MIB《高级计量经济学》
2026 年春季学期第一次作业

黄嘉平

2026.4.16

针对单变量回归模型

$$y_i = \alpha + \beta x_i + u_i, \qquad i = 1, 2, \dots, n$$

回答下列问题。

1. 推导系数 β 的最小二乘估计量 $\hat{\beta}$ 的表达式（不需要写成矩阵形式）。
2. 推导 $\hat{\beta}$ 和样本相关系数 r_{xy} 之间的关系。
3. 在什么时候 $\hat{\beta}$ 的估计值和样本相关系数是一致的? 利用蒙特卡洛仿真确认你的结论。

参考答案

1. **推导系数 β 的最小二乘估计量 $\hat{\beta}$ 的表达式。**

 令残差 $u_i = y_i - \alpha - \beta x_i$ 的平方和为

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

SSE 是 α 和 β 的函数，对二者分别求偏导可得一阶条件

$$\frac{\partial\,\mathrm{SSE}}{\partial \alpha} = 2n\alpha - 2\sum_{i=1}^{n}(y_i - \beta x_i) = 0$$

$$\frac{\partial\,\mathrm{SSE}}{\partial \beta} = 2\beta\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n}(y_i - \alpha)x_i = 0$$

针对 α 解第一个等式可得 $\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta x_i) = \bar{y} - \beta\bar{x}$，将其代入第二个等式，整理可得

$$\beta\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i + \bar{y}\sum_{i=1}^{n} x_i - \beta\bar{x}\sum_{i=1}^{n} x_i = 0$$

针对 β 求解可得

$$\hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n} y_i x_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^{n}\left(x_i^2 - \bar{x}^2\right)}$$

2. **推导 $\hat{\beta}$ 和样本相关系数 r_{xy} 之间的关系。**

 首先将 $\hat{\beta}$ 变形

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^{n}(x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^{n}\left(x_i^2 - \bar{x}^2\right)} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \\
&= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2} = \frac{\sum_{i=1}^{n}(x_i y_i - \bar{x} y_i - x_i\bar{y} + \bar{x}\bar{y})}{\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)} \\
&= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2}
\end{aligned}$$

最后一个表达式中的 s_{xy} 是 x 和 y 的样本协方差，s_x^2 是 x 的样本方差。根据样本相关系数 r_{xy} 的定义，可知

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{s_{xy}}{s_x^2} \cdot \frac{s_x}{s_y} = \hat{\beta} \cdot \frac{s_x}{s_y}$$

因此 $\hat{\beta} = r_{xy} \cdot \frac{s_y}{s_x}$。

3. **在什么时候 $\hat{\beta}$ 的估计值和样本相关系数是一致的？利用蒙特卡洛仿真确认你的结论。**

从上一问的结果可知，当且仅当 $s_x = s_y \neq 0$ 时，$\hat{\beta} = r_{xy}$。然而 x 和 y 样本标准差是否相等完全取决于样本，在现实中是基本不可能发生的。但是，虽然我们无法干预 s_x 和 s_y，我们可以想办法对样本中的观测值进行线性变换，使变换后的标准差相等。最方便的做法是将 x 和 y 的观测值分别除以各自的样本标准差（前提是它们不为零），这样变换后的标准差都等于 1，此时再进行 OLS 估计，斜率的估计值就等于样本相关系数。换句话说，样本相关系数反映了 x 增加一个标准差时，对应 y 增加的标准差数量。

下面进行仿真验证。

```
n <- 1000
set.seed(12345)
x <- runif(n,0,10)
y <- 2 + 5*x + rnorm(n,0,2)

lm(y ~ x) |> coef()
(Intercept)           x
   1.980021    4.997353
```

斜率的 OLS 估计值为 4.997353，非常接近真实值 5。

```
cor(x,y)
[1] 0.9901136
```

样本相关系数和 OLS 估计值差别很大。下面对数据进行处理后再进行 OLS 估计。

```
xx <- x / sd(x)
yy <- y / sd(y)
lm(yy ~ xx) |> coef()
(Intercept)          xx
  0.1379578   0.9901136
```

此时估计值和样本相关系数一致。