REGRESSION MODELS IN RESEARCH SYNTHESIS

BY

LARRY V. HEDGES and INGRAM OLKIN

TECHNICAL REPORT NO. 176
OCTOBER 1981

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MCS 78-07736

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
REGRESSION MODELS IN RESEARCH SYNTHESIS

By

Larry V. Hedges and Ingram Olkin
The University of Chicago and Stanford University

1. Introduction.

In many areas of behavioral, social and medical sciences, research studies are replicated. Indeed, in some instances, so many studies have been conducted that novel techniques to synthesize the results of these studies are needed. The focus to date has been on the determination of measure of effect magnitude from each study. This measure is usually an index or statistic, such as a correlation coefficient. The task then is to analyze the collection of indices.

Glass (1976) was among the first to suggest the use of quantitative indices of effect magnitude in research synthesis. He suggested the standardized mean difference and the correlation coefficient as the two most useful indices of effect magnitude.

The present study was stimulated by an analysis of six independent studies in the form of a 2 × 3 table of sex (male, female) by grade (7, 9, 11) in which the observation in each cell is the correlation between an aptitude test score and a formal reasoning task. The problem is to assess whether there is a systematic effect of sex or grade on the values of the population correlations.

Although the above example falls into an analysis of variance framework, it is apparent that other more complicated analyses might be required.
With this in mind, we state our results in the more general framework of regression.

The main ideas of the method are developed in Section 2. The principles are then applied to deal with correlation coefficients (Section 3), and standardized mean differences (Section 4). In each case an application is provided.

2. Model, Notation and Methodology.

Consider a series of $k$ independent studies in which the $i$-th study yields an estimate $T_i$ of the population effect magnitude $\theta_i$, $i = 1, \ldots, k$. Denote the $k$-dimensional column vector of estimates and parameters by

$$T = (T_1, \ldots, T_k)', \quad \theta = (\theta_1, \ldots, \theta_k)'$$

Further, we assume a regression model

$$\theta = X\beta,$$  \hspace{1cm} (2.1)

where $X$ is a $k \times p$ design matrix and $\beta$ is a $p \times 1$ matrix of parameters. For the moment assume that $X$ is of full rank. This is not essential, since analogous procedures can be used in the less than full rank case. In the less than full rank case, inverses are replaced by generalized inverses. However, since the assumption that $X$ is of full rank simplifies the procedures and focuses on the essentials, we prefer to avoid the unnecessary complications of the more general situation.

The problem is to estimate $\beta$ or to test a hypothesis about $\beta$.

Suppose that each $T_i$ is based on a sample of size $n_i$, $i = 1, \ldots, k$, and that as $n = \sum n_j \to \infty$ with $n_i/n$, $i = 1, \ldots, k$ fixed, the asymptotic
joint distribution of \( T_1, \ldots, T_k \) is multivariate normal, i.e.,

\[
\sqrt{n} \ (T-\theta) \sim \mathcal{N}(0, \Sigma) ,
\]  

(2.2)

where the covariance matrix \( \Sigma = (\sigma_{ij}) \) may depend on the sample sizes \( n_1, \ldots, n_k \) but not on the parameters.

It then follows that the maximum likelihood estimator (and generalized least squares estimator) of \( \beta \) is

\[
\hat{\beta} = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}T ,
\]

(2.3)

with covariance matrix \( \Psi = (X'\Sigma^{-1}X)^{-1} \). Consequently,

\[
\sqrt{n} \ (\hat{\beta} - \beta) \sim \mathcal{N}(0, \Psi) .
\]

(2.4)

Since \( \Psi \) is a known matrix, a confidence ellipsoid for \( \beta \) can be obtained from the fact that

\[
\sqrt{n} (\hat{\beta} - \beta) \Psi^{-1/2} \sim \chi^2_p .
\]

(2.5)

Simultaneous \( 100(1-\alpha)\% \) confidence intervals for the \( \beta \)'s can be obtained from Bonferroni inequalities:

\[
\hat{\beta} - c \sqrt{\psi_{ii}/n} \leq \beta_i \leq \hat{\beta} + c \sqrt{\psi_{ii}/n} , \ i = 1, \ldots, p,
\]

where \( c \) is the critical value of the standard normal distribution corresponding to the \( \alpha/2p \) point.

To test that \( \beta = 0 \) we can use the statistic
\[ q = \hat{\beta}' \psi^{-1} \hat{\beta} = n^T \Sigma^{-1} X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} T, \]  

(2.6)

which has a chi-square distribution with \( p \) degrees of freedom.

