ordinary least squares estimates

TIPS FOR SERVICE

Let's assume that you are a small restaurant owner or a very business minded server / waiter at a nice restaurant. Here in the U.S. "tips" are a very important part of a waiter's pay. Most of the time the dollar amount of tip is related to the dollar amount of the total bill.

As the waiter or owner, you would like to develop a model that will allow you to make a prediction about what amount of tip to expect. Therefore one evening, you collect data for six meals.

TIPS FOR SERVICE

Let's assume that you are a small restaurant owner or a very business minded server / waiter at a nice restaurant. Here in the U.S. "tips" are a very important part of a waiter's pay. Most of the time the dollar amount of tip is related to the dollar amount of the total bill.

As the waiter or owner, you would like to develop a model that will allow you to make a prediction about what amount of tip to expect. Therefore one evening, you collect data for six meals.

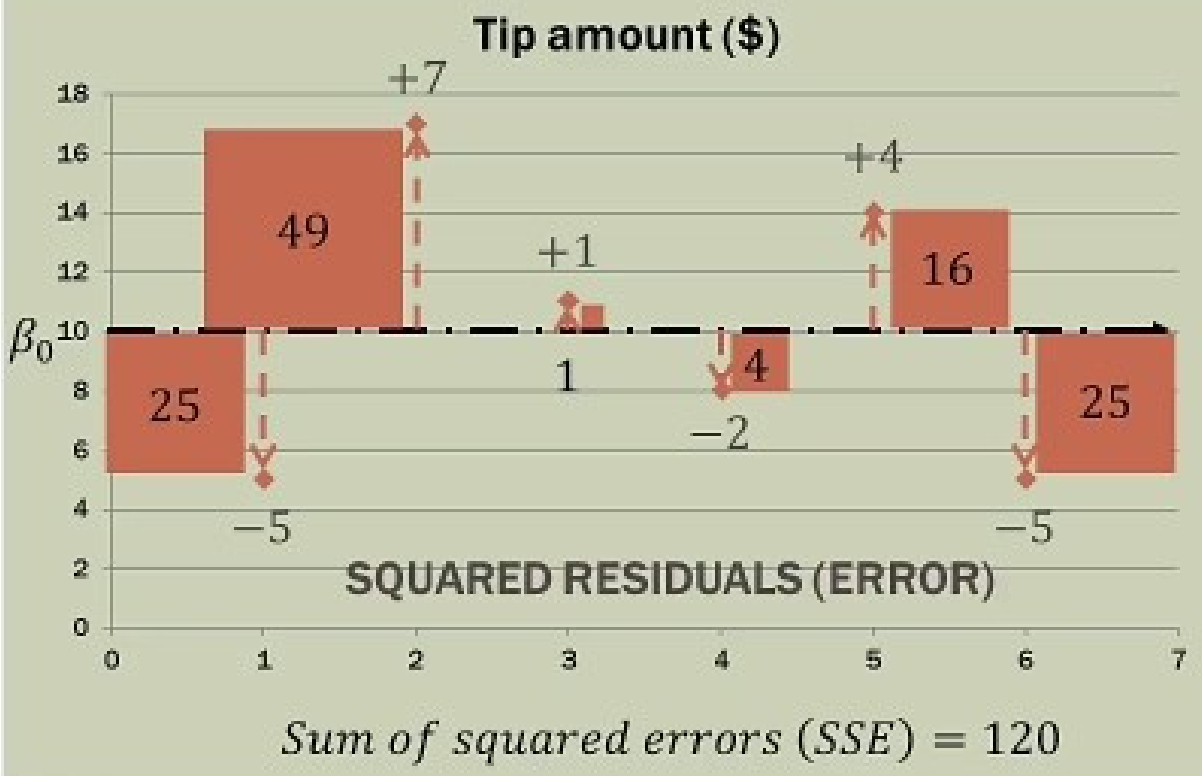
TIPS FOR SERVICE

In the last video, you only had the tip data. But now you were able to go back and get the actual bill data as well. So now you are working with two variables that are matched pairs.

You want to know to what degree the tip amount can be predicted by the bill. So the tip is the DEPENDENT variable; bill is the INDEPENDENT variable.

