

12.2.4: Prediction

We can now use the least squares regression equation for prediction.

Use the regression equation to predict the grade for a student who has studied for 18 hours for their exam using the previous data.

Hours Studied for Exam 20 16 20 18 17 16 15 17 15 16 15 17 16 17 14 Grade on Exam 89 72 93 84 81 75 70 82 69 83 80 83 81 84 76

Solution

We found the regression equation $\hat{y} = 26.742 + 3.216346x$. The x -variable is the hours studied so let $x = 18$ hours. \hat{y} is the symbol for the predicted y .

Substitute $x = 18$ into the regression equation and you get: $\hat{y} = 26.742 + 3.216346 \cdot 18 = 84.636228$. We would estimate the student's grade to be 84.6 when they studied 18 hours. This is a point estimate for the grade on the exam for a student that studied 18 hours.

Prediction Interval

We can find a special type of confidence interval to estimate the true value of y called the **prediction interval**.

The prediction interval is the confidence interval for the actual value of y :

$$\hat{y} \pm t_{\alpha/2} \cdot s \sqrt{\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{SS_{xx}}\right)}$$

where \hat{y} is the predicted value of y for the given value of x .

Using the previous data, find and interpret the 95% prediction interval for a student who studies 18 hours.

Solution

From the question, $x = 18$. From previous examples we found that $\hat{y} = 26.742 + 3.216346 \cdot 18 = 84.636228$ and $s = 3.935892$. Find the critical value from the invT using $df = n - 2 = 13$; we get $t_{\alpha/2} = 2.160369$. Make sure to go out at least 6 decimal places in between steps. Ideally, never round between steps. Use the 2-Var Stats from your calculator to find the sums and then substitute values back into the equation to get

$$\begin{aligned} & 84.636228 \pm 2.160369 \cdot 3.935892 \sqrt{\left(1 + \frac{1}{15} + \frac{(18 - 16.6)^2}{41.6}\right)} \\ \Rightarrow & 84.636228 \pm 8.9723 \\ \Rightarrow & 75.6639 < y < 93.6085 \end{aligned}$$

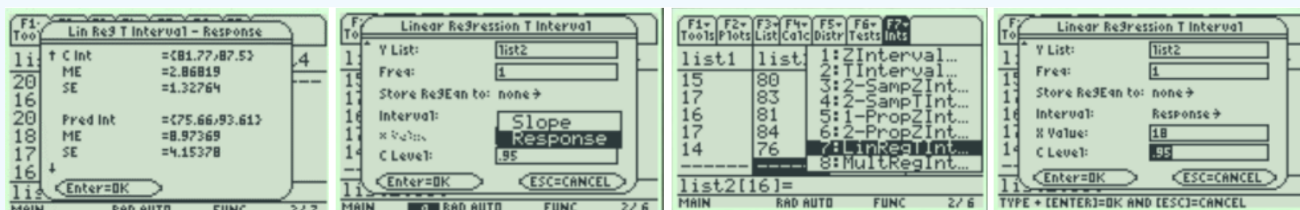
We are 95% confident that the predicted exam grade for a student that studies 18 hours is between 75.6639 and 93.6085.

A confidence interval can be more accurate (narrower) when you increase the sample size. Note that in the last example, the predicted grade for an individual student could have been anywhere from a C to an A grade. If you wanted to predict y with more accuracy, then you would want to sample more than 15 students to get a smaller margin of error. The confidence interval for a mean will have a smaller margin of error than for an individual's predicted value.

Excel, the TI-83 and 84 do not have built in prediction intervals.

TI-89: Enter the x -values in list1 and the y -values in list2, select [F7] Intervals, then select option 7:LinRegTInt... Use the Var-Link button to enter in list1 and list2 for the X List and Y List. Select Response in the drop-down menu for Interval. Enter in the x -value given in the question. Change the confidence level (C-Level) to match what was in the question, the [Enter].

Scroll down to Pred Int for the prediction interval. The calculator does not round between steps so if you rounded b_0 and b_1 , for instance, when doing hand calculations, your answer may be slightly different than the calculator results.



Extrapolation is the use of a regression line for prediction far outside the range of values of the independent variable x . As a general rule, one should not use linear regression to estimate values too far from the given data values. The further away you move from the center of the data set, the more variable results become. For instance, we would not want to estimate a student's grade for someone that studied way less than 14 hours or more than 20 hours.

This page titled [12.2.4: Prediction](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Rachel Webb](#).