

NUMB3RS Activity: Dialing Ben Episode: "Contenders"

Topic: Scatter plots, least-squares regression, and residuals

Grade Level: 10 - 12

Objective: Students will interpret plots over time to find change points.

Time: 20 - 25 minutes

Materials: TI-83 Plus/TI-84 Plus graphing calculator.

Introduction

In "Contenders," one of David's friends, Ben Ellis, is a leading suspect in the murder of two contenders in the All-Out Fighting League. The evidence against Ben increases as the story develops, and Ben stops using his cell phone after David tries to arrest him. Charlie explains that it is highly likely that he may now be using a friend's cell phone. To determine which friend's cell phone Ben might be using, Charlie looks at the cell phone records of Ben and his friends. Charlie states that Ben has a unique calling pattern and if he can identify the calling pattern when it appears in another friend's records, then he may be able to determine Ben's location. In this activity, students will use scatter plots and regression models to find patterns in data.

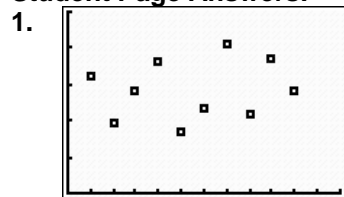
There are many ways to study the calling patterns. Charlie explains that he will be using a "change-point detector" algorithm. Change-point detection looks for abrupt changes in data over time. Change-point algorithms typically involve several variables and a large number of calculations. Due to the large number of calculations involved in carrying out change-point detection algorithms, this activity will instead search for change points in data using a variety of statistical plots that can be done on a graphing calculator.

Discuss with Students

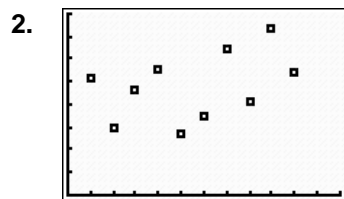
Questions 1 and 2 ask the students to make scatter plots using their graphing calculators. Make sure the students know how to enter data and set up Stat Plots. Help the students make appropriate windows. You should discuss the window settings for each scatter plot before graphing to give them a feel for the domain and range of the data. You may also want to emphasize the difference between independent and dependent variables.

The calculations on Question 5 may differ from student to student due to rounding. The residuals should sum to 0 but do not in the answer key for Question 5 due to rounding.

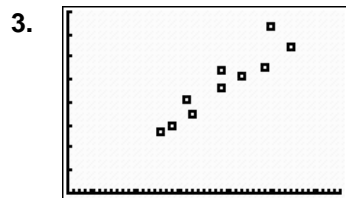
Student Page Answers:



There do not appear to be any noticeable change-points.

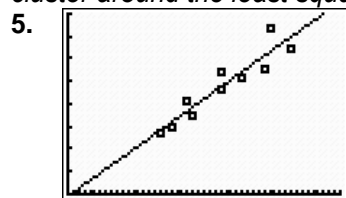


There do not appear to be any noticeable change-points in this scatter plot.



There seems to be a positive relationship between the time talking and total distances. This makes sense because an increased amount of time talking may mean an increased number of calls which may increase the total distance of calls made. No change points appear in this graph.

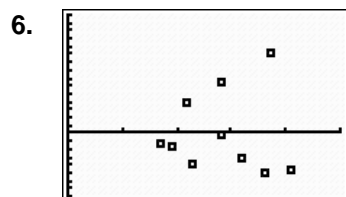
4. The least squares line seems to fit the data. The pattern of the data appears linear and seems to cluster around the least squares line.



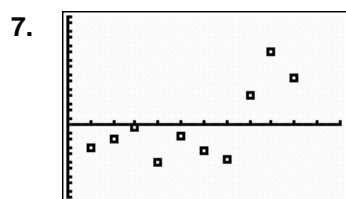
The predicted values were calculated using the formula $\hat{y} = 1.73x - 0.96$ and small differences between these results and the calculator's results will be due to rounding.

Friend A's Cell Phone Records

Day	1	2	3	4	5	6	7	8	9	10
Total Time (minutes)	32	19	28	36	17	23	41	22	37	28
Total Distance of Calls (miles)	51	30	47	56	27	35	65	41	73	54
Predicted Total Distance	54.4	31.9	47.5	61.3	28.5	38.8	70.0	37.1	63.1	47.5
Residuals (observed - predicted)	-3.4	-1.9	-0.5	-5.3	-1.5	-3.8	-5	3.9	9.9	6.5



The residual plot seems to be somewhat random which provides support for the least-squares model.



