lecture 2 : 1

Galton slope
$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$
$$= \frac{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2+n\bar{x}^2-2\bar{x}\sum_{i=1}^{n}x_i} = \frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2}$$

$$\hat{\beta} = \sum_{i,j} w_{ij}b_{ij} = \frac{\sum_{i,j}(x_i-x_j)(y_i-y_j)}{\sum_{i',j'}(x_{i'}-x_{j'})^2} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_iy_i-x_iy_j-x_jy_i+x_jy_j)}{\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i^2-2x_jx_i+x_j^2)}$$
$$= \frac{\sum_{i=1}^{n}(x_iy_i-x_i\cdot n\bar{y}-n\bar{x}y_i+\sum_{j=1}^{n}x_jy_j)}{\sum_{i=1}^{n}(x_i^2-2x_i\cdot n\bar{x}+\sum_{j=1}^{n}x_j^2)} = \frac{\cancel{2n}(\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y})}{\cancel{2n}(\sum_{i=1}^{n}x_i^2-n\bar{x}^2)}$$
$$= \frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2} = \text{Galton slope}$$

lecture 3 : 4

$X=QR$, $X^TX = R^TQ^TQR = R^TR$
$$\hat{\beta} = (X^TX)^{-1}X^TY = (R^TR)^{-1}R^TQ^TY$$
$$= R^{-1}\cancel{(R^T)^{-1}R^T}QY = R^{-1}QY \Rightarrow R\hat{\beta} = QY$$
$$H = X(X^TX)^{-1}X^T = QR(R^TR)^{-1}R^TQ^T = Q\cancel{R}\cancel{R^{-1}}\cancel{(R^T)^{-1}R^T}Q^T = QQ^T$$
Therefore, $h_{ii} = \sum_{j=1}^{p} q_{ij}^2$

lecture 4 : 4

Since $\beta_2=0$, our model is : $Y = X_1\beta_1+\varepsilon$

$\hat{\beta}_1 = (I_k \vdots 0)\hat{\beta}$, $E(\hat{\beta}_1) = (I_k \vdots 0)E\hat{\beta} = (I_k \vdots 0)\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} = \beta_1$

Therefore, we know : $\hat{\beta}_1 = (I_k \vdots 0)HY := \hat{H}_kY$, where $\hat{H}_k$ not depending on Y

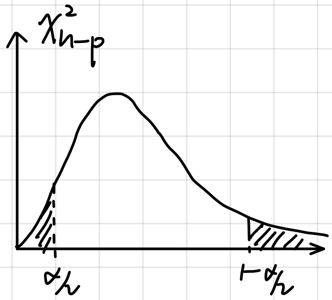
$\hat{\beta}_1$ is unbiased for β_1.

$\tilde{\beta}_1$ is the ols estimator of Y on X_1

By Gauss-Markov Theorem, we know: $\mathrm{cov}(\tilde{\beta}_1) \preceq \mathrm{cov}(\hat{\beta}_1)$

lecture 5 : 7

By theorem 1, we know $\varepsilon^T\varepsilon/\sigma^2 \sim \chi^2_{n-p}$



$$P(\chi^2_{n-p,\alpha/2} < \varepsilon^T\varepsilon/\sigma^2 < \chi^2_{n-p,1-\alpha/2}) = \alpha$$

$\Rightarrow 1-\alpha$ confidence Interval for σ^2 :

$$\chi^2_{n-p,\alpha/2} < \varepsilon\varepsilon^T/\sigma^2 < \chi^2_{n-p,1-\alpha/2}$$

$\Rightarrow \left[\frac{\varepsilon^T\varepsilon}{\chi^2_{n-p,1-\alpha/2}}, \frac{\varepsilon^T\varepsilon}{\chi^2_{n-p,\alpha/2}} \right]$, and we know $\hat{\sigma}^2 = \frac{\varepsilon^T\varepsilon}{n-p}$

$\Rightarrow \left[\frac{(n-p)\hat{\sigma}^2}{\chi^2_{n-p,1-\alpha/2}}, \frac{(n-p)\hat{\sigma}^2}{\chi^2_{n-p,\alpha/2}} \right]$

lecture 6 : 4

$Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\varepsilon}$, $\hat{\beta}_1 = (\tilde{X}_1'\tilde{X}_1)^{-1}\tilde{X}_1'Y$

$X_1 = X_2\hat{\delta} + \hat{V}$, $\hat{\delta} = (X_2'X_2)^{-1}X_2'X_1$, $\tilde{X}_1 = \hat{V} \Rightarrow \hat{\beta}_1 = (\hat{V}'\hat{V})^{-1}\hat{V}'Y$

$Y = X_2\tilde{\beta}_2 + \tilde{\varepsilon}$, $\tilde{\beta}_2 = (X_2'X_2)^{-1}X_2'Y$

