University of Waterloo W. H. Cherry

RESPONSE MODELS IN STAT 231: Least Squares Estimating of Parameters

This Statistical Highlight #73 summarizes the least squares estimating process, using both calculus and linear algebra, for the parameters of the common STAT 231 response models; a more evocative name than *least squares* would be *minimum sum of squared residuals*. Matters omitted from some derivations in this Highlight #73, which the reader should supply, include:

- second derivative demonstrations of *minima* (Step 6);
- for Model 2, extension from the case $\mathrm{i} = 1, 2$ to the case $\mathrm{i} = 1, 2, \ldots, \mathrm{q}$;
- definitions of vector notation (Step 3);
- vector diagrams (Step 4).

A **general** form for a response model – a response variate expressed as a **structural component** plus **noise** – is shown in two versions at the right;

- in the upper version, the structural component μ_j is a linear combination of q (unknown) parameters θ_k and (known) explanatory variate values x_{jk};
- in the lower version, the explanatory variate values form the $\mathrm{n} \times \mathrm{q}$ matrix $\mathbf{X}$ and the parameters are the $\mathrm{q} \times 1$ column vector $\underset{\sim}{\Theta}$.

$Y_j = \mu_j + R_j$, $j = 1, 2, \ldots, \mathrm{n}$,
$\underset{\sim}{\boldsymbol{Y}} = \mathbf{X}\underset{\sim}{\Theta} + \underset{\sim}{\boldsymbol{R}}$ $R_j \sim G(0, \sigma)$, independent, EPS;
$\mu_j = \theta_1 x_{j1} + \ldots + \theta_k x_{jk} + \ldots + \theta_q x_{jq}$
[explanatory variates $X_1, \ldots, X_k, \ldots, X_q$]

– $\underset{\sim}{\boldsymbol{Y}}$ and $\underset{\sim}{\boldsymbol{R}}$ are $\mathrm{n} \times 1$ column vectors with $\mathrm{n} \gg \mathrm{q}$; *i.e.*, there are *many* more observations than parameters to be estimated.

The response models given below are special cases of this general form; their numbering (1, 1a, 1b, 2, 3, 3a, 4, 4a) is only for convenience in this Highlight #73 (and in #72) and does *not* carry over to the Course Notes; 'EPS' denotes *equiprobable selecting*.

Model 1: $Y_j = \mu + R_j$, $j = 1, 2, \ldots, \mathrm{n}$; $R_j \sim G(0, \sigma)$; independent, EPS.

Step 1: $Y_j = \mu + R_j$, $j = 1, 2, \ldots, \mathrm{n}$.

Step 2: $y_j = \mu + r_j$, $j = 1, 2, \ldots, \mathrm{n}$.

Step 3: $r_j = y_j - \mu$, $j = 1, 2, \ldots, \mathrm{n}$.

Step 4: $g(\mu) = \sum_{j=1}^{\mathrm{n}} r_j^2 = \sum_{j=1}^{\mathrm{n}} [y_j - \mu]^2$.

Step 5: $\frac{dg}{d\mu} = -2\sum_{j=1}^{\mathrm{n}} [y_j - \mu]$,

which is zero when: $\sum_{j=1}^{\mathrm{n}} [y_j - \hat{\mu}] = 0$;

i.e., $\sum_{j=1}^{\mathrm{n}} y_j - \mathrm{n}\hat{\mu} = 0$,

or: $\hat{\mu} = \frac{1}{\mathrm{n}}\sum_{j=1}^{\mathrm{n}} y_j = \bar{y}$, the **estimate** of μ.

Step 6: $\frac{d^2 g}{d\mu^2} = -2\sum_{j=1}^{\mathrm{n}} [-1] = 2\mathrm{n}$,

which is **positive**, as required for a **minimum**.

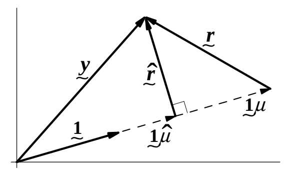
Also: $\hat{\sigma} = \sqrt{\frac{\sum_{j=1}^{\mathrm{n}} \hat{r}_j^2}{\mathrm{n} - \mathrm{q}}} = \sqrt{\frac{\sum_{j=1}^{\mathrm{n}} [y_j - \hat{\mu}]^2}{\mathrm{n} - 1}} = \sqrt{\frac{\sum_{j=1}^{\mathrm{n}} [y_j - \bar{y}]^2}{\mathrm{n} - 1}}$.

