Scatter Plots and Line of Best Fit

A little vocab...

- The line of best fit is the line that lies as close as possible to all the data points.
- Linear Regressions a method used to find the equation of the best fitting line or curve.
- Extrapolation is the use of the regression curve to make predictions outside the domain of values of the independent variable.
- Interpolation is used to make predictions within the domain of values of the independent variable.

Line of Best Fit by Hand:

1. The environment club is interested in the relationship between the number of canned beverages sold in the cafeteria and the number of cans that are recycled. The data they collected are listed in this chart.

									4
Beverag	e Can	Recy	cling		and the	1	States		20
Number of Canned Beverages Sold	18	15	19	8	10	13	9	14	4 . 2 18
Number of Cans Recycled	3	6	10	6	3	7	5	4	4 × 16
									8 14
lot the points to make a		ttor	pla						\$ 12
ion me points to make a	sco	mer	pio	1.					# 10
									8
lse a straightedge to ap	pro)	kimo	ate t	he	line	oft	pest	fit k	t by hand.
									A A A A A A A A A A A A A A A A A A A
nd an equation of the line	of b	est f	it for	the	e da	ta.	19,1	5).), (13,7)
1=7-5 = 2=1		5:	:1	19	1+	b			
13-9 4 2		5	- 4	5	+6				2 4 6 8 10 12 14 16 18 20
		-	1				4=	1	X+1 + AC - II - II
	Beverage Number of Cauned Beverages Sold Number of Caus Recycled Number of Caus Recycled Not the points to make a use a straightedge to ap and an equation of the line $N = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$	Beverage CanNumber of Canned Beverages Sold18Number of Cans Recycled8lot the points to make a scalot the points to make a scalse a straightedge to approximation of the line of b $n = 1 - 5$ $= 2 + = 1$ $13 - 9$ $4 = 2$	Beverage Can RecyNumber of Canned Beverages Sold1815Number of Cans Recycled86Iot the points to make a scatterIse a straightedge to approximationInd an equation of the line of best to $N = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ 5 $3 = \frac{7}{13-9} = \frac{2}{4} = \frac{1}{2}$ 5	Beverage Can RecyclingNumber of Canned Beverages Sold181519Number of Cans Recycled8610Iot the points to make a scatter ploIse a straightedge to approximate toInd an equation of the line of best fit for $N = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ $5 = -\frac{1}{2}$ $5 = 4$	Beverage Can RecyclingNumber of Canned Beverages Sold1815198Number of Cans Recycled86106lot the points to make a scatter plot.lot the points to make a scatter plot.lot a straightedge to approximate thean equation of the line of best fit for the $N = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ $5 = -\frac{1}{2} \begin{pmatrix} 9 \\ 5 = 4.5 \end{pmatrix}$	Beverage Can RecyclingNumber of Canned Beverages Sold181519810Number of Cans Recycled801063Iot the points to make a scatter plot.Ise a straightedge to approximate the linean equation of the line of best fit for the day $N = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ $5 = -\frac{1}{2}(9) + 5 = 4.5 + b$	Beverage Can RecyclingNumber of Canned Beverages Sold18151981013Number of Cans Recycled8610637Iot the points to make a scatter plot.Ise a straightedge to approximate the line of bInd an equation of the line of best fit for the data. $A = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ $5 = -\frac{1}{2} \begin{pmatrix} 9 \\ 9 \end{pmatrix} + b$ $5 = 4.5 + b$	Beverage Can RecyclingNumber of Canned Beverages Sold181519810139Number of Cans Recycled86106375Iot the points to make a scatter plot.Ise a straightedge to approximate the line of bestan equation of the line of best fit for the data.(9, 10) $A = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ $5 = -\frac{1}{2}(9) + b$ $5 = 4.5 + b$ $4 = 2$	Beverage Can RecyclingNumber of Canned Beverages Sold1815198101391Number of Cans Recycled86106375Iot the points to make a scatter plot.Ise a straightedge to approximate the line of best fitan equation of the line of best fit for the data.(9.5) $A = \frac{7-5}{13-9} = \frac{2}{4} = \frac{1}{2}$ $5 = -\frac{1}{2} \begin{pmatrix} 9 \\ 9 \end{pmatrix} + b$ $5 = 4.5+b$ $u = 1$

Line of Best Fit using the calculator:

- 1. DATA DATA 4 (this will clear all data already in the tables)
- 2. DATA (type in data)
- 3. 2nd DATA
- LinReg ax + b (for linear regression)
 ExpReg ab^x (for exponential regression)
 L1 L2 ONE YES CALC
- 5. a = b =
 - r =
- 6. The equation of the line is y = ax + b
- (, 7. Correlation Coefficient is r.

