Section 4.3

Diagnostics on the Least-Square Regression Line

Coefficient of Determination (R^2)

- Measures the proportion of total variation in the response variable that is explained by the least-squares regression line
- Note: R is in the range: $0 \leq R^2 \leq 1$. If it is equal to 0, the least square regression line has no explanatory value. if it is equal to 1, the least square regression line explains 100% of the variation in the response variable

Deviations

$$\text{Explained Deviation} = \hat{y} - \bar{y}$$

$$\text{Unexplained Deviation} = y - \hat{y}$$

$$\text{Total Deviation} = y - \bar{y}$$

Interpretation

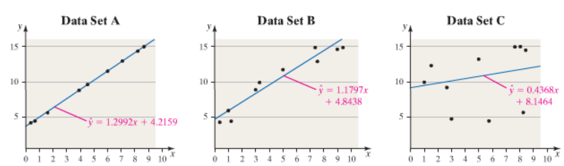
- Consider if R^2 = 90%, we would say 90% of the variation in distance is explained by the least-squares regression line and 10% of the variation in distance is explained by other factors.
- The smaller the sum of squared residuals, the smaller the unexplained variation and therefore the larger R^2

Understanding R^2

TABLE 7

Data Set A		Data Set B		Data Set C	
x	y	x	y	x	y
3.6	8.9	3.1	8.9	2.8	8.9
8.3	15.0	9.4	15.0	8.1	15.0
0.5	4.8	1.2	4.8	3.0	4.8
1.4	6.0	1.0	6.0	8.3	6.0
8.2	14.9	9.0	14.9	8.2	14.9
5.9	11.9	5.0	11.9	1.4	11.9
4.3	9.8	3.4	9.8	1.0	9.8
8.3	15.0	7.4	15.0	7.9	15.0
0.3	4.7	0.1	4.7	5.9	4.7
6.8	13.0	7.5	13.0	5.0	13.0

Understanding R^2



Understanding R^2

Data Set	Coefficient of Determination, R^2	Interpretation
A	99.99%	99.99% of the variability in y is explained by the least-squares regression line.
B	94.7%	94.7% of the variability in y is explained by the least-squares regression line.
C	9.4%	9.4% of the variability in y is explained by the least-squares regression line.

Finding Coefficient of Determination (R^2)

1. Put x values into L1
2. Put y values into L2
3. “Stat” button
4. Right arrow to CALC
5. Down arrow to LinReg (ax + b)
6. “enter” button

* Make sure Diagnostics is on

1. Find the coefficient of determination (by hand and TI-83/84)

X	Y
2	17
4	30
7	41
11	44

Residual Plot

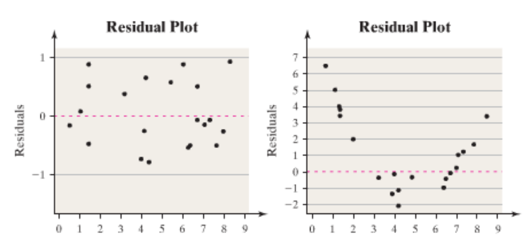
Residual Plot = a scatter diagram with the residuals on the vertical axis and the explanatory variable on the horizontal axis.

Note: If a plot of the residuals against the explanatory variable shows a discernible pattern, such as a curve, then explanatory and response variable may not be linearly related

Note on Linear Models

r by itself does not indicate whether you can use a linear model! Have to look at the residual plot also.

Examples



(a) Linear Model Appropriate

(b) Linear Model Not Appropriate: Patterned Residuals

Residual Plots (TI-83/84)

1. Put x values in L1, y values in L2
2. Stat->Calc->4:LinReg
3. “2nd” “Y=“, choose PLOT1, then choose:

ON

First Graph

Ylist: RESID (2nd Stat->NAMES->RESID)

4. Zoom->ZoomStat

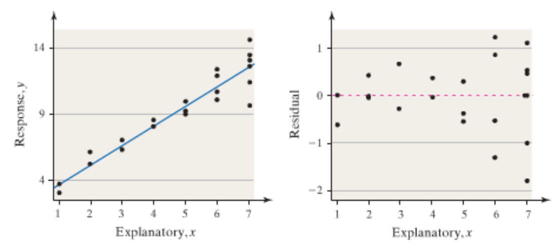
2. Find the residual plots (by hand and TI-83/84)

X	Y
2	17
4	30
7	41
11	50
13	70
17	92

Constant Error Variance

If a plot of the residuals against the explanatory variable shows the spread of the residuals increasing or decreasing as the explanatory variable increases, then a strict requirement of the linear model is violated. This requirement is called constant error variance.

Example



Influential Observations

- An observation that significantly affects the least-squares regression line’s slope and/or y-intercept, or the value of the correlation coefficient.

Note: Influential observations typically exist when the point is an outlier relative to the values of the explanatory variable

Example