Estimable Functions of β: An Example

		movie 1	movie 2	movie 3
customer	1	4	1	?
	2	?	3	5
	3	?	?	3
	4	3	1	?

Can we guess ratings for customer/movie combinations not in the dataset?

y_{ij} = customer i's rating of movie j

Which movie is best?

$$y_{ij} = \mu + c_i + m_j + \epsilon_{ij}$$

The Linear Model in Vector/Matrix Form

$$
\begin{bmatrix} 4 \\ 1 \\ 3 \\ 5 \\ 3 \\ 3 \\ 1 \end{bmatrix}
=
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix} \mu \\ c_1 \\ c_2 \\ c_3 \\ c_4 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix}
+
\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \end{bmatrix}
$$

$$
\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}
$$

Can We Estimate the Means that Underly the Missing Table Entries?

$$X\beta = \begin{bmatrix} \mu + c_1 + m_1 \\ \mu + c_1 + m_2 \\ \mu + c_2 + m_2 \\ \mu + c_2 + m_3 \\ \mu + c_3 + m_3 \\ \mu + c_4 + m_1 \\ \mu + c_4 + m_2 \end{bmatrix}$$

Can we estimate

$\mu + c_1 + m_3$?

$\mu + c_2 + m_1$?

$\mu + c_3 + m_1$?

$\mu + c_3 + m_2$?

$\mu + c_4 + m_3$?

$m_1 - m_2$ is estimable because

$$[1, -1, 0, 0, 0.0.0] \, X\beta = m_1 - m_2.$$

Can We Estimate the Means that Underly the Missing Table Entries?

$$X\boldsymbol{\beta} = \begin{bmatrix} \mu + c_1 + m_1 \\ \mu + c_1 + m_2 \\ \mu + c_2 + m_2 \\ \mu + c_2 + m_3 \\ \mu + c_3 + m_3 \\ \mu + c_4 + m_1 \\ \mu + c_4 + m_2 \end{bmatrix}$$

Can we estimate

$\mu + c_1 + m_3$?
$\mu + c_2 + m_1$?
$\mu + c_3 + m_1$?
$\mu + c_3 + m_2$?
$\mu + c_4 + m_3$?

Likewise, $m_2 - m_3$ is estimable because

$$[0, 0, 1, -1, 0, 0, 0]\, X\boldsymbol{\beta} = m_2 - m_3.$$

We can estimate all pairwise differences between movie effects.

Because $m_1 - m_2$ and $m_2 - m_3$ are estimable, we can also estimate

$$(m_1 - m_2) + (m_2 - m_3) = m_1 - m_3.$$

This follows because any linear combination of estimable functions is also estimable.

We can estimate the mean underlying the rating for any combination of customer and movie.

It follows that any linear combination of the form

$$\mu + c_i + m_j$$

can be estimated $\forall\ i = 1, 2, 3, 4$ and $j = 1, 2, 3$ because

$$\mu + c_i + m_j = (\mu + c_i + m_{j'}) + (m_j - m_{j'})$$

$$\forall\ i = 1, 2, 3, 4 \text{ and } j, j' = 1, 2, 3.$$

Movie *lsmeans*

If our goal is to compare movies to see which is most highly rated, we can accomplish that by estimating the pairwise differences between movie effects.

However, if we want to retain information about the mean rating rather than the difference between mean ratings, it is natural to consider estimating the average (across *all* customers) of the mean rating for each movie.

Movie *lsmeans*

This average for the jth movie is

$$\frac{1}{4}\sum_{i=1}^{4}(\mu + c_i + m_j) = \mu + \frac{1}{4}\sum_{i=1}^{4} c_i + m_j = \mu + \bar{c}_{\cdot} + m_j.$$

This average is estimable for each movie in our example because it is a linear combination of estimable functions.

SAS calls estimates of $\mu + \bar{c}_{\cdot} + m_j \ \forall\, j = 1, 2, 3$ movie *lsmeans*, where *ls* presumably denotes least squares to indicate that these estimates are based on OLS estimators.

Suppose we consider a different model.

		movie 1	movie 2	movie 3
customer	1	4	1	?
	2	?	3	5
	3	?	?	3
	4	3	1	?

Can we guess ratings for customer/movie combinations not in the dataset?

y_{ij} = customer i's rating of movie j

Which movie is best?

$$y_{ij} = \mu_{ij} + \epsilon_{ij}$$

The Linear Model in Vector/Matrix Form

$$
\begin{bmatrix} 4 \\ 1 \\ 3 \\ 5 \\ 3 \\ 3 \\ 1 \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{41} \\ \mu_{42} \\ \mu_{43} \end{bmatrix}
+
\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{33} \\ \epsilon_{41} \\ \epsilon_{42} \end{bmatrix}
$$

Can we estimate the means that underly the missing table entries? No.

$$X\beta = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{22} \\ \mu_{23} \\ \mu_{33} \\ \mu_{41} \\ \mu_{42} \end{bmatrix}$$

Can we estimate

μ_{13}?
μ_{21}?
μ_{31}?
μ_{32}?
μ_{43}?

None of the means underlying missing table entries are estimable under this cell-means model.