**2.1** Consider the following five observations. You are to do all the parts of this exercise using only a calculator.

| $x$ | $y$ | $x-\bar{x}$ | $(x-\bar{x})^2$ | $y-\bar{y}$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|
| 3 | 4 | 2 | 4 | 2 | 4 |
| 2 | 2 | 1 | 1 | 0 | 0 |
| 1 | 3 | 0 | 0 | 1 | 0 |
| –1 | 1 | -2 | 4 | -1 | 2 |
| 0 | 0 | -1 | 1 | -2 | 2 |
| $\sum x_i =$ 5 | $\sum y_i =$ 10 | $\sum(x_i-\bar{x}) =$ 0 | $\sum(x_i-\bar{x})^2 =$ 10 | $\sum(y_i-\bar{y}) =$ 0 | $\sum(x_i-\bar{x})(y_i-\bar{y}) =$ 8 |

**a.** Complete the entries in the table. Put the sums in the last row. What are the sample means $\bar{x}$ and $\bar{y}$?
**b.** Calculate $b_1$ and $b_2$ using (2.7) and (2.8) and state their interpretation.
**c.** Compute $\sum_{i=1}^{5} x_i^2$, $\sum_{i=1}^{5} x_i y_i$. Using these numerical values, show that $\sum(x_i-\bar{x})^2 = \sum x_i^2 - N\bar{x}^2$ and $\sum(x_i-\bar{x})(y_i-\bar{y}) = \sum x_i y_i - N\bar{x}\bar{y}$.
**d.** Use the least squares estimates from part (b) to compute the fitted values of $y$, and complete the remainder of the table below. Put the sums in the last row.
Calculate the sample variance of $y$, $s_y^2 = \sum_{i=1}^{N}(y_i-\bar{y})^2/(N-1)$, the sample variance of $x$, $s_x^2 = \sum_{i=1}^{N}(x_i-\bar{x})^2/(N-1)$, the sample covariance between $x$ and $y$, $s_{xy} = \sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})/(N-1)$, the sample correlation between $x$ and $y$, $r_{xy} = s_{xy}/(s_x s_y)$ and the coefficient of variation of $x$, $CV_x = 100(s_x/\bar{x})$. What is the median, 50th percentile, of $x$?

$y_i-\hat{y}_i$ $(y_i-\hat{y}_i)^2$

| $x_i$ | $y_i$ | $\hat{y}_i$ | $\hat{e}_i$ | $\hat{e}_i^2$ | $x_i\hat{e}_i$ |
|---|---|---|---|---|---|
| 3 | 4 | 3.6 | 0.4 | 0.16 | 1.2 |
| 2 | 2 | 2.8 | -0.8 | 0.64 | -1.6 |
| 1 | 3 | 2 | 1 | 1 | 1 |
| –1 | 1 | 0.4 | 0.6 | 0.36 | -0.6 |
| 0 | 0 | 1.2 | -1.2 | 1.44 | 0 |
| $\sum x_i =$ 5 | $\sum y_i =$ 10 | $\sum \hat{y}_i =$ 10 | $\sum \hat{e}_i =$ 0 | $\sum \hat{e}_i^2 =$ 3.6 | $\sum x_i\hat{e}_i =$ 0 |

**e.** On graph paper, plot the data points and sketch the fitted regression line $\hat{y}_i = b_1 + b_2 x_i$.
**f.** On the sketch in part (e), locate the point of the means $(\bar{x}, \bar{y})$. Does your fitted line pass through that point? If not, go back to the drawing board, literally.
**g.** Show that for these numerical values $\bar{y} = b_1 + b_2\bar{x}$.
**h.** Show that for these numerical values $\bar{\hat{y}} = \bar{y}$, where $\bar{\hat{y}} = \sum \hat{y}_i/N$.
**i.** Compute $\hat{\sigma}^2$.
**j.** Compute $\widehat{\text{var}}(b_2|\mathbf{x})$ and $\text{se}(b_2)$.

a.

$\bar{x} = 1$ , $\bar{y} = 2$

b.

$$b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = 0.8$$

$$b_1 = \bar{y} - b_2 \bar{x} = 2 - 0.8 \times 1 = 1.2$$

c.

$$\sum x_i^2 = 15$$

$$\sum x_i y_i = 18$$

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - N\bar{x}^2 = 10$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - N\bar{x}\bar{y} = 8$$

d.

