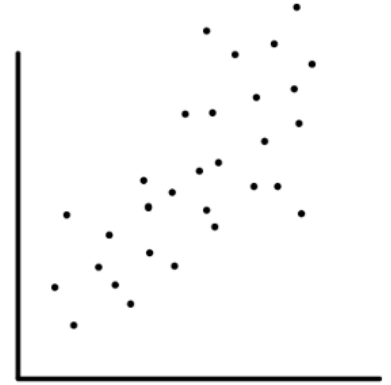


“Best Fit Line” = “Regression Line” = “Line of Best Fit” = “Least Squares Regression Line”

The regression line makes the sum of the squared vertical distances from each data point to the line as small as possible.

Ex: The following information gives the amount of time (min) students studied for a test and the grade they received.

Time	15	45	20	60	30	75	40	90
Grade	67	86	70	92	74	98	81	96



- (a) Make a scatterplot on the calculator. (b) Describe the form, direction, and strength.
- (c) Find the correlation. What does it mean?
- (d) Find the equation of the “best fit line”.
- (e) What are the slope and y-intercept? What do they mean in terms of the problem?
- (f) If another student was added to the data who studied for 10 minutes and earned an 86, what would happen to the correlation? Why? (Delete this value before doing the next part!)
- (g) If a student studies for 50 minutes, predict her score on the test.
- (h) If a student earned a 78, estimate how long she studied.
- (i) If a student studies for 140 minutes, predict her grade. What seems wrong about this prediction.

** You can also find the regression equation by using the following information. These formulas are given on the AP Exam (and on your test). Use these formulas and the summary statistics below to find the regression equation.

$$\hat{y} = a + bx$$

$$b = r \frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

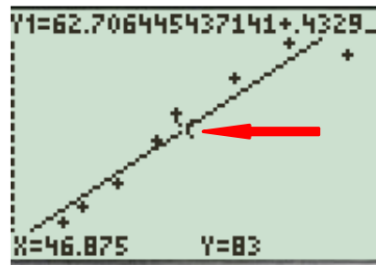
1-Var Stats
$\bar{x}=46.875$
$\Sigma x=375$
$\Sigma x^2=22475$
$Sx=26.44907506$
$\sigma x=24.74084427$
$\downarrow n=8$

Study Time (x)

1-Var Stats
$\bar{x}=83$
$\Sigma x=664$
$\Sigma x^2=56106$
$Sx=11.91637529$
$\sigma x=11.1467484$
$\downarrow n=8$

Grade (y)

** The least squares regression line ALWAYS passes through the point (\bar{x}, \bar{y}) , which is (46.875, 83) for this problem.

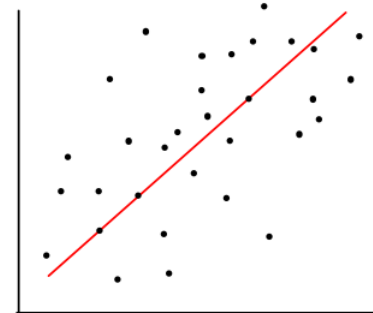


R^2 – the proportion/percent of variation in the y values that is explained by the regression model with the x values.

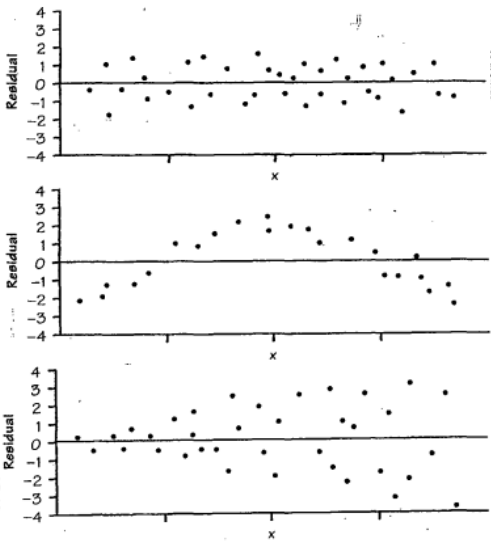
- Square the correlation (r) to obtain this statistic.
- Values will only range from 0 to 1.
- Measures the accuracy of our predictions.

Residual – the difference between the actual value and the predicted value.

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



Residual Plots



Ex: Mean Height of Kalama Children Problem

Age x in months	Height y in centimeters
18	76.1
19	77.0
20	78.1
21	78.2
22	78.8
23	79.7
24	79.9
25	81.1
26	81.2
27	81.8
28	82.8
29	83.5

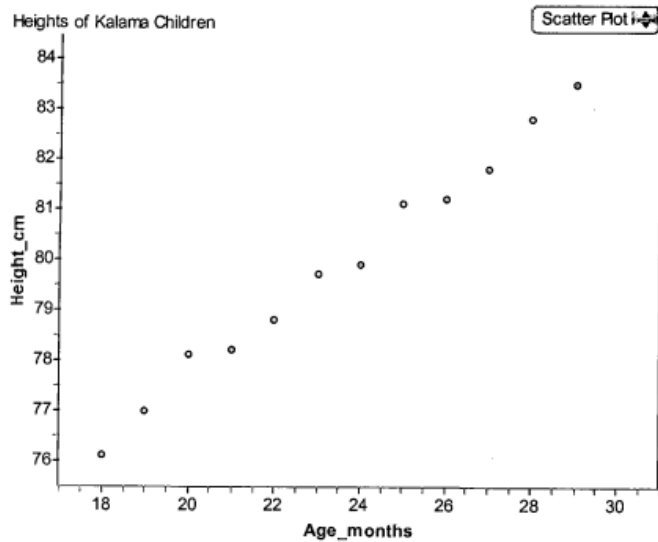


FIGURE 2.12 Mean height of children in Kalama, Egypt, plotted against age from 18 to 29 months, from Table 2.5.