Assignment 1

Rin Meng
Student ID: 51940633

September 23, 2024

1. a. The linear regression model can be written as,

$$P = \beta_0 + \beta_1 C + \epsilon$$

where,

- P is the prime interest rate
- C is the core inflation rate
- β_0 is the intercept of the regression line
- β_1 is the slope of the regression line
- ϵ is the error term

Assumptions of the linear regression model are,

   i. **Linearity:** The relationship between P and C is linear
   ii. **Independence:** The error term ϵ is independent of C
   iii. **Normality:** The error term ϵ is normally distributed
   iv. **Homoscedasticity:** The error term ϵ has a constant variance across all levels of C

2. **Remark:** from ch2A.pdf slide 9 and 13,

- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$
- $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
- $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$

Let e_i be the ith residual term. Then for each observation, we can see that the residual for each observation i is defined as:

$$e_i = y_i - \hat{y}_i$$

then we can say that the predicted value $\hat{y}_i$ is equivalent to

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

(a) $\sum_{i=1}^{n} e_i = 0$

Proof. LHS:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(y_i - \hat{y}_i)$$
$$= \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$
$$= \sum_{i=1}^{n}(y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)$$
$$= \sum_{i=1}^{n} y_i - n\bar{y} + \hat{\beta}_1 n\bar{x} - \hat{\beta}_1 \sum_{i=1}^{n} x_i$$
$$= n\bar{y} - n\bar{y} + \hat{\beta}_1 n\bar{x} - \hat{\beta}_1 n\bar{x}$$
$$= 0$$

$$LHS = 0 = RHS$$

$\therefore$ The sum of residuals e_i is zero. $\qquad\square$

(b) $\sum_{i=1}^{n} x_i e_i = 0$

Proof. LHS:

$$\sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x_i (y_i - \hat{y}_i)$$

$$= \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$= \sum_{i=1}^{n} (x_i y_i - x_i(\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i^2)$$

$$= \sum_{i=1}^{n} (x_i y_i - x_i \bar{y} + x_i \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i^2)$$

$$= \sum_{i=1}^{n} (x_i(y_i - \bar{y}) + \hat{\beta}_1(\bar{x} x_i - x_i^2))$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - \bar{x} \sum_{i=1}^{n} x_i \right)$$

$$= \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \right) - \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right)$$

$$= \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \right) - \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}\bar{x} - n\bar{x}\bar{x} + n\bar{x}\bar{x} \right)$$

$$= \left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} \right) - \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}\bar{x} + n\bar{x}\bar{x} \right)$$

$$= \left(\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i - \bar{x} \sum_{i=1}^{n} y_i + n\bar{x}\bar{y} \right) - \hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + n\bar{x}^2 \right)$$

$$= \sum_{i=1}^{n} (x_i y_i - \bar{y} x_i - \bar{x} y_i + \bar{x}\bar{y}) - \hat{\beta}_1 \sum_{i=1}^{n} (x_i^2 - 2\bar{x} x_i + \bar{x}^2)$$

$$= \sum_{i=1}^{n} (x_i(y_i - \bar{y}) + \bar{x}(y_i - \bar{y})) - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} (x_i + \bar{x})(y_i - \bar{y}) - \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= S_{xy} - \hat{\beta}_1 S_{xx}$$

$$= S_{xy} - \frac{S_{xy}}{S_{xx}} S_{xx}$$

$$= S_{xy} - S_{xy}$$

$$= 0$$

$$LHS = 0 = RHS$$

$$\square$$

$\therefore$ The residuals and the covariate are uncorrelated in simple linear regression.

(c) $\sum_{i=1}^{n} \hat{y}_i e_i = 0$

Proof. LHS:

$$\sum_{i=1}^{n} \hat{y}_i e_i = \sum_{i=1}^{n} \hat{y}_i (y_i - \hat{y}_i)$$

$$= \sum_{i=1}^{n} (y_i \hat{y}_i - \hat{y}_i^2)$$

$$= \sum_{i=1}^{n} (y_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_i)^2)$$

$$= \sum_{i=1}^{n} (y_i(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i) - (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i)^2)$$

$$= \sum_{i=1}^{n} (y_i \bar{y} - y_i \hat{\beta}_1 \bar{x} + y_i \hat{\beta}_1 x_i$$

$$- (\bar{y}^2 - 2\hat{\beta}_1 \bar{y}\bar{x} + \hat{\beta}_1^2 \bar{x}^2 + 2\hat{\beta}_1 \bar{y}x_i - 2\hat{\beta}_1^2 \bar{x}x_i + \hat{\beta}_1^2 x_i^2))$$

$$= (\sum_{i=1}^{n} y_i \bar{y} - \sum_{i=1}^{n} y_i \hat{\beta}_1 \bar{x} + \sum_{i=1}^{n} y_i \hat{\beta}_1 x_i$$

$$- \sum_{i=1}^{n} (\bar{y}^2 - 2\hat{\beta}_1 \bar{y}\bar{x} + \hat{\beta}_1^2 \bar{x}^2 + 2\hat{\beta}_1 \bar{y}x_i - 2\hat{\beta}_1^2 \bar{x}x_i + \hat{\beta}_1^2 x_i^2))$$

$$= (n\bar{y}^2 - \hat{\beta}_1 \bar{y}\bar{x} + \hat{\beta}_1 \bar{y}\hat{\beta}_1 x_i$$

$$- (\sum_{i=1}^{n} \bar{y}^2 - \sum_{i=1}^{n} 2\hat{\beta}_1 \bar{y}\bar{x} + \sum_{i=1}^{n} \hat{\beta}_1^2 \bar{x}^2 + \sum_{i=1}^{n} 2\hat{\beta}_1 \bar{y}x_i - \sum_{i=1}^{n} 2\hat{\beta}_1^2 \bar{x}x_i + \sum_{i=1}^{n} \hat{\beta}_1^2 x_i^2))$$

$$= (n\bar{y}^2 - (n\bar{y}^2 - 2\hat{\beta}_1 n\bar{y}\bar{x} + \hat{\beta}_1^2 n\bar{x}^2 + 2\hat{\beta}_1 n\bar{y}\bar{x} - 2\hat{\beta}_1^2 n\bar{x}\bar{x} + \hat{\beta}_1^2 \sum_{i=1}^{n} x_i^2))$$

proof continued.

$$\sum_{i=1}^{n} \hat{y}_i e_i = (n\bar{y}^2 - (n\bar{y}^2 - \hat{\beta}_1^{\,2} n\bar{x}^2 + \hat{\beta}_1^{\,2} n\bar{x}^2))$$
$$= (n\bar{y}^2 - n\bar{y}^2)$$
$$= 0$$

$\square$