Additional Sum of Squares Principle

We're used to testing hypotheses of the form:

$$H_0: \beta_j = 0 \quad \text{vs.} \quad H_A: \beta_j \neq 0$$

But what about testing a hypothesis about β's simultaneously?

$$H_0: \beta_3 = \beta_4 = \beta_5 = 0 \quad \text{vs.} \quad H_A: \beta_j \neq 0 \text{ for some } j = 3,4,5$$

Hypothesis tests like these are useful for testing:
(i) Nested models
(ii) Equality of means at different levels of a categorical variable.
(iii) The test of overall significance in a linear regression

The general approach to simultaneously testing a system of hypotheses like this is to define a constraint matrix A such that $A\vec{\beta} = \vec{0}$ corresponds to the statement of the null hypothesis H_0.

Example: Consider the full model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$$

Under the null hypothesis $H_0: \beta_3 = \beta_4 = \beta_5 = 0$ the reduced model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Note, this null hypothesis can be stated as:

$$H_0: A\vec{\beta} = \vec{0} \qquad \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

where $\underbrace{A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{3 \times 6}$

For a test of $A\vec{\beta} = \vec{0}$ we compare the full model to the reduced model to decide whether they're significantly different.

So how do we do this?

Consider the model

$$\vec{y} = \beta_0 \vec{1} + \beta_1 \vec{x}_1 + \beta_2 \vec{x}_2 + \cdots + \beta_p \vec{x}_p + \vec{\varepsilon}$$

What we want to do is test a hypothesis that a certain linear combination(s) of the βs is equal to zero:

$$A\vec{\beta} = \vec{0}$$

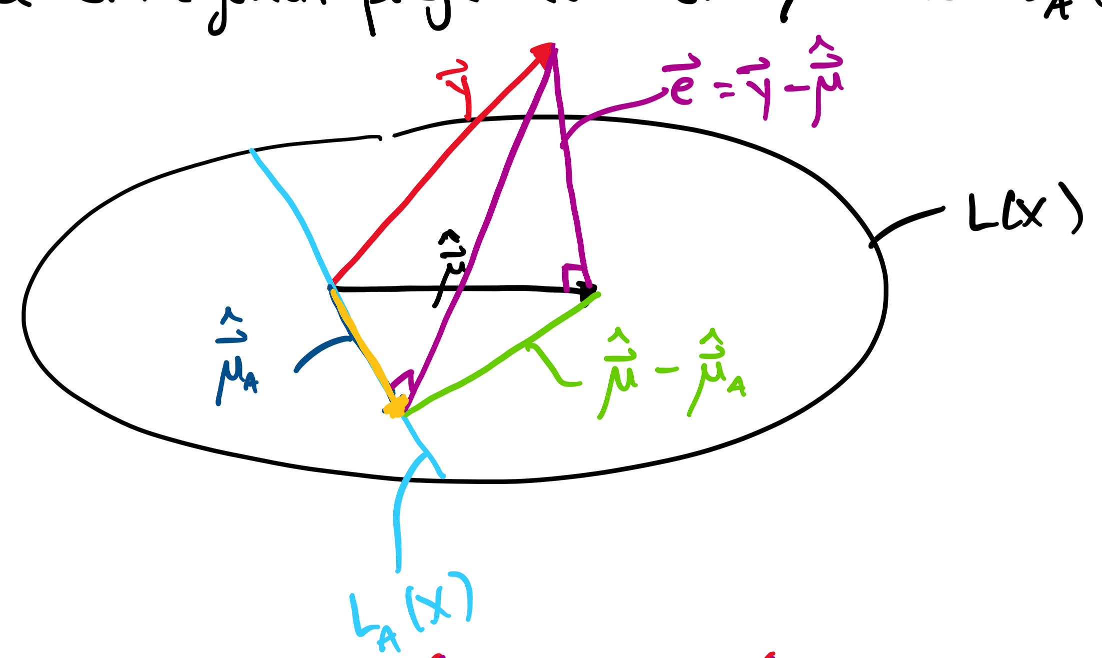
where A is an $\ell \times (p+1)$ matrix of rank ℓ (# restrictions). The test we do relies on the additional sum of squares principle. To understand this we approach the problem geometrically:

Let $L(X)$ be the span of the columns of $X = [\vec{1}\ \vec{x}_1\ \vec{x}_2 \cdots \vec{x}_p]$. Now let $L_A(X)$ be the span of the vectors $\vec{1}, \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_p$ with coefficients satisfying $A\vec{\beta} = \vec{0}$:

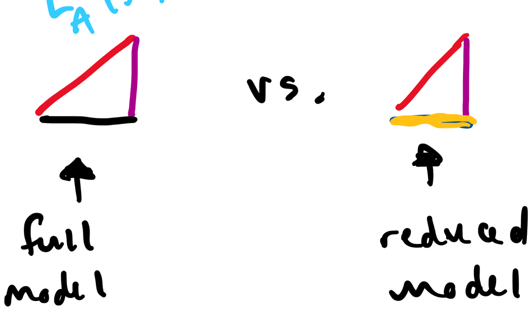
$$L_A(X) = \left\{ \beta_0 \vec{1} + \beta_1 \vec{x}_1 + \cdots + \beta_p \vec{x}_p \mid A\vec{\beta} = \vec{0} \right\}$$

Clearly $L_A(X)$ is a subspace of $L(X)$ since every linear combination in $L_A(X)$ is also in $L(X)$.

Let $\hat{\vec{\mu}}$ be the orthogonal projection of $\vec{y}$ onto $L(X)$ and let $\hat{\vec{\mu}}_A$ be the orthogonal projection of $\vec{y}$ onto $L_A(X)$:



We have two triangles:



If $H_0: A\vec{\beta} = \vec{0}$ is true then these two triangles should be exactly the same because $L(X) = L_A(X)$. We evaluate the difference between the full and reduced models by the magnitude of the $\hat{\vec{\mu}} - \hat{\vec{\mu}}_A$ vector. Small values of this magnitude give evidence in favor of H_0 and large values give evidence against H_0.

The test statistic is based on the squared distance:

$$\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|_2^2 = (\hat{\vec{\mu}} - \hat{\vec{\mu}}_A)^T (\hat{\vec{\mu}} - \hat{\vec{\mu}}_A)$$

L_2-norm

and the following properties:

(i) $$U = \frac{\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|^2}{\sigma^2} \sim \chi^2_{(\ell)}$$

$$\hat{\sigma}^2 = \frac{\vec{e}^T \vec{e}}{n-p-1}$$

(ii) $$V = \frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} = \frac{\vec{e}^T \vec{e}}{\sigma^2} \sim \chi^2_{(n-p-1)}$$

(iii) U and V are independent.

(iv) $$\frac{U/\ell}{V/(n-p-1)} \sim F_{(\ell, n-p-1)}$$

Putting all of this together, we define the test statistic (which tests $H_0: A\vec{\beta} = \vec{0}$) as:

$$t = \left(\frac{\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|^2}{\sigma^2} \Big/ \ell \right) \div \left(\frac{\vec{e}^T \vec{e}}{\sigma^2} \Big/ (n-p-1) \right)$$

$$= \frac{\|\hat{\vec{\mu}} - \hat{\vec{\mu}}_A\|^2 / \ell}{SSE/(n-p-1)} \sim F_{(\ell, n-p-1)}$$

null distribution

MSE

Note: Large values of t provide evidence against H_0, while small values (close to 0) provide evidence in favour of H_0. We formalize this by calculating a p-value:

$$\text{p-value} = P(T \geq t) \quad \text{where} \quad T \sim F_{(\ell, n-p-1)}$$