Consider the model
\[ Y_i = \mu + \beta x_i + \epsilon_i \]
where \( x_i = x_i - \bar{x}, i=1, \ldots, n \), and \( \epsilon_i \) is normally distributed with zero mean.

The complete set of observations may be written as
\[ Y = X(\mu) + R \]
where \( Y \) and \( R \) are \( n \times 1 \) matrices and \( X \) is an \( n \times 2 \) matrix. Let the covariance matrix of the \( \epsilon_i \)'s, i.e. \( RR' \), be \( M \).

In the usual regression, we assume \( M = \sigma^2 I \), i.e. the \( \epsilon \)'s are independent and have a common variance. A fairly common regression problem arises when
\[
M = \begin{pmatrix}
  w_1 & 0 & \cdots & 0 \\
  0 & \ddots & \ddots & \vdots \\
  \vdots & \ddots & \ddots & 0 \\
  0 & \cdots & 0 & w_n
\end{pmatrix} \sigma^2,
\]
i.e. when the \( \epsilon \)'s are assumed to be independent but are multiples of a common variance; a regression of means, based on different numbers of observations, on an independent variable exemplifies this case. The least common case, where \( M \) is a non-singular, covariance matrix, arises when the assumptions of independence are not justified.

When normality is assumed, the joint distribution of the \( Y_i \)'s requires the quadratic form
\[
Q = \frac{1}{2} \sum_{i,j} \sigma^{ij}(Y_i - \mu - \beta x_i)(Y_j - \mu - \beta x_j)
\]

---

*Cornell University, Ithaca, New York*
where \( \sigma^{ij} \) is an element of \( M^{-1} \). It is required to minimize this expression by appropriately estimating \( \mu \) and \( \beta \).

The partial derivatives are:

\[
\frac{\partial Q}{\partial \mu} = - \sum_i (Y_i - \mu - x_i \beta) \sum_j \sigma^{ij} \\
\frac{\partial Q}{\partial \beta} = - \sum_i (Y_i - \mu - x_i \beta) \sum_j \sigma^{ij} x_j
\]

Equating the derivatives to zero, we obtain

\[
(\sum_i \sigma^{ij}) \hat{\mu} + (\sum_i x_i \sum_j \sigma^{ij}) \hat{\beta} = \sum_i Y_i \sum_j \sigma^{ij} \\
(\sum_i x_i \sum_j \sigma^{ij} x_j) \hat{\mu} + (\sum_i x_i \sum_j \sigma^{ij} x_j) \hat{\beta} = \sum_i Y_i \sum_j \sigma^{ij} x_j
\]

Solution of these equations is given by

\[
\hat{\mu} = \frac{\sum \sigma^{ij} \sum_i x_i \sum_j \sigma^{ij} \sum_i Y_i \sum_j \sigma^{ij}}{\sum \sigma^{ij} \sum_i x_i \sum_j \sigma^{ij}} \\
\hat{\beta} = \frac{(\sum \sigma^{ij} \sum_i x_i \sum_j \sigma^{ij})/ \sum \sigma^{ij}}{(\sum \sigma^{ij} \sum_i x_i \sum_j \sigma^{ij})/ \sum \sigma^{ij}}
\]

If we define \( \bar{x} \) as

\[
\frac{(\sum_i x_i \sum_j \sigma^{ij})/ \sum \sigma^{ij}}{(\sum \sigma^{ij} \sum_i x_i \sum_j \sigma^{ij})/ \sum \sigma^{ij}}
\]
then \[ \frac{\sum (\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j})}{\sigma_{i} \sigma_{j}} = 0 \]

\[
\hat{\mu} = \frac{\sum_{i} \sigma_{i} \sigma_{j}}{\sigma_{i} \sigma_{j}}.
\]

When the \( \epsilon \)'s are independent and their variances are multiples of a common variance, equations (1) and (2) reduce to

\[
\hat{\mu} = \begin{bmatrix}
\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j} & \sum_{i} \epsilon_{i} \sigma_{j} \\
\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j} & \sum_{i} \epsilon_{i} \sigma_{j} \\
\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j} & \sum_{i} \epsilon_{i} \sigma_{j}
\end{bmatrix}
\]

\[
\hat{\beta} = \frac{\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j}}{\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j}}
\]

since the \( \sigma_{i}^{2} \)'s now cancel. In this case, it is customary to define \( \bar{x} \) as \( \frac{\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j}}{\sum_{i} \epsilon_{i} \sigma_{j}} \). Then \( \sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j} = \hat{\mu} \frac{\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j}}{\sum_{i} \epsilon_{i} \sigma_{j}} \).

