Regression Analysis Tutoring1

Seung Bong Jung

Seoul National University

November 1, 2021

Linear Regression Model

Linear Regression Model

The model: $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$, where Y is a real-valued response, $(x_1, \ldots, x_p)$ is a set of p predictors assumed to be nonrandom, the error ϵ is a random variable with mean $\mathrm{E}(\epsilon) = 0$ and finite variance $\sigma^2 = \mathrm{Var}(\epsilon)$

- Assume that we observe the responses Y_i at the preselected predictor values $x_{i1}, \ldots x_{ip}$, respectively, for $1 \leq i \leq n$, such that $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$, where $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$
- Then the parameter is $\theta = (\boldsymbol{\beta}, \sigma^2)$ and the likelihood is given by

$$L(\theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^t$

Linear Regression Model

- Thus, if we are to get MLE of $\boldsymbol{\beta}$, we have to minimize $||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$ with respect to β, where $\mathbf{Y} = (Y_1, \ldots, Y_n)^t, \mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^t$, and $\mathbf{X} = (\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p)$. We call such minimizer as least squared estimator denoted by $\hat{\boldsymbol{\beta}}$.

Simple Linear Regression

Consider the case when $p = 1$.

- $(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\arg\min} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$.
- $\hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{x}$ is the projection of $\mathbf{Y} = (Y_1, \ldots, Y_n)^t$ onto the linear subspace in $\mathbb{R}^n$ spanned by $\mathbf{1} = (1, \ldots, 1)^t$ and $\mathbf{x} = (x_1, \ldots, x_n)^t$.

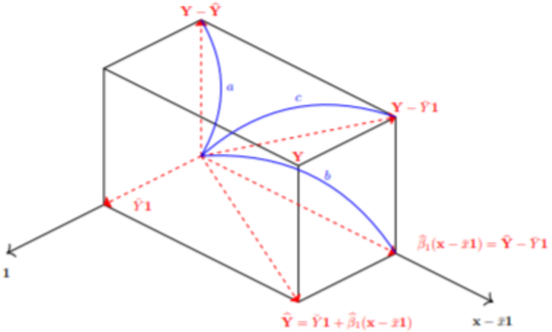
$$\hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{x} = \Pi(\mathbf{Y} | \mathcal{C}_{\mathbf{1},\mathbf{x}})$$

  , where $\mathcal{C}_{\mathbf{1},\mathbf{x}} = \{\beta_0 \mathbf{1} + \beta_1 \mathbf{x} : \beta_0, \beta_1 \in \mathbb{R}\}$
- Orthogonal decomposition of the column space: $\mathcal{C}_{\mathbf{1},\mathbf{x}} = \mathcal{C}_{\mathbf{1},\mathbf{x}-\bar{\mathbf{x}}\mathbf{1}} = \mathcal{C}_{\mathbf{1}} \oplus \mathcal{C}_{\mathbf{x}-\bar{\mathbf{x}}\mathbf{1}}$ so that

$$\hat{\beta}_0 \mathbf{1} + \hat{\beta}_1 \mathbf{x} = (\hat{\beta}_0 + \hat{\beta}_1 \bar{x}) \mathbf{1} + \hat{\beta}_1 (\mathbf{x} - \bar{\mathbf{x}}\mathbf{1})$$

$$\begin{aligned} \Pi(\mathbf{Y} | \mathcal{C}_{\mathbf{1},\mathbf{x}}) &= \Pi(\mathbf{Y} | \mathcal{C}_{\mathbf{1}}) + \Pi(\mathbf{Y} | \mathcal{C}_{\mathbf{x}-\bar{\mathbf{x}}\mathbf{1}}) \\ &= \bar{Y} \mathbf{1} + \frac{S_{xy}}{S_{xx}} (\mathbf{x} - \bar{\mathbf{x}}\mathbf{1}) \end{aligned}$$

Decomposition of Sums of Squares



- $c^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2(\mathsf{SST}), a^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2(\mathsf{SSE})$
  $b^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2(\mathsf{SSR})$

