UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
CURSO DE GRADUAÇÃO EM ESTATÍSTICA

ANTÔNIO ARTHUR SILVA DE LIMA
FRANCISCO GUSTAVO BRAGA BATISTA
RÔMULO BARROS DE FREITAS

TRABALHO 3 - MODELOS DE REGRESSÃO I

FORTALEZA
2023

ANTÔNIO ARTHUR SILVA DE LIMA
FRANCISCO GUSTAVO BRAGA BATISTA
ROMULO BARROS DE FREITAS

TRABALHO 3 - MODELOS DE REGRESSÃO I

Trabalho apresentado ao curso de Bacharelado em Estatística do Centro de Ciências da Universidade Federal do Ceará, como parte dos requisitos para a aprovação na disciplina de Modelos de Regressão I no semestre de 2023.2.
Prof.: Ronald Targino Nojosa.

FORTALEZA
2023

Sumário

1 Tradução

1.0.1 Caso em que a variável regressora x é aleatória

O modelo de regressão linear que apresentamos neste capítulo supõe que os valores da variável regressora x são constantes conhecidas. Essa suposição faz com que os níveis de confiança, e os erros de tipo I e II, se refiram a amostras repetidas de y para os mesmos níveis de x. Há várias situações nas quais supor que os x's sejam constantes fixas é inapropriado. Por exemplo, considere os dados do tempo de entrega de refrigerantes do Capítulo 1 (Figura 1.1). Já que os pontos de venda visitados pelo entregador são selecionados ao acaso, não é realista acreditar que podemos controlar o volume x de entregas. É mais racional supor que tanto y quanto x são variáveis aleatórias. Felizmente, sob certas circunstâncias, todos os nossos resultados prévios de estimação, testes e predições dos parâmetros, são válidos. Agora, discutiremos essas situações.

1.0.2 x e y são conjuntamente distribuídas

Suponha que x e y são variáveis aleatórias conjuntamente distribuídas, mas a forma dessa distribuição conjunta é desconhecida. Pode ser mostrado que todos os nossos resultados da regressão anteriores sustentam-se se as seguintes condições são satisfeitas:

1. A distribuição condicional de y dado x é normal, com média condicional $\beta_0 + \beta_1 x$ e variância condicional σ^2.

2. Os x's são variáveis aleatórias independentes, cuja distribuição de probalidade não envolve β_0, β_1 e σ^2.

Apesar de todos os procedimentos da regressão permanecerem os mesmos sob tais condições, os níveis de confiança e os erros têm uma interpretação diferente. Quando o regressor é uma variável aleatória, essas quantidades aplicam-se a amostras repetidas dos valores de (x_i, y_i), e não a amostras repetidas de y_i para níveis fixos de x_i.

2 Demonstrações

Queremos provar que a distribuição condicional de y dado x é da forma:

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sigma_{1.2}} \exp\left[-\frac{1}{2}\left(\frac{y - \beta_0 - \beta_1 x}{\sigma_{1.2}}\right)^2\right],$$

onde $\sigma_{1.2} = \sqrt{\sigma_1^2(1 - \rho^2)}$.

E sabemos que, por definição, possui a seguinte forma:

$$f(y|x) = \frac{f(y,x)}{f(x)},$$

sendo $f(y,x)$ a função densidade conjunta, e $f(x)$ a função densidade marginal de x.

Temos que

$$f(y,x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]\right\}.$$

Portanto, para encontrar $f(x)$, iremos integrar a densidade conjunta em relação à y:

$$
\begin{aligned}
f(x) &= \int_{-\infty}^{\infty} f(y,x)dy \\
&= \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]\right\}dy \\
&= \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \\
&\quad \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right) + \rho^2\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \rho^2\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right\}dy \\
&= \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
&= \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right) - \rho\left(\frac{x-\mu_2}{\sigma_2}\right)\right]^2 \left[-\rho^2\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right\}dy \\
&= \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} e^{-\frac{\rho^2}{2(1-\rho^2)}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\
&= \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right) - \rho\left(\frac{x-\mu_2}{\sigma_2}\right)\right]^2\right\}dy \\
&= \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_2}{\sigma_2}\right)^2[1-\rho^2]}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right) - \rho\left(\frac{x-\mu_2}{\sigma_2}\right)\right]^2\right\}dy \\
&= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right) - \rho\left(\frac{x-\mu_2}{\sigma_2}\right)\right]^2\right\}dy
\end{aligned}
$$

