text{cov}(\hat{\pi}_h, \hat{\delta}_{hh}) = \frac{2\pi_h \left[ \pi_h(1-\pi_h)H + D(\delta_{hh} - \pi_h^2) \right]}{nH}
\]

Note that the order of convergence to zero of \( \text{var}(\hat{\pi}_h), \text{var}(\hat{\delta}_{hh}), \) and \( \text{cov}(\hat{\pi}_h, \hat{\delta}_{hh}) \) is \( o(a^{-1}) \). Substituting (A9), (A10), (A11) along with the partial derivatives (A2) and (A3) into (A8), we obtain, using the notation \( p_{hh} = \pi_h(1-\pi_h) \), \( \hat{\rho}_{hh} = \frac{\pi_h(1-\pi_h)H + D(\delta_{hh} - \pi_h^2)}{nH} \),

\[
\text{var}(\hat{\rho}_{hh}) = \frac{\pi_h(1-\pi_h) + H(\delta_{hh} - \pi_h^2)}{n^2} + \frac{4\pi_h^2 \left[ \pi_h(1-\pi_h)D + (L-D)(\delta_{hh} - \pi_h^2) \right]}{nH}
\]

Substituting (A9), (A10), (A11) along with the partial derivatives (A2) and (A3) into (A8), we obtain, using the notation \( p_{hh} = \pi_h(1-\pi_h) \),

\[
\text{var}(\hat{\rho}_{hh}) = \frac{\pi_h(1-\pi_h) + H(\delta_{hh} - \pi_h^2)}{n^2} + \frac{4\pi_h^2 \left[ \pi_h(1-\pi_h)D + (L-D)(\delta_{hh} - \pi_h^2) \right]}{nH}
\]

The order of convergence to zero of \( \text{var}(\hat{\rho}_{hh}) \) is also \( o(a^{-1}) \).

**Appendix B: Variance of \( \tilde{\rho} \) shown in (10)**

\( \tilde{\rho} \) is now seen as function of \( \hat{\theta} = (\hat{\delta}_{h1}, \ldots, \hat{\delta}_{hr} \) and \( \hat{\pi}_1, \ldots, \hat{\pi}_r \).

The variance of average intraclass correlation over the \( r \) response categories is obtained by linear approximation as

\[
\text{var}(\tilde{\rho}) = \frac{\pi_h(1-\pi_h) + H(\delta_{hh} - \pi_h^2)}{n^2} + \frac{4\pi_h^2 \left[ \pi_h(1-\pi_h)D + (L-D)(\delta_{hh} - \pi_h^2) \right]}{nH}
\]
where \( V \) is the 2r x 2r covariance matrix of 
\( \hat{\theta}' = (\hat{\theta}_{11}, \ldots, \hat{\theta}_{rr}, \hat{\pi}_{1}, \ldots, \hat{\pi}_{r}) \),
and \( F' = (F_{61}, \ldots, F_{6r}, F_{\pi_{1}}, \ldots, F_{\pi_{r}}) \) is 1x2r
partial derivative vector of \( \hat{\rho} \) with respect to \( \hat{\theta} \),
evaluated at \( \Theta = (\theta_{11}, \ldots, \theta_{rr}, \pi_{1}, \ldots, \pi_{r}) \); that is,
\[
F_{\pi_{h}} = \frac{\partial \rho}{\partial \pi_{h}} = \frac{2m_{h}(\sum_{h} \pi_{hh} - 1)}{[1 - \pi_{hh}^2]^{2}}
\] (B2)
and
\[
F_{6_{hh}} = \frac{\partial \rho}{\partial 6_{hh}} = \frac{1}{(1 - \pi_{hh})}
\] (B3)

