EXPLORATIONS

Activity 3

Fitting an Equation to Bivariate Data

In this activity, you will start by fitting a linear least-squares regression line in Topic 11 to the U.S. Census data given on the next page. This will set the stage for the activities that follow.

The shape of the resulting curve looks like part of a parabola (a quadratic equation), which is one of the polynomial regression fits discussed in Topic 12 (cubic and quadrinomial fits being the others).

Population models suggest exponential or logistic growth as possible fits. Exponential growth will be discussed in Topic 13 as a fit that uses a transformation of data to make it more linear (logarithmic and power fits are the others). The logistic fit (which is our selection for the best fit) is covered in Topic 14.

Topic 15 returns to fitting a straight line to data , but by a technique that is more resistant to unusual values (median-median fit) than the least-squares fit of Topic 11.

Topic 16 fits a trigonometric sine curve to periodic data.

Note that if your fit display screens are different from those shown in this activity (do not show **r**, $\mathbf{r}^2$ or $\mathbf{R}^2$ when this handbook does) your diagnostic flag is off. Topic 8 shows how to turn it on.

📖 *Read Topic 11 before reading other topics in Activity 3.*

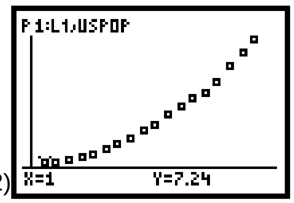
Setting Up

The main data set for this activity is the U.S. Census data (in millions of people) given on the next page. Store it in list **USPOP** with a coded year value of 1 to 18 for the years 1810 to 1980 in list $\mathbf{L_1}$. The value for 1990 is 249.63 million people, but you do not include this in the list because you will use it to check how well the fit equation can predict it.

YEAR:		1810	1820	1830	1840	1850	1860	1870	1880	1890
L_1	**X:**	1	2	3	4	5	6	7	8	9
USPOP	**Y:**	7.24	9.64	12.87	17.07	23.19	31.44	39.82	50.16	62.95
YEAR:		1900	1910	1920	1930	1940	1950	1960	1970	1980
	X:	10	11	12	13	14	15	16	17	18
	Y:	75.99	91.97	105.71	122.78	131.67	151.33	179.32	203.21	226.5

(1)

1. Set up **Plot1** for a **Scatter** plot, as shown in Topic 7 and in screen 1.
2. Press [ZOOM] **9: ZoomStat** [TRACE] to produce the plot of your data, as shown in screen 2.



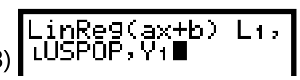
(2)

The top of the plot screen shows the setup with **P1:L_1,USPOP**. The points seem to lie more on a curve than a straight line, but you will start with fitting the linear least-squares regression line (Topic 11) to the data to set the stage and understand the notation for the activities that follow.

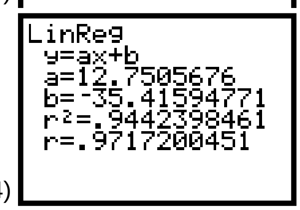
Topic 11—Linear Least Squares Regression Line

The following procedure obtains a linear least squares regression line.

1. Calculate the fit equation.
   a. Press [STAT] **<CALC> 4:LinReg(ax + b) L_1** [,] **∟USPOP** [,] **Y1**, as shown in screen 3, with **Y1** pasted from [VARS] **<Y-VARS> 1:Function 1:Y1**.
   b. Press [ENTER] for screen 4 showing your linear fit **Y1 = 12.751x - 35.416** stored in **Y1** in the Y= editor.
2. Plot data scatter and fit equation.
   a. Keeping **Y1** turned on (this was done automatically in step 1), turn on **Plot1** as a **Scatter** plot (as shown in Topic 7) with all other Y= functions and stat plots turned off.
   b. Press [ZOOM] **9:ZoomStat** [TRACE] for both a **Scatter** plot of the data and a plot of the regression line, as shown in screen 5.

