Before:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$i = 1, \dots, n$$

Now:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_p x_{i,p} + \varepsilon_i$$

$$i = 1, \dots, n$$

alt.

$$y_i = \underline{x_i}^T \underline{\beta} + \varepsilon_i$$

$$i = 1, \dots, n$$

$$\underline{x_i} = \begin{pmatrix} 1 \\ x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,p} \end{pmatrix} \qquad \underline{a}^T \underline{b} = \underline{a} \cdot \underline{b}$$

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

$$\underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$\underline{x_i}^T \underline{\beta} = \begin{pmatrix} 1 & x_{i,1}, & \dots, & x_{i,p} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_p x_{i,p}$$

alt.

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

$\underline{Y}$ is vector of length n

$\underline{\beta}$ is vector of length $(p+1)$

$\underline{\varepsilon}$ is vector of length n

X is matrix $(n \times (p+1))$

$$\underset{n \times m}{A} \; \underset{m \times p}{B}$$

What is the X matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} \text{first} & p \end{array} \\
n & \begin{pmatrix}
1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\
1 & & & & \\
1 & & & & \\
1 & & & & \\
1 & & & & \\
\vdots & & & & \\
1 & & & &
\end{pmatrix}
\end{array}
\qquad \underline{x_i}^T
$$

$$
\underline{x_i} = \begin{pmatrix} 1 \\ x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,p} \end{pmatrix}
$$

$$
\begin{array}{cc}
 & \begin{array}{cc} 1 & p \end{array} \\
n & \begin{pmatrix}
1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\
1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\
1 & & & & \\
1 & & \vdots & & \\
1 & & & & \\
\vdots & & & & \\
1 & x_{n,1} & \cdots & & x_{n,p}
\end{pmatrix}
\end{array}
\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}
$$

$$= \begin{pmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} \\ \vdots \\ \\ \end{pmatrix}$$

$$y_i = \beta_0 + \beta_1 x_{11} + \dots + \beta_p x_{1p} + \varepsilon_i$$

SLR

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & x_{21} & x_{22} \\ & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$$

$\text{weight}_i = \text{int} + \text{height}_i + \text{age}_i$

Brief norm Reviews

L_2 norm : $\|\underline{a}\|$

$: \|\underline{a}\|_2$

$= \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$

$\underline{a}^T \underline{b} = \underline{a} \cdot \underline{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$

$\underline{a}^T \underline{a} = \underline{a} \cdot \underline{a} = a_1^2 + a_2^2 + \dots + a_n^2$

$\sqrt{\underline{a}^T \underline{a}} = \|\underline{a}\|_2$

$\underline{a}^T \underline{a} = \|\underline{a}\|_2^2$

$\boxed{\text{LEAST SQUARES}}$

$\hat{\underline{\beta}}_{LS} = \underset{\underline{b}}{\operatorname{argmin}} \|\underline{y} - X\underline{b}\|_2^2$

$$f(\underline{b}) = ||\underline{y} - X\underline{b}||_2^2$$

$$f: \mathbb{R}^n \to \mathbb{R}$$

$$\underset{\underline{b}}{\arg\min}\ ||\underline{y} - X\underline{b}||_2^2$$

$$= \underset{b}{\arg\min}\ (\underline{y} - X\underline{b})^T(\underline{y} - X\underline{b})$$

$$= \underset{b}{\text{ar}}\ (\underline{y}^T - (X\underline{b})^T)(\underline{y} - X\underline{b})$$

in general

$$(AB)^T = B^T A^T$$

$$= \underset{b}{\arg\min} \left(\underline{y}^T - \underline{b}^T X^T \right)\left(\underline{y} - X\underline{b} \right)$$

$$= \underset{b}{\arg\min}\ \underline{y}^T\underline{y} - \underline{y}^T X \underline{b} - \underline{b}^T X^T \underline{y}$$

$$+ \underline{b}^T X^T X \underline{b}$$

$$\underline{y}^T X \underline{b} = \underline{b}^T X^T \underline{y}$$

for $a \in \mathbb{R}$ $a^T = a$

$$\underline{y}^T X \underline{b} \in \mathbb{R}$$

$$\left(\underline{y}^T X \underline{b}\right)^T = \underline{b}^T X^T \underline{y}$$

$$= \underset{\underline{b}}{\operatorname{argmin}}\ \underline{y}^T\underline{y} - 2\,\underline{b}^T X^T \underline{y} + \underline{b}^T X^T X\,\underline{b}$$