The residual plot over time seems to show a change-point at day 8. All of the residuals are negative before day 8 and the rest are positive. The pattern in the data for days 1 to 7 appear to be consistent while the patterns for days 8 to 10 show a different pattern. This indicates a change in cell-phone usage over time and could mean that Ben is using Friend A's phone. The residual plot over time shows these changes because the plot is looking at the relationship between time talking and total distance. The cell phone usage for the final 3 days was within the range of the entire 10 days so the scatter plot showing day versus time talking (L_1 vs. L_2) does not show any change points. The same is true for the day versus total distance plot (L_1 vs. L_3).

Name: _____

Date: _____

NUMB3RS Activity: Dialing Ben

In "Contenders," one of David's friends, Ben Ellis, is a leading suspect in the murder of two contenders in the All-Out Fighting League. The evidence against Ben increases as the story develops, and Ben stops using his cell phone after David tries to arrest him. Charlie explains that it is highly likely that he may now be using a friend's cell phone. To determine which friend's cell phone Ben might be using, Charlie looks at the cell phone records of Ben and his friends. Charlie states that Ben has a unique calling pattern and if he can identify the calling pattern when it appears in another friend's records, then he may be able to determine Ben's location. In this activity, you will use scatter plots and regression models to find patterns in data.

Change-Point Detection

Charlie uses a change-point detection algorithm to find Ben's calling patterns in his friends' cell phone records. Change-point detection looks for abrupt changes in data over time. In this activity you will learn ways to look at data that help you locate these changes.

One variable that may be helpful in finding abrupt changes in cell phone records is the number of calls made in a given day. If this amount consistently goes up or down after a given day, this may provide evidence that a different person is using the phone. Another variable that may be of interest is the distance between the two people talking on a given call. Many cell phone companies keep track of the location of the cell phone when a call is made as well as the location of the phone that receives the call. Ben's pattern of calling distances may be different than that of his friend.

Suppose the data below tracks the cell phone use for one of Ben's friends for 10 days. The first variable is the total time (in minutes) Friend A's cell phone is used for a given day. The second variable is the sum of the distances (in miles) of each phone call in a given day. This sum is found by calculating the distance between the two phones when each call is made then adding these distances.

Friend A's Cell Phone Records

Day	1	2	3	4	5	6	7	8	9	10
Total Time (minutes)	32	19	28	36	17	23	41	22	37	28
Total Distance of Calls (miles)	51	30	47	56	27	35	65	41	73	54

Enter the data into your graphing calculator. Put the day into L_1 , time talking into L_2 , and total distances into L_3 .

1. Make a scatter plot of L_1 vs. L_2 , where day (L_1) is the independent variable and time talking is the dependent variable (L_2). Explain any abrupt changes you see that may signify a change point.

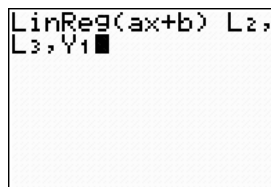
2. Make another scatter plot of L_1 vs. L_3 , where day (L_1) is the independent variable and total distance (L_3) is the dependent variable. Describe any patterns you see in the data that may signify changes in the calling patterns of Friend A's phone.

Create a new scatter plot of L_2 vs. L_3 , where time talking (L_2) is the independent variable and total distance (L_3) is the dependent variable. Make sure to adjust the window settings for this plot.

3. Explain the relationship between time talking and total distance. Explain why this relationship might make sense.

Charlie uses sophisticated techniques to look for change points, but his methods do use some of the same statistical ideas you will use in the following problems.

Calculate the least-squares regression line between time talking (L_2) and total distance (L_3). (To enter the command **LinReg(ax+b)**, press **STAT**, go to the **Calc** menu, and select **4:LinReg(ax+b)**.) The Y_1 argument (optional) pastes the results of the regression calculation into Y_1 so the line will be graphed on the scatter plot.