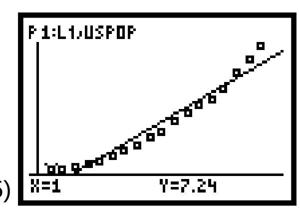


(3)



(4)

***Note:** You would get the same results with [STAT] **<CALC> 8:LinReg(a + bx) L_1** [,] **∟USPOP** [,] **Y1** but the slope would be **b** instead of **a**.*



(5)

3. Plot residuals.

   Step 1 automatically stores the residuals in list **RESID**.

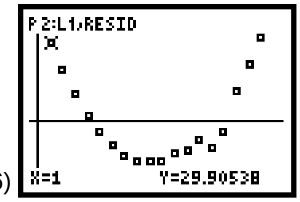
   a. Set up **Plot2** as a **Scatter** plot with **Xlist:L1** and **Ylist:RESID** (making sure all other stat plots and Y= plots are off).

   b. Press [ZOOM] **9:ZoomStat** [TRACE] for screen 6.

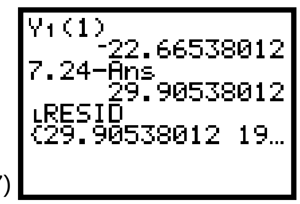
   The non-random pattern of the **Scatter** plot of residuals confirms that the linear least squares regression line does not fit the data very well. Note that the pattern looks quadratic.

   The *residual* is the difference from the actual y-value and the value obtained by plugging the x-value that goes with the y-value into the regression equation. When $x = 1$, you have **Y1** = $12.751x - 35.416$, which becomes **Y1**(1) = 12.751(1) − 35.416 = − **22.665**. The difference from the actual value of 7.24 is **7.24** minus -**22.665**, or **29.905**.

   When you paste list **RESID** to the home screen (as shown in the last two lines of screen 7), you confirm this calculation.

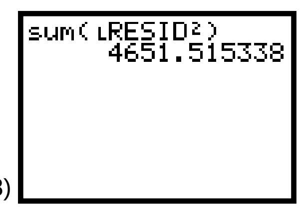
(6)


***Note**: For a perfect fit, the residuals will be all zero and [ZOOM] **9:ZoomStat** will result in a WINDOW RANGE error since Ymin = 0 Ymax = 0. If you still wish to see the plot, change Ymin = -1 and Ymax = 1 and then press [TRACE].*

(7)


4. Measure the fit (SSE).

   With some residuals positive, some negative, and some possibly zero, you will use the Sum of the Squared Residual Errors (SSE) as your measure of how close the points fit the curve. (If all the points are on the curve, this would be zero.) SSE is calculated in screen 8, where SSE = **4651.51534**, with **sum** pasted from [2nd] [LIST] **<MATH> 5:sum.**

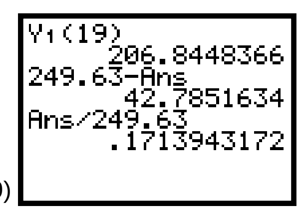
(8)


5. Predict the population in 1990 ($X = 19$).

   a. Paste **Y1** to the home screen, and then type [(] 19 [)], as shown on the first line in screen 9.

   b. Press [ENTER] for the next line, which is the predicted value of **Y**, or **206.845**.

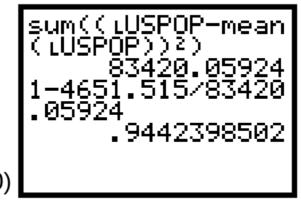
      Because you know the actual census value was 249.63, you can calculate the difference. The difference is **42.78516**, or 17 percent, a fairly large error. (See the calculations in screen 9. Note **Ans** is from [2nd] [ANS] in the last row of the keyboard.)

(9)


6. Calculate **r** and $\mathbf{r}^2$ as measure of linearity.

   $\mathbf{r}^2$ is related to SSE in the current case. To show this, you need to calculate the Sum of Squares Total (SST); that is, the sum to squared differences for each y data value and the mean of the complete **Y** list. (SST depends only on the data list and is independent of the fit equation used.)

   SST = **sum(** [(] **ʟUSPOP** [−] **mean** [(] **ʟUSPOP** [)] [)] [x^2] [)] = **83420.06** with the mean pasted from [2nd] [LIST] **<MATH> 3:mean**. $\mathbf{r}^2$ = 1 − SSE/SST = **0.9442** as before and in screen 10.