I am now going to take the gradient of this

$$\nabla_{\underline{b}}\, \underline{y}^T\underline{y} = \underline{0}$$

$$\nabla_{\underline{b}}\, -2\,\underline{b}^T X^T \underline{y}$$

$f(x) = -2cx$

$\frac{df}{dx} = -2c$

$$= -2\,\nabla_{\underline{b}}\, \underline{b}^T \underline{c} \qquad \underline{c} = X^T \underline{y}$$

$$= -2\,\nabla_{\underline{b}} \left(b_0 c_0 + b_1 c_1 + \dots + b_p c_p \right)$$

$$= -2X^T\underline{y}$$

$$\nabla_{\underline{x}} f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

$$\nabla_{\underline{b}}\ \underline{b}^T X^T X \underline{b}$$

$$\underset{(p+1)\times n \cdot n \times (p+1)}{X^T X} = A$$

$$= \nabla_{\underline{b}}\ \underline{b}^T A \underline{b} \qquad A = \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_{p+1} \end{pmatrix}$$

$$= \nabla_{\underline{b}}\ \underline{b}^T \begin{pmatrix} \underline{a}_1^T \underline{b} \\ \underline{a}_2^T \underline{b} \\ \vdots \\ \underline{a}_{p+1}^T \underline{b} \end{pmatrix}$$

$$= \nabla_{\underline{b}} \Big(b_0(\underline{a}_1^T \underline{b}) + b_1(\underline{a}_2^T \underline{b}) + \dots + b_p(\underline{a}_{p+1}^T \underline{b}) \Big)$$

$$= \nabla_b b^T X^T X \underline{b} = 2 X^T X \underline{b}$$

$$\nabla_{\underline{b}} \underline{b}^T A \underline{b} = 2 A \underline{b}$$

$$b^T X^T X b = (Xb)^T (Xb)$$

$$= \|Xb\|_2^2$$

$$= (\underline{x}_1^T \underline{b})^2 + (\underline{x}_2^T \underline{b})^2 + \dots + (\underline{x}_{p+1}^T \underline{b})^2$$

$$\nabla_{\underline{b}} = -\cancel{2} X^T \underline{y} + \cancel{2} X^T X \underline{b}$$

$$= 0$$

$$X^T X \underline{b} = X^T \underline{y}$$

$$\underline{b} = (X^T X)^{-1} X^T \underline{y}$$

$$\underline{\hat{\beta}}_{LS} = (X^T X)^{-1} X^T \underline{y}$$

$$(AB)^{-1} = B^{-1} A^{-1}$$

only holds if A, B square invertible

$$\left(b_0 \left(\underline{a}_1^T \underline{b} \right) + b_1 \left(\underline{a}_2^T \underline{b} \right) + \dots + b_p \left(\underline{a}_{p+1}^T \underline{b} \right) \right)$$

$$= b_0 \left(a_{11} b_0 + a_{12} b_1 + \dots + a_{1(p+1)} b_p \right)$$

$$+ \dots +$$

$$\frac{\partial}{\partial b_0} = 2 a_{11} b_0 + a_{12} b_1 + \dots + a_{1p+1} b_p$$

$$+ a_{21} b_1$$

$$\vdots$$

$$+ a_{p+1\,1} b_p$$

b/c $X^T X$ is symmetric

A is symmetric

$$\frac{\partial}{\partial b_0} = 2a_{11} b_0 + 2a_{12} b_1 + \dots + 2a_{1,p+1} b_p$$

$$\Rightarrow \nabla_{\underline{b}} = 2A\underline{b} = 2X^T X \underline{b}$$

$$\underline{\hat{\beta}}_{LS} = (X^T X)^{-1} X^T \underline{y}$$

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

$$E(\underline{\varepsilon}) = \underline{0} \quad \text{(Assumption 1)}$$

$$E(\hat{\beta}_{LS})$$

$$= E\left((X^T X)^{-1} X^T \underline{y}\right)$$

$$= (X^T X)^{-1} X^T E(\underline{y})$$

$$= (X^T X)^{-1} X^T E(X\underline{\beta} + \underline{\varepsilon})$$

$$= (X^T X)^{-1} X^T (X\underline{\beta} + \cancel{E(\underline{\varepsilon})}^{\to 0}) \quad A1$$

