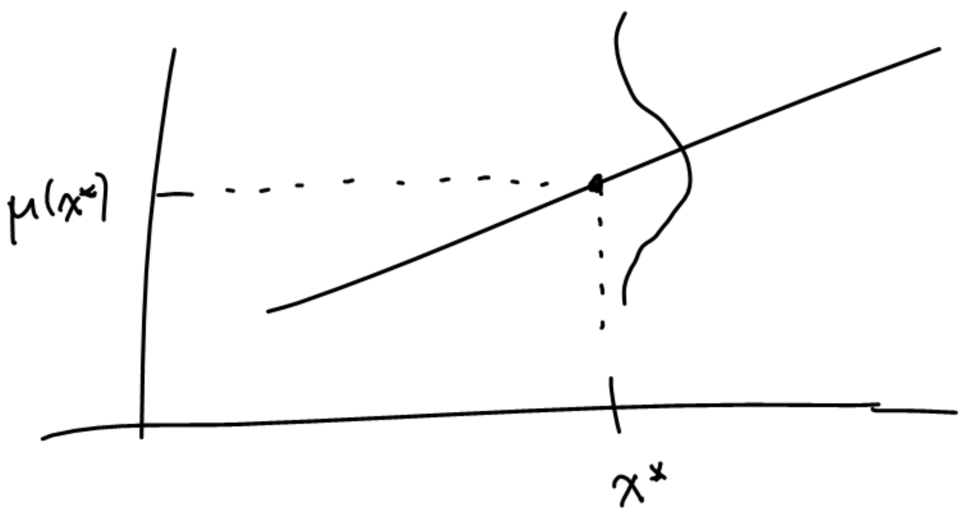
Lecture 10 October 11

Prediction Interval

Model $Y_i \mid x_i \sim N(\mu(x_i), \sigma^2)$

where $\mu(x_i) = \beta_0 + \beta_1 x_i$



If β_0, β_1 are known ($\Rightarrow \mu(x^*)$ known)

and σ^2 known

$$\Rightarrow P\left(\mu(x^*) - 2\sigma < Y^* < \mu(x^*) + 2\sigma\right) \cong 0.95$$

95% PI for Y^* is $[\mu(x^*) - 2\sigma, \ \mu(x^*) + 2\sigma]$

when $\beta_0, \beta_1, \sigma^2$ are not known (hence they need to be estimated), the prediction interval must take into account the uncertainty due to estimating these unknown quantities.

Let $\{(x_i, Y_i),\ i = 1, \ldots, n\}$ = DATASET = $\mathcal{D}$

The estimator from $\mathcal{D}$

$$\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{Y}$$

$$\hat{\underline{\beta}} \sim N\left(\underline{\beta},\ \operatorname{cov}(\hat{\underline{\beta}}) = (X'X)^{-1} \otimes \sigma^2\right)$$

Goal form a prediction interval for Y_* when $x = x^*$,

$(x_0, Y_0) \in \mathcal{D}$

$$Y_* = \mu(x_*) + \varepsilon_*, \quad \varepsilon_* \sim N(0, \sigma^2)$$

$$\hat{Y}_* = \hat{\mu}(x^*) + \varepsilon_*$$

$$= (\hat{\beta}_0 + \hat{\beta}_1 x^*) + \varepsilon_*$$

$$= \boxed{\underbrace{(1 \quad x^*)\, \hat{\beta}}} + \boxed{\underline{\varepsilon_*}}$$

$\hat{\beta}$ depends on $\mathcal{D}$

$\hat{\beta} \perp \varepsilon_*$

$$\hat{\beta} \sim N(\beta, V_{\hat{\beta}})$$

$$\Rightarrow \underline{s}\hat{\beta} \sim N(\underline{s}\beta, \underline{s} V_{\hat{\beta}} \underline{s}') \checkmark$$

Since $\varepsilon^* \sim N(0, \sigma^2)$ $\checkmark$

$$\Rightarrow \hat{Y}_* \sim N(\underline{s}\beta, \underline{s} V_{\hat{\beta}} \underline{s}' + \sigma^2)$$

$$\sim N\left(\underline{c}\beta,\ \underbrace{\underline{c}(X'X)^{-1}\sigma^2}\underline{c}' + \sigma^2\right)$$

$$\sim N\left(\underline{c}\beta,\ \left(\underline{\underline{\underline{c}(X'X)^{-1}\underline{c}'}} + \overset{\}}{1}\right)\sigma^2\right)$$

<u>In contrast</u>

$$\hat{\mu}(x^*) = \underline{c}\hat{\beta} \sim N\left(c\beta,\ \underline{\underline{\underline{c}(X'X)^{-1}\underline{c}'}}\sigma^2\right)$$

The 95% CI for $\mu(x^*)$ if σ^2 known:

$$\underline{c}\hat{\beta} \pm 2\sqrt{\underline{c}(X'X)^{-1}\underline{c}'\sigma^2}$$

The 95% CI for $\mu(x^*)$ if σ^2 not known:

$$\underline{c}\hat{\beta} \pm t(.975;\ df = n-2)\sqrt{\underline{c}(X'X)^{-1}\underline{c}'\hat{\sigma}^2}$$

