

**LEAST SQUARES REGRESSION LINE****6.1.4**

A **least squares regression line** (LSRL) is a line of best fit that minimizes the sum of the square of the residuals.

For additional information, see the Math Notes box in Lesson 6.2.1 and the narrative in Checkpoint 8 at the back of the textbook. Also see problem 6-34 in the textbook for a demonstration of finding the least squares regression line. For additional examples and more practice on the topics from this chapter, see the Checkpoint 8 problems.

**CALCULATORS**

Statistical calculators or software can find the LSRL and residuals with ease. Students are welcome to use a TI-83/84+ calculator if they have one available. See details about using calculators in the Introduction section of this *Parent Guide with Extra Practice*.

An eTool is also available at [www.cpm.org/technology](http://www.cpm.org/technology) for students who do not have a graphing calculator.

**TI-83/84+ NOTES**

More detailed instructions for using a TI-83/84+ calculator can be found under “Technology Resources” at [www.cpm.org](http://www.cpm.org). Basic instructions follow.

For part (a) of the example:

- Enter the data points in two lists.
- For data tables in which a *checksum* is given, the *checksum* is used to verify that data has been entered into the calculator correctly. Use  $\text{>[TBLSET]}$   $[i][F4][\text{TRACE}] \Sigma [F4] [\text{GRAPH}]$  and verify that  $\Sigma x$  on the calculator screen matches the “checksum” given at the bottom of the table in certain problems.  $\Sigma x$  stands for “the sum of  $x$ ” or the sum of all the values in the list.
- Use  $\text{>[TBLSET]}[i][F4][\text{TRACE}] \Sigma [F4] [\text{GRAPH}]$  to find the maximum and the minimum of both lists, then set the window appropriately.
- Use  $[2\text{nd}] [\text{STAT PLOT}]$  to set up a graph, and then press  $[\text{GRAPH}]$ .

For part (b) of the example:

- Find the LSRL with  $[\text{STAT}] [\text{“CALC”}] [\text{“8:LinReg(a+bx)”}] [2\text{nd}] [L1] [,] [2\text{nd}] [L2] [,] [\text{VARS}] [\text{“Y-VARS”}] [\text{“1:Function”}] [\text{“1:Y}_1\text{”}] [\text{ENTER}]$ .
- Use  $[2\text{nd}] [\text{WINDOW}]$  to set the scale of the graph, then press  $[\text{GRAPH}]$ .

For part (c) of the example:

Each time you perform a LSRL regression, the calculator will automatically store the residuals in a list named “RESID”. You can display the residuals in List 3 as follows.

- Highlight the label for List 3 and press  $[\text{ENTER}]$ .
- Press  $[2\text{nd}] [\text{LIST}] [\text{“RESID”}]$  and then  $[\text{ENTER}]$  to copy the list.

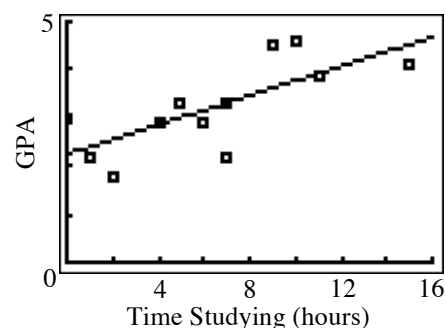
## Example

It seems reasonable that there would be a relationship between the amount of time a student spends studying and their GPA. Suppose you were interested in predicting a student's GPA based on the hours they study per week. You were able to randomly select 12 students and obtain this information from each student.

Hours	4	5	11	1	15	2	10	6	7	0	7	9
GPA	2.9	3.3	3.9	2.2	4.1	1.8	4.6	2.9	2.2	3	3.3	4.5

- a. Use your calculator to graph the scatterplot. Choose a reasonable viewing window on your calculator, then make a sketch of the scatterplot on your paper and label the axes.

Since the "hours" range from 0 to 15, an appropriate window might have an  $x$ -axis scale from 0 to 16 with an interval of 2. The GPA data ranges from 1.8 to 4.6, so an appropriate  $y$ -axis might range from 0 to 5 with an interval of 1. Note that the axes are labeled with both the variable name and with the units.



- b. Use your calculator to find the equation of the LSRL.

