

MULTIVARIATE LINEAR REGRESSION MODEL

Consider a data set where we have n observations (cases) on m outcomes (dependent variables) with p explanatory variables (independent variables).

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$n \times m$ $n \times (p+1)$ $(p+1) \times m$ $n \times m$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X} \boldsymbol{\beta}$$

$$\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) \quad \leftarrow \text{error SSCP}$$

The OLS task in this model is to minimize $\text{trace}[(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})]$

Also, OLS estimates of $\hat{\boldsymbol{\beta}}$ minimize the generalized variance: $|(\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})|$

$$\hat{\mathbf{Y}} = \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}$$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = [\mathbf{I} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{Y}$$

$$\mathbf{Y}' \mathbf{Y} = (\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}})' (\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}) = \hat{\mathbf{Y}}' \hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} + \mathbf{0} + \mathbf{0} = \text{SSR} + \text{SSE} = \text{SST}$$