4. How well does the least squares line model the pattern in the data?

Residual Plots

One way you can assess how well a mathematical model fits data is to create a residual plot. When you calculated the least squares line above, your calculator also calculated the residuals. The residuals are the differences between the observed value of the dependent variable and the predicted value of the dependent variable from the model (residual = observed y – predicted y).

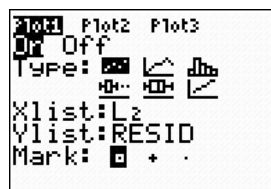
5. Find the remaining residuals in the table on the next page. The predicted value of total distance is calculated for day 1 using the least-squares equation. The time talking value for day 1 is 32 minutes. Plug this into the regression equation $\hat{y} = 1.73(32) - 0.96 = 54.4$ to get the predicted total distance for day 1. Note that these values are rounded and may be slightly different than your calculator results.

Friend A's Cell Phone Records

Day	1	2	3	4	5	6	7	8	9	10
Total Time (minutes)	32	19	28	36	17	23	41	22	37	28
Total Distance of Calls (miles)	51	30	47	56	27	35	65	41	73	54
Predicted Total Distance	54.4	31.9								
Residuals (observed - predicted)	-3.4									

A residual plot can be used to further analyze the fit of the regression line. The independent variable in a residual plot is the independent variable of the least-squares line (time talking) and the dependent variable is the residual.

To graph the residuals, press **2nd** [STAT PLOT] and select a plot. Enter L_2 as the **Xlist**. Place the cursor after **Ylist**, press **2nd** [LIST] and under the **NAMES** menu select **RESID** (which stands for residuals). Make appropriate changes to the window settings for this plot. Press **TRACE** to match the residuals you calculated. Note there are some differences due to rounding.

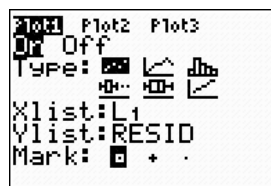


- Patterns in a residual plot can reveal limitations in the least-squares model and can indicate that a different model may better describe the relationship between the variables. On the other hand, support for the least-squares line is increased if the residual plot appears somewhat random with points both above and below the axis. What patterns do you see in this plot? Does the residual plot provide evidence for or against a linear model?

Residual Plots Over Time

It may appear that Ben is not using Friend A's cell phone. A residual plot over time is a technique used to look further into the relationship between variables and possibly locate change-points if they exist.

The residual plot over time graph uses the residuals as the dependent variable and the day (L_1) as the independent variable. Make appropriate changes in the window settings to see the plot.



- A change-point is an abrupt change in the data at some given time. Explain how the residual plot over time shows a change-point. On what day does the change occur? Why does the residual plot over time show a change point while the other plots do not?

The goal of this activity is to give your students a short and simple snapshot into a very extensive math topic. TI and NCTM encourage you and your students to learn more about this topic using the extensions provided below and through your own independent research.

Extensions

Change-Point Detection

- Look at the data below and try to determine if Ben is using any one of the friend's cell phones below. If you detect a change-point try to determine the day Ben began using the friend's phone.

Friend B's Cell Phone Records

Day	1	2	3	4	5	6	7	8	9	10
Total Time (minutes)	32	19	28	36	17	23	41	22	37	28
Total Distance of Calls (miles)	39	29	35	44	28	31	53	31	46	35

Friend C's Cell Phone Records

Day	1	2	3	4	5	6	7	8	9	10
Total Time (minutes)	32	19	28	36	17	23	41	22	37	28
Total Distance of Calls (miles)	57	31	49	65	31	42	78	39	60	53

Friend D's Cell Phone Records

Day	1	2	3	4	5	6	7	8	9	10
Total Time (minutes)	32	19	28	36	17	23	41	22	37	28
Total Distance of Calls (miles)	95	63	85	99	74	84	115	82	108	92

- A description of change-point analysis can be found at <http://www.variation.com/cpa/tech/changepoint.html>.

Data Analysis

- Go to <http://illuminations.nctm.org/LessonDetail.aspx?id=L673> for a lesson (Impact of a Superstar) on using linear regression to find outliers.
- A lesson finding lines of best fit can be found at <http://illuminations.nctm.org/LessonDetail.aspx?id=L298>.
- Go to <http://illuminations.nctm.org/LessonDetail.aspx?id=U135> for a lesson on regression and correlation.