Hypothesis Testing for the Slope

- Due to the projection interpretation, for the residual vector $\mathbf{e} = (e_1, \ldots, e_n)^t$,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} e_i x_i = \sum_{i=1}^{n} e_i \hat{Y}_i = 0$$

- Pythagorean theorem: $c^2 = a^2 + b^2$
- Suppose we want to test

$$H_0 : \beta_1 = 0 \text{ versus } H_1 : \beta_1 \neq 0$$

- t-statistic:

$$T \equiv \frac{\frac{\hat{\beta}_1 - 0}{\sqrt{\sigma^2 / S_{xx}}}}{\sqrt{\frac{\mathsf{SSE}}{(n-2)\sigma^2}}} \sim t(n-2)$$

- Reject H_0 if $|T| > t_{\alpha/2}(n-2)$.

Hypothesis Testing for the Slope

- F-statistic:

$$F \equiv \frac{\frac{\mathsf{SSR}}{1\sigma^2}}{\frac{\mathsf{SSE}}{(n-2)\sigma^2}} \sim F(1, n-2)$$

- Reject H_0 if $F > F_\alpha(1, n-2)$.

Interval Estimation/Prediction

- Regression coefficients:
  $\beta_1: \quad \hat{\beta}_1 \pm t_{\alpha/2}(n-2)\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$; $\beta_0: \quad \hat{\beta}_1 \pm t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}})}$
- Mean Response:
  $\hat{\mu}_x = \hat{\beta}_0 + \hat{\beta}_1 x; \quad \mu_x: \quad \hat{\mu}_x \pm t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2(\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}})}$
- Out-of-sample prediction of a response :
  $\hat{Y}_x = \hat{\beta}_0 + \hat{\beta}_1 x; \quad Y_x: \quad \hat{Y}_x \pm t_{\alpha/2}(n-2)\sqrt{\hat{\sigma}^2(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}})}$

Multiple Linear Regression Models

Consider the case when $p \geq 2$.

- The model: $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$
- $\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2 = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$
  where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^t, \mathbf{X} = (\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_p})$ and $\mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^t$.
- Denote the column space of $\mathbf{X}$ by $\mathcal{C}_{\mathbf{X}}$(the linear space in $\mathbb{R}^n$ spanned by the columns of the matrix $\mathbf{X}$.
- Then $\mathbf{X}\hat{\boldsymbol{\beta}} = \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}})$
- Least squared estimator: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y}$
- Note $\hat{\boldsymbol{\beta}}$ is also BLUE in multivariate case.

Least Squares Estimation

- Orthogonal decomposition of the column space: Write $\mathbf{X} = (\mathbf{1}, \mathbf{X_1})$. Then,

$$\mathcal{C}_{\mathbf{X}} = \mathcal{C}_{\mathbf{1}, \mathbf{x}_1 - \Pi_1 \mathbf{x}_1} = \mathcal{C}_{\mathbf{1}} \oplus \mathcal{C}_{\mathbf{x}_1 - \Pi_1 \mathbf{x}_1}$$

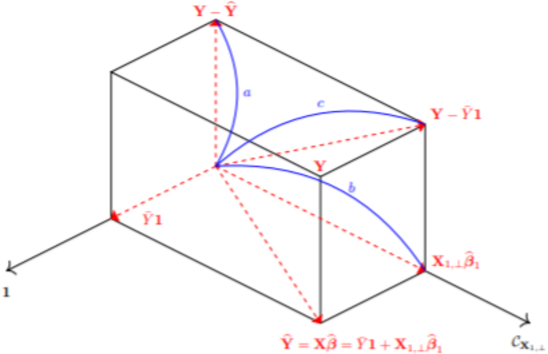
  where $\Pi_1 = \mathbf{1}(\mathbf{1}^t \mathbf{1})^{-1} \mathbf{1}^t$.