Step 1: $Y_j = \mu + R_j$, $j = 1, 2, \ldots, \mathrm{n}$.

Step 2: $y_j = \mu + r_j$, $j = 1, 2, \ldots, \mathrm{n}$.

Step 3: $\underset{\sim}{\boldsymbol{y}} = \underset{\sim}{\mathbf{1}}\mu + \underset{\sim}{\boldsymbol{r}}$, where: $\underset{\sim}{\boldsymbol{y}} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_\mathrm{n} \end{pmatrix}$, $\underset{\sim}{\mathbf{1}} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, $\underset{\sim}{\boldsymbol{r}} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_\mathrm{n} \end{pmatrix}$.

Step 4: The least squares estimate of μ is determined by the vector $\underset{\sim}{\hat{\boldsymbol{r}}}$, the vector $\underset{\sim}{\boldsymbol{r}}$ of **minimum** length; this occurs when $\underset{\sim}{\boldsymbol{r}}$ is **perpendicular** to the vector $\underset{\sim}{\mathbf{1}}$, so we set the dot (or inner) product $\underset{\sim}{\mathbf{1}} \cdot \underset{\sim}{\hat{\boldsymbol{r}}}$ to zero.



Step 5: $\underset{\sim}{\mathbf{1}}^{\mathrm{T}}\underset{\sim}{\hat{\boldsymbol{r}}} = 0$, *i.e.*, $\underset{\sim}{\mathbf{1}}^{\mathrm{T}}(\underset{\sim}{\boldsymbol{y}} - \underset{\sim}{\mathbf{1}}\hat{\mu}) = 0$,

i.e., $\underset{\sim}{\mathbf{1}}^{\mathrm{T}}\underset{\sim}{\boldsymbol{y}} - \underset{\sim}{\mathbf{1}}^{\mathrm{T}}\underset{\sim}{\mathbf{1}}\hat{\mu} = 0$,

or: $\hat{\mu} = (\underset{\sim}{\mathbf{1}}^{\mathrm{T}}\underset{\sim}{\mathbf{1}})^{-1}(\underset{\sim}{\mathbf{1}}^{\mathrm{T}}\underset{\sim}{\boldsymbol{y}}) = \mathrm{n}^{-1}\sum_{j=1}^{\mathrm{n}} y_j = \bar{y}$, the **estimate** of μ.

Also: $\hat{\sigma} = \sqrt{\frac{\underset{\sim}{\hat{\boldsymbol{r}}}^{\mathrm{T}}\underset{\sim}{\hat{\boldsymbol{r}}}}{\mathrm{n} - \mathrm{q}}} = \sqrt{\frac{(\underset{\sim}{\boldsymbol{y}} - \underset{\sim}{\mathbf{1}}\hat{\mu})^{\mathrm{T}}(\underset{\sim}{\boldsymbol{y}} - \underset{\sim}{\mathbf{1}}\hat{\mu})}{\mathrm{n} - 1}} = \sqrt{\frac{\sum_{j=1}^{\mathrm{n}} [y_j - \bar{y}]^2}{\mathrm{n} - 1}}$.

Model 1a: ${}_{\mathrm{M}}Y_j = \tau + \delta + R_j$, $j = 1, 2, \ldots, \mathrm{m}$; $R_j \sim G(0, \sigma)$; independent, EPS.

If we take the response variate as $Y_j = {}_{\mathrm{M}}Y_j - \tau$, the **difference** between the **measured** value and the **true** value:

Model 1b: $Y_j = \delta + R_j$, $j = 1, 2, \ldots, \mathrm{m}$; $R_j \sim G(0, \sigma)$; independent, EPS;

thus, Model 1a rewritten as Model 1b is equivalent to Model 1, except the structural component is δ instead of μ. Least squares estimating of δ and σ in Model 1b is therefore equivalent to that for μ and σ in Model 1 above.

Model 3: $Y_{ij} = \mu_\mathrm{i} + \gamma_j + R_{ij}$, $\mathrm{i} = 1, 2$, $j = 1, 2, \ldots, \mathrm{n}$; $R_{ij} \sim G(0, \sigma)$; independent, EPS.

Using this model which is appropriate for a **blocked** Plan, we work with the **differences** of the responses of the pairs of units within the n blocks:

$Y_{1j} - Y_{2j} = \mu_1 - \mu_2 + \gamma_j - \gamma_j + R_{1j} - R_{2j}$, which becomes:

Model 3a: $Y_j = \mu_\mathrm{d} + R_j$, $j = 1, 2, \ldots, \mathrm{n}$; $R_j \sim G(0, \sigma_\mathrm{d})$; independent, EPS;

thus, Model 3a for the **differences** is equivalent to Model 1, but with parameters μ_d and σ_d instead of μ and σ. Least squares estimating of μ_d and σ_d in Model 3a is therefore equivalent to that for μ and σ in Model 1 above.

