Econometrics Problem Sets
William M. Boal

Problem Set 2
"Introduction: Fitting Lines to Data"

(2.1) [Meaning of slope] Suppose the slope of the relationship between x and y is given by $\Delta y/\Delta x = 3$.

a. If x increases by 5 units, does y increase or decrease? By how many units?.
b. If x decreases by 2 units, does y increase or decrease? By how many units?
c. If we wish to increase y by 12 units, must x increase or decrease? By how many units?

(2.2) [Linear relationship] Suppose we have the relationship $y = 4 - 3x$. First suppose we graph this relationship with y on the vertical axis and x on the horizontal axis.

a. Give the slope.
b. Give the y-intercept.

Now suppose instead we graph this relationship with x on the vertical axis and y on the horizontal axis.

c. Give the new slope.
d. Give the x-intercept.

(2.3) [Least-squares calculation] Suppose the following three observations on x_i and y_i are given.

Observation (i)	x_i	y_i
1	2	10
2	1	2
3	3	6

a. Compute the sample means of x and y .
b. Compute the following sums.

$$\sum_{i=1}^{3}(x_i - \bar{x})^2 \quad , \quad \sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})$$

c. Compute $\hat{\beta}_2$, the least-squares estimate of the slope of the line $y_i = \beta_1 + \beta_2 x_i$.
d. Compute $\hat{\beta}_1$, the least-squares estimate of the intercept of the same line.
e. Compute the three fitted values $\hat{y}_i$ and the three residuals $\hat{\varepsilon}_i$ of this estimated least-squares regression line.

(2.4) [Least-squares calculation] Suppose the following three observations on x_i and y_i are given.

Observation (i)	x_i	y_i
1	3	3
2	2	7
3	4	5

a. Compute the sample means of $\bar{x}$ and $\bar{y}$.
b. Compute the following sums.

$$\sum_{i=1}^{3}(x_i - \bar{x})^2 \quad , \qquad \sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})$$

c. Compute $\hat{\beta}_2$, the least-squares estimate of the slope of the line $y_i = \beta_1 + \beta_2 x_i$.
d. Compute $\hat{\beta}_1$, the least-squares estimate of the intercept of the same line.
e. Compute the three fitted values $\hat{y}_i$ and the three residuals $\hat{\varepsilon}_i$ of this estimated least-squares regression line.

(2.5) [Least-squares calculation] Suppose the following three observations on x_i and y_i are given.

Observation (i)	x_i	y_i
1	3	12
2	1	6
3	2	0

a. Compute the sample means of $\bar{x}$ and $\bar{y}$.
b. Compute the following sums.

$$\sum_{i=1}^{3}(x_i - \bar{x})^2 \quad , \qquad \sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})$$

c. Compute $\hat{\beta}_2$, the least-squares estimate of the slope of the line $y_i = \beta_1 + \beta_2 x_i$.
d. Compute $\hat{\beta}_1$, the least-squares estimate of the intercept of the same line.
e. Compute the three fitted values $\hat{y}_i$ and the three residuals $\hat{\varepsilon}_i$ of this estimated least-squares regression line.

(2.6) [Least-squares calculation] Suppose the following four observations on x_i and y_i are given.

Observation (i)	x_i	y_i
1	2	5
2	4	10
3	4	8
4	6	13

a. Compute the sample means of $\bar{x}$ and $\bar{y}$.
b. Compute the following sums.

$$\sum_{i=1}^{4}(x_i - \bar{x})^2 \quad , \quad \sum_{i=1}^{4}(x_i - \bar{x})(y_i - \bar{y})$$

c. Compute $\hat{\beta}_2$, the least-squares estimate of the slope of the line $y_i = \beta_1 + \beta_2 x_i$.
d. Compute $\hat{\beta}_1$, the least-squares estimate of the intercept of the same line.
e. Compute the four fitted values $\hat{y}_i$ and the four residuals $\hat{\varepsilon}_i$ of this estimated least-squares regression line.

(2.7) [Least squares calculation] Suppose the following eight observations on x_i and y_i are given.

Observation (i)	x_i	y_i
1	0	1
2	1	2
3	1	4
4	2	4
5	2	6
6	3	5
7	3	9
8	4	9

a. Compute the sample means of $\bar{x}$ and $\bar{y}$.
b. Compute the following sums.

$$\sum_{i=1}^{8}(x_i - \bar{x})^2 \quad , \quad \sum_{i=1}^{8}(x_i - \bar{x})(y_i - \bar{y})$$

c. Compute $\hat{\beta}_2$, the least-squares estimate of the slope of the line $y_i = \beta_1 + \beta_2 x_i$.
d. Compute $\hat{\beta}_1$, the least-squares estimate of the intercept of the same line.
e. Compute the eight fitted values $\hat{y}_i$ and the eight residuals $\hat{\varepsilon}_i$ of this estimated least-squares regression line.

(2.8) [Least-squares principle] Suppose the line $x_i = \beta_1 + \beta_2 x_i$ were estimated by least squares by mistake. Note that x_i appears on both sides of the equation!

a. What would be the least-squares estimate of β_1 ?
b. What would be the least-squares estimate of β_2 ?
c. What would be the value of each residual? Why?