- Writing $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}_1^t)^t = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^t = \hat{\boldsymbol{\beta}}^t$ and $\mathbf{X}_{1,\perp} - \mathbf{X}_1 - \Pi_1 \mathbf{X}_1$,

$$\begin{aligned} \mathbf{X}\hat{\boldsymbol{\beta}} &= \hat{\beta}_0 \mathbf{1} + \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 \\ &= (\hat{\beta}_0 + n^{-1} \mathbf{1}^t \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1) \mathbf{1} + \mathbf{X}_{1,\perp} \hat{\boldsymbol{\beta}}_1 \end{aligned}$$

- This gives

$$\hat{\beta}_0 = \bar{Y} - n^{-1} \mathbf{1}^t \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1, \quad \hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})^{-1} \mathbf{X}_{1,\perp}^t \mathbf{Y}$$



- $c^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2(\text{SST}), a^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2(\text{SSE})$
  $b^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2(\text{SSR})$

Testing for Significance of Regression

- Due to the projection interpretation, the residual vector $\mathbf{e}$ is orthogonal to $\mathbf{1}$, all columns of $\mathbf{X}_1$ and $\hat{\mathbf{Y}}$, for all $1 \leq j \leq p$,
$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} e_i x_{ij} = \sum_{i=1}^{n} e_i \hat{Y}_i = 0$$
- Pythagorean theorem: $c^2 = a^2 + b^2$(SST=SSE+SSR)
- Want to test $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_1 \neq \mathbf{0}$
- Distribution of $\hat{\boldsymbol{\beta}}_1$: $\hat{\boldsymbol{\beta}}_1 \sim N_p(\boldsymbol{\beta}_1, \sigma^2(\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})^{-1})$, so that $(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)^t(\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) \sim \sigma^2\chi^2(p)$
- SSE $\sim \sigma^2\chi^2(n-p-1)$ and independent with $\hat{\boldsymbol{\beta}}_1$.
- Under H_0, we have $\hat{\boldsymbol{\beta}}_1^t(\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})\hat{\boldsymbol{\beta}}_1 \sim \sigma^2\chi^2(p)$

Testing for Significance of Regression

- F-statistic: $F \equiv \frac{\frac{\mathsf{SSR}}{p\sigma^2}}{\frac{\mathsf{SSE}}{(n-p-1)\sigma^2}} \sim F(p, n-p-1)$.
- Reject H_0 if $F > F_\alpha(p, n-p-1)$

Testing for Individual β_j

- Want to test $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$, $1 \leq j \leq p$
- Distribution of $\hat{\beta}_j$: $\hat{\beta}_j \sim N(\beta_j, \sigma^2 \mathbf{1}_j^t (\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})^{-1} \mathbf{1}_j)$, where $\mathbf{1}_j$ is the unit vector of dimension p with its jth entry being equal to one.
- It follows that $\mathbf{1}_j^t (\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})^{-1} \mathbf{1}_j = \mathbf{1}_{j+1}^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{1}_{j+1}$
- Thus, writing $C_{jj} = \mathbf{1}_j^t (\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})^{-1} \mathbf{1}_j$, we have

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}} \sim N(0,1)$$

- Letting, $D_{kk} = \mathbf{1}_k^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{1}_k$, we get $C_{jj} = D_{j+1,j+1}, 1 \leq j \leq p$ and $D_{11} = \frac{1}{n} + (\frac{\mathbf{1}^t \mathbf{X}_1}{j})(\mathbf{X}_{1,\perp}^t \mathbf{X}_{1,\perp})^{-1}(\frac{\mathbf{X}_1^t \mathbf{1}}{n})$

Testing for Individual β_j

- Under H_0,

$$T \equiv \frac{\frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}}}{\sqrt{\frac{\mathsf{SSE}}{(n-p-1)\sigma^2}}} \sim t(n-p-1)$$

- Reject H_0 if $|T| > t_{\alpha/2}(n-p-1)$
- For $j = 0$, under H_0,

$$T \equiv \frac{\frac{\hat{\beta}_0}{\sqrt{\sigma^2 D_{11}}}}{\sqrt{\frac{\mathsf{SSE}}{(n-p-1)\sigma^2}}} \sim t(n-p-1)$$