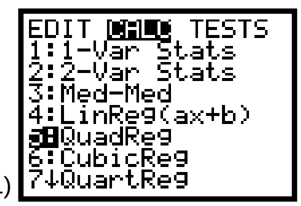
(10)

Note: *Small residuals (SEE) give an* $\mathbf{r}^2$ *close to one. Large residuals (SEE) give an* $\mathbf{r}^2$ *close to zero.*

Topic 12—Polynomial Regression: Quadratic, Cubic, and Quadrinomial

Press [STAT] **<CALC>** to reveal screen 11. This topic covers the last three functions shown.

5:QuadReg fits $Y = ax^2 + bx + c$
6:CubicReg fits $Y = ax^3 + bx^2 + cx + d$
7:QuartReg fits $Y = ax^4 + bx^3 + cx^2 + dx + e$



(11)

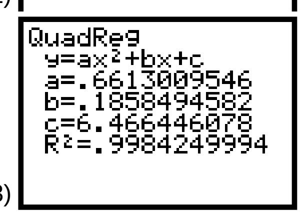
Quadratic Fit

In this procedure, you will fit the quadratic equation to the population census data. The procedure for the other fits is the same.

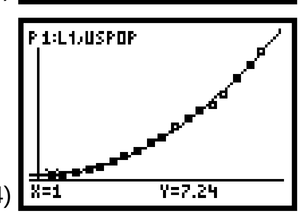
Note that the numbers of the steps below refer to the steps in Topic 11 that present more detail.

Step in Topic 11	Display
1. Calculate the fit equation.	Press [STAT] **<CALC> 5:QuadReg L1** [,] **ʟUSPOP** [,] **Y1** (with **Y1** pasted from [VARS] **<Y-VARS> 1:Function 1:Y1**) for screen 12. Press [ENTER] for screen 13. Note that $\mathbf{R}^2$ = **0.9984** compared to $\mathbf{r}^2$ = **0.9442** for the linear regression in Topic 11.
2. Plot data scatter and fit equation	The regression plot through the data appears to fit very well. (See screen 14.)

(12)

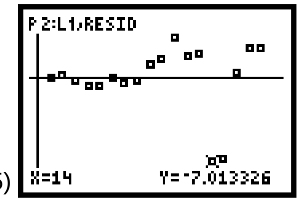


(13)

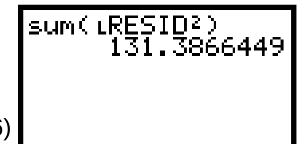


(14)

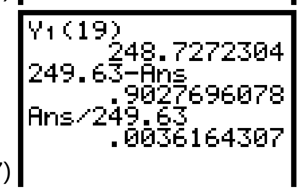
(cont.) **Step in Topic 11**	**Display**
3. *Plot residuals.*	The residual plot appears more random than that in Topic 11. (See screen 15.)
4. *Measure the fit (SSE).*	As shown in screen 16, SSE = **131.387** for the quadratic fit compared to **4651.515** in Topic 11.
5. *Predict population in 1990 (X = 19).*	The prediction for 1990 (X = 19) is off by less than 1 percent (**0.36** percent). (See screen 17.)
6. *Calculate* **r** *and* $\mathbf{r}^2$ *as measure of linearity.*	Note that SSE is directly related to $\mathbf{R}^2$ for this multiple linear least squares fit as it was to $\mathbf{r}^2$ before for the “simple” linear least squares fit. ($Y = ax^2 + bx + c = a\,X_2 + bX_1 + c$ is linear in the coefficient a, b, and c with $X_2 = (X_1)^2$.) (See screen 18.)



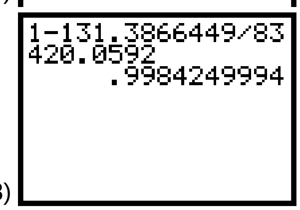
(15)



(16)



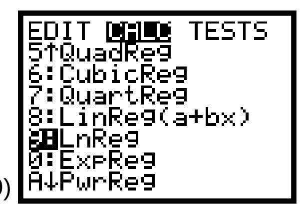
(17)



(18)

Topic 13— Fits Linear by Transformations: Logarithmic, Exponential, and Power Regression

Press [STAT] **<CALC>** and then [▼] a few times to reveal screen 19. This section discusses the last three functions shown.