From the calculator,  $y = 2.249 + 0.152x$ . It is essential when giving an LSRL equation to define the variables and units. In this case,  $x$  is the numbers of hours spent studying each week, and  $y$  is the predicted GPA for that student.

- c. Use your calculator to create a residual list. What is the residual for the student who spent 10 hours studying? Interpret the residual in context.

A partial residual follows (in List L3):

From the residual list, the residual for the student who studied 10 hours is 0.83 grade points. That means the student earned a GPA that was 0.83 points higher than was predicted by the least squares regression line.

L1	L2	L3	3
4	2.9	0.04249	
5	3.3	.29041	
11	3.9	-.0219	
1	2.2	-.2014	
15	4.1	-.4301	
2	1.8	-.7534	
10	4.6	.83014	
L3(1)=		.0424657534...	

## Problems

1. It seems reasonable that the horsepower of a car is related to its gas mileage. Suppose a random sample of 9 car models is selected and the engine horsepower and city gas mileage is recorded for each one.

HP	197	170	166	230	381	170	326	451	290
MPG	16	24	19	15	13	21	11	10	15

- a. Use your calculator to graph the scatterplot. Choose a reasonable viewing window on your calculator, then make a sketch of the scatterplot on your paper and label the axes.
- b. Use your calculator to find the equation of the LSRL. What do you predict the gas mileage will be for a car with 400 horsepower?
- c. Use your calculator to create a residual list. What is the residual for the 230 horsepower car?
2. Many people believe that students who are strong in music are also strong in mathematics. But the principal at University High School wonders if that same connection exists between music students and English students. The principal went through the records for the past year and found 10 students who were enrolled in both Advanced Placement Music and Advanced Placement English. He compared their final exam scores.

Final Exam Scores	
AP Music	AP English
88	63
74	96
82	86
64	90
97	68
90	90
82	78
72	74
78	96
62	79
<i>checksum 789</i>	<i>checksum 820</i>

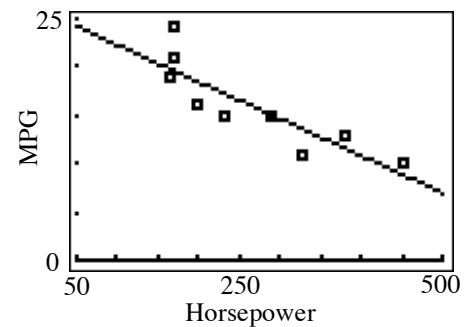
- a. Use your calculator to graph the scatterplot. Choose a reasonable viewing window on your calculator, then make a sketch of the scatterplot on your paper and label the axes. Did you remember to use the *checksum* to verify you entered the data correctly?
- b. Use your calculator to find the equation of the LSRL.
- c. Use your calculator to create a residual list. Would a student prefer a positive or a negative residual in this situation?

## Answers

1. a. See scatterplot at right.

b.  $y = 26.11 - 0.0382x$ .  $y$  is the predicted gas mileage (in mpg), and  $x$  is the horsepower. A car with 400 horsepower is predicted to have gas mileage of  $26.11 - 0.0382(400) = 10.8$  mpg.

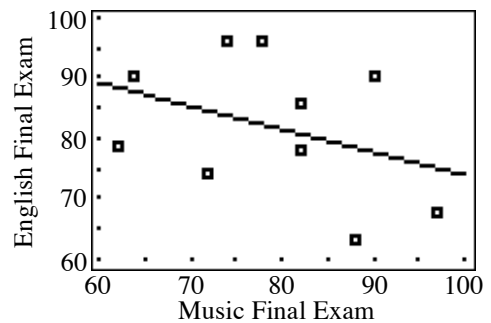
c. The car with 230 horsepower had residual of  $-2.3$  mpg. That means the car had gas mileage that was 2.3 mpg less than was predicted by the LSRL.



2. a. See scatterplot at right.

b.  $y = 112.1 - 0.3813x$  where  $y$  is the predicted AP English final exam score, and  $x$  is the AP Music final exam score.

c. Residual list follows in List L3 below. A student would prefer a positive residual. That means their AP English score is higher than predicted by the LSRL.



L1	L2	L3	3
88	63	-13.53	
74	96	12.132	
82	86	5.1821	
64	90	2.3183	
97	68	-7.098	
90	90	12.233	
82	78	-2.818	
L3(1) = -15.5299673...			