Estatística II

1º Sem. 2015/2016

Lic. Economia e Finanças

Previsão

Previsão

- Anteriormente, definimos os valores ajustados (ou previstos) de y como $\hat{y} = x\hat{\beta}$.
- Isto dá-nos uma estimativa do valor esperado condicional de y para um certo valor de x.
- Naturalmente, os valores de $\hat{y}$ são variáveis aleatórias.
- Como fazer inferência sobre $E\left[y|x\right]$ tendo por base $\hat{y}$?

Previsão para a média vs previsão para uma unidade particular

- Considere-se o seguinte exemplo:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$$

- Como de costume:

$$E(u|educ, exper) = 0$$

$$E(wage|educ, exper) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2$$

Previsão para a média

Pretende-se prever o **salário médio dos indivíduos** com

- 15 anos de escolaridade
- 5 anos de experiência

(não tem de ser observado na amostra), isto é, prever o valor

$$\begin{aligned} &E(wage^0|educ = 15, exper = 5) \\ &\qquad = \beta_0 + \beta_1 \times 15 + \beta_2 \times 5 + \beta_3 \times 25 = \theta_0 \end{aligned}$$

Previsão individual (para uma unidade particular)

Pretende-se prever o **salário de um indivíduo** com

- 15 anos de escolaridade
- 5 anos de experiência

isto é, prever o valor

$$
\begin{aligned}
& wage^0 \,|\, (educ = 15, exper = 5) \\
& \quad = \beta_0 + \beta_1 \times 15 + \beta_2 \times 5 + \beta_3 \times 25 + u^0 = \theta_0 + u^0
\end{aligned}
$$

Previsão pontual para a média

- No modelo de regressão mútlipla usual, considere-se

  $x_1 = x_1^0, \ldots, x_k = x_k^0$; seja então

$$\theta_0 = \beta_0 + \beta_1 x_1^0 + \cdots + \beta_k x_k^0 = x^0 \beta = E\left(y | x^0\right)$$

- Pelo corolário do teorema de Gauss-Markov, o estimador de variância mínima (BLUE) para θ_0 é dado por

$$\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1^0 + \cdots + \hat{\beta}_k x_k^0 = x^0 \hat{\beta}$$

Previsão pontual para a média: exemplo

Exemplo:

$$\widehat{wage} = -3.965 + 0.595 educ + 0.268 exper - 0.0046 exper^2$$
$$n = 526, \quad \widehat{\sigma}^2 = 10.024$$

A previsão para o **salário médio dos indivíduos** com 15 anos de escolaridade e 5 anos de experiência é dada por:

$$\begin{aligned} E(\widehat{wage} | educ = 15, exper = 5) = \\ &= -3.965 + 0.595 \times 15 + 0.268 \times 5 - 0.0046 \times 25 \\ &= 6.19\$/h \end{aligned}$$

Previsão por intervalo para a média

Se se verificarem MLR.1–MLR.5 então:

- **Valor esperado de $\hat{\theta}_0$:**

$$E(\hat{\theta}_0|x_1 = x_1^0, \ldots, x_k = x_k^0) = E(x^0\hat{\beta}|x^0) = x^0\beta = \theta_0$$

- **Variância de $\hat{\theta}_0$:**

$$Var(\hat{\theta}_0) = Var(x^0\hat{\beta}) = x^0 Var(\hat{\beta})x^{0\prime} = \sigma^2 x^0(X'X)^{-1}x^{0\prime}$$

$$\text{se}(\hat{\theta}_0) = \hat{\sigma}\sqrt{x^0(X'X)^{-1}x^{0\prime}}$$

Previsão por intervalo para a média

- **Distribuição de $\widehat{\theta}_0 = x^0\hat{\beta}$:** da normalidade assimptótica de $\hat{\beta}$ segue

$$\hat{\theta}_0 \overset{a}{\sim} N(\theta_0, \sigma^2 x^0 (X'X)^{-1} x^{0\prime})$$

- **Variável fulcral:**

$$\frac{\hat{\theta}_0 - \theta_0}{\text{se}(\hat{\theta}_0)} \overset{a}{\sim} N(0,1)$$

- **Intervalo de Confiança a** $(1-\alpha) \times 100\%$

$$\hat{\theta}_0 \pm z_{\alpha/2}\,\text{se}(\hat{\theta}_0)$$

Obtenção da variância da previsão

- Reparametrização do modelo $y = x\beta + u$

$$\begin{aligned} \theta_0 &= \beta_0 + \beta_1 x_1^0 + \cdots + \beta_k x_k^0 \\ \beta_0 &= \theta_0 - \beta_1 x_1^0 - \cdots - \beta_k x_k^0 \end{aligned}$$

- Substituindo:

$$y = \theta_0 + \beta_1(x_1 - x_1^0) + \cdots + \beta_k(x_k - x_k^0) + u$$

- Fazendo a regressão de y nos valores dos regressores centrados no ponto para o qual queremos fazer previsão, obtemos directamente $\hat{\theta}_0$ e se$(\hat{\theta}_0)$

Observações

- Se existir heterocedasticidade então usa-se estimador robusto de White para as variâncias para obter se$(\hat{\theta}_0)$.
- Demonstra-se que $Var(\hat{\theta}_0)$ é tanto menor quanto mais próximo x^0 estiver de $\bar{x}$.
- $Var(\widehat{\theta_0}) \to 0$ quando $n \to \infty$.

