STATS 100C: Linear Models **Spring 2017**

Lecture 9: May 2

Lecturer: Arash Amini *Scribe: Eric Chuu*

1.1 Recap

Recall the following definitions:

$$\mathrm{X} = \begin{bmatrix} | & | & | & & | \\ \mathbf{1} & \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_p \\ | & | & | & & | \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad A = (X'X)^{-1}X', \quad H = X(X'X)^{-1}X'$$

Then we can define the following:

$$\mathbf{y} = X\boldsymbol{\beta} + \varepsilon, \quad \mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2 I_n), \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n),$$

where $\boldsymbol{\mu} = X\boldsymbol{\beta}$. We also consider $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\mu}}$, where

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(X'X)^{-1}), \quad \hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}} = H\mathbf{y},$$

where $\hat{\boldsymbol{\mu}}$ is the projection of $\mathbf{y}$ onto the image of X. Note that $\hat{\boldsymbol{\mu}}$ is an unbiased estimate of $\boldsymbol{\mu}$, and the covariance can be calculated:

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}) = H\mathrm{Cov}(\mathbf{y})H' = \sigma^2 HH' = \sigma^2 H$$

1.2 Residuals

We can define the residual as follows,

$$\mathbf{e} = \mathbf{y} - \hat{\boldsymbol{\mu}} = \mathbf{y} - H\mathbf{y} = (I - H)\mathbf{y}$$

where $\mathbf{e}$ can be seen as the projection of $\mathbf{y}$ onto the orthogonal complement of the image of X. Since the residual as a function of y, we can easily calculate its distribution:

$$\begin{aligned} \mathbb{E}(\mathbf{e}) &= (I - H)\mathbb{E}(\mathbf{y}) = (I - H)\boldsymbol{\mu} = 0, \qquad \text{since } \boldsymbol{\mu} \in \mathrm{Im}(X) \\ \mathrm{Cov}(\mathbf{e}) &= (I - H)\mathrm{Cov}(\mathbf{y})(I - H)' = \sigma^2(I - H)(I - H)' = \sigma^2(I - H) \end{aligned}$$

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2(I - H))$$

We are interested in the joint behavior of $\left(\hat{\boldsymbol{\beta}}, \mathbf{e}\right)'$. We know that they are marginally normal, but what about jointly?

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} A\mathbf{y} \\ (I - H)\mathbf{y} \end{bmatrix} = \begin{bmatrix} A \\ I - H \end{bmatrix} \mathbf{y} \Rightarrow \mathbb{E}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \mathbf{e} \end{bmatrix} \right) = \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}$$

Before calculating the covariance matrix, consider the matrix product AH. Since $A' = X(X'X)^{-1}$, it follows that $\operatorname{Im}(A') \subseteq \operatorname{Im}(X)$. Thus, applying the projection matrix to A' just gives back A', i.e.,

$$HA' = A' \Leftrightarrow AH = A \tag{1.1}$$

$$\begin{aligned}
\operatorname{Cov}\left(\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \mathbf{e}\end{bmatrix}\right) = \begin{bmatrix} A \\ I-H\end{bmatrix} \operatorname{Cov}(\mathbf{y}) \begin{bmatrix} A' & (I-H)'\end{bmatrix} &= \sigma^2 \begin{bmatrix} A \\ I-H\end{bmatrix} \begin{bmatrix} A' & (I-H)'\end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} AA' & A(I-H)' \\ (I-H)A' & (I-H)\end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & (I-H)\end{bmatrix}
\end{aligned}$$

where the last equality holds by applying the result in equation (1.1) to the off diagonal matrix elements. The distribution of $\left(\hat{\boldsymbol{\beta}}, \mathbf{e}\right)'$ is then given by

$$\begin{bmatrix}\hat{\boldsymbol{\beta}} \\ \mathbf{e}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\beta} \\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\sigma^2(X'X)^{-1} & 0 \\ 0 & \sigma^2(I-H)\end{bmatrix}\right) \tag{1.2}$$

Since $\left(\hat{\boldsymbol{\beta}}, \mathbf{e}\right)'$ is multivariate normal, then uncorrelatedness is equivalent to independence, so its joint distribution given in (1.2) implies that $\hat{\boldsymbol{\beta}}$ is independent of $\mathbf{e}$. More generally, any function of $\hat{\boldsymbol{\beta}}$ is independent of any function of $\mathbf{e}$. For example, the sample variance is independent of $\hat{\boldsymbol{\beta}}$,

$$s^2 = \frac{\|\mathbf{e}\|^2}{n-p-1} \quad \perp \quad \hat{\boldsymbol{\beta}}$$

Recall that $\frac{|\mathbf{e}|^2}{\sigma^2} \sim \chi^2_{n-p-1}$. The expectation of a chi-square random variable is equal to the degrees of freedom. Then,

$$\mathbb{E}\left(\frac{\|\mathbf{e}\|^2}{\sigma^2}\right) = n-p-1$$

$$\mathbb{E}(s^2) = \mathbb{E}\left(\frac{\|\mathbf{e}\|^2}{n-p-1}\right) = \frac{\sigma^2}{n-p-1}\mathbb{E}\left(\frac{\|\mathbf{e}\|^2}{\sigma^2}\right) = \frac{\sigma^2}{n-p-1}(n-p-1) = \sigma^2$$

so s^2 is an unbiased estimate for σ^2. We can still calculate the expectation of s^2 without knowing the distribution $\frac{\|\mathbf{e}\|^2}{\sigma^2}$. We use the following claim:

Claim 1: If $A \in \mathbb{R}^{n\times p}$ is symmetric and has the spectral decomposition $A = U\Lambda U'$, then $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$, where λ_i is the i-th eigenvalue of A.

$$\begin{aligned}
\mathbb{E}\left(\|\mathbf{e}\|^2\right) = \mathbb{E}\left(\sum_{i=1}^n e_i^2\right) = \sum_{i=1}^n \mathbb{E}(e_i^2) = \sum_{i=1}^n \mathrm{Var}(e_i) &= \sum_{i=1}^n [\mathrm{Cov}(\mathbf{e})]_{ii} \\
&= \mathrm{tr}\left(\mathrm{Cov}(\mathbf{e})\right) \\
&= \mathrm{tr}\left(\sigma^2(I-H)\right) \\
&= \sigma^2 \sum_{i=1}^n \lambda_i, \qquad \text{(Claim 1)} \\
&= \sigma^2 \cdot \mathrm{rank}(I-H) \\
&= \sigma^2 \cdot \mathrm{Im}(I-H) \\
&= \sigma^2 \cdot \dim\left([\mathrm{Im}(X)]^\perp\right) \\
&= \sigma^2(n-p-1)
\end{aligned}$$

It follows that $\mathbb{E}\left(s^2\right) = \sigma^2$.

Proof of Claim 1: Consider the trace of the spectral decomposition given in the claim. We then use the circular property of the trace:

$$\mathrm{tr}(A) = \mathrm{tr}\left(U\Lambda U'\right) = \mathrm{tr}\left(U'U\Lambda\right) = \mathrm{tr}(\Lambda) = \sum_{i=1}^n \lambda_i$$

$\square$