The 95% Prediction interval for Y_* when $x = x^*$:

If σ^2 known, :

$$\underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x_*)}_{\substack{\mu(x^*) \\ c\hat{\beta}}} \pm \begin{matrix} 2.00 \\ 1.96 \end{matrix} \sqrt{(c(X'X)^{-1}c' + 1)\sigma^2}$$

If σ^2 is not known:

$$(\hat{\beta}_0 + \hat{\beta}_1 x_*) \pm t_{(.975, n-2)} \sqrt{\hat{\sigma}^2}$$

ANALYSIS OF COVARIANCE

Males : $\{(x_i, y_i),\ i = 1, \dots, n_0\}$

Females : $\{(x_i, y_i),\ i = n_0+1, \dots, n_0+n_1\}$

Let $\mu^M(x_i)$ be the mean function (trend) for the male popn.

$\mu^F(x_i)$ ———————— female popn.

Example $\mu^M(x_i) = \beta_0^M + \beta_1^M x_i$

$$\mu^F(x_i) = \beta_0^F + \beta_1^F x_i$$

Let $M_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ subject is male} \\ 0, & \text{o/w} \end{cases}$

$F_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ subject is female} \\ 0, & \text{o/w} \end{cases}$

General model

$$Y_i = \mu(x_i, M_i, F_i) + \varepsilon_i$$

$$= \boxed{(\beta_0^M + \beta_1^M x_i) M_i + (\beta_0^F + \beta_1^F x_i) F_i} + \varepsilon_i$$

In particular, for $i = 1, \ldots, n_0 \Rightarrow M_i = 1, \; F_i = 0$

$$\Rightarrow Y_i = (\beta_0^M + \beta_1^M x_i) + \varepsilon_i$$

For $i = n_0 + 1, \ldots, n_0 + n_1 \Rightarrow M_i = 0, \; F_i = 1$

$$\Rightarrow Y_i = (\beta_0^F + \beta_1^F x_i) + \varepsilon_i$$

A linear model $\underline{Y} = X \underline{\beta} + \underline{\varepsilon}$

$$\begin{array}{c} \underline{Y}^M \\ \\ \\ \underline{Y}^F \end{array}
\begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_0} \\ \hline Y_{n_0+1} \\ \vdots \\ Y_{n_0+n_1} \end{bmatrix}
= \left[\begin{array}{cc|cc} \overbrace{1 \quad\;\; x_1}^{X^M} & & \underline{0} & \underline{0} \\ \vdots \quad\;\; \vdots & & & \\ 1_{n_0} \;\; x_{n_0} & & & \\ \hline \underline{0} & \underline{0} & 1 & x_{n_0+1} \\ & & \vdots & \vdots \\ & & 1_{n_1} & x_{n_0+n_1} \end{array}\right]_{X^F}
\begin{pmatrix} \beta_0^M \\ \beta_1^M \\ \beta_0^F \\ \beta_1^F \end{pmatrix}
+ \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n_0+n_1} \end{pmatrix}$$

The LSE for β

$$\hat{\beta} = (X'X)^{-1} X' \underline{Y}$$

$$(X'X) = \left[\begin{array}{c|c} (X^{m'}X^{m}) & 0 \\ \hline 0 & (X^{F'}X^{F}) \end{array}\right]$$

$$(X'X)^{-1} = \left[\begin{array}{c|c} (X^{m'}X^{m})^{-1} & 0 \\ \hline 0 & (X^{F'}X^{F})^{-1} \end{array}\right]$$

$$(X'Y) = \left(\begin{array}{c} X^{m'}\underline{Y}^{m} \\ \hline X^{F'}\underline{Y}^{F} \end{array}\right)$$

$$\Rightarrow \quad \hat{\beta} = \begin{pmatrix} \hat{\beta}^{m} \\ \hat{\beta}^{F} \end{pmatrix} = \left(\begin{array}{c} (X^{m'}X^{m})^{-1}X^{m'}\underline{Y}^{m} \\ \hline (X^{F'}X^{F})^{-1}X^{F'}\underline{Y}^{F} \end{array}\right)$$

An estimate for σ^2: combine residuals from the two groups.

$$Y_i = \mu(x_i) + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

If $\{\varepsilon_i\text{'s}\}$ observed $\Rightarrow$ we can use $\{\varepsilon_i\}$ to est σ^2

because

$$E(\varepsilon_i^2) = \mathrm{Var}(\varepsilon_i) = \sigma^2$$

$$\Rightarrow E\left\{\frac{(\varepsilon_1^2 + \dots + \varepsilon_n^2)}{n}\right\} = \sigma^2$$

Since $\{\varepsilon_i\}$ not observed:

$$\varepsilon_i = Y_i - \mu(x_i)$$

$$R_i = Y_i - \widehat{\mu}(x_i) \quad \text{residual}$$

$$p = \dim(\underline{\beta})$$

<u>Result</u>:

