NOTE ON THE STANDARD ERROR
OF THE DIFFERENCE BETWEEN COEFFICIENTS
OF VARIATION OF CORRELATED VARIABLES

JOSEPH ZUBIN
Teachers College, Columbia University

Offprinted from The Journal of Applied Psychology, Vol. 18, No. 3,
June, 1934, pp. 491-492
NOTE ON THE STANDARD ERROR OF THE DIFFERENCE BETWEEN COEFFICIENTS OF VARIATION OF CORRELATED VARIABLES

JOSEPH ZUBIN

Psychiatric Institute, Columbia University

In many experiments in Education and Psychology the crucial comparison is made between variabilities rather than between central tendencies. One of the measures of variability in common use, especially when the central tendencies of the compared variables are unequal, is the coefficient of variation, \( V = \frac{\sigma}{M} \). The standard error of this coefficient is given by Yule\(^4\) but the standard error of the difference between the coefficients of two correlated variables has not been developed yet.\(^2\) The purpose of this note is to derive the equation for this standard error.

The general expression for the standard error of two correlated variables is

\[
\sigma^2_{\beta_1 - \beta_2} = \sigma^2_{\beta_1} + \sigma^2_{\beta_2} - 2r_{\beta_1 \beta_2} \sigma_{\beta_1} \sigma_{\beta_2}
\]

(1)

Utilizing the equations derived by Pearson\(^3\) under the single assumption that quantities involving \( \frac{1}{N} \) are negligible in comparison with values involving \( \frac{1}{\sqrt{N}} \),

\[
\sigma^2_{\beta_1} = \frac{\beta_1}{N} \left( (\beta_1 - 1)/4 + \beta_1 - \beta_1 \sqrt{\beta_1} \right)
\]

(2) and a similar expression can be written for \( \sigma^2_{\beta_2} \).

Now \( r_{\beta_1 \beta_2} \sigma_{\beta_1} \sigma_{\beta_2} = E \left\{ \delta \left[ \frac{\sigma_{\beta_1}}{M_1} \right] \delta \left[ \frac{\sigma_{\beta_2}}{M_2} \right] \right\} \) where \( E \) represents


\(^{2}\text{The equation to be derived has appeared without proof in—Some Effects of Incentives, A Study of Individual Differences in Rivalry, p. 33, Teachers College Contributions to Education no. 532., 1932, by the present writer.}\)

\(^{3}\text{Pearson, K. On the probable errors of frequency constants, Biometrika vol. 9, p. 14.}\)
the sum of all cross product values in all samples, divided by the number of samples. Utilizing Pearson’s equations again, we obtain

\begin{align}
(3) \quad r_{mn} & = n_2 n_3 \left\{ \frac{P_{mn} P_{m} - P_{m} P_{n}}{P_{mm} P_{nn} - P_{mn}^2} \right\} \\
& - \frac{4 - 2 r_{m} r_{n}}{M_1 P_{mn} + P_{m} P_{n} - P_{mm} P_{nn}} \\
& \quad \text{Substituting in (1)}
\end{align}

\begin{align}
(4) \quad \sigma^2_{(n-m,n)} & = \frac{1}{N} \cdot \left\{ \nu_1 \nu_2 \left[ (\beta_1 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_1} \right] \\
& + \nu_2^2 \left[ (\beta_2 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_2} \right] \\
& - 2 \nu_1 \nu_2 \left[ (\beta_1 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_1} \right] \\
& - \nu_1 \sqrt{\beta_1} \right\} \\
& \quad \text{For linear and homoscedastic regression,}
\end{align}

\begin{align}
P_{mn} & = r_{mn} \sigma^2_{n} \sigma^2_{m} \sqrt{\beta_1} \\
P_{mn} & = r_{nm} \sigma^2_{n} \sigma^2_{m} \sqrt{\beta_1} \quad \text{and} \quad \beta_1 = \beta^*_1 \\
P_{mn} P_{m} P_{n} & = r_{nm} (\beta_1 - 1) + 1 \\
& \quad \text{Substituting}
\end{align}

\begin{align}
(5) \quad \sigma^2_{(n-m,n)} & = \frac{1}{N} \cdot \left\{ \nu_1 \left[ (\beta_1 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_1} \right] \\
& + \nu_2^2 \left[ (\beta_2 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_2} \right] \\
& - 2 \nu_1 \nu_2 \left[ (\beta_1 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_1} \right] \\
& - \nu_1 \sqrt{\beta_1} \right\} \\
& \quad \text{For symmetrical distributions,} \quad \beta_1 = \beta^*_1 = 0.
\end{align}

\begin{align}
(6) \quad \sigma^2_{(n-m,n)} & = \frac{1}{N} \cdot \left\{ \nu_1 \left[ (\beta_1 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_1} \right] \\
& + \nu_2^2 \left[ (\beta_2 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_2} \right] \\
& - 2 \nu_1 \nu_2 \left[ (\beta_1 - 1)/4 + \nu_2^2 - \nu_1 \sqrt{\beta_1} \right] \\
& - \nu_1 \sqrt{\beta_1} \right\} \\
& \quad \text{and for normal distributions}
\end{align}

\begin{align}
(7) \quad \sigma^2_{(n-m,n)} & = \frac{1}{2N} \cdot \left\{ \nu_1 \left[ 1 + 2 \nu_2^2 \right] + \nu_2 \left[ 1 + 2 \nu_1^2 \right] \\
& \quad - 2 \nu_1 \nu_2 \left[ \nu_2 \left[ 1 + 2 \nu_2^2 \right] \right] \\
& \quad \text{Following the usual convention of taking} \quad n_1 = 100 \sigma_1 / M_1, \\
& \quad \text{instead of} \quad \sigma_1 / M_1, \quad \text{as we have done,}
\end{align}

\begin{align}
(8) \quad \sigma^2_{(n-m,n)} & = \frac{1}{2N} \cdot \left\{ \nu_1 \left[ 1 + 2 \nu_2^2 \right] + \nu_2 \left[ 1 + 2 \nu_1^2 \right] \\
& \quad - 2 \nu_1 \nu_2 \left[ \nu_2 \left[ 1 + 2 \nu_2^2 \right] \right] \\
& \quad \text{Discarding the values that are divided by (100)c, if} \\
& \quad n_1 \text{and} \quad n_2 \text{are small,}
\end{align}

\begin{align}
(9) \quad \sigma^2_{(n-m,n)} & = \frac{1}{2N} \cdot \left\{ \nu_1^2 + \nu_2^2 - 2 \nu_1 \nu_2 \nu_3 \right\}
\end{align}