Model 4a: $Y_j = \beta x_j + R_j$, $j = 1, 2, \ldots, \mathrm{n}$; $R_j \sim G(0, \sigma)$; independent, EPS;

establish the results at the right using **both** calculus and linear algebra.

$$\hat{\beta} = \frac{\sum_{j=1}^{\mathrm{n}} x_j y_j}{\sum_{j=1}^{\mathrm{n}} x_j^2} \qquad \hat{\sigma} = \sqrt{\frac{\sum_{j=1}^{\mathrm{n}} y_j^2 - \hat{\beta}\sum_{j=1}^{\mathrm{n}} x_j y_j}{\mathrm{n} - 1}}$$

(*continued overleaf*)

University of Waterloo W. H. Cherry

Model 2: $Y_{ij} = \mu_i + R_{ij}$, $i = 1, 2$, $j = 1, 2, ..., n_i$; $R_{ij} \sim G(0, \sigma)$; independent, EPS. [think of $n_1 + n_2$ as n]

Step 1: $Y_{ij} = \mu_i + R_{ij}$, $i = 1, 2$, $j = 1, 2, ..., n_i$.

Step 2: $y_{ij} = \mu_i + r_{ij}$, $i = 1, 2$, $j = 1, 2, ..., n_i$.

Step 3: $r_{ij} = y_{ij} - \mu_i$, $i = 1, 2$, $j = 1, 2, ..., n_i$.

Step 4: $g(\mu_1, \mu_2) = \sum_{i=1}^{2}\sum_{j=1}^{n_i} r_{ij}^2 = \sum_{j=1}^{n_1}[y_{1j} - \mu_1]^2 + \sum_{j=1}^{n_2}[y_{2j} - \mu_2]^2.$

Step 5: $\dfrac{\partial g}{\partial \mu_1} = -2\sum_{j=1}^{n_1}[y_{1j} - \mu_1] + 0,$

which is zero when: $\sum_{j=1}^{n_1}[y_{1j} - \hat{\mu}_1] = 0;$

i.e., $\sum_{j=1}^{n_1} y_{1j} - n_1\hat{\mu}_1 = 0,$

or: $\hat{\mu}_1 = \dfrac{1}{n_1}\sum_{j=1}^{n_1} y_{1j} = \bar{y}_1$, the **estimate** of μ_1.

Similarly: $\hat{\mu}_2 = \dfrac{1}{n_2}\sum_{j=1}^{n_2} y_{2j} = \bar{y}_2$, the **estimate** of μ_2.

Step 6: Show the two estimates are **minima** using the four second partial derivatives.

Also: $$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{2}\sum_{j=1}^{n_i}\hat{r}_{ij}^2}{n - q}} = \sqrt{\frac{\sum_{j=1}^{n_1}[y_{1j} - \hat{\mu}_1]^2 + \sum_{j=1}^{n_2}[y_{2j} - \hat{\mu}_2]^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}}.$$

Step 1: $Y_{ij} = \mu_i + R_{ij}$, $i = 1, 2$, $j = 1, 2, ..., n_i$.

Step 2: $y_{ij} = \mu_i + r_{ij}$, $i = 1, 2$, $j = 1, 2, ..., n_i$.

Step 3: $\mathbf{y} = \mathbf{a}\mu_1 + \mathbf{b}\mu_2 + \mathbf{r}$,
where: $\mathbf{y}$, $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{r}$ are $n_1 + n_2 \times 1$ column vectors,
with $\mathbf{a}^T\mathbf{a} = n_1$, $\mathbf{b}^T\mathbf{b} = n_2$, $\mathbf{a}^T\mathbf{b} = \mathbf{b}^T\mathbf{a} = 0$.

Step 4: The least squares estimates of μ_1 and μ_2 are determined by the vector $\hat{\mathbf{r}}$, the vector $\mathbf{r}$ of **minimum** length; this occurs when $\mathbf{r}$ is **perpendicular** to the plane which is *span*($\mathbf{a}$, $\mathbf{b}$), so we set the dot (or inner) products $\mathbf{a}\cdot\hat{\mathbf{r}}$ and $\mathbf{b}\cdot\hat{\mathbf{r}}$ to zero. (Illustrate this matter with a **diagram**.)