Total bill ($)	Tip amount ($)
34.00	5.00
108.00	17.00
64.00	11.00
88.00	8.00
99.00	14.00
51.00	5.00

WHEN THE SLOPE, $\beta_1 = 0$



Sum of squared errors (SSE) = 120

When conducting simple linear regression with TWO variables, we will determine how good the regression line "fits" the data **by comparing it to THIS TYPE; where we pretend the second variable does not even exist; $\beta_1 = 0$.**

In this situation, the value of $\hat{y}$ is 10 for every value of x.

$\hat{y} = b_0 + b_1 x$ $\quad b_0 = 10$

$\hat{y} = b_0 + (0)x$ $\quad \hat{y} = 10$

$\hat{y} = b_0$

LEAST SQUARES CRITERION

$$\min \sum (y_i - \hat{y}_i)^2$$

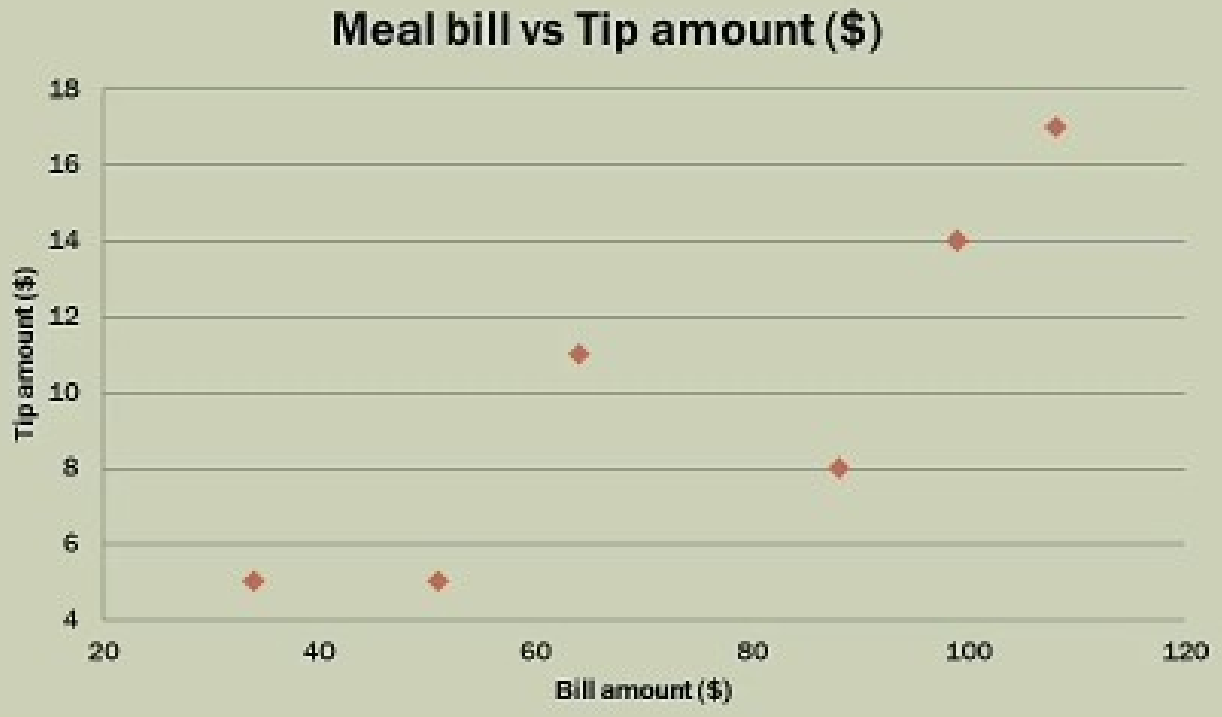
y_i = observed value of dependent variable (tip amount)

$\hat{y}_i$ = estimated(predicted)value of the dependent variable (predicted tip amount)

Plain English. The goal is to minimize the sum of the squared differences between the observed value for the dependent variable (y_i) and the estimated/predicted value of the dependent variable $(\hat{y}_i)$ that is provided by the regression line. Sum of the squared residuals.