To test that the model \( \theta = X\beta \) adequately fits the data we use the statistic

\[ Q = n(T^T \Sigma^{-1} T - \hat{\beta}' \psi^{-1} \hat{\beta}) \]

\[ = n(T^T \Sigma^{-1} T - T^T \Sigma^{-1} X (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} T) \]

(2.7)

which is to be compared to a chi-squared statistic with \( k - p \) degrees of freedom.

Although the above development is standard, it has not been used in the analysis of correlations or standardized mean differences. Note that each of these applications requires a variance stabilizing transformation to put it into the present framework.

3. The Correlation Coefficient.

3.1 Methodology. Consider a series of \( k \) independent studies each of which produces a pair of measurements. Although the studies may not actually be replications, they do measure the same characteristics in the sense that they are linearly equatable or have a single common factor and no unique factors.

The index used in each study is a correlation. If \( r_{ij} \) and \( \rho_{ij} \) are the sample and population correlation coefficients for the \( i \)-th study, then the asymptotic distribution of \( \sqrt{n_i}(r_{ij} - \rho_{ij}) \) is normal with zero mean and
variance \((1-\rho_i^2)^2\), where \(n_i\) is the sample size of the \(i\)-th study.

Since the asymptotic variance is not independent of \(\rho_i\), we cannot use the methodology of Section 2. However, if we transform \(r\) by Fisher's z-transformation:

\[
z = f(r) = \frac{1}{2} \log \left( \frac{1+r}{1-r} \right),
\]

\[
\zeta = f(\rho) = \frac{1}{2} \log \left( \frac{1+\rho}{1-\rho} \right),
\]

then the methodology does apply.

Let

\[
z = (z_1, \ldots, z_k)', \quad \zeta = (\zeta_1, \ldots, \zeta_k)',
\]

be the vector of \(z\)'s and \(\zeta\)'s for the \(k\) studies. Then, as \(n = \Sigma n_j \to \infty\) with \(n_i/n, i = 1, \ldots, k\), fixed, the asymptotic joint distribution of \(z\) is

\[
\sqrt{n} \left( z - \zeta \right) \sim \mathcal{N}(0, \Psi),
\]

where \(\Psi = (\psi_{ij})\) is a diagonal matrix with \(\psi_{ii} = n/n_i, i = 1, \ldots, k\).

3.2 An Application. A formal reasoning task (requiring prediction of displaced volume) and an aptitude test were administered to males and females in each of three grade levels (grades 7, 9, 11) at the Lawrence Hall of Sciences, Linn (1980). The design yielded six groups differentiated by sex and grade level. The measures obtained were correlations, as given in Table 1.
Table 1

Correlations Between Formal Reasoning and Aptitude Test Scores
for Males and Females at Three Grade Levels

<table>
<thead>
<tr>
<th>Grade</th>
<th>Sex</th>
<th>n</th>
<th>r</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Male</td>
<td>145</td>
<td>.19</td>
<td>.19</td>
</tr>
<tr>
<td>7</td>
<td>Female</td>
<td>159</td>
<td>-.08</td>
<td>-.08</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>136</td>
<td>.31</td>
<td>.32</td>
</tr>
<tr>
<td>9</td>
<td>Female</td>
<td>77</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td>11</td>
<td>Male</td>
<td>122</td>
<td>.16</td>
<td>.16</td>
</tr>
<tr>
<td>11</td>
<td>Female</td>
<td>139</td>
<td>.21</td>
<td>.21</td>
</tr>
</tbody>
</table>

Note: These data were obtained from M. Linn, Director, Adolescent Reasoning Project, Laurence Hall of Science, University of California, Berkeley.

To analyze the data in a regression context, let \( \beta_1 \) denote the grand mean, \( \beta_2 \) the sex effect, and \( \beta_3, \beta_4 \) be the grade level effects. The design matrix is

\[
X = \begin{bmatrix}
1 & 1 & 1 & 0 \\
1 & -1 & 1 & 0 \\
1 & 1 & 0 & 1 \\
1 & -1 & 0 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & -1 \\
\end{bmatrix}
\]

the transformed correlations are

\[
z = \begin{bmatrix}
.19 \\
-.08 \\
.32 \\
.10 \\
.16 \\
.21 \\
\end{bmatrix}
\]

with diagonal covariance matrix \( \Sigma = (\sigma_{ij}) \)