Step 5: $\mathbf{a}^T\hat{\mathbf{r}} = 0$, *i.e.*, $\mathbf{a}^T(\mathbf{y} - \mathbf{a}\hat{\mu}_1 - \mathbf{b}\hat{\mu}_2) = 0,$

i.e., $\mathbf{a}^T\mathbf{y} - \mathbf{a}^T\mathbf{a}\hat{\mu}_1 = 0$ [because: $\mathbf{a}^T\mathbf{b} = 0$],

or: $\hat{\mu}_1 = (\mathbf{a}^T\mathbf{a})^{-1}(\mathbf{a}^T\mathbf{y}) = n_1^{-1}\sum_{j=1}^{n_1} y_{1j} = \bar{y}_1$, the **estimate** of μ_1.

Similarly: $\hat{\mu}_2 = 3s11(\mathbf{b}^T\mathbf{b})^{-1}(\mathbf{b}^T\mathbf{y}) = n_2^{-1}\sum_{j=1}^{n_2} y_{2j} = \bar{y}_2$, the **estimate** of μ_2.

Also: $$\hat{\sigma} = \sqrt{\frac{\hat{\mathbf{r}}^T\hat{\mathbf{r}}}{n - q}} = \sqrt{\frac{(\mathbf{y} - \mathbf{a}\hat{\mu}_1 - \mathbf{b}\hat{\mu}_2)^T(\mathbf{y} - \mathbf{a}\hat{\mu}_1 - \mathbf{b}\hat{\mu}_2)}{n_1 + n_2 - 2}} = \text{etc.}$$

Model 4: $Y_j = \alpha + \beta_1(x_j - \bar{x}) + R_j$, $j = 1, 2, ..., n$; $R_j \sim G(0, \sigma)$; independent, EPS.

Step 1: $Y_j = \alpha + \beta_1(x_j - \bar{x}) + R_j$, $j = 1, 2, ..., n$.

Step 2: $y_j = \alpha + \beta_1(x_j - \bar{x}) + r_j$, $j = 1, 2, ..., n$.

Step 3: $r_j = y_j - \alpha - \beta_1(x_j - \bar{x})$, $j = 1, 2, ..., n$.

Step 4: $g(\alpha, \beta_1) = \sum_{j=1}^{n} r_j^2 = \sum_{j=1}^{n}[y_j - \alpha - \beta_1(x_j - \bar{x})]^2.$

Step 5: $\dfrac{\partial g}{\partial \alpha} = -2\sum_{j=1}^{n}[y_j - \alpha - \beta_1(x_j - \bar{x})],$

which is zero when:

$\sum_{j=1}^{n}[y_j - \hat{\alpha} - \hat{\beta}_1(x_j - \bar{x})] = 0;$

i.e., $\sum_{j=1}^{n} y_j - n\hat{\alpha} - \hat{\beta}_1\sum_{j=1}^{n}(x_j - \bar{x}) = 0,$

i.e., $\sum_{j=1}^{n} y_j - n\hat{\alpha} = 0$ [because: $\sum_{j=1}^{n}(x_j - \bar{x}) \equiv 0$],

or: $\hat{\alpha} = \dfrac{1}{n}\sum_{j=1}^{n} y_j = \bar{y}$, the **estimate** of α.

Also: $\dfrac{\partial g}{\partial \beta_1} = -2\sum_{j=1}^{n}[y_j - \alpha - \beta_1(x_j - \bar{x})](x_j - \bar{x}),$

which is zero when:

$\sum_{j=1}^{n}[y_j - \hat{\alpha} - \hat{\beta}_1(x_j - \bar{x})](x_j - \bar{x}) = 0;$

i.e., $\sum_{j=1}^{n} y_j(x_j - \bar{x}) - \hat{\alpha}\sum_{j=1}^{n}(x_j - \bar{x}) - \hat{\beta}_1\sum_{j=1}^{n}(x_j - \bar{x})^2 = 0,$

i.e., $\sum_{j=1}^{n} y_j(x_j - \bar{x}) - \hat{\beta}_1\sum_{j=1}^{n}(x_j - \bar{x})^2 = 0$ [because: $\sum_{j=1}^{n}(x_j - \bar{x}) \equiv 0$],

or: $\hat{\beta}_1 = \dfrac{\sum_{j=1}^{n} y_j(x_j - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \equiv \dfrac{\sum_{j=1}^{n} y_j(x_j - \bar{x})}{SS_x}$, the **estimate** of β_1.