Not only that, but the sum of the squared residuals should be much smaller than when we just used the dependent variable alone; $\beta_1 = 0, \hat{y} = 10$ for all values of x. That sum of squared residuals was 120.

STEP 1: SCATTER PLOT



Bill ($)	Tip ($)
34.00	5.00
108.00	17.00
64.00	11.00
88.00	8.00
99.00	14.00
51.00	5.00

STEP 2: LOOK FOR A VISUAL LINE

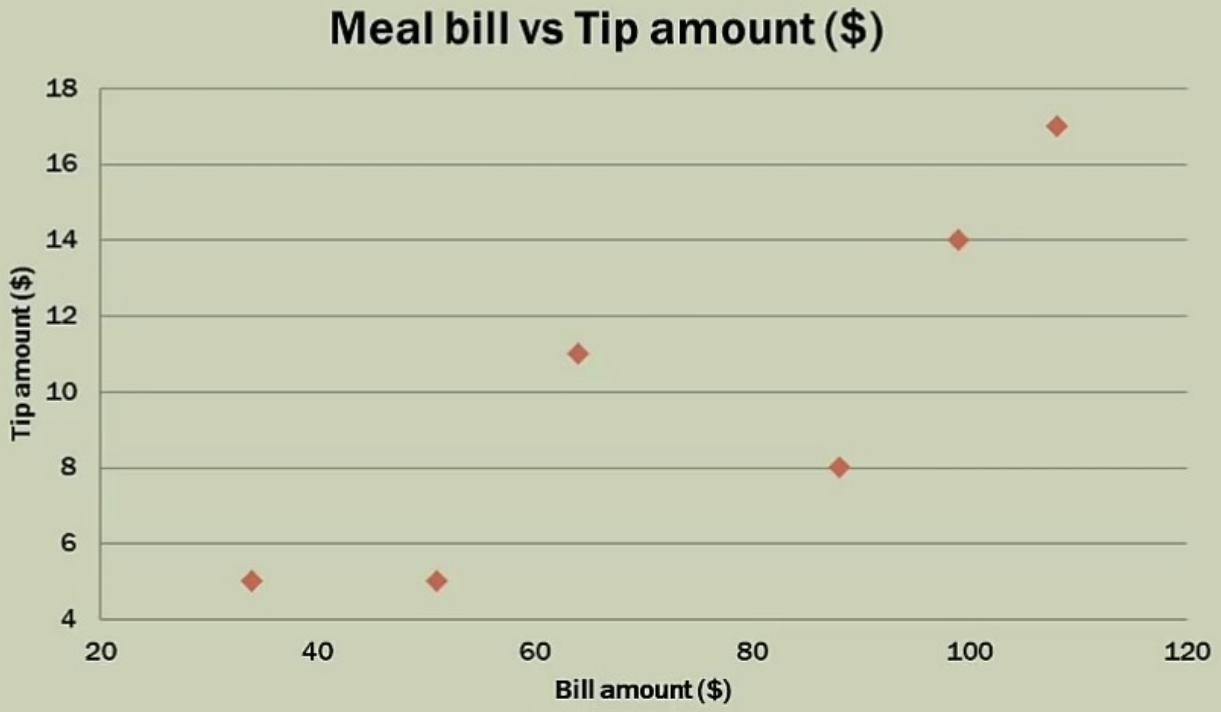


Does the data seem to fall along a line?

In this case, YES! Proceed.

If not…if it's a BLOB with no linear pattern, then stop.

STEP 3: CORRELATION (OPTIONAL)



