Making

Predictions

How do we make **prediction**?

- When **linear correlation is significant,** use $\hat{y} = a + bx$.

  Plug in the given x value to find the prediction value y.

- When **linear correlation is not significant,** use $\overline{y}$.

Example:

Eight pairs of data yield the regression equation
$\hat{y} = 55.6 + 2.8x$ with $\bar{y} = 71.5$.
What is the best predicted value for y for $x = 5.5$ if we assume the linear correlation is significant?

Solution:

Since the linear correlation coefficient is assumed to be significant,

we use the equation of the regression line $\hat{y} = 55.6 + 2.8x$, and plug in $x = 5.5$ to find the prediction value.

$$\begin{aligned} \hat{y} &= 55.6 + 2.8x \\ &= 55.6 + 2.8(5.5) \\ &= 55.6 + 15.4 \\ &= 71 \end{aligned}$$

So, our prediction value is 71.

Example:

Ten pairs of data yield the regression equation
$\hat{y} = 73.5 - 4.5x$ with $\bar{y} = 58.5$.

What is the best predicted value for y for $x = 4.5$ if we assume the linear correlation is not significant?

Solution:

Since the linear correlation is assumed to be not significant,

we use $\bar{y}$ as the prediction value regardless of the value of x.

So, our prediction value is 58.5.

Example:

The study time and midterm exam score for a random sample of 10 students in a statistics course are shown in the following table.

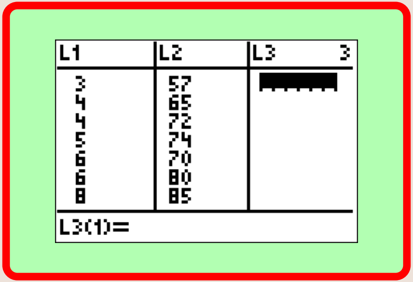
Study Time; x (Hours)	3	4	4	5	6	6	8	9	10	9
Midterm Score; y	57	65	72	74	70	80	85	90	97	92

Use TI command **LinRegTTest** to find

1. CTS t
2. P-Value p
3. Linear Correlation Coefficient r

Solution:

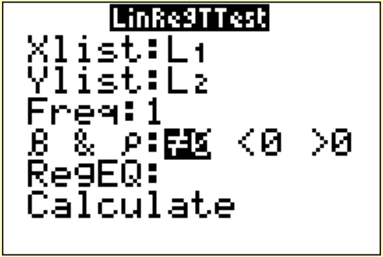
We store all x values in L_1 and corresponding y values in L_2,



then we perform the following steps.

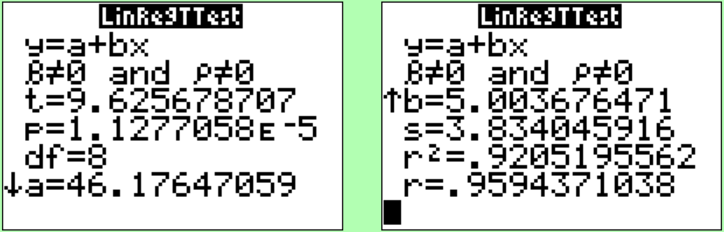
STAT $\rightarrow$ TESTS $\downarrow$ LinRegTTest **Xlist:** L_1 **Ylist:** L_2
Freq: 1 $\rho :\neq 0$ RegEQ: blank

Solution Continued:



Now press **Calculate** key to view the results.

Solution Continued:

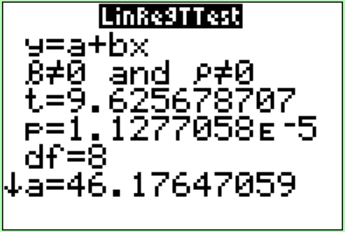


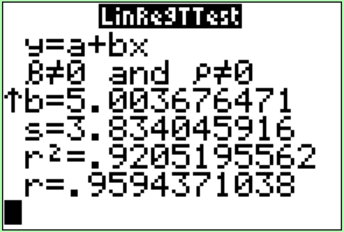
Here are the answers:

1. CTS $t = 9.626$
2. P-Value $p = 1.1 \times 10^{-5}$
3. Linear Correlation Coefficient $r = .959$

Solution Continued:

We can also find the equation of the regression line $\hat{y} = a + bx$.





Here are the answers:

1. $a = 46.176 \approx 46$
2. $b = 5.004 \approx 5$
3. Equation of the regression line $\hat{y} \approx 46 + 5x$

Example:

Use the study time and midterm exam score for a random sample of 10 students from the last example, determine whether the linear correlation is significant or not at 0.05 significance level, then make the prediction of exam score for study time of 7 hours.

Solution:

We first set up H_0 and H_1,

- $H_0 : \rho = 0 \Rightarrow$ **Linear Correlation is not significant**
- $H_1 : \rho \neq 0 \Rightarrow$ **Linear Correlation is significant**

Since $1.1 \times 10^{-5} \leq 0.05 \Rightarrow P - value \leq \alpha$, then H_1 is valid which implies

Linear Correlation is significant.

Solution Continued:

Since the linear correlation coefficient is significant, we use the equation of the regression line $\hat{y} = 46 + 5x$. and plug in $x = 7$ to find the prediction value.

$$\begin{aligned} \hat{y} &= 46 + 5x \\ &= 46 + 5(7) \\ &= 46 + 35 \\ &= 81 \end{aligned}$$

So, our prediction value is 81.

If we were to assume that the linear correlation coefficient is not significant, then the predicted value is
$\bar{y} = \frac{\sum y}{n} = \frac{782}{10} = 78.2 \approx 78.$