$X_2(\hat{\delta}\hat{\beta}_1 + \hat{\beta}_2) = X_2\hat{\delta}\cdot\hat{\beta}_1 + X_2\hat{\beta}_2 = (X_1 - \hat{V})\hat{\beta}_1 + X_2\hat{\beta}_2$

$= X_1\hat{\beta}_1 + X_2\hat{\beta}_2 - \hat{V}\hat{\beta}_1 = HY - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'Y$

$= HY - \tilde{H}_1Y$

From last homework, we know: $H = \tilde{H}_1 + H_2$, where $H_2 = X_2(X_2'X_2)^{-1}X_2'$

$\therefore X_2(\hat{\delta}\hat{\beta}_1 + \hat{\beta}_2) = H_2Y = X_2(X_2'X_2)^{-1}X_2'Y = X_2\tilde{\beta}_2$

$\therefore \hat{\delta}\hat{\beta}_1 + \hat{\beta}_2 = \tilde{\beta}_2$

lecture 7: 1

The projection matrix $H_1 = X_1(X_1'X_1)^{-1}X_1'$

$$X_1'X_1 = \begin{pmatrix} \overbrace{1 \cdots 1}^{n_1} & \overbrace{0 \cdots 0}^{n_2} & \cdots & \overbrace{0 \cdots 0}^{n_k} \\ 0 \cdots 0 & 1 \cdots 1 & \cdots & 0 \cdots 0 \\ \vdots & \vdots & & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & 1 \cdots 1 \end{pmatrix} \begin{array}{c} \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \end{array} \begin{array}{l} \}n_1 \\ \vdots \\ \}n_k \end{array} = \begin{pmatrix} n_1 & & & \\ & n_2 & & \\ & & \ddots & \\ & & & n_k \end{pmatrix}$$

$$H_1 = X_1(X_1'X_1)^{-1}X_1' = \begin{pmatrix} n_1^{-1} & \cdots & 0 \\ \vdots & & \\ n_1^{-1} & \cdots & 0 \\ 0 & \cdots & n_k^{-1} \\ \vdots & & \vdots \\ 0 & \cdots & n_k^{-1} \end{pmatrix} \begin{pmatrix} \overbrace{1 \cdots 1}^{n_1} & \overbrace{0 \cdots 0}^{n_2} & \cdots & \overbrace{0 \cdots 0}^{n_k} \\ 0 \cdots 0 & 1 \cdots 1 & \cdots & 0 \cdots 0 \\ \vdots & \vdots & & \vdots \\ 0 \cdots 0 & 0 \cdots 0 & \cdots & 1 \cdots 1 \end{pmatrix} = \begin{pmatrix} n_1^{-1} \cdots n_1^{-1} & & \\ \vdots \quad \vdots & & \\ n_1^{-1} \cdots n_1^{-1} & & \\ & \ddots & \\ & & n_k^{-1} \cdots n_k^{-1} \\ & & \vdots \quad \vdots \\ & & n_k^{-1} \cdots n_k^{-1} \end{pmatrix}_{n \times n}$$

$$X_1\hat{\beta} = H_1 Y = \begin{pmatrix} n_1^{-1} \cdots n_1^{-1} & & \\ \vdots \quad \vdots & & \\ n_1^{-1} \cdots n_1^{-1} & & \\ & \ddots & \\ & & n_k^{-1} \cdots n_k^{-1} \\ & & \vdots \quad \vdots \\ & & n_k^{-1} \cdots n_k^{-1} \end{pmatrix} \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{pmatrix} \begin{array}{l} \}n_1 \\ \vdots \\ \}n_k \end{array} = \begin{pmatrix} \frac{1}{n_1}\sum_{i=1}^{n} y_{1i} \\ \vdots \\ \frac{1}{n_1}\sum_{i=1}^{n} y_{1i} \\ \vdots \\ \frac{1}{n_k}\sum_{i=1}^{n} y_{ki} \\ \vdots \\ \frac{1}{n_k}\sum_{i=1}^{n} y_{ki} \end{pmatrix} \begin{array}{l} \}n_1 \\ \vdots \\ \}n_k \end{array} = \begin{pmatrix} \bar{y}_{[1]} \\ \vdots \\ \bar{y}_{[1]} \\ \vdots \\ \bar{y}_{[k]} \\ \vdots \\ \bar{y}_{[k]} \end{pmatrix} \begin{array}{l} \}n_1 \\ \vdots \\ \}n_k \end{array}$$

$$\therefore \hat{\varepsilon} = Y - X_1\hat{\beta} = Y - \begin{pmatrix} \bar{y}_{[1]} \\ \vdots \\ \bar{y}_{[1]} \\ \vdots \\ \bar{y}_{[k]} \\ \vdots \\ \bar{y}_{[k]} \end{pmatrix}$$

lecture 8: 5

$$\hat{V}_0 = \left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2 x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$$\hat{V}_2 = \left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \frac{\hat{\varepsilon}_i^2}{1-h_{ii}} x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$$E(\hat{V}_2) = \left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \frac{E(\hat{\varepsilon}_i)^2}{1-h_{ii}} x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$$= \left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \frac{\sigma^2\cancel{(1-h_{ii})}}{\cancel{1-h_{ii}}} x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$$= \sigma^2\left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1} = \sigma^2(X'X)^{-1}$$

