# Line Up

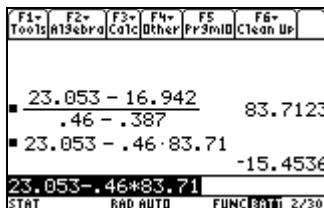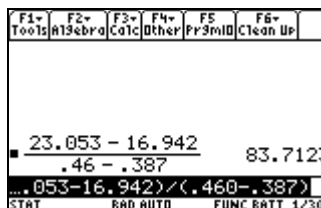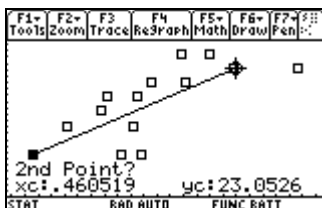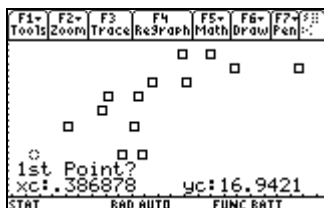## An Exploration of the Least Squares Regression Line with the TI-89

## Teacher Notes

An activity in support of the NCTM Data Analysis and Probability Standard: To select and use appropriate statistical methods to analyze data, with expectation: for bivariate measurement data...determine regression coefficients, (and) regression equations using technological tools.

Statistics Scope and Sequence Topic: Bivariate Data, least-squares regression line.

This exploration uses baseball's slugging average and number of games won, to explore a bivariate association. Students estimate the slope and intercept of the regression line and compute the sum of the squared residuals. The calculator is used to compute the actual coefficients.

Sample Solutions:

Solutions

# Line Up

## An Exploration of the Least Squares Regression Line

In a previous exploration, we examined the statistic slugging average, used in the game of baseball. To review, slugging average is defined as the number of total bases divided by the number of at-bats. If a player reaches first base safely, a single, the player would be credited with one base. A two base hit, a double is worth two total bases, a triple is worth three and a home run is worth four. An at-bat is each attempt for a hit.

Teams that have a high slugging average tend to score more runs, which should mean they have a better chance of winning. However, winning also depends on the number of runs they allow the other team to score. Let's examine the association between slugging average and wins.

The following table shows the slugging average and the number of wins for each of 16 teams for one season of 40 games.

| Team | Avg. | Wins | Team | Avg. | Wins |
|------|------|------|------|------|------|
| Bears | .412 | 20 | Manatees | .426 | 17 |
| Bulls | .442 | 22 | Monkeys | .396 | 16 |
| Cougars | .430 | 22 | Panthers | .387 | 17 |
| Dolphins | .419 | 17 | Ravens | .414 | 21 |
| Eagles | .483 | 23 | Rhinos | .393 | 15 |
| Hawks | .423 | 19 | Sharks | .399 | 19 |
| Hornets | .451 | 24 | Tigers | .460 | 23 |
| Lions | .425 | 21 | Vultures | .441 | 24 |

Type the 16 values for slugging in a list named slugg. Type the 16 values for



wins into a list named wins.

Set up the STAT PLOT and press F5 for a window that will display the
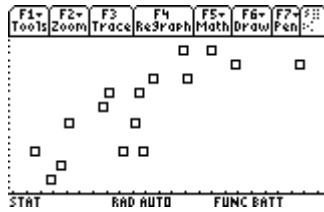


scatterplot.
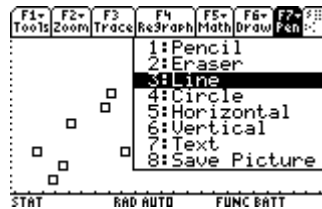


Now press 2nd F7 and select 3:Line

Imagine a line that is close to the points of this scatterplot. Use ◁ ▽ to move the cursor to a point that appears to be on or at least very close to the line. Record the coordinates of the point.

Use ▷△ to move the cursor to another point that appears to be on or at least very close to the line. Record the coordinates of this point.

Compute the slope of the line that passes through the two points you have selected.

Write the equation of the line in slope-intercept form.

Type the equation of the line in $Y_1$ and ♦ F3. The graph should look something
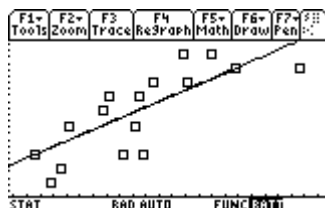


like this:

Now we will try to determine which line is closest to the points. Each student should define list1 as: wins-y1(slugg). These values represent the directed

distance from each point to the line, when the distance is measured vertically. Some of these values are positive, and some are negative, so the sum of these values is not a good way to determine which line is closest to the points. To obtain a better measure, each student should define list2 as list1^2 and compute sum(list2).

What is the smallest sum in the class?

What is the equation of the line?

Now let's find the line that has the absolute smallest sum.



Press F4 3 2 and complete the list names:

Press ENTER twice to display the result.

Compare this equation to the one that had the smallest sum. How close are the coefficients?

The line we have computed is the line with the minimum sum of the directed distances we computed before. These distances are called residuals, and the sum is the sum of the squared residuals. This line is often referred to as the least-squares regression line.

We could compute the sum in the same manner as before. However, there is an easier way, since the residuals are automatically computed and stored when a

regression line is computed. The list is stored to the folder m\named statvars so



enter the command as: 

How does this sum compare to the smallest sum discovered by your class?