(19)

9: LnReg Fits $y = a + b(\ln x) = a + b\,X$ (linear in a and b). Calculates a and b using linear least squares on lists of $\ln x$ and y instead of x and y.

0: ExpReg Fits $y = a * b^x = a * b^x$.
Transforms to $(\ln y) = (\ln a) + (\ln b)x = A + Bx$ (not linear in a and b).
Calculates A and B using linear least squares on list of x and $\ln y$ instead of x and y, then $a = e^A$ and $b = e^B$.

A: PwrReg Fits $y = a * x^b = a * x^b$
Transforms to $(\ln y) = (\ln a) + b(\ln x) = A + bX$ (not linear in a and b).
Calculates A and b using linear least squares on list of $\ln x$ and $\ln y$ instead of x and y, then $a = e^A$.

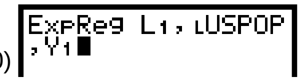
Activity 3, Fitting an Equation to Bivariate Data (cont.)

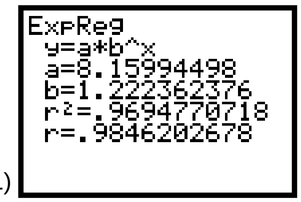
Exponential Fit

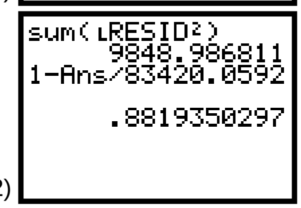
You will fit the exponential equation to the population census data; however, the procedure for the other fits is the same.

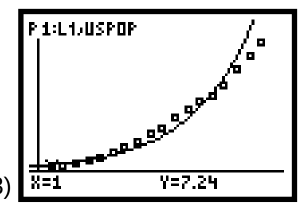
Note that the numbers of the steps below refer to the steps in Topic 11 that present more detail.

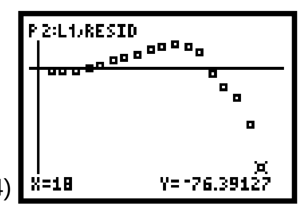
Step in Topic 11	Display
1. *Calculate the fit equation.*	Press [STAT] **<CALC> 0:ExpReg L1** [,] **ʟUSPOP** [,] **Y1** [ENTER] for screens 20 and 21 It is best to turn off the diagnostic flag (see Topic 8) here because **r** and $\mathbf{r}^2$ pertain to the transformed equation above and not to the fit equation in screen 21.
4. *Measure the fit (SSE).* 6. *Calculate* **r** *and* $\mathbf{r}^2$ *as measure of linearity.*	As shown in screen 22, SSE **= 9848.987**, but there is no relationship between this and $\mathbf{r}^2$ (**.8819 ≠ .9695**)
2. *Plot data scatter and fit equation.* 3. *Plot residuals.*	As you can see in screens 22 and 24, the exponential curve does not fit well. The residual plot shows the unfortunate pattern of larger errors as time progresses.
5. *Predict population in 1990 (X = 19).*	Screen 25 shows an error of **-48.3** percent for 1990.

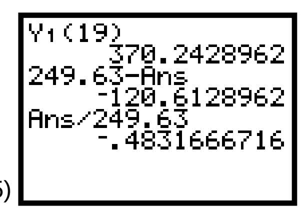
(20)


(21)


(22)


(23)


(24)


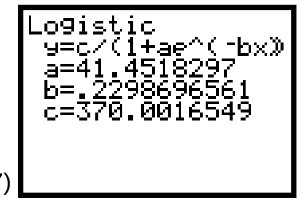
(25)


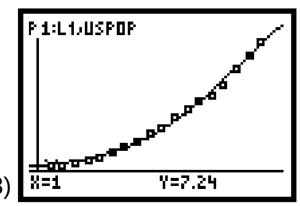
Topic 14—Logistic Fit

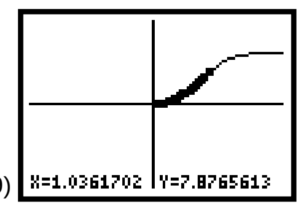
Note that the numbers of the steps below refer to the steps in Topic 11 that present more detail.