(2.9) [Least-squares principle] Suppose we estimate the line $y_i = \beta_1 + \beta_2 x_i$ using a sample of a dozen or so observations. Answer each question below and explain your reasoning.

a. Suppose all the observations had the same value for x. What would be the value of the least-squares estimate of the slope β_2 ? [Hint: In this case, $x_i = \bar{x}$ for the entire sample.] What would be the value of the least-squares estimate of the intercept β_1? Why?
b. Suppose all the observations had the same value for y. What would be the value of the least-squares estimate of the slope β_2 ? [Hint: In this case, $y_i = \bar{y}$ for the entire sample.] What would be the value of the least-squares estimate of the intercept β_1? Why?

(2.10) [Least-squares principle] Consider the problem of estimating the line $y_i = \beta_1 + \beta_2 x_i$ using a small sample. Answer each question below and explain your reasoning.

a. Suppose the sample consisted of only *one* observation. Could the method of least squares estimate the intercept and slope? If so, describe what a graph of the estimated regression line would look like. If not, explain why not.
b. Suppose the sample consisted of only *two* observations. Could the method of least squares estimate the intercept and slope? If so, describe what a graph of the estimated regression line would look like. If not, explain why not.

(2.11) [Least-squares principle] Suppose we estimate the line $y_i = \beta_1 + \beta_2 x_i$ using a sample of several dozen observations. Suppose we graph our sample as a scatter-plot with y on the vertical axis and x on the horizontal axis, and then graph our least-squares estimated regression line on the same scatter-plot. Answer each question below and explain your reasoning.

a. Could the least-squares estimated regression line ever pass *below* all the observations? If not, why not?
b. Could the least-squares estimated regression line ever pass *above* all the observations? If not, why not?

(2.12) [Least-squares principle] True or false? Explain your reasoning.

a. The least-squares estimate of the intercept is necessarily positive.
b. The least-squares estimate of the slope is necessarily positive.
c. The least-squares estimated regression line necessarily passes through the point ($\bar{x}$, $\bar{y}$). [Hint: Review the definition of least-squares.]

(2.13) [Least-squares principle] Look at the formulas for the least-squares estimate of the slope and the intercept. Consider what would happen if the order of the observations were changed. For example, suppose observation #1 became observation #2 in the dataset, observation #2 became observation #3, and so forth until observation n became observation #1. Note that the number of observation has not changed and the data have not changed—just the order of observations.

a. Would the least-squares estimate of the slope change? Why or why not?
b. Would the least-squares estimate of the intercept change? Why or why not?

(2.14) [Alternatives to least-squares] You have asked your research assistant to compute least-squares estimates, reverse least-squares estimates, and least-absolute-deviation estimates for the following data.

x_i	y_i
1	2
2	6
3	4

Your research assistant sends you three estimated equations, but does not tell you which is which. The three equations are

(i) y = 1 + x.
(ii) y = -4 + 4x.
(iii) y = 2 + x.

a. For each equation, substitute the three values of x in the table above to compute the three fitted values $\hat{y}$. Then compute the sum of squared residuals in the vertical direction $\sum_{i=1}^{3}(y_i - \hat{y})^2$. Which equation was apparently fitted by the method of *least squares*?
b. For each equation, use the three fitted values you computed in part (a) to compute the sum of the absolute deviations in the vertical direction $\sum_{i=1}^{3}|y_i - \hat{y}|$. Which equation was apparently fitted by the method of *least-absolute-deviation*?
c. For each equation, solve for x on the left-hand side. Then substitute the three values of y in the table above to compute the three fitted values $\hat{x}$. Then compute the sum of squared residuals in the horizontal direction $\sum_{i=1}^{3}(x_i - \hat{x})^2$. Which equation was apparently fitted by the method of *reverse least squares*?

(2.15) [Alternatives to least-squares] Suppose the line $x_i = \beta_1 + \beta_2 x_i$ were estimated mistake. Note that x_i appears on both sides of the equation!

a. If the equation were estimated by the method of reverse least squares, what would be the estimates of β_1 and β_2 ? Why?
b. If the equation were estimated by the method of least absolute devation, what would be the estimates of β_1 and β_2 ? Why?

(2.16) [Alternatives to least-squares—trick question!] Suppose the following five observations on x_i and y_i are given.

Observation (i)	x_i	y_i
1	1	5
2	3	9
3	2	7
4	0	3
5	4	11

Suppose we wish to fit a line of the form $y_i = \beta_1 + \beta_2 x_i$ to these data. [Hint: Plot the data before answering the questions below.]

a. If possible, plot the data on a graph, with x on the horizontal axis and y on the vertical axis.
b. Compute the least-squares estimates of β_1 and β_2 .
c. Compute the least-absolute-deviation estimates of β_1 and β_2 .
d. Compute the reverse-least-squares estimates of β_1 and β_2 .
e. What is strange about these data that causes the peculiar relationship between the least-squares, least-absolute deviation, and reverse-least squares estimates?

[end of problem set]