- Similarly, reject H_0 if $|T| > t_{\alpha/2}(n-p-1)$

Interval Estimation/Prediction

- Regression coefficients:

$$\beta_j : \hat{\beta}_j \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 C_{jj}},\ 1 \leq j \leq p$$
$$\beta_0 : \hat{\beta}_0 \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 D_{11}}$$

- Mean response: Writing

$$C_{\mathbf{z}} = \frac{1}{n} + (\mathbf{z}^t - \frac{\mathbf{1}^t\mathbf{X}_1}{n})(\mathbf{X}_{1,\perp}^t\mathbf{X}_{1,\perp})^{-1}(\mathbf{z}^t - \frac{\mathbf{X}_1^t\mathbf{1}}{n})$$

  we get

$$\mu_{\mathbf{z}} : \quad \hat{\mu}_{\mathbf{z}} \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2 C_{\mathbf{z}}}$$

- Out-of-sample prediction of a response:

$$Y_{\mathbf{z}} : \quad \hat{Y}_{\mathbf{z}} \pm t_{\alpha/2}(n-p-1)\sqrt{\hat{\sigma}^2(1 + C_{\mathbf{z}})}$$

Estimator of Sub-vector of Regression Coefficients

Write $\boldsymbol{\beta}^t = (\boldsymbol{\beta}_1^t, \boldsymbol{\beta}_2^t)$ and $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, so that $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$.

- Orthogonal decomposition of the column space:

$$\mathcal{C}_{\mathbf{X}} = \mathcal{C}_{\mathbf{X}_1, \mathbf{X}_2 - \Pi_1 \mathbf{X}_2} = \mathcal{C}_{\mathbf{X}_1} \oplus \mathcal{C}_{\mathbf{X}_2 - \Pi_1 \mathbf{X}_2}$$

  where $\Pi_1 = \mathbf{X}_1(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t$.

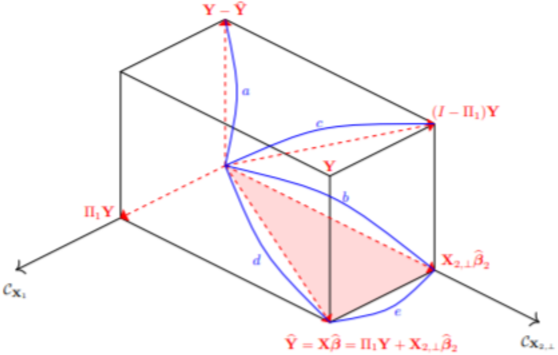
- Writing $\mathbf{X}_{2,\perp} = \mathbf{X}_2 - \Pi_1\mathbf{X}_2$, we get

$$\begin{aligned} \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}_1[\hat{\boldsymbol{\beta}}_1 + (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{X}_2\hat{\boldsymbol{\beta}}_2] + \mathbf{X}_{2,\perp}\hat{\boldsymbol{\beta}}_2 \\ \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}}) &= \Pi_1\mathbf{Y} + \mathbf{X}_{2,\perp}(\mathbf{X}_{2,\perp}^t\mathbf{X}_{2,\perp})^{-1}\mathbf{X}_{2,\perp}^t\mathbf{Y} \end{aligned}$$

- This gives

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t(\mathbf{Y} - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2), \hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_{2,\perp}^t\mathbf{X}_{2,\perp})^{-1}\mathbf{X}_{2,\perp}^t\mathbf{Y}.$$