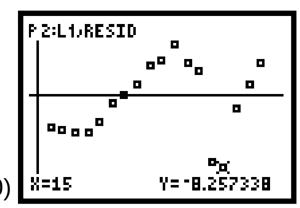
Step in Topic 11	Display
1. *Calculate the fit equation.*	Press [STAT] **<CALC>** **B:Logistic L1** [,] **ʟUSPOP** [,] **Y1** for screen 26. Press [ENTER], and notice the busy symbol in the upper-right corner of the display screen as calculations are being crunched out. The results are shown in screen 27. The technique used attempts to recursively estimate **a**, **b**, and **c** to make SSE as small as possible, and this takes some time.
2. *Plot data scatter and fit equation.* 3. *Plot residuals.*	The logistic fit curve seems to snake through the data as seen in screen 28 and confirmed by the residual plot shown in screen 30. If after making the first plot (screen 28) you press [ZOOM] **3: Zoom Out** [ENTER], you get the view that is shown in screen 29. The logistic curve levels off. It does not continue to grow as fast as the quadratic and exponential curves.
4. *Measure the fit (SSE).* 5. *Predict population in 1990 (X = 19).*	Screens 31 and 32 show SSE = **307.184**. The prediction for 1990 is 2.85 percent off the actual value.

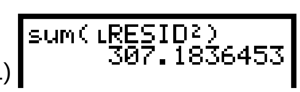
(26)

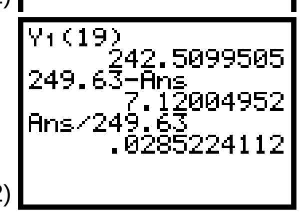

(27)


(28)


(29)


(30)


(31)


(32)


Comparison of Fits Used (Topics 11 to 14)

The quadratic fit seemed best in the short run, but the logistic fit is not far behind and, hopefully, has the advantage of a more realistic long-run projection.

Fit	SSE	% Error X = 19	Resid Plot	Long Run
Linear	4652	17.1	clear pattern	grows linearly
Quadratic	131	0.4	seems random	grows prop to x^2
Exponential	9849	-48.3	clear pattern	grows exponentially
Logistic	307	2.9	screen 30	levels off

Topic 15—Median-Median Linear Fit

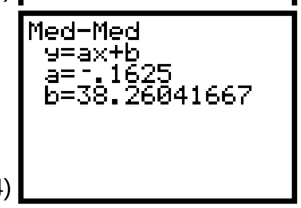
Xlist in **L1**: (horsepower)	75	80	85	100	125	135	160	175
Ylist in **L2**: (milles per gal.)	27	25	15	22	19	16	10	12

The data set above was selected to show the advantage of the median-median fit. Because the medians of batches of data are used, the fit is resistant to unusual data points.

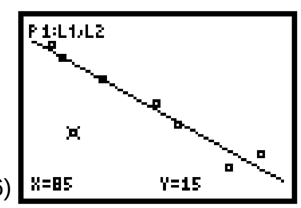
Note that the numbers of the steps below refer to the steps in Topic 11 that present more detail.

Step in Topic 11	Display
1. *Calculate the fit equation.*	Press [STAT] **<CALC>** **3:Med-Med L1** [,] **L2** [,] **Y1** for screen 33. Press [ENTER] for screen 34.
4. *Measure the fit (SSE).*	As shown in screen 35, SSE = **101.25**.
2. *Plot data scatter and fit equation*	Press [ZOOM] **9:ZoomStat** [TRACE] [▸] [▸] for both a **Scatter** plot of the data (with point **x = 85** and **y = 15** highlighted) and of the **Med-Med** line, as shown in screen 36.

(33)


(34)


(35)

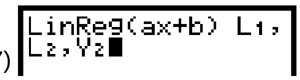

(36)


Comparison with Least-Squares Fit Line

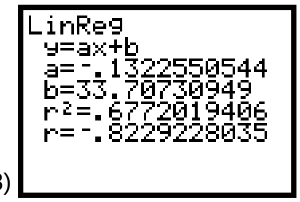
Note that the numbers of the steps below refer to the steps in Topic 11 that present more detail.