When the \( \epsilon \)'s are independent and have a common variance, equations (1) and (2) become

\[
\hat{\mu} = \begin{bmatrix}
\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j} & \sum_{i} \epsilon_{i} \sigma_{j} \\
\frac{\sum_{i} \epsilon_{i} \sigma_{i} \sigma_{j}}{n} & \frac{\sum_{i} \epsilon_{i} \sigma_{j}}{n}
\end{bmatrix}
\]
\[ \hat{\beta} = \frac{\begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i Y_i \end{bmatrix}}{(\text{as for } \hat{\mu})} \]

and the \( \sigma^2 \)'s again cancel. Here we define \( \bar{x} \) as \( (\Sigma X_i)/n \). Now \( \Sigma X_i = 0 \) and \( \bar{\mu} = \Sigma X_i/n \).

In the general case, i.e. correlated \( \epsilon_i \)'s, we have assumed that \( M \) was at our disposal. Since \( \hat{\mu} \) and \( \hat{\beta} \) turn out to be linear combinations of the \( Y \)'s, we are able to compute their variances and covariance. In the two special cases, the same is true. However if \( \sigma^2 \) is unknown, it is now possible and relatively easy to estimate \( \sigma^2 \) and, in turn, compute an estimate of the variances and covariance. Known sampling distributions apply to these two special cases.

**Example**

Consider the following set of data

<table>
<thead>
<tr>
<th>i</th>
<th>x</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1</td>
<td>114</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>124</td>
</tr>
<tr>
<td>3</td>
<td>+1</td>
<td>143</td>
</tr>
</tbody>
</table>

with

\[
M = \begin{pmatrix}
7/24 & 1/6 & 1/12 \\
1/6 & 4/6 & 2/6 \\
1/12 & 2/6 & 7/6
\end{pmatrix}
\]
and

\[ M^{-1} = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 2 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}; \quad \Sigma_1^{1j} = 3; \quad \Sigma_2^{2j} = 1/2; \quad \Sigma_3^{3j} = 1/2 \]

Applying equations (1) and (2), we find

\[
\hat{\mu} = \begin{bmatrix} 3(114) + \frac{1}{2}(124) + \frac{1}{2}(143) \\ -4(114) + (1-\frac{1}{2})(124) + 1(143) \\ 3(-1) + 1(\frac{3}{2}) \end{bmatrix} \begin{bmatrix} -1(3) + 1(\frac{3}{2}) \\ -1(-4) + 1(1) \\ 5 \end{bmatrix} = 127.3 \]

\[
\hat{\beta} = \begin{bmatrix} 4 \\ -2.5 \end{bmatrix} \begin{bmatrix} 475.5 \\ 55/4 \end{bmatrix} = 13.4 \]

If there were zero covariances and the variances were $7/24$, $4/6$ and $7/6$, then

\[
M = \begin{pmatrix} 7/24 & 0 & 0 \\ 0 & 4/6 & 0 \\ 0 & 0 & 7/6 \end{pmatrix} \quad \text{and} \quad M^{-1} = \begin{pmatrix} 24/7 & 0 & 0 \\ 0 & 6/4 & 0 \\ 0 & 0 & 6/7 \end{pmatrix} \]

Now

\[
\bar{x} = \frac{\Sigma w_i x_i}{\Sigma w_i} = \frac{\frac{24}{7}(-1) + \frac{6}{4}(0) + \frac{6}{7}(1)}{\frac{24}{7} + \frac{6}{4} + \frac{6}{7}} = -\frac{4}{9} \]
and \( x_1^* = -\frac{5}{9}, \ x_2^* = \frac{4}{9}, \ x_3^* = 1 \frac{4}{9} \).

\[
\hat{\mu} = \frac{\sum w_i y_i}{\sum w_i}
\]

\[
= \frac{\frac{24}{7}(114) + \frac{6}{4}(124) + \frac{6}{7}(143)}{\frac{24}{7} + \frac{6}{4} + \frac{6}{7}}
\]

\[
= 126.8
\]

\[
\hat{\beta} = \frac{\sum w_i x_i y_i}{\sum w_i x_i^2}
\]

\[
= \frac{\frac{24}{7}(-\frac{5}{9})114 + \frac{6}{4}(\frac{4}{9})124 + \frac{6}{7}(\frac{13}{9})143}{\frac{24}{7}(\frac{5}{9})^2 + \frac{6}{4}(\frac{4}{9})^2 + \frac{6}{7}(\frac{13}{9})^2}
\]

\[
= 13.5
\]

If the variances were equal, then we would find

\[
\hat{\mu} = \frac{\sum y}{3}
\]

\[
= 127
\]

\[
\hat{\beta} = \frac{\sum xy}{\sum x^2}
\]

\[
= 14.5
\]

References