Exemplo

- Retome-se o exemplo inicial:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$$

- Considere-se agora a **regressão auxiliar** seguinte:

$$\widehat{wage} = \underset{(0.219)}{6.19} + 0.595(educ - 15) + 0.268(exper - 5) - 0.0046(exper^2 - 25)$$

- Com 95% de confiança, o intervalo para θ_0 é dado por:

$$6.19 \pm 1.96 \times 0.219 = (5.76, 6.62)$$

Previsão pontual para uma realização de y

- Considere-se $x_1 = x_1^0, \ldots, x_k = x_k^0$, e

$$y^0 = \beta_0 + \beta_1 x_1^0 + \cdots + \beta_k x_k^0 + u^0$$

- Assumindo que $E(u^0|x^0) = 0$ e $Var(u^0|x^0) = \sigma^2$, prova-se que o melhor previsor pontual de y^0 é

$$\hat{y}^0 = \hat{\beta}_0 + \hat{\beta}_1 x_1^0 + \cdots + \hat{\beta}_k x_k^0 = x^0 \hat{\beta} = \hat{\theta}_0$$

Erro de previsão

- **Erro de previsão:**

$$\begin{aligned} e^0 &= y^0 - \hat{y}^0 \\ &= x^0\beta + u^0 - x^0\hat{\beta} = u^0 + x^0\left(\beta - \hat{\beta}\right) \end{aligned}$$

- **Valor esperado do erro de previsão:**

$$E(e^0|X) = 0$$

pois, supondo MLR.1–MLR.4, têm-se

$$E\left(\hat{\beta}|X\right) = \beta$$

Erro de previsão

- **Variância do erro de previsão:**

  Assumindo MLR.1–MLR.5 têm-se

$$
\begin{aligned}
Var(e^0|X) &= \sigma^2 + Var(\hat{\theta}_0) \\
&= \sigma^2 + \sigma^2 x^0 (X'X)^{-1} x^{0\prime} \\
&= \sigma^2 \left(1 + x^0 (X'X)^{-1} x^{0\prime}\right)
\end{aligned}
$$

- Note-se que a variância tem duas componentes com origens diferentes: uma desaparece assimptoticamente e a outra não.

Erro de previsão

- **Distribuição do erro de previsão:**

  Se os erros do modelo forem homocedásticos e normais, segue que a distribuição do erro de previsão $e^0 = y^0 - \hat{y}^0$ é dada por:

$$\frac{y^0 - \hat{y}^0}{\sigma\sqrt{1 + x^0(X'X)^{-1}x^{0\prime}}} \sim t(n-k-1)$$

Previsão por intervalo para uma realização de y

Para calcular um intervalo de confiança para uma realização de y podemos então recorrer ao resultado anterior e obter:

$$\hat{y}^0 \pm t_{\alpha/2}(n-k-1)\sqrt{\hat{\sigma}^2 + \left(\mathrm{se}(\hat{\theta}_0)\right)^2}$$

Exemplo

- Previsão por intervalo do salário de um indivíduo com $educ = 15$, $exper = 5$

- Retomem-se os resultados da regressão auxiliar:

$$\widehat{wage} = \underset{(0.219)}{6.19} + 0.595(educ - 15) + 0.268(exper - 5) - 0.0046(exper^2 - 25)$$

$$n = 526, \quad \hat{\sigma}^2 = 10.024$$

Exemplo

- Então:

$$\begin{aligned} \widehat{wage}^0 &= -3.965 + 0.595 \times 15 + 0.268 \times 5 - 0.0046 \times 25 \\ &= 6.19\$/h \end{aligned}$$

$$\begin{aligned} \mathrm{se}(\widehat{wage}^0) &= \sqrt{\hat{\sigma}^2 + \left(\mathrm{se}(\hat{\theta}_0)\right)^2} \\ &= \sqrt{10.024 + 0.219^2} = 3.1736 \end{aligned}$$

- O intervalo de confiança é dado por:

$$6.19 \pm 3.1736 \times 1.96 = (-0.03026, 12.41026)$$

Previsão de y nos modelos logarítmicos

- Como prever y quando se estima um modelo para o logaritmo?

$$ln(y) = x\beta + u$$

- Como $E[y|x] \neq \exp E[ln(y)|x]$, em geral não é possível estimar $E[y|x]$ partindo de $E[ln(y)|x]$.

Previsão de y nos modelos logarítmicos

- No entanto, **caso u seja independente de x,** o $E[y|x]$ é proporcional a $\exp E[ln(y)|x] = \exp(x\beta)$:

$$E[y|x] = \alpha_0 \ e^{x\beta}$$

- Portanto, pode estimar-se $E[y|x]$ fazendo uma regressão auxiliar de y em $\exp\left[\widehat{ln(y)}\right]$, sem constante, e construir previsões de y.

Método de estimação com u independente de x

1. Regressão de $ln(y)$ sobre x e obtenção dos valores ajustados $\widehat{ln(y)} = x\hat{\beta}$.

2. Regressão de y sobre $\hat{m} = \exp(x\hat{\beta})$ sem constante:
$$y = \alpha_0 \ \hat{m} + erro$$

3. Obtido $\hat{\alpha}_0$ da regressão anterior, calcula-se
$$\hat{y} = \hat{\alpha}_0 \ \exp(x\hat{\beta})$$

Os intervalos de confiança podem ser feitos exponenciando os limites dos intervalos para $\widehat{ln(y)}$.

Previsão quando u é Normal

- Se $u \sim N(0, \sigma^2)$, segue que $ln(y) \sim N(x\beta, \sigma^2)$, ou seja y é Lognormal.

- Pelas propriedades da distribuição Lognormal

$$E(y|x) = e^{\sigma^2/2} e^{x\beta}$$

- Portanto a previsão de y é obtida como

$$\hat{y} = e^{\hat{\sigma}^2/2} e^{x\hat{\beta}} = e^{\hat{\sigma}^2/2} e^{\widehat{ln(y)}}$$