Step in Topic 11	Display
1. *Calculate the fit equation.*	Press [STAT] **<CALC>** **4:LinReg(ax + b)** **L1** [,] **L2** [,] **Y2** for screen 37 (be sure that you use **Y2** and not **Y1** as before). Press [ENTER] for screen 38. Note that **r** = **-.823**.
4. *Measure the fit (SSE).*	As shown in screen 39, SSE **= 83.77** is less than the **101.25** of **Med-Med** as theory guarantees. But having the smallest SSE does not always guarantee the better fit, as you can observe in the plots that follow in screen 40.
2. *Plot data scatter and fit equation*	With **Plot1**, **Y1**, and **Y2** on, press [ZOOM] **9:ZoomStat** for screen 40, which shows both the **Med-Med** and **LinReg** fit lines. Note that the least-squares line is pulled toward point **x** = 85, **y** = 15.

(37)

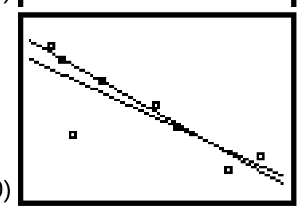


(38)



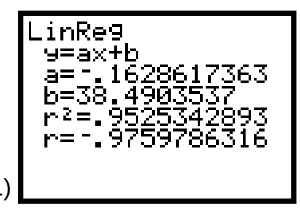
(39)

(40)

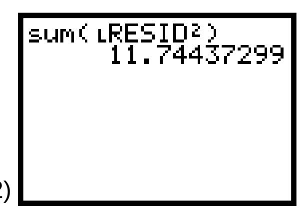


If the point **x = 85**, **y = 15** is deleted from **L1** and **L2** and a **LinReg** line plotted to the data, we obtain the results shown in screens 41 and 42. The slope and the intercept are about the same as the **Med-Med** fit without deleting the data point (see screen 34); **r = -.976** compared to **-.8229** in screen 38, and SSE = **11.74**, reduced from **83.77** in screen 39. The **Med-Med** fit is a good check on how influential such points are.

(41)



(42)



Topic 16—Trigonometric Sine Fit

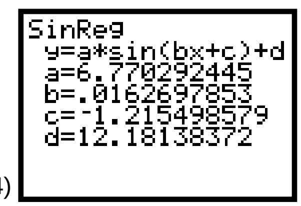
Those who deal with periodic data, in Physics experiments, for example, will want to read about the **SinReg** (sinusoidal regression) function in the Statistics chapter of the *TI-83 Guidebook* for more information on this topic. The following data is from the example in the *Guidebook* with x representing the day of the year (equal intervals of every 30th day) and y the number of daylight hours in Alaska.

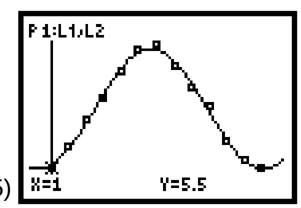
x (day) **L1**:	1	31	61	91	121	151	181	211	241	271	301	331	361
y (hrs) **L2**:	5.5	8	11	13.5	16.5	19	19.5	17	14.5	12.5	8.5	6.5	5.5

Note that the numbers of the steps below refer to the steps in Topic 11 that present more detail.

Step in Topic 11	Display
1. *Calculate the fit equation.*	Press [STAT] **<CALC> C:SinReg L1** [,] **L2** [,] **Y1** for screen 43. Press [ENTER] for screen 44.
2. *Plot data scatter and fit equation.*	Enter [2nd] [STAT PLOT] **<Plot1>** with **Xlist: L1** and **Ylist: L2**. (**Y1** is on from step 1, but all other **Y=** or stat plots must be off.)
	Press [ZOOM] **9:ZoomStat** [TRACE], and both a **Scatter** plot of the data and the sine fit will show as shown in screen 45.
	Press [ZOOM] **3: Zoom Out** [ENTER] to get the view shown in screen 46. This view better shows the periodic nature of the fit.

(43)


(44)


(45)


(46)