Lecture 18

§1 $P=1$ 时关于 β 的一些假设检验

假设 observations 为 (x_{i1}, y_i)，$i=1,2,\cdots,n$，我们有 linear model：

$$y_i = \beta_0 + x_{i1}\beta_1 + \varepsilon_1$$

根据 Lecture 15 中的结论，有

$$(X^TX)^{-1} = \frac{1}{S_{xx}}\begin{bmatrix} \frac{T_{xx}}{n} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

1、t-test

- 我们有

$$\hat{\beta}_0 \sim N\left(\beta_0,\ \frac{\sigma^2}{n}\left(1+\frac{n\bar{x}^2}{S_{xx}}\right)\right) \quad \left(\frac{T_{xx}}{S_{xx}} = 1+\frac{n\bar{x}^2}{S_{xx}}\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_0,\ \frac{\sigma^2}{S_{xx}}\right)$$

$$S^2 = \frac{SSE}{n-2}$$

- 由于 $\hat{\beta}_0$ 与 $\hat{\beta}_1 \perp S^2$，有

$$T_0 = \frac{\hat{\beta}_0-\beta_0}{\sqrt{\frac{S^2}{n}\left(1+\frac{n\bar{x}^2}{S_{xx}}\right)}} \sim t_{n-2}$$

$$T_1 = \frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{S^2}{n}/S_{xx}}} \sim t_{n-2}$$

- 将 β_0 与 β_1 替换为 H_0 下的值，即可求出检验量并进行 t-test

2、变式：检验 $\beta_0+\beta_1$

- 检验 $H_0: \beta_0+\beta_1=0$
- 若 $\hat{\beta} \sim N(\beta, Cov(\hat{\beta}))$，则我们有

$$W = a^T\hat{\beta} \sim N(a^T\beta,\ a^T Cov(\hat{\beta})a) \quad (Cov(a^T\hat{\beta}) = a^T Cov(\hat{\beta})a)$$

因此

$$\hat{\beta}_0+\hat{\beta}_1 = (1,1)\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix} \sim N\left(\beta_0+\beta_1,\ \frac{\sigma^2}{S_{xx}}[1,1]\begin{bmatrix} \frac{T_{xx}}{n} & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right)$$

$$\sim N\left(\beta_0+\beta_1,\ \frac{\sigma^2}{S_{xx}}\left(\frac{T_{xx}}{n}+1-2\bar{x}\right)\right)$$

- 由于 $\hat{\beta}_0$ 与 $\hat{\beta}_1 \perp S^2$，有 $\hat{\beta}_0+\hat{\beta}_1 \perp S^2$，因此

$$T_{0+1} = \frac{\hat{\beta}_0+\hat{\beta}_1-(\beta_0+\beta_1)}{\sqrt{\frac{S^2}{S_{xx}}\left(\frac{T_{xx}}{n}+1-2\bar{x}\right)}} \sim t_{n-2}$$

- 将 $\beta_0+\beta_1$ 替换为 H_0 下的值 0，即可求出检验量并进行 t-test

  $\beta_0+\beta_1$ 的 $1-\alpha$ CI 为：

$$\hat{\beta}_0+\hat{\beta}_1 \pm t_{\alpha/2,n-2}\sqrt{\frac{S^2}{S_{xx}}\left(\frac{T_{xx}}{n}+1-2\bar{x}\right)}$$

3、变式：检验 $\beta_0+x_{01}\beta_1$

- 根据上述分析，我们可以对 β_0 与 β_1 的任意线性组合进行假设检验

这里不妨对 $\beta_0+x_{01}\beta_1$ 进行假设检验（预测 new observation 的均值）

$H_0: \beta_0+x_{01}\beta_1=0$

· 同上，我们有

$$\hat\beta_0+x_{01}\hat\beta_1=(1,x_{01})\begin{pmatrix}\hat\beta_0\\ \hat\beta_1\end{pmatrix}\sim N\left(\beta_0+x_{01}\beta_1,\ \frac{\sigma^2}{S_{xx}}[1,x_{01}]\begin{bmatrix}\frac{T_{xx}}{n} & -\bar x\\ -\bar x & 1\end{bmatrix}\begin{bmatrix}1\\ x_{01}\end{bmatrix}\right)$$

$$\sim N\left(\beta_0+x_{01}\beta_1,\ \frac{\sigma^2}{S_{xx}}\left(\frac{S_{xx}+n\bar x^2}{n}+x_{01}^2-2\bar x x_{01}\right)\right)$$

$$\sim N\left(\beta_0+x_{01}\beta_1,\ \frac{\sigma^2}{S_{xx}}\left(\frac{S_{xx}}{n}+(x_{01}-\bar x)^2\right)\right)$$

$$\sim N\left(\beta_0+x_{01}\beta_1,\ \sigma^2\left(\frac{1}{n}+\frac{(x_{01}-\bar x)^2}{S_{xx}}\right)\right)$$

· 重复先前步骤即可进行假设检验

$$T=\frac{\hat\beta_0+x_{01}\hat\beta_1-(\beta_0+x_{01}\beta_1)}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(x_{01}-\bar x)^2}{S_{xx}}\right)}}\sim t_{n-2}$$

将 $\beta_0+x_{01}\beta_1$ 替换为 H_0 下的值 0，即可求出检验量

$\beta_0+x_{01}\beta_1$ 的 $1-\alpha$ CI 为：

$$\hat\beta_0+x_{01}\hat\beta_1\pm t_{\alpha/2,n-2}\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(x_{01}-\bar x)^2}{S_{xx}}\right)}$$

§2 Prediction via linear model

对于 linear model $y_0=\beta_0+\beta_1x_{01}+\beta_2x_{02}+\cdots+\beta_px_{0p}+\varepsilon_0$，我们通常进行两种 prediction：

① estimate mean $E[y_0]$

② estimate random variable y_0

1. Estimate mean $E[y_0]$

令 X 表示 historical data，$\mathbf{x}_0=[1,x_{01},\cdots,x_{0p}]^T$ 表示 new regressor，则类似之前对 $\beta_0+x_{01}\beta_1$ 的检验，我们有以下结论：

① The predictor $\hat y_0=\mathbf{x}_0^T\hat{\boldsymbol\beta}=\beta_0+\sum_{j=1}^{p}x_{0j}\hat\beta_j$ 为 the mean of y_0 的一个 reasonable estimate.