已知 $\hat{y}_i = 1.2 + 0.8 x_i$

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{N-1} = 2.5$$

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{N-1} = 2.5$$

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N-1} = 2$$

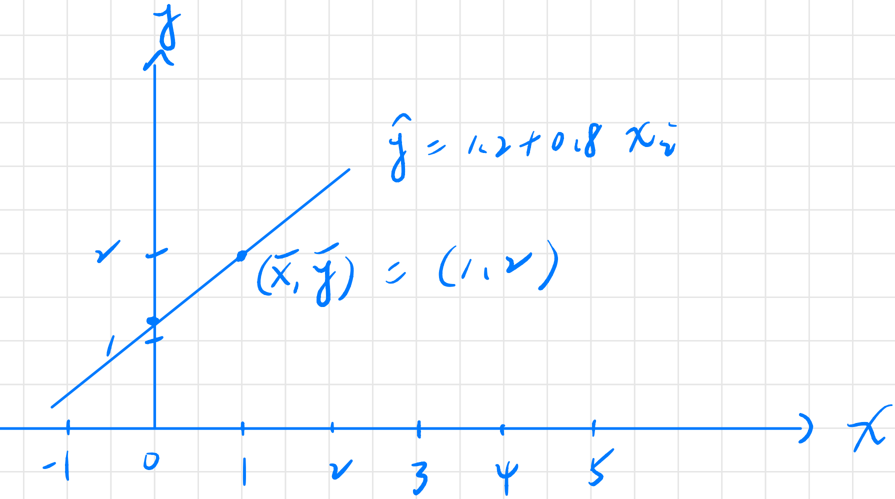
$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{2}{\sqrt{2.5 \times 2.5}} = 0.8$$

$$CV_x = 100\frac{s_x}{\bar{x}} = 100 \times \frac{\sqrt{2.5}}{1} = 50\sqrt{10}$$

$$\doteq 158.1$$

中位数 $= \bar{x} = 1$

e,f,




g,

$$\bar{y} = 2 = 1.2 + 0.8 \times 1 = b_1 + b_2\bar{x}$$

h,

$$\bar{\hat{y}} = \frac{\sum \hat{y}_i}{N} = \frac{3.6 + 2.8 + 2 + 0.4 + 1.2}{5}$$

$$= 2 = \bar{y}$$

i. $\hat{\sigma}^2 = \dfrac{\sum e_i^2}{N-2} = \dfrac{\sum (y_i - \hat{y}_i)^2}{3}$

$= 1.2$

j. $\widehat{Var}(b_2 | x) = \dfrac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = \dfrac{1.2}{10}$

$= 0.12$

$SE(b_2) = \sqrt{0.12} \doteq 0.346$

**2.14** Consider the regression model $WAGE = \beta_1 + \beta_2 EDUC + e$, where $WAGE$ is hourly wage rate in U.S. 2013 dollars and $EDUC$ is years of education, or schooling. The regression model is estimated twice using the least squares estimator, once using individuals from an urban area, and again for individuals in a rural area.

$$\text{Urban} \quad \underset{(\text{se})}{\widehat{WAGE}} = \underset{(2.27)}{-10.76} + \underset{(0.16)}{2.46}\, EDUC, \quad N = 986$$

$$\text{Rural} \quad \underset{(\text{se})}{\widehat{WAGE}} = \underset{(3.29)}{-4.88} + \underset{(0.24)}{1.80}\, EDUC, \quad N = 214$$

**a.** Using the estimated rural regression, compute the elasticity of wages with respect to education at the "point of the means." The sample mean of $WAGE$ is \$19.74.



**b.** The sample mean of $EDUC$ in the urban area is 13.68 years. Using the estimated urban regression, compute the standard error of the elasticity of wages with respect to education at the "point of the means." Assume that the mean values are "givens" and not random.

**c.** What is the predicted wage for an individual with 12 years of education in each area? With 16 years of education?

a. 已知 $E(\widehat{WAGE})_{rural} = 19.74$

$$\Rightarrow E(\widehat{WAGE}_{rural}) = -4.88 + 1.80\,E(EDUC)$$

$$\Rightarrow E(EDUC) = \frac{19.74 + 4.88}{1.80} = 13.68$$

$$\hat{\varepsilon}_{rural} = \text{斜率} \times \frac{\bar{x}}{\bar{y}} = 1.80 \times \frac{13.68}{19.74}$$

$$\doteqdot 1.25$$

b. 已知 $E(EDUC)_{urban} = 13.68$

$$E(\widehat{WAGE}_{urban}) = -10.76 + 2.46 \times 13.68$$

$$\doteqdot 22.89$$

$$SE(\hat{\varepsilon}_{urban}) = \frac{\sqrt{Var(\hat{\varepsilon}_{urban})}}{\sqrt{N}} = \frac{\sqrt{Var(b_{2_{urban}} \frac{\bar{x}}{\bar{y}})}}{\sqrt{986}}$$

$$= \frac{\bar{x}}{\bar{y}} SE(b_{r_{urban}}) = \frac{13.68}{22.89} \times 0.16$$

$$\doteq 0.096$$

c.

① $EDUC = 12$

$$\widehat{E(WAGE)}_{urban} = -10.76 + 2.46 \times 12$$

$$= 18.76$$

$$\widehat{E(WAGE)}_{rural} = -4.88 + 1.8 \times 12$$

$$= 16.72$$

② $EDUC = 16$

$$\widehat{E(WAGE)}_{urban} = -10.76 + 2.46 \times 16$$

$$= 28.6$$

$$\widehat{E(WAGE)}_{rural} = -4.88 + 1.8 \times 16$$

$$= 23.92$$