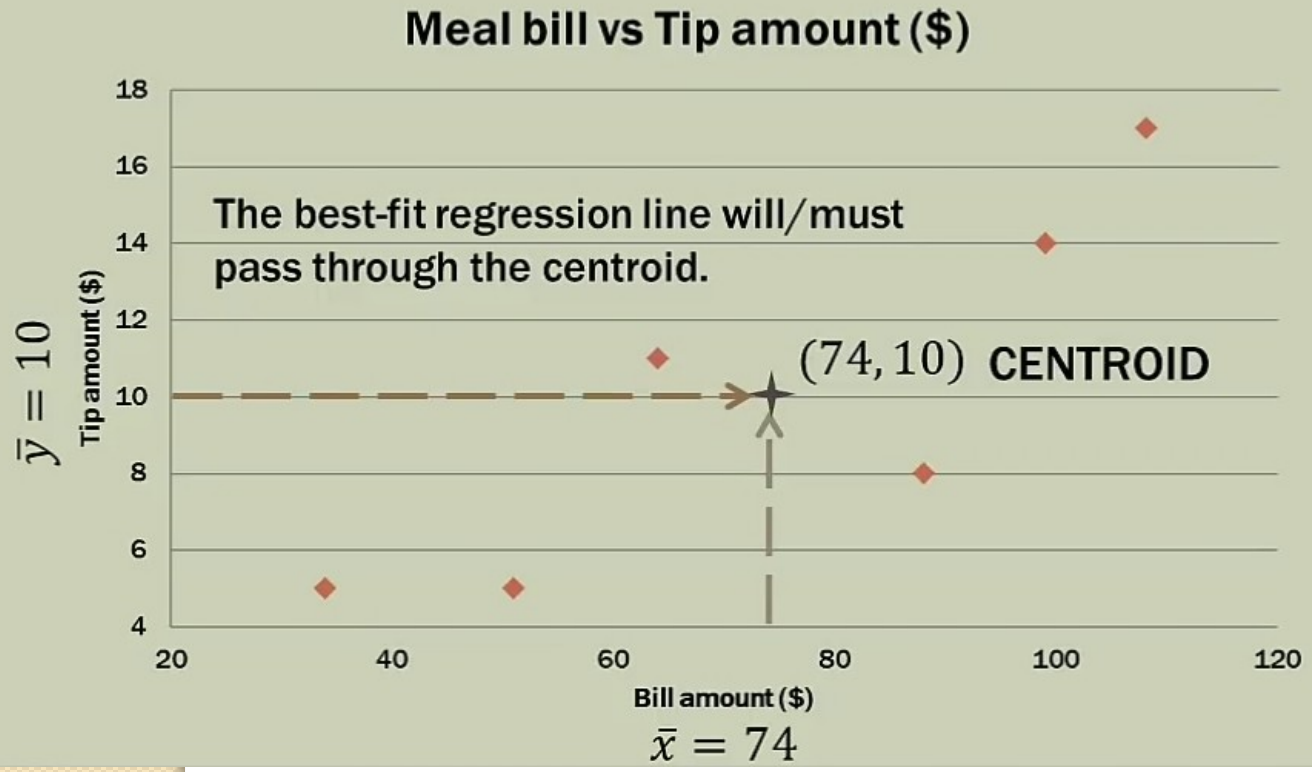
What is the correlation coefficient, r?

In this case, $r = 0.866$.

Is the relationship strong?

In this case, YES!

STEP 4: DESCRIPTIVE STATISTICS / CENTROID



Bill ($)	Tip ($)
34.00	5.00
108.00	17.00
64.00	11.00
88.00	8.00
99.00	14.00
51.00	5.00
$\bar{x} = 74$	$\bar{y} = 10$

STEP 5: CALCULATIONS

Intercept

$\hat{y}_i = b_0 + b_1 x_i$

Slope

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$\bar{x}$ = mean of the independent variable

$\bar{y}$ = mean of the dependent variable

x_i = value of independent variable

y_i = value of dependent variable

STEP 5: CALCULATIONS

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

1. For each data point.
2. Take the x-value and subtract the mean of x.
3. Take the y-value and subtract the mean of y.
4. Multiply Step 2 and Step 3
5. Add up all of the products.

---

1. For each data point.
2. Take the x-value and subtract the mean of x.
3. Square Step 2
4. Add up all the products.

$$b_0 = \bar{y} - b_1\bar{x}$$

STEP 5: CALCULATIONS

Meal	Total bill ($)	Tip amount ($)	Bill deviation	Tip Deviations	Deviation Products	Bill Deviations Squared
	x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	34	5				
2	108	17				
3	64	11				
4	88	8				
5	99	14				
6	51	5				
	$\bar{x} = 74$	$\bar{y} = 10$				