② 由于 $\hat{\boldsymbol\beta}\sim N(\boldsymbol\beta,\sigma^2(X^TX)^{-1})$，我们有

$$\mathbf{x}_0^T\hat{\boldsymbol\beta}\sim N(\mathbf{x}_0^T\boldsymbol\beta,\ \sigma^2\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0)$$

③ 由于 $\hat{\boldsymbol\beta}\perp S^2$，有 $\mathbf{x}_0^T\hat{\boldsymbol\beta}\perp S^2$，因此

$$T=\frac{\mathbf{x}_0^T\hat{\boldsymbol\beta}-\mathbf{x}_0^T\boldsymbol\beta}{\sqrt{S^2\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0}}\sim t_{n-p-1}$$，其中 $S^2=\frac{SSE}{n-p-1}=\frac{\sum_i(y_i-\hat y)^2}{n-(p+1)}$

重复先前步骤即可进行假设检验

④ C.I. for $E[y_0]$ given $\mathbf{x}_0=[1,x_{01},\cdots,x_{0p}]^T$ 为

$$\mathbf{x}_0^T\hat{\boldsymbol\beta}\pm t_{\alpha/2,n-p-1}\sqrt{S^2\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0}$$

2. Definition: consistent $1-\alpha$ prediction interval

我们称集合 $C=C(\mathbf{x}_0;X,\mathbf{y})$ 为 consistent $1-\alpha$ prediction interval，若

$$\lim_{n\to\infty} P(y_0 \in C) = 1-\alpha$$

注: ① 该概率 is taken w.r.t. y_0 & $\hat{\boldsymbol\beta}$ (由于 ε_0 与 $\boldsymbol\varepsilon$ 的存在均为随机变量)

② 由于 y_0 与 historical data 独立，而 $\hat{\boldsymbol\beta}$ 由 X 与 $\mathbf{y}$ 决定，因此 $y_0 \perp \hat{\boldsymbol\beta}$ ($\varepsilon_0 \perp \hat{\boldsymbol\beta}$)

3. y_0 的分布

我们已经得出

$$\mathbf{x}_0^T\hat{\boldsymbol\beta} \sim N(\mathbf{x}_0^T\boldsymbol\beta,\ \sigma^2\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0)$$

由于 $y_0 = \mathbf{x}_0^T\boldsymbol\beta + \varepsilon_0$，且 $\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_n \overset{iid}{\sim} N(0,\sigma^2)$，我们有

$$y_0 = \mathbf{x}_0^T\hat{\boldsymbol\beta} + \varepsilon_0 \sim N(\mathbf{x}_0^T\boldsymbol\beta,\ \sigma^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0))$$

4. Estimate random variable y_0

- 由于 $y_0 = \mathbf{x}_0^T\hat{\boldsymbol\beta} + \varepsilon_0 \sim N(\mathbf{x}_0^T\boldsymbol\beta,\ \sigma^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0))$,

  我们同时有

  $$\mathbf{x}_0^T\hat{\boldsymbol\beta} - \varepsilon_0 \sim N(\mathbf{x}_0^T\boldsymbol\beta,\ \sigma^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0))$$

  考虑 fitted value $\hat{y}_0 = \mathbf{x}_0^T\hat{\boldsymbol\beta}$，则有

  $$\begin{aligned} y_0 - \hat{y}_0 &= \mathbf{x}_0^T\boldsymbol\beta + \varepsilon_0 - \mathbf{x}_0^T\hat{\boldsymbol\beta} \\ &= \mathbf{x}_0^T\boldsymbol\beta - (\mathbf{x}_0^T\hat{\boldsymbol\beta} - \varepsilon_0) \quad (\mathbf{x}_0^T\boldsymbol\beta \text{ 可视作常数}) \\ &\sim N(0,\ \sigma^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0)) \end{aligned}$$

- 因此我们有

  $$\frac{y_0 - \hat{y}_0}{\sqrt{S^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0)}} \sim t_{n-p-1}$$

- 因此 y_0 的 $1-\alpha$ CI 为

  $$\hat{y}_0 \pm t_{\alpha/2,n-p-1}\sqrt{S^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0)}$$

注: 对比两种 predictions 的 CI：

① estimate mean: $\mathbf{x}_0^T\hat{\boldsymbol\beta} \pm t_{\alpha/2,n-p-1}\sqrt{S^2\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0}$ (*)

② estimate random variable: $\mathbf{x}_0^T\hat{\boldsymbol\beta} \pm t_{\alpha/2,n-p-1}\sqrt{S^2(1+\mathbf{x}_0^T(X^TX)^{-1}\mathbf{x}_0)}$ (#)

发现 (#) 永远大于 (*)，因为 (#) 中要考虑 ε_0 的随机性，该随机性不会随 n 增大而减小。且当 $n\to\infty$ 时，

width of (*) $\to 0$

width of (#) $\to 2t_{\alpha/2,n-p-1}\sigma$ $(2z_{\alpha/2}\sigma)$