Step 6: Show the two estimates are **minima** using the four second partial derivatives.

Step 1: $Y_j = \alpha + \beta_1(x_j - \bar{x}) + R_j$, $j = 1, 2, ..., n$.

Step 2: $y_j = \alpha + \beta_1(x_j - \bar{x}) + r_j$, $j = 1, 2, ..., n$.

Step 3: $\mathbf{y} = \mathbf{1}\alpha + \mathbf{x}\beta_1 + \mathbf{r}$, where $\mathbf{x} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$

Step 4: The least squares estimates of α and β_1 are determined by the vector $\hat{\mathbf{r}}$, the vector $\mathbf{r}$ of **minimum** length; this occurs when $\mathbf{r}$ is **perpendicular** to the plane which is *span*($\mathbf{1}$, $\mathbf{x}$), so we set the dot (or inner) products $\mathbf{1}\cdot\hat{\mathbf{r}}$ and $\mathbf{x}\cdot\hat{\mathbf{r}}$ to zero. (Illustrate this matter with a **diagram**.)

Step 5: $\mathbf{1}^T\hat{\mathbf{r}} = 0$, *i.e.*, $\mathbf{1}^T(\mathbf{y} - \mathbf{1}\hat{\alpha} - \mathbf{x}\hat{\beta}_1) = 0,$

i.e., $\mathbf{1}^T\mathbf{y} - \mathbf{1}^T\mathbf{1}\hat{\alpha} - \mathbf{1}^T\mathbf{x}\hat{\beta}_1 = 0,$

i.e., $\mathbf{1}^T\mathbf{y} - \mathbf{1}^T\mathbf{1}\hat{\alpha} = 0$ [because: $\mathbf{1}^T\mathbf{x} = \sum_{j=1}^{n}(x_j - \bar{x}) \equiv 0$],

or: $\hat{\alpha} = (\mathbf{1}^T\mathbf{1})^{-1}(\mathbf{1}^T\mathbf{y}) = n^{-1}\sum_{j=1}^{n} y_j = \bar{y}$, the **estimate** of α.

Also: $\mathbf{x}^T\hat{\mathbf{r}} = 0$, *i.e.*, $\mathbf{x}^T(\mathbf{y} - \mathbf{1}\hat{\alpha} - \mathbf{x}\hat{\beta}_1) = 0,$

i.e., $\mathbf{x}^T\mathbf{y} - \mathbf{x}^T\mathbf{1}\hat{\alpha} - \mathbf{x}^T\mathbf{x}\hat{\beta}_1 = 0,$

i.e., $\mathbf{x}^T\mathbf{y} - \mathbf{x}^T\mathbf{x}\hat{\beta}_1 = 0$ [because: $\mathbf{x}^T\mathbf{1} = \mathbf{1}^T\mathbf{x} \equiv 0$],

or: $\hat{\beta}_1 = (\mathbf{x}^T\mathbf{x})^{-1}(\mathbf{x}^T\mathbf{y}) = [\sum_{j=1}^{n}(x_j - \bar{x})^2]^{-1}[\sum_{j=1}^{n}(x_j - \bar{x})y_j]$

$= \dfrac{\sum_{j=1}^{n} y_j(x_j - \bar{x})}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \equiv \dfrac{\sum_{j=1}^{n} y_j(x_j - \bar{x})}{SS_x}$, the **estimate** of β_1.

Also: $$\hat{\sigma} = \sqrt{\frac{\hat{\mathbf{r}}^T\hat{\mathbf{r}}}{n - q}} = \sqrt{\frac{(\mathbf{y} - \mathbf{1}\hat{\alpha} - \mathbf{x}\hat{\beta}_1)^T(\mathbf{y} - \mathbf{1}\hat{\alpha} - \mathbf{x}\hat{\beta}_1)}{n - 2}} = \text{etc.}$$

Also: $$\hat{\sigma} = \sqrt{\frac{\sum_{j=1}^{n}\hat{r}_j^2}{n - q}} = \sqrt{\frac{\sum_{j=1}^{n}[y_j - \bar{y} - \hat{\beta}_1(x_j - \bar{x})]^2}{n - 2}} = \sqrt{\frac{SS_y - \hat{\beta}_1 SS_{xy}}{n - 2}};$$ $SS_y = \sum_{j=1}^{n}(y_j - \bar{y})^2$, $SS_{xy} = \sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y}) \equiv \sum_{j=1}^{n}(x_j - \bar{x})y_j$.