STEP 5: CALCULATIONS

Meal	Total bill ($)	Tip amount ($)	Bill deviation	Tip Deviations	Deviation Products	Bill Deviations Squared
	x	y	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	34	5	-40	-5	200	1600
2	108	17	34	7	238	1156
3	64	11	-10	1	-10	100
4	88	8	14	-2	-28	196
5	99	14	25	4	100	625
6	51	5	-23	-5	115	529
	$\bar{x} = 74$	$\bar{y} = 10$			$\sum = 615$	$\sum = 4206$

b_1 CALCULATIONS (SLOPE)

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$b_1 = \frac{615}{4206}$$

$$b_1 = 0.1462$$

Deviation Products	Bill Deviations Squared
$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
200	1600
238	1156
-10	100
-28	196
100	625
115	529
$\sum = 615$	$\sum = 4206$

b_0 CALCULATIONS(Y-INTERCEPT)

$$b_0 = \bar{y} - b_1\bar{x} \qquad b_1 = 0.1462$$

$$b_0 = 10 - 0.1462(74)$$

$$b_0 = 10 - 10.8188$$

$$b_0 = -0.8188$$

Total bill ($)	Tip amount ($)
x	y
34	5
108	17
64	11
88	8
99	14
51	5
$\bar{x} = 74$	$\bar{y} = 10$

YOUR REGRESSION LINE

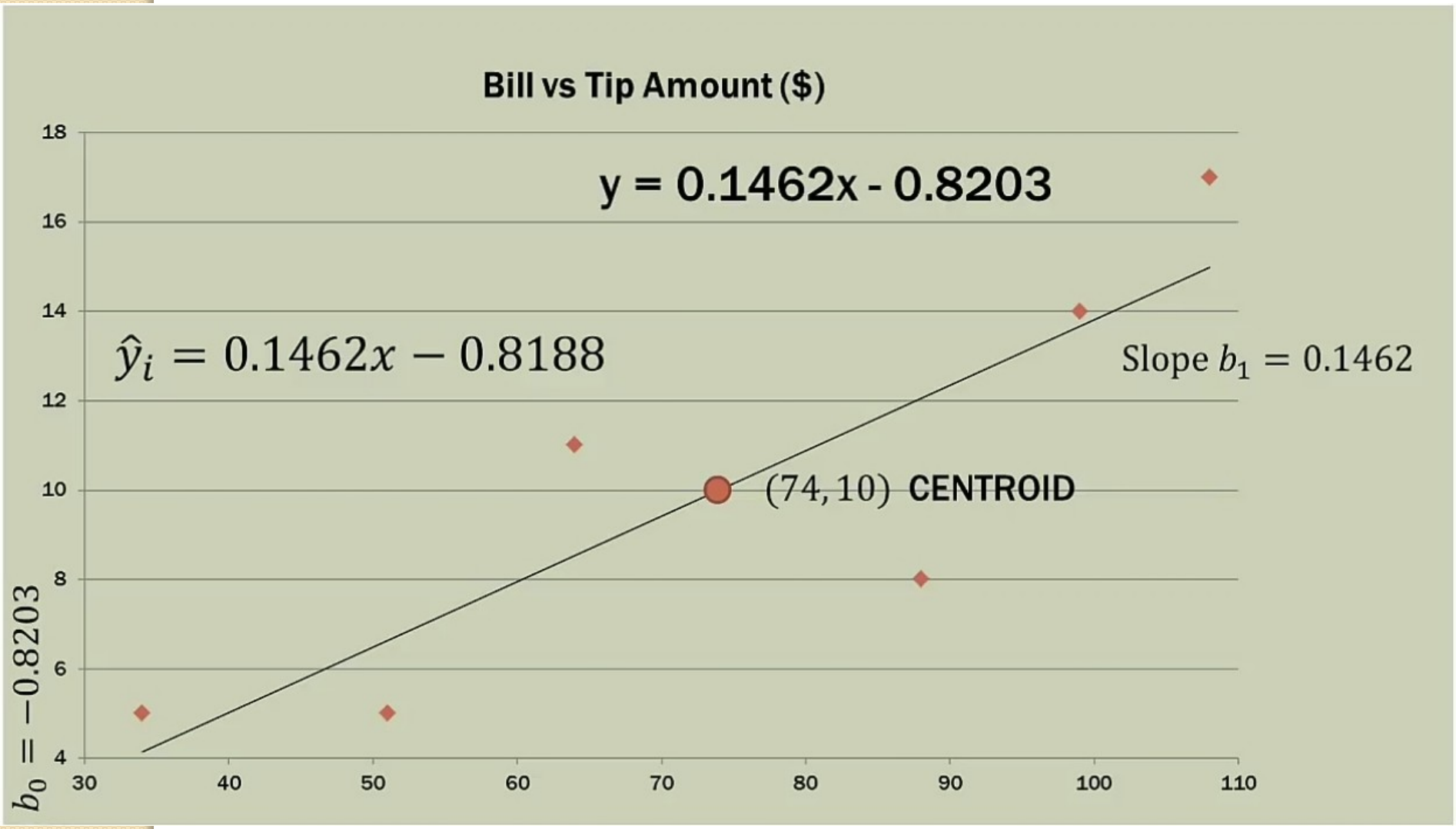
$\hat{y}_i = b_0 + b_1 x_i$ $\quad b_0 = -0.8188$ $\quad b_1 = 0.1462$