Since the covariance matrix \( V \) is now for the
2r estimates, each estimate based on the n
variables of \( y_{11h}', y_{12h}', \ldots, y_{nh} \) (n = \( \sum b_{i} \)) for
\( h = 1, \ldots, r \), we may use the linear
approximation for the second time to
find this covariance matrix \( V = JWJ' \), where the
covariance matrix \( W \) is of \( y_{11h}', y_{12h}', \ldots, y_{nh} \) for
\( h = 1, \ldots, r \), expressed as
\[
W = \begin{bmatrix}
W_{11} & W_{21} & W_{31} & \cdots & W_{r1} \\
W_{12} & W_{22} & W_{32} & \cdots & W_{r2} \\
W_{13} & W_{23} & W_{33} & \cdots & W_{r3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
W_{1r} & W_{2r} & W_{3r} & \cdots & W_{rr}
\end{bmatrix}
\] (B4)
where each submatrix is the \( n \times n \) covariance
matrix for corresponding category \( h, h' = 1, \ldots, r \).
The submatrices \( W_{hh} \) (\( h = h' \)) on the diagonal are
the covariance matrices of \( y_{11h}', y_{12h}', \ldots, y_{nh} \)
for \( h = 1, \ldots, r \), and \( W_{hh} \) (\( h \neq h' \)) on the
off-diagonal are the covariance matrices between
\( y_{11h}', y_{12h}', \ldots y_{nh} \) and \( y_{11h}', \ldots, y_{nh} \) for \( h \neq h' \)
according to the definition (2).

Denote the partial derivatives of \( \hat{\rho} \) with
respect to \( y \) by the \( (r \times nr) \) diagonal matrix
\( J_{\pi} \) of submatrices \( J_{\pi_{h}} \) as
\[
J_{\pi} = \begin{bmatrix}
J_{\pi_{1}} & 0 & 0 & \cdots & 0 \\
0 & J_{\pi_{2}} & 0 & \cdots & 0 \\
0 & 0 & J_{\pi_{3}} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & J_{\pi_{r}}
\end{bmatrix}
\] (B5)
where the \( (1 \times n) \) submatrix \( J_{\pi_{h}} \) represents the
partial derivatives of \( \hat{\pi}_{h} \) with respect to \( y_{h} \),
evaluated at \( \pi_{h} (h = 1, \ldots, r) \) as shown in (A5),
and \( 0 \) is the \( (1 \times n) \) null matrix of zeros.

Denote the partial derivatives of \( \hat{\delta} \) with
respect to \( y \) by
\[
\begin{bmatrix}
J_{\delta_{1}} & 0 & 0 & \cdots & 0 \\
0 & J_{\delta_{2}} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & J_{\delta_{r}}
\end{bmatrix}
\] (B6)
where the \( (1 \times n) \) submatrix \( J_{\delta_{hh}} \) represents the
partial derivatives of \( \hat{\delta}_{hh} \) with respect to \( y_{h} \),
evaluated at \( \delta_{hh} (h = 1, \ldots, r) \) as shown in (A6).

We now take the linear approximation of the
covariance \( V \) under usual constraints as
\[
V = \begin{bmatrix}
J_{\pi}' & W & J_{\delta}
\end{bmatrix} = J W J' = J_{\pi}W_{\pi}J_{\pi} + J_{\pi}W_{\delta}J_{\delta} + J_{\pi}W_{\delta_{hh}}J_{\delta_{hh}}
\] (B7)
where \( J' = [J_{\pi}, J_{\delta}] \).

Substituting (B7) into (B1), the variance
of \( \hat{\rho} \) is given by
\[
\text{var}(\hat{\rho}) = F' (J W J') F
\] (B8)
The double linear estimation is often useful when
a direct partial derivative is not easily
obtainable. The partial derivation can be repeated
further if necessary. The equivalence of the
repeated partial derivatives to a single partial
derivative can be seen from the fact that
\[
\frac{\partial \rho}{\partial \pi_{h}} = \frac{\partial \rho}{\partial \delta_{hh}} \frac{\partial \delta_{hh}}{\partial \pi_{h}}
\]
(B9)
(B8) can be rewritten explicitly as
\[
\text{var}(\hat{\rho}) = F' \sum_{h} F_{\pi_{h}}\text{var}(\hat{\pi}_{h}) + \sum_{h \neq h'} F_{\pi_{h}}F_{\pi_{h'}}\text{cov}(\hat{\pi}_{h}, \hat{\pi}_{h'})
\]
\[
+2\sum_{h} F_{\pi_{h}}F_{\delta_{hh}}\text{cov}(\hat{\delta}_{hh}, \hat{\pi}_{h}) + \sum_{h \neq h'} F_{\pi_{h}}F_{\delta_{hh}}\text{cov}(\hat{\delta}_{hh}, \hat{\delta}_{h'})
\]
\[
+ \sum_{h \neq h'} F_{\delta_{hh}}\text{var}(\hat{\delta}_{hh}) + \sum_{h \neq h'} F_{\pi_{h}}F_{\delta_{hh}}\text{cov}(\hat{\delta}_{hh}, \hat{\delta}_{h'})
\] (A10)
where we have seen \( \text{var}(\hat{\pi}_{h}), \text{var}(\hat{\delta}_{hh}), \) and
\( \text{cov}(\hat{\pi}_{h}, \hat{\pi}_{h'}), \text{cov}(\hat{\delta}_{hh}, \hat{\delta}_{h'}) \) previously in (A9), (A10) and (A11),
respectively. We can also obtain the linear
approximation of \( \text{cov}(\hat{\pi}_{h}, \hat{\delta}_{h'}) \),
and \( \text{cov}(\hat{\pi}_{h}, \hat{\delta}_{h'}) \) for \( h \neq h' \) as
\[
\text{cov}(\hat{\pi}_{h}, \hat{\delta}_{h'}) = -\pi_{h}\pi_{h'}/n + H (6_{hh}, - \pi_{h}\pi_{h'})/ n^2
\] (B11)
cov(\hat{\pi}_h, \hat{\pi}_h) = 2\pi_h \left[ \frac{\pi_h(1 - \pi_h)}{n^2} + D(\pi_h, \pi_h, \pi_h) \right] \tag{B12}

