

The Least-Squares Criterion Now that we have obtained what we call the “best fit” line for describing the relationship between our two variables, we need to determine by what criterion it is considered best. Before the criterion is stated, let us examine Figure 9.3.3. We note that generally the least-squares line does not pass through the observed points that are plotted on the scatter diagram. In other words, most of the observed points *deviate* from the line by varying amounts.

The line that we have drawn through the points is best in this sense:

The sum of the squared vertical deviations of the observed data points (y_i) from the least-squares line is smaller than the sum of the squared vertical deviations of the data points from any other line.

In other words, if we square the vertical distance from each observed point (y_i) to the least-squares line and add these squared values for all points, the resulting total will be smaller than the similarly computed total for any other line that can be drawn through the points. For this reason the line we have drawn is called the least-squares line.

EXERCISES

- 9.3.1** Plot each of the following regression equations on graph paper and state whether X and Y are directly or inversely related.

(a) $\hat{y} = -3 + 2x$

(b) $\hat{y} = 3 + 0.5x$

(c) $\hat{y} = 10 - 0.75x$

- 9.3.2** The following scores represent a nurse’s assessment (X) and a physician’s assessment (Y) of the condition of 10 patients at time of admission to a trauma center.

X :	18	13	18	15	10	12	8	4	7	3
Y :	23	20	18	16	14	11	10	7	6	4

- (a) Construct a scatter diagram for these data.

- (b) Plot the following regression equations on the scatter diagram and indicate which one you think best fits the data. State the reason for your choice.

(1) $\hat{y} = 8 + 0.5x$

(2) $\hat{y} = -10 + 2x$

(3) $y = 1 + 1x$

For each of the following exercises (a) draw a scatter diagram and (b) obtain the regression equation and plot it on the scatter diagram.

- 9.3.3** Methadone is often prescribed in the treatment of opioid addiction and chronic pain. Krantz et al. (A-2) studied the relationship between dose of methadone and the corrected QT (QTc) interval for 17 subjects who developed *torsade de pointes* (ventricular tachycardia nearly always due to medications). QTc is calculated from an electrocardiogram and is measured in mm/sec. A higher QTc value indicates a higher risk of cardiovascular mortality. A question of interest is how well one can predict and estimate the QTc value from a knowledge of methadone dose. This question is typical of those that can be answered by means of regression analysis. Since QTc is the variable

about which we wish to make predictions and estimations, it is the dependent variable. The variable methadone dose, knowledge of which will be used to make the predictions and estimations, is the independent variable.

Methadone Dose		Methadone Dose	
(mg/day)	QTc (mm/sec)	(mg/day)	QTc (mm/sec)
1000	600	650	785
550	625	600	765
97	560	660	611
90	585	270	600
85	590	680	625
126	500	540	650
300	700	600	635
110	570	330	522
65	540		

Source: Mori J. Krantz, Ilana B. Kutinsky, Alastair D. Roberston, and Philip S. Mehler, "Dose-Related Effects of Methadone on QT Prolongation in a Series of Patients with Torsade de Pointes," *Pharmacotherapy*, 23 (2003), 802–805.

- 9.3.4** Reiss et al. (A-3) compared point-of-care and standard hospital laboratory assays for monitoring patients receiving a single anticoagulant or a regimen consisting of a combination of anticoagulants. It is quite common when comparing two measuring techniques, to use regression analysis in which one variable is used to predict another. In the present study, the researchers obtained measures of international normalized ratio (INR) by assay of capillary and venous blood samples collected from 90 subjects taking warfarin. INR, used especially when patients are receiving warfarin, measures the clotting ability of the blood. Point-of-care testing for INR was conducted with the CoaguChek assay product. Hospital testing was done with standard hospital laboratory assays. The authors used the hospital assay INR level to predict the CoaguChek INR level. The measurements are given in the following table.

CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)
1.8	1.6	2.4	1.2	3.1	2.4
1.6	1.9	2.3	2.3	1.7	1.8
2.5	2.8	2.0	1.6	1.8	1.6
1.9	2.4	3.3	3.8	1.9	1.7
1.3	1.5	1.9	1.6	5.3	4.2
2.3	1.8	1.8	1.5	1.6	1.6
1.2	1.3	2.8	1.8	1.6	1.4
2.3	2.4	2.5	1.5	3.3	3.3
2.0	2.1	0.8	1.0	1.5	1.5
1.5	1.5	1.3	1.2	2.2	2.8
2.1	2.4	3.7	1.4	1.1	1.6
1.5	1.5	2.4	1.6	2.6	2.6

(Continued)

CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)
1.5	1.7	4.1	3.2	6.4	5.0
1.8	2.1	2.4	1.2	1.5	1.4
1.0	1.2	2.3	2.3	3.0	2.8
2.1	1.9	3.1	1.6	2.6	2.3
1.6	1.6	1.5	1.4	1.2	1.2
1.7	1.6	3.6	2.1	2.1	1.9
2.0	1.9	2.5	1.7	1.1	1.1
1.8	1.6	2.1	1.7	1.0	1.0
1.3	4.1	1.8	1.2	1.4	1.5
1.5	1.9	1.5	1.3	1.7	1.3
3.6	2.1	2.5	1.1	1.2	1.1
2.4	2.2	1.5	1.2	2.5	2.4
2.2	2.3	1.5	1.1	1.2	1.3
2.7	2.2	1.6	1.2	2.5	2.9
2.9	3.1	1.4	1.4	1.9	1.7
2.0	2.2	4.0	2.3	1.8	1.7
1.0	1.2	2.0	1.2	1.2	1.1
2.4	2.6	2.5	1.5	1.3	1.1

Source: Data provided courtesy of Curtis E. Haas, Pharm.D.

- 9.3.5** Digoxin is a drug often prescribed to treat heart ailments. The purpose of a study by Parker et al. (A-4) was to examine the interactions of digoxin with common grapefruit juice. In one experiment, subjects took digoxin with water for 2 weeks, followed by a 2-week period during which digoxin was withheld. During the next 2 weeks subjects took digoxin with grapefruit juice. For seven subjects, the average peak plasma digoxin concentration (Cmax) when taking water is given in the first column of the following table. The second column contains the percent change in Cmax concentration when subjects were taking the digoxin with grapefruit juice [GFJ (%) change]. Use the Cmax level when taking digoxin with water to predict the percent change in Cmax concentration when taking digoxin with grapefruit juice.

Cmax (ngl/ml) with Water	Change in Cmax with GFJ (%)
2.34	29.5
2.46	40.7
1.87	5.3
3.09	23.3
5.59	-45.1
4.05	-35.3
6.21	-44.6
2.34	29.5

Source: Data provided courtesy of Robert B. Parker, Pharm.D.

- 9.3.6** Evans et al. (A-5) examined the effect of velocity on ground reaction forces (GRF) in dogs with lameness from a torn cranial cruciate ligament. The dogs were walked and trotted over a force

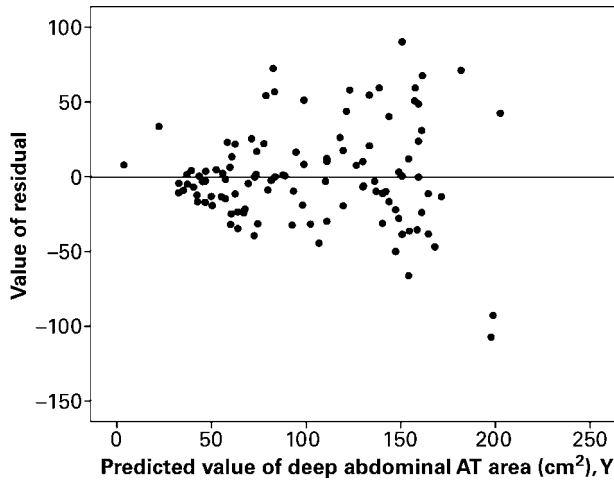


FIGURE 9.4.8 Residual plot of data from Example 9.3.1.

Many computer packages will provide residual plots automatically. These plots often use standardized values (i.e., e_i/\sqrt{MSE}) of the residuals and predicted values, but are interpreted in the same way as are plots of unstandardized values.

EXAMPLE 9.4.3

Refer to Example 9.3.1. We wish to use residual plots to test the assumptions of linearity and equal variances in the data.

Solution: A residual plot is shown in Figure 9.4.8.

Since there is a relatively equal and random scatter of points above and below the residual $(y_i - \hat{y}_i) = 0$ line, the linearity assumption is presumed to be valid. However, the funneling tendency of the plot suggests that as the predicted value of deep abdominal AT area increases, so does the amount of error. This indicates that the assumption of equal variances may not be valid for these data. ■

EXERCISES

9.4.1 to 9.4.5 Refer to Exercises 9.3.3 to 9.3.7, and for each one do the following:

- Compute the coefficient of determination.
- Prepare an ANOVA table and use the F statistic to test the null hypothesis that $\beta_1 = 0$. Let $\alpha = .05$.
- Use the t statistic to test the null hypothesis that $\beta_1 = 0$ at the .05 level of significance.
- Determine the p value for each hypothesis test.
- State your conclusions in terms of the problem.
- Construct the 95 percent confidence interval for β_1 .