intercept slope

$$\hat{y}_i = -0.8188 + 0.1462x$$

OR

$$\hat{y}_i = 0.1462x - 0.8188$$


Bill vs Tip Amount ($)
y = 0.1462x - 0.8203
$\hat{y}_i = 0.1462x - 0.8188$
Slope $b_1 = 0.1462$
(74, 10) CENTROID
$b_0 = -0.8203$
18
16
14
12
10
8
6
4
30
40
50
60
70
80
90
100
110

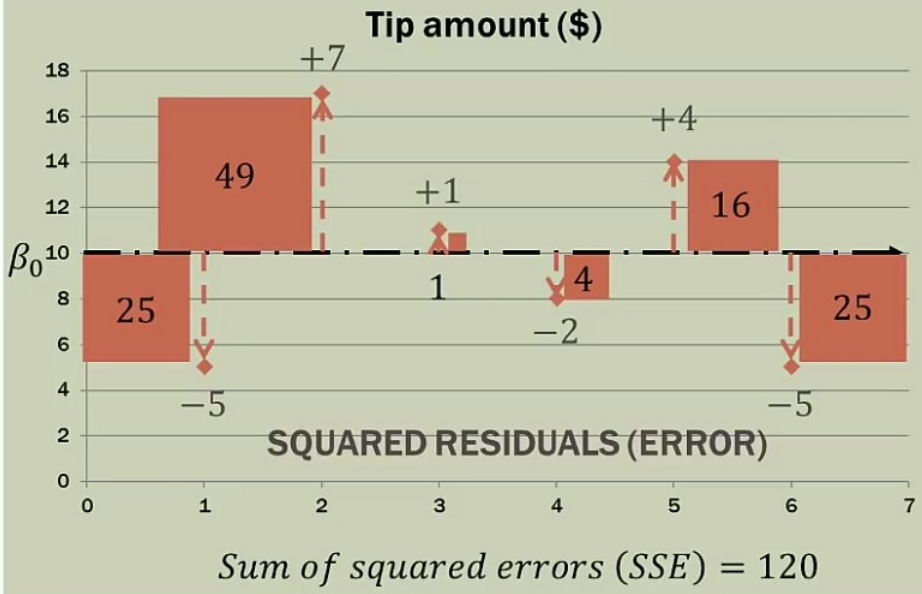
QUICK INTERPRETATION

$$\hat{y}_i = 0.1462x - 0.8188$$

For every \$1 the bill amount (x) increases, we would expect the tip amount to increase by \$0.1462 or about 15-cents.

If the bill amount (x) is zero, then the expected/predicted tip amount is \$-0.8188 or negative 82-cents! Does this make sense? NO. The intercept may or may not make sense in the "real world."

WHEN THE SLOPE, $\beta_1 = 0$



$Sum\ of\ squared\ errors\ (SSE) = 120$

Having only the DV, the best prediction for the tip of the next meal is the mean of the tips.

$\hat{y} = 10$

Since the "mean" line is flat, its slope is zero.