$$= \cancel{(X^T X)^{-1} X^T X}^{\to I} \beta$$

$$= \beta$$

$$\mathrm{var}(\underline{\varepsilon}) = \sigma^2 I \quad A2$$

$$\hat{\beta}_{LS} = (X^T X)^{-1} X^T \underline{y}$$

$$= (X^T X)^{-1} X^T (X\underline{\beta} + \underline{\varepsilon})$$

$$= \cancel{(X^T X)^{-1} (X^T X)}^{\to I} \underline{\beta} + (X^T X)^{-1} X^T \underline{\varepsilon}$$

$$= \beta + (X^T X)^{-1} X^T \underline{\varepsilon}$$

$$\mathrm{var}(\hat{\beta}_{LS})$$

$$= \mathrm{var}\left((X^T X)^{-1} X^T \underline{\varepsilon}\right)$$

$$= (X^T X)^{-1} X^T \mathrm{var}(\underline{\varepsilon}) \left((X^T X)^{-1} X^T\right)^T$$

$$= \sigma^2 (X^T X)^{-1} X^T \left((X^T X)^{-1} X^T\right)^T \qquad \text{A2}$$

$$= \sigma^2 \underbrace{(X^T X)^{-1} X^T X}_{I} (X^T X)^{-1}$$

$$= \sigma^2 (X^T X)^{-1}$$

$$\mathrm{var}(\hat{\beta}_{1LS}) = \frac{\sigma^2}{S_{xx}} = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

Problems

$$\underline{X} \sim N_n(\underline{\mu}, \sigma^2 I)$$

$\underline{X} - \bar{\underline{x}}$ uncorr $\bar{X}$

$$\underline{X} - \bar{\underline{x}} = \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$$

$$\operatorname{cov}(\bar{X}, X - \bar{\underline{x}}) \overset{\text{WTS}}{=} 0$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n}(1^T X)$$

$$\operatorname{cov}\left(\frac{1^T}{n} X, \left(I - \frac{1 1^T}{n}\right) X\right)$$

Thm c (cntd)

$$= \frac{1}{n}^T \operatorname{Cov}(X, x)\left(I - \frac{11^T}{n}\right)^T$$

$$= \frac{1}{n}^T \sigma^2 I \left(I - \frac{11^T}{n}\right)^T$$

$$= \sigma^2 \frac{1}{n}^T \left(I - \frac{11^T}{n}\right)$$

$$= \sigma^2 \left(\frac{1}{n}^T - \frac{\cancel{1^T 1}\, 1^T}{n^{\cancel{2}}}\right)$$

$$= 0$$

Suppose $\underline{X}$ multivariate Gaussian and

$\operatorname{cov}(X_i, X_j) = 0.$

$\Rightarrow X_i \perp\!\!\!\perp X_j$

Q52 Ch 14

e) Test whether the slopes for males and females are equal

$H_0: \beta_{1,F} = \beta_{1,M}$

$H_0: \beta_{1,F} - \beta_{1,M} = 0$

$$\hat{\beta}_{1,F} - \hat{\beta}_{1,m} \sim N\left(\beta_{1,F} - \beta_{1,m},\ S_{\hat{\beta}diff}\right)$$

$S_{\hat{\beta},diff}$

we will need

$$\hat{var}(\hat{\beta}_{1,F}) \quad \hat{var}(\hat{\beta}_{1,m})$$

std error

CI:

$$(\hat{\beta}_{1,m} - \hat{\beta}_{1,F}) \pm t_{m+n-2}\left(\frac{\alpha}{2}\right) \cdot S_{\hat{\beta}diff}$$

$$S_{\hat{\beta}_{diff}} = \frac{se(\hat{\beta}_{1,m})}{n} + \frac{se(\hat{\beta}_{1,F})}{m}$$

df = Complicated welch.