Agora, faremos a seguinte transformação na integral:

$$
u = \frac{1}{\sqrt{1-\rho^2}}\left[\left(\frac{y-\mu_1}{\sigma_1}\right) - \rho\left(\frac{x-\mu_2}{\sigma_2}\right)\right] \quad \therefore \quad du = \frac{1}{\sigma_1\sqrt{1-\rho^2}}dy
$$

Mas $y \in (-\infty, \infty) \Longrightarrow u \in (-\infty, \infty)$, e $\sigma_1\sqrt{1-\rho^2}du = dy$. Logo, ficamos com:

$$
\begin{aligned}
f(x) &= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}\sigma_1\sqrt{1-\rho^2}du \\
&= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}du \\
&= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{2\pi\sigma_2}\sqrt{2\pi} \\
&= \frac{e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}}{\sqrt{2\pi}\sigma_2}\mathbb{I}_{(-\infty,\infty)}(X) \sim N\left(\mu_2;\sigma_2^2\right)
\end{aligned}
$$

Agora, iremos demonstrar a distribuição condicional de y dado x:

$$
\begin{aligned}
f(y|x) &= \frac{f(y,x)}{f(x)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2+\left(\frac{x-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \\
&\quad \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2+\left(\frac{x-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)\right]+\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right\}
\end{aligned}
$$

Fazendo $k = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}$

$$
\begin{aligned}
f(y|x) &= k \\
& \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2+\left(\frac{x-\mu_2}{\sigma_2}\right)^2-2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)-(1-\rho^2)\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right. \\
&= k\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{y-\mu_1}{\sigma_1}\right)\left(\frac{x-\mu_2}{\sigma_2}\right)+\rho^2\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]\right\} \\
&= k\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_1}{\sigma_1}\right)^2-\rho\left(\frac{x-\mu_2}{\sigma_2}\right)\right]^2\right\} \\
&= k\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{y\sigma_2-\mu_1\sigma_2-\rho x\sigma_1+\rho\sigma_1\mu_2}{\sigma_1\sigma_2}\right)^2\right\} \\
&= k\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{y}{\sigma_1}-\frac{\mu_1}{\sigma_1}-\frac{\rho x\sigma_1}{\sigma_1\sigma_2}+\frac{\rho\sigma_1\mu_2}{\sigma_1\sigma_2}\right)^2\right\} \\
&= k\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(y-\mu_1-\frac{\rho x\sigma_1}{\sigma_2}+\frac{\rho\sigma_1\mu_2}{\sigma_2}\right)^2\right\} \\
&= k\exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[y-\left(\mu_1-\mu_2\rho\frac{\sigma_1}{\sigma_2}\right)-\frac{\sigma_1}{\sigma_2}\rho x\right]^2\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\ \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(y-\beta_0-\beta_1 x\right)^2\right\}
\end{aligned}
$$

Onde $\beta_0 = \mu_1 - \mu_2\rho\frac{\sigma 1}{\sigma 2}$ e $\beta_1 = \frac{\sigma 1}{\sigma 2}\rho$

$$
\begin{aligned}
f(y|x) &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2(1-\rho^2)}}\ \exp\left\{-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sqrt{\sigma_1^2(1-\rho^2)}}\right)^2\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_{1.2}}\ \exp\left\{-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma_{1.2}}\right)^2\right\} \sim N(\beta_0+\beta_1 x\,;\sigma_{1.2}^2)
\end{aligned}
$$

Onde $\sigma_{1.2} = \sqrt{\sigma_1^2(1-\rho^2)}$