$= \mathrm{cov}(\hat{\beta})$, $\hat{V}_2$ is unbiased

$$E(\hat{V}_0) = \left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} \sigma^2(1-h_{ii}) x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$$= \sigma^2\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} (1-h_{ii}) x_i x_i^T\right)\left(n^{-1}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

$\neq \mathrm{cov}(\hat{\beta})$, $\hat{V}_0$ is biased lecture 10: 6

$\hat{\beta}_{[-i]} = \hat{\beta} - (1-h_{ii})^{-1}(X^TX)^{-1}x_i\hat{\varepsilon}_i$, $w_i = \frac{1-h_{ii}}{n-p}$

Therefore, $\sum_{i=1}^{n} w_i \hat{\beta}_{[-i]} = \sum_{i=1}^{n} \frac{1-h_{ii}}{n-p}\hat{\beta} - \sum_{i=1}^{n} \frac{1-h_{ii}}{n-p} \frac{(X^TX)^{-1}x_i\hat{\varepsilon}_i}{1-h_{ii}}$

$= \hat{\beta} - \frac{(X^TX)^{-1}(\sum_{i=1}^{n} x_i\hat{\varepsilon}_i)}{1-h_{ii}} = \hat{\beta}$

Each w_i's, $w_i = \frac{1-h_{ii}}{n-p} > 0$, $\sum_{i=1}^{n} w_i = \frac{\sum_{i=1}^{n}(1-h_{ii})}{n-p} = \frac{n-p}{n-p} = 1$

$\hat{\beta} = n^{-1}\sum_{i=1}^{n}\hat{\beta}_{[-i]}$ isn't hold in general, because $\frac{1}{n} \neq w_i = \frac{1-h_{ii}}{n-p}$,

unless $h_{ii} = \frac{p}{n}$

lecture 11: 8

$X \sim f(X)$, $\varepsilon \sim N(0,1)$, $y = x^2 + \varepsilon$

$\hat{\beta} = \frac{cov(X,y)}{varX} = \frac{EXy - EXEy}{varX} = \frac{EX\cdot x^2 - EXEX^2}{varX} = \frac{EX^3 - EXEX^2}{varX}$

$\hat{\alpha} = E(y) - E(X)\hat{\beta} = E(X^2) - E(X)\hat{\beta}$

(1) $EX = 0$, $EX^2 = varX + (EX)^2 = \frac{(1+1)^2}{12} + 0 = \frac{1}{3}$

$EX^3 = 0$, $\hat{\beta} = \frac{EX^3 - EXEX^2}{varX} = 0$, $\hat{\alpha} = E(y) - \hat{\beta}E(X) = \frac{1}{3}$

$\therefore BLP = 0\cdot x + \frac{1}{3}$, if $X \sim unif(-1,1)$

(2) $EX = \frac{1}{2}$, $EX^2 = varX + (EX)^2 = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$

$EX^3 = \frac{1}{4}\sum_{k=0}^{3} 1^k \cdot 0^{n-k} = \frac{1}{4} \times 1 = \frac{1}{4}$

$\hat{\beta} = \frac{\frac{1}{4} - \frac{1}{2}\times\frac{1}{3}}{\frac{1}{12}} = 1$, $\hat{\alpha} = E(y) - \hat{\beta}E(X) = \frac{1}{3} - 1\times\frac{1}{2} = -\frac{1}{6}$

$\therefore BLP = X - \frac{1}{6}$, if $X \sim unif(0,1)$

(3) $EX = -\frac{1}{2}$, $EX^2 = varX + (EX)^2 = \frac{1}{3}$, $EX^3 = -\frac{1}{4}$

$\hat{\beta} = \frac{-\frac{1}{4} + \frac{1}{2}\times\frac{1}{3}}{\frac{1}{12}} = -1$, $\hat{\alpha} = E(y) - \hat{\beta}E(X) = \frac{1}{3} - \frac{1}{2} = -\frac{1}{6}$

$BLP = -X - \frac{1}{6}$, if $X \sim unif(-1,0)$

lecture 13-2

$$\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = \left[(X^T, \sqrt{\lambda} I_p)\begin{pmatrix} X \\ \sqrt{\lambda} I_p \end{pmatrix}\right]^{-1}(X^T, \sqrt{\lambda} I_p)\begin{pmatrix} Y \\ 0 \end{pmatrix}$$

$$= (X^T X + \lambda I_p)^{-1} X^T Y$$

$$= \hat{\beta}^{ridge}(\lambda)$$