Decomposition of Sums of Squares



- Pythagorean theorem: $c^2 = a^2 + b^2$ and $d^2 = b^2 + e^2$

Testing for subsets of Regression Coefficients

- Extra sum of squares due to $\boldsymbol{\beta}_2$:

$$R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = R(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - R(\boldsymbol{\beta}_1)$$

  where $R(\boldsymbol{\beta}_1)$ and $R(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, respectively, denote the squared norms of $\Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}_1})$ and $\Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}_1,\mathbf{X}_2}) = \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}})$.
- $R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)$ tends to get larger if $\boldsymbol{\beta}_2 \neq 0$ is true.
-

$$\begin{aligned} R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) &= d^2 - e^2 = c^2 - a^2 = b^2 \\ &= \mathsf{SSE}(\boldsymbol{\beta}_1) - \mathsf{SSE}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = b^2 = ||\mathbf{X}_{2,\perp}\hat{\boldsymbol{\beta}}_2||^2 \end{aligned}$$

  where $\mathsf{SSE}(\boldsymbol{\beta}_1) = ||\mathbf{Y} - \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}_1})||^2$,
  $\mathsf{SSE}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = ||\mathbf{Y} - \Pi(\mathbf{Y}|\mathcal{C}_{\mathbf{X}})||^2$

Testing for subsets of Regression Coefficients

Set $\boldsymbol{\beta}_2 \in \mathbb{R}^q$ and assume that $\epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$. We want test

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ versus } H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

- Distribution of $\hat{\boldsymbol{\beta}}_2$: $(\mathbf{X}_{2,\perp}^t \mathbf{X}_{2,\perp})^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \sim N_q(\mathbf{0}_q, \sigma^2 \mathbf{I}_q)$, so that $(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)^t (\mathbf{X}_{2,\perp}^t \mathbf{X}_{2,\perp})(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \sim \sigma^2 \chi^2(q)$
- Distribution of SSE: SSE $\sim \sigma^2 \chi^2(n - p - 1)$
- SSE and $(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ are independent.
- Under H_0,

$$\hat{\boldsymbol{\beta}}_2^t (\mathbf{X}_{2,\perp}^t \mathbf{X}_{2,\perp}) \hat{\boldsymbol{\beta}}_2 = R(\boldsymbol{\beta}_2 | \boldsymbol{\beta}_1) \sim \sigma^2 \chi^2(q)$$

Testing for subsets of Regression Coefficients

- F-statistic:

$$F \equiv \frac{\frac{R(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1)}{q\sigma^2}}{\frac{\mathsf{SSE}}{(n-p-1)\sigma^2}} \sim F(q, n-p-1)$$

- Reject H_0 if $F > F_\alpha(q, n-p-1)$

General Linear Hypothesis Tests

Let $\mathbf{A}$ be a $(p+1) \times q$ matrix of full column rank and $\mathbf{c}$ be q-dimensional vector($q < p+1$). We want to test

$$H_0 : \mathbf{A}^t\boldsymbol{\beta} = \mathbf{c} \text{ versus } H_1 : \mathbf{A}^t\boldsymbol{\beta} \neq \mathbf{c}$$

- $\hat{\boldsymbol{\beta}}_r = \underset{\mathbf{A}^t\boldsymbol{\beta}=\mathbf{c}}{\arg\min} ||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_r &= (\mathbf{X^tX})^{-1}\mathbf{X^tY} \\ &\quad - (\mathbf{X^tX})^{-1}\mathbf{A}[\mathbf{A^t}(\mathbf{X^tX})^{-1}\mathbf{A}]^{-1}(\mathbf{A}^t(\mathbf{X^tX})^{-1}\mathbf{X^tY} - \mathbf{c})\end{aligned}$$

- Let $R(\boldsymbol{\beta}) = ||\mathbf{X}\hat{\boldsymbol{\beta}}||^2$ and $R(\boldsymbol{\beta}_r) = ||\mathbf{X}\hat{\boldsymbol{\beta}}_r||^2$. Then $R(\boldsymbol{\beta}) - R(\boldsymbol{\beta}_r) \sim \sigma^2\chi^2(q)$.
- F-statistic:

$$F \equiv \frac{\frac{R(\boldsymbol{\beta})-R(\boldsymbol{\beta}_r)}{q\sigma^2}}{\frac{\text{SSE}}{(n-p-1)\sigma^2}} \sim F(q, n-p-1)$$

- Reject H_0 if $F > F_\alpha(q, n-p-1)$.