\( = \text{diag}(1/145, 1/159, 1/136, 1/77, 1/22, 1/139) \). The estimated regression
coefficients are

\[ \hat{\beta} = (0.154, 0.074, -0.102, 0.065) \]

with asymptotic covariance matrix \( \psi = (X'\Sigma^{-1}X)^{-1} \):

\[
(1/n)\psi = 10^{-4} \\
\begin{bmatrix}
13.17 & -7.3 & -2.24 & 2.68 \\
-7.3 & 13.16 & 1.33 & 2.92 \\
-2.24 & 1.33 & 24.22 & -13.78 \\
2.68 & 2.92 & -13.78 & 29.42
\end{bmatrix}
\]

To test the hypothesis \( H: \beta = 0 \) we compute the statistic (2.6):

\[ q = n\hat{\beta}'\psi^{-1}\hat{\beta} = 26.271 \]

which is to be compared with the critical value 9.49 of a chi-square distribution with \( p = 4 \) degrees of freedom. Since we reject the hypothesis that \( \beta = 0 \) (at the 5% level of significance). We wish to obtain confidence bounds for the \( \beta \)'s. Simultaneous 95% confidence bounds using Bonferroni inequalities are obtained from

\[ \hat{\beta}_i - \sqrt{\psi_{ii}/n} \leq \beta_i \leq \hat{\beta}_i + \sqrt{\psi_{ii}/n} \]

where \( c \) is the critical value of the normal distribution corresponding to the \( \alpha/2p = .00625 \) point.

This process yields the following 95% asymptotic confidence intervals:
.049 \leq \beta_1 \leq .258, \quad -.030 \leq \beta_2 \leq .178, \\
-.244 \leq \beta_3 \leq .040, \quad -.091 \leq \beta_4 \leq .221.

To test that the model fits the data compute the statistic \( Q \) given by (2.7):

\[
Q = n (z' \Sigma^{-1} z - \hat{\beta}' \hat{\psi}^{-1} \beta) = 3.931
\]

which is to be compared with the critical value 5.99 obtained from the chi-square distribution with \( k - p = 6 - 4 = 2 \) degrees of freedom.

4. The Standardized Mean Difference.

4.1 Methodology. Consider a series of \( k \) independent studies each of which consists of a sample of size \( n \) from an experimental (E) and a control (C) group. The outcome variable \( w_i \) from the \( i \)-th study is assumed to be normally distributed with means \( \mu^E_i \) and \( \mu^C_i \) for the experimental and control groups, respectively, and with common standard deviation \( \sigma^E_i = \sigma^C_i = \sigma_i \). The respective sample means are denoted by \( \bar{w}^E_i \) and \( \bar{w}^C_i \), and the pooled within-group sample variance is denoted by \( s_i^2 \). The size of effect for the \( i \)-th study is the standardized mean difference:

\[
\gamma_i = \frac{\mu^E_i - \mu^C_i}{\sigma_i},
\]

which is estimated by

\[
\hat{\gamma}_i = \frac{\bar{w}^E_i - \bar{w}^C_i}{s_i}.
\]
The estimators $g_i$ are consistent estimators of $\gamma_i$, i.e., the $g_i$ approach $\gamma_i$ in probability as $n_i \to \infty$. Hedges (1981) has shown that the bias of the $g_i$ is less than two percent when $n_i > 20$, and that the asymptotic variance of $\sqrt{n_i}(g_i - \gamma_i)$ is $(8 + \gamma_i^2)/4$. A variance stabilizing transformation is given by Hedges and Olkin (1981):

$$d = \sinh^{-1}\left(\frac{g}{2\sqrt{2}}\right) = \log\left(\frac{g + \sqrt{g^2 + 8}}{2\sqrt{2}}\right)$$

(4.1)

for the sample estimator $g$. A similar transformation from $\gamma$ to $\delta$ is used for the population parameters, i.e., $\delta = \sinh^{-1}(\gamma/2\sqrt{2})$. Let

$$d = (d_1, \ldots, d_k)' , \quad \delta = (\delta_1, \ldots, \delta_k)'$$

denote the vectors of transformed sample estimators and parameter values.

If $n = \sum_{i=1}^{k} n_i \to \infty$ with $n_i/n_i$, $i = 1, \ldots, k$, constant, then the asymptotic distribution of $d$ is multivariate normal:

$$\sqrt{n} (d - \delta) \sim n(0, \Omega) ,$$

(4.2)

where $\Omega = (\omega_{ij})$ is a diagonal matrix with $\omega_{ii} = n_i/4n_i$, $i = 1, \ldots, n$.