cov(\hat{\pi}_{1hh}, \hat{\pi}_{1hh}) = 4\pi_h \left[ \frac{\pi_h(1 - \pi_h)}{n^2} + D(\pi_h, \pi_h, \pi_h) \right] \tag{B13}

where H, D, and L are given in (12).

Substituting (A8), (A10), (A11), (B12), and (B13), along with partial derivatives (B2) and (B3), \text{var}(\hat{\rho}) can be written as,

\begin{align*}
\text{var}(\hat{\rho}) &= \frac{1}{H^2} \left[ \frac{1}{2} \sum_{i=1}^{r} \left[ \sum_{h} \pi_h \left( \frac{\pi_h(1 - \pi_h)}{n^2} + D(\pi_h, \pi_h, \pi_h) \right) \right] \right] \\
&+ \frac{1}{2} \sum_{i=1}^{r} \left[ \sum_{h} \pi_h \left( \frac{\pi_h(1 - \pi_h)}{n^2} \right) \right] \\
&+ \frac{1}{2} \sum_{i=1}^{r} \left[ \sum_{h} \pi_h \left( \frac{\pi_h(1 - \pi_h)}{n^2} \right) \right] \\
&+ \frac{1}{2} \sum_{i=1}^{r} \left[ \sum_{h} \pi_h \left( \frac{\pi_h(1 - \pi_h)}{n^2} \right) \right]
\end{align*}

where D = \sum_i b_i(b_i - 1)^2, L = \sum_i \left[ b_i(b_i - 1) \right]^2, and H = \sum_i b_i(b_i - 1). Note that (B14) approaches zero in the order of o(a^{-1}).

\begin{table}[h]
\centering
\caption{Intracluster Correlations and their Standard Deviations from the three methods (MANOVA, Direct, kappa)}
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
\textbf{Category} & \textbf{MANOVA (SE)} & \textbf{biased (SE)} & \textbf{unbiased (SE)} & \textbf{*Z-value} & \textbf{Kappa (SE)} \\
\hline
1 & 0.254(0.1062) & 0.245(0.055) & 0.254(0.0532) & \texttt{*4.780} & 0.248 (0.1140) \\
2 & 0.254(0.0994) & 0.245(0.055) & 0.254(0.0532) & \texttt{*4.780} & 0.248 (0.1140) \\
3 & 0.350(0.0719) & 0.350(0.132) & 0.350(0.1272) & \texttt{*4.166} & 0.517 (0.1166) \\
4 & 0.451(0.0742) & 0.471(0.054) & 0.481(0.0525) & \texttt{*9.165} & 0.470 (0.1396) \\
5 & 0.574(0.1263) & 0.566(0.101) & 0.576(0.0978) & \texttt{*5.886} & 0.565 (0.1277) \\
\hline
\textbf{Overall} & 0.440(0.0541) & 0.430(0.037) & 0.440(0.1072) & \texttt{*4.108} & 0.430 (0.0275) \\
\hline
\end{tabular}
\end{table}

* Significant at \( \alpha = 0.01 \) level of Z value for testing \( \rho = \rho_{hh} = 0 \).