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 i = 1, 2 to the case i = 1, 2, ..., 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 $n \times q$ matrix $\mathbf{X}$ and the parameters are the $q \times 1$ column vector $\underset{\sim}{\Theta}$.

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

– $\underset{\sim}{Y}$ and $\underset{\sim}{R}$ are $n \times 1$ column vectors with $n \gg q$; *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: $Y_j = \mu + R_j, \quad j = 1, 2, ..., n; \quad R_j \sim G(0, \sigma);$ independent, EPS.

Step 1: $Y_j = \mu + R_j, \quad j = 1, 2, ..., n.$

Step 2: $y_j = \mu + r_j, \quad j = 1, 2, ..., n.$

Step 3: $r_j = y_j - \mu, \quad j = 1, 2, ..., n.$

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

Step 5: $\frac{dg}{d\mu} = -2\sum_{j=1}^{n}[y_j - \mu],$
which is zero when: $\sum_{j=1}^{n}[y_j - \hat{\mu}] = 0;$
i.e., $\sum_{j=1}^{n} y_j - n\hat{\mu} = 0,$
or: $\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} y_j = \overline{y},$ the **estimate** of μ.

Step 6: $\frac{d^2 g}{d\mu^2} = -2\sum_{j=1}^{n}[-1] = 2n,$
which is **positive**, as required for a **minimum**.

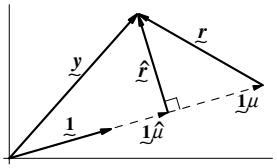
Also: $\hat{\sigma} = \sqrt{\frac{\sum_{j=1}^{n} \hat{r}_j^2}{n-q}} = \sqrt{\frac{\sum_{j=1}^{n}[y_j - \hat{\mu}]^2}{n-1}} = \sqrt{\frac{\sum_{j=1}^{n}[y_j - \overline{y}]^2}{n-1}}.$

Step 1: $Y_j = \mu + R_j, \quad j = 1, 2, ..., n.$

Step 2: $y_j = \mu + r_j, \quad j = 1, 2, ..., n.$

Step 3: $\underset{\sim}{y} = \underset{\sim}{1}\mu + \underset{\sim}{r},$ where: $\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \underset{\sim}{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \underset{\sim}{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}.$

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



Step 5: $\underset{\sim}{1}^T\underset{\sim}{\hat{r}} = 0, \quad$ *i.e.*, $\underset{\sim}{1}^T(\underset{\sim}{y} - \underset{\sim}{1}\hat{\mu}) = 0,$
i.e., $\underset{\sim}{1}^T\underset{\sim}{y} - \underset{\sim}{1}^T\underset{\sim}{1}\hat{\mu} = 0,$
or: $\hat{\mu} = (\underset{\sim}{1}^T\underset{\sim}{1})^{-1}(\underset{\sim}{1}^T\underset{\sim}{y}) = n^{-1}\sum_{j=1}^{n} y_j = \overline{y},$ the **estimate** of μ.

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

Model 1a: ${}_M Y_j = \tau + \delta + R_j, \quad j = 1, 2, ..., m; \quad R_j \sim G(0, \sigma);$ independent, EPS.
If we take the response variate as $Y_j = {}_M Y_j - \tau$, the **difference** between the **measured** value and the **true** value:

Model 1b: $Y_j = \delta + R_j, \quad j = 1, 2, ..., m; \quad R_j \sim G(0, \sigma);$ 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: $Y_{ij} = \mu_i + \gamma_j + R_{ij}, \quad i = 1, 2, \quad j = 1, 2, ..., n; \quad R_{ij} \sim G(0, \sigma);$ 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: $Y_j = \mu_d + R_j, \quad j = 1, 2, ..., n; \quad R_j \sim G(0, \sigma_d);$ 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: $Y_j = \beta x_j + R_j, \quad j = 1, 2, ..., n; \quad R_j \sim G(0, \sigma);$ independent, EPS;
establish the results at the right using **both** calculus and linear algebra.

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

(*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: $\frac{\partial g}{\partial \mu_1} = -2\sum_{j=1}^{n_1}[y_{1j} - \mu_1] + 0$,

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

Similarly: $\hat{\mu}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} y_{2j} = \overline{y}_2$, 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: $\underset{\sim}{Y}_{ij} = \mu_i + \underset{\sim}{R}_{ij}$, $i = 1, 2$, $j = 1, 2, ..., n_i$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which is zero when:

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

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

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

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

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

which is zero when:

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

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

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

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

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

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

Step 1: $\underset{\sim}{Y}_j = \alpha + \beta_1(x_j - \overline{x}) + \underset{\sim}{R}_j$, $j = 1, 2, ..., n$.

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

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

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

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

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

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

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

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

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

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

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

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

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