The structural model relating $\delta$ to $X$ is

$$\delta = X\beta ,$$

which is the same as our model of Section 2.
4.2 An Application. Eleven studies of the effects of open education on student independence were examined by Hedges and Gage (1981). Some of these studies defined open education via teaching practices while others defined open education via the type of school architecture. The fidelity of the open teaching treatment presumably varied across studies. Some studies used classroom observations to verify that open teaching practices occurred and these studies are thought to have the highest treatment fidelity. The studies were conducted at several grade levels, and are summarized in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Definition of Openness</th>
<th>Classroom Observations</th>
<th>Grade Level</th>
<th>( n_1 )</th>
<th>( g_1 )</th>
<th>( d_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Teaching</td>
<td>No</td>
<td>5</td>
<td>15</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>No</td>
<td>3</td>
<td>18</td>
<td>.701</td>
<td>.245</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>No</td>
<td>2</td>
<td>22</td>
<td>2.016</td>
<td>.663</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>Yes</td>
<td>2</td>
<td>54</td>
<td>-.540</td>
<td>-.190</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>No</td>
<td>4</td>
<td>30</td>
<td>.669</td>
<td>.234</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>No</td>
<td>5</td>
<td>30</td>
<td>-.235</td>
<td>-.083</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>No</td>
<td>6</td>
<td>30</td>
<td>-.079</td>
<td>-.028</td>
</tr>
<tr>
<td>Open Space</td>
<td>No</td>
<td>6</td>
<td>30</td>
<td>-.494</td>
<td>-.174</td>
</tr>
<tr>
<td>Open Space</td>
<td>No</td>
<td>7</td>
<td>30</td>
<td>-.058</td>
<td>-.020</td>
</tr>
<tr>
<td>Open Space</td>
<td>No</td>
<td>8</td>
<td>30</td>
<td>-.587</td>
<td>-.206</td>
</tr>
<tr>
<td>Open Teaching</td>
<td>Yes</td>
<td>8</td>
<td>114</td>
<td>-.737</td>
<td>-.258</td>
</tr>
</tbody>
</table>
To analyze the data using a regression model, let $\beta_1$ be the effect of open space versus open teaching practices, $\beta_2$ be the effect of verification of treatment by classroom observations, and $\beta_3$ be the effect of grade level. Here we group grades into sets and assign a predictor value of one for grades one to three, two for grades four to six, and three for grades seven and eight. The design matrix is

$$
X = 
\begin{bmatrix}
1 & 0 & 2 \\
1 & 0 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
1 & 0 & 2 \\
1 & 0 & 2 \\
1 & 0 & 2 \\
0 & 0 & 2 \\
0 & 0 & 3 \\
0 & 0 & 3 \\
1 & 1 & 3
\end{bmatrix}
$$

the transformed standardized mean differences are

$$d = (.000, .245, .663, -.190, .234, -.033, -.028, -.174, -.020, -.206, -.258)'$$

The covariance matrix is

$$\Sigma = (\sigma_{ij}) =
\text{diag}(1/60, 1/72, 1/88, 1/216, 1/120, 1/120, 1/120, 1/120, 1/120, 1/120, 1/456).$$

The estimated regression coefficients are

$$\hat{\beta} = (.257, -.356, -.058)$$

with estimated covariance matrix $\Omega = (X'\Sigma^{-1}X)^{-1}$:
\[
\begin{pmatrix}
2.614 & -1.398 & -.516 \\
10^{-3} & -1.398 & 3.332 \\
-.516 & .189 & .299
\end{pmatrix}
\]

To test the hypothesis \( H: \beta = 0 \), we compute the statistic (2.6):
\[ q = n\beta'\Omega^{-1}\beta = 62.892 \]
which is compared with 7.81, the critical value of the chi-square distribution with \( p = 3 \) degrees of freedom. Since we reject the hypothesis that \( \beta = 0 \) at the 5% level of significance, we wish to obtain confidence bounds for \( \beta_1, \beta_2, \) and \( \beta_3 \). Simultaneous 95% confidence bounds are obtained using the Bonferroni inequality as in the previous example, yielding the following 95% asymptotic confidence intervals:
\[
.115 \leq \beta_1 \leq .399 \\
-.516 \leq \beta_2 \leq -1.96 \\
-.106 \leq \beta_3 \leq -.010.
\]

A test of the specification of the model uses the statistic \( Q \) given in (2.7):
\[ Q = nd'\bar{\epsilon}^{-1}d - q = 34.483, \]
which is compared to 15.5, the critical value of the chi-square distribution with \( k - p = 8 \) degrees of freedom.
REFERENCES


Linn, M. Personal Communication.