Section 4.2

Least Squares Regression

Finding Linear Equation that Relates x and y values together Based on Two Points (Algebra)

1. Pick two data points (first and last maybe), these will be (x_1, y_1) and (x_2, y_2)
2. Find m using the formula: $m = \frac{y_2 - y_1}{x_2 - x_1}$
3. Into y = mx + b, plug in either point for x, y and m from step 2 and solve for b
4. Write answer (plug in m from step 2 and the b from step 3 into y = mx + b)

1. Create the scatter diagram and pick two "good" points and find the equation of the line containing them

X	Y
1	13
2	20
3	35
4	41
5	40

Definitions

residual = Difference between the observed and predicted values of y (aka "error")

Formula:

Residual = observed y – predicted y

Least-Squares Regression Criterion

The least - squares regression line is the line that minimizes the sum of the squared errors (or residuals). This line minimizes the sum of the squared vertical distance between the observed values of y and those predicted by the line, $\hat{y}$ (read "y - hat")

We represent this as "minimize Σ residuals^2"

Least-Squares Regression Line (By Hand)

$$\hat{y} = b_1 x + b_0$$

where

$$b_1 = r \cdot \frac{s_y}{s_x} \qquad \text{(slope)}$$

$$b_0 = \bar{y} - b_1 \bar{x} \qquad \text{(y - intercept)}$$

Finding Regression Equation (TI-83/84)

1. Put x values (explanatory) into L1
2. Put y values (response) 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

2. Find the least-squares regression equation (by hand and TI-83/84)

X	Y
1	10
2	15
8	35
13	44

Using Regression Equation for Predictions

1. Find the regression equation : $\hat{y}$
2. Plug in the x value that you are wanting to predict and simplify

3. Using the following data and its corresponding regression equation, predict y when x is equal to 21

X	Y
3	17
5	23
7	41
9	50

Residuals

The difference between the observed value of y and the predicted value of y aka error.

formula:
Residual = Observed - Predicted

$$\text{residual} = y - \hat{y}$$

$$\text{sum of squared residuals} = \Sigma(y - \hat{y})^2$$

4. Based on the least-squares regression line below, find the residual at x = 3 given the actual data point below:

$$\hat{y} = 3.1x - 0.05$$

$$\text{actual data point} : (3, 9)$$

5. Find the sum of the squared residuals for the least-squares regression line using the following data

X	Y
3	17
5	23
7	41
9	50