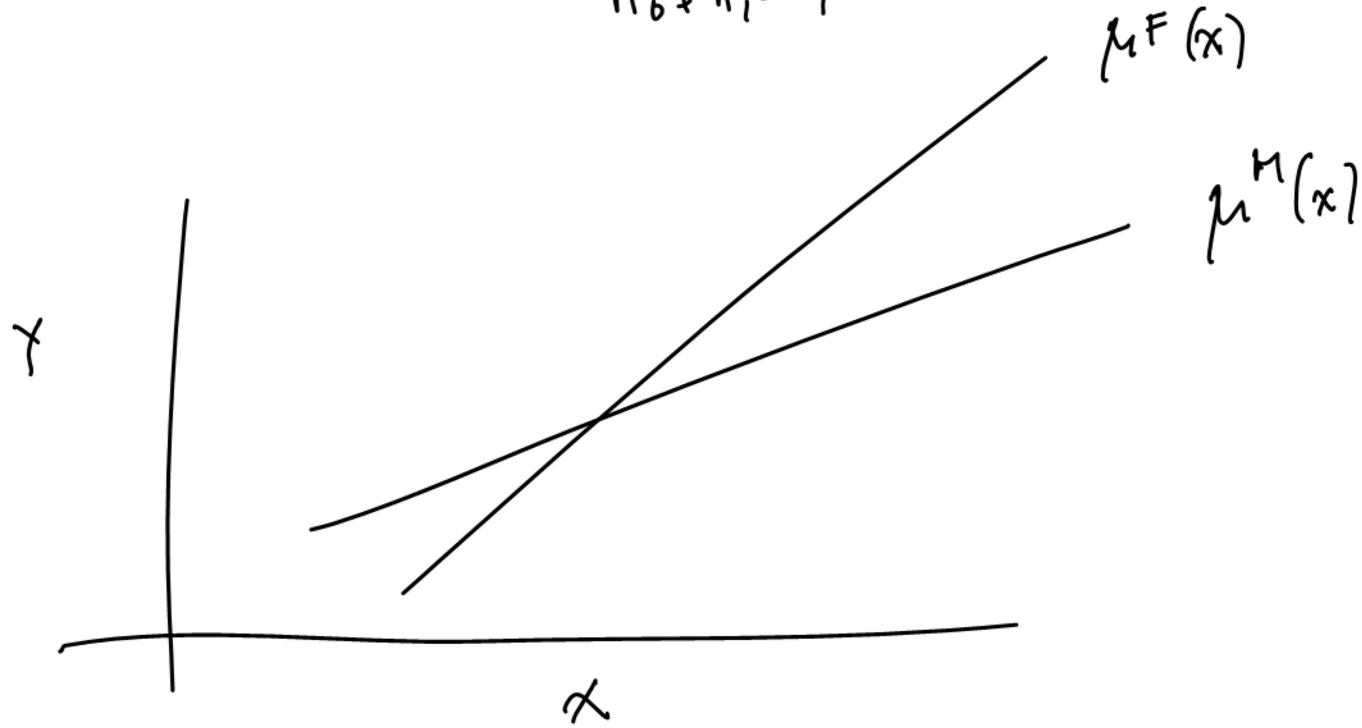
$$E\left\{\underbrace{\frac{\sum_{i=1}^{n} R_i^2}{n-p}}_{\hat{\sigma}^2}\right\} = \sigma^2$$

Specifically:

$$\underline{R} = \begin{pmatrix} R_1 \\ \vdots \\ R_{n_0+n_1} \end{pmatrix} = \begin{pmatrix} Y_1 - (\hat{\beta}_0^M + \hat{\beta}_1^F x_1) \\ \vdots \\ Y_{n_0+n_1} - (\hat{\beta}_0^F + \hat{\beta}_1^F x_{n_0+n_1}) \end{pmatrix}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_0+n_1} R_i^2}{n_0+n_1-4}$$

$$= \frac{\| \underline{R} \|_2}{n_0+n_1-4} = \frac{\underline{R}^T \underline{R}}{n_0+n_1-4}$$



Recall:

$$\mu(x_i) = (\beta_0^M + \beta_1^M x_i) M_i + (\beta_0^F + \beta_1^F x_i) F_i$$

Instead of expressing the female group intercept as β_0^F, we write it as $\beta_0^M + \delta_0$

The female group slope is:

$$\beta_1^M + \delta_1$$

$$\mu(x_i) = (\beta_0^M + \beta_1^M x_i) M_i + \left[(\beta_0^M + \delta_0) + (\beta_1^M + \delta_1) x_i\right] F_i$$

Under the new parameterization

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_0} \\ \hline Y_{n_0+1} \\ \vdots \\ Y_{n_0+n_1} \end{bmatrix} = \left[\begin{array}{cc|cc} \mathbb{1} & x_1 & 0 & 0 \\ & \vdots & & \\ & x_{n_1} & & \\ \hline \mathbb{1} & x_{n_0+1} & \mathbb{1} & x_{n_0+1} \\ & \vdots & & \vdots \\ & x_{n_0+n_1} & & x_{n_0+n_1} \end{array}\right] \begin{bmatrix} \beta_0^M \\ \beta_1^M \\ \delta_0 \\ \delta_1 \end{bmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n_0+n_1} \end{pmatrix}$$