		L1: 0	5 10	15	18987
2. The table shows	the total outstan	ding consumer d	ebt (excluding h	ome mortaages)	in billions of
dollars in selected	years. (Data is fro	om the Federal Re	eserve Bulletin.)	0.0.1	
(et x = 0 correspon	nd to 1985.				
Year, t	1985 (D)	1990 (5)	1995 (10)	2000 (15)	2003 (18)
Consumer Debt	585	789	1096	1693	1987
a) Find the linear places,	regression equati	ion. Also, find the	correlation coeff $y = \underline{79}$	86 x +463	two decimal
b) Find and interp	ret the slope of the	ne regression line.	The consu	mer debt	Increases
J		ionars per	year.		
c) Find and interpo	4 63.35 bil	t of the regression lion dalars.	line. At yea	ar o, the c	consumer
d) Find the approx	kimate consumer	debt in 1998.	79.86(13) + 463.34	5= \$ 1501.50 billia
e) Find the approx	imate consumer M 1985)	debt in 2008.	19.86(23)+463.35:	= 2300.10 billin
3. The table below	shows the numb	per of deaths per	100,000 people f	rom heart disease	e in selected
ears. (Data is fro	m the U.S. Nation	al Center for Hec	alth Statistics.)		
Let x = 0 correspon	nd to 1960.				
Year	1960 (0)	1970 (IL) 1	980 (20) 199	20 (30) 2000	(40) 2002 (42)
Deaths	559	483	412 32	2 230	
a)) Find the linea places.	ar regression equ	ation. Also, find th	the correlation correlation $y = -7$	efficient r. Round	to two decimal 25 (=999
b) Find and interpr 100,000 per	et the slope of the	he regression line	the num decrease	ber of dra	reach year
c) Find and interpr	et the y-intercer	ot of the regression	on line. At ye	ar o, the r	number of
acamp wow	imate number o	of deaths due to h	neart disease in i	995 using the line	ear regression
d) Find the approx equation.	62(35)+5	59.25=		293 deat	hs
			and disagra in '	2008 using the line	ear regression
e) Find the approx	imate number o	of deaths due to l	nearr disease in .	194 dea	ths

((equation. -7.62(48)+559.25 =

3

4. Mike is riding his bike home from his grandmother's house. In the table below, x represents the number of hours Mike has been biking and y represents the number of miles Mike is away from home. Make a scatter plot for this data on the grid below.

Hours	(x)	1	2	3	4	5	6	7	8
Miles (y)	35	29	26	20	16	9	6	0

b. Find the linear regression equation for the data. Round answers to 2 decimal places. $Y = -4.94 \times + 39.86$

- c. What does the slope represent in the context of the problem? The decrease in miles each how
- d. What does the y-intercept represent in the context of the problem? At time=Ohrs, Mike is 39.86 miles a Way from his house.



e. Could you use your equation to predict how far Mike would be after 10 hours? Use mathematics to justify your answer. -4.94(10) + 39.86 = -9.54

NO, he's already 4.94(10)+39.86 =

5. Use the table below to answer the questions about the population p (in millions) in Florida.

Let x = 0 correspond to 2002

Or

Year, t	2002 (0)	2003 (1)	2004 (2)	2005 (3)
Population (millions)	16.4	17.0	17.4	17.8

a. Find the linear regression equation for the data. y = -44, x + 16.46

c. Using the regression equation, what will be the population in 2020?

X=18

42.46(18) HG.46= 24.74 million

Use the table below to answer the questions about the U.S. residential carbon dioxide emissions from 1993 to 2002. Emissions are measured in million metric tons.

 $f_1 + x = 0$ correspond to 1993

Year, t	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Emissions	1027.6	1020.9	1026.5	1086.1	1077.5	1083.3	1107.1	1170.4	1163.3	1193.9

- a. Find the linear regression equation for the data. $y = 19.89 \times \pm 1000.16$
- c. Using this model, how many residential tons were emitted in 1990? V = -3 U = 10.96(-3) + 100(-1) = 0(1) + 100(-1)
- y=19.89(-3)+100(.1)=946.49 millim Md d. Using this model, how many residential tons were emitted in 2010?

x=17 y=19.89(17)+1006.16=1344.29 millin Incharctone

7. Use the table below to answer the questions about the operating costs in thousands of a small business from 2000 to 2007.

Let x = 0 correspond to 2000

Year	,t	2000 [6]	2001	2002 2)	2003 (3)	2004 (4)	2005 (5)	2006 (4)	2007 (7	1)
Ope	rating ts	2.3	2.6	3.1	3.3	4.0	5.2	5.9	7.0	1

- a. Find the linear regression equation for the data. $y = -67 \times +1.83$
- b. Identify and describe the correlation coefficient. r = -97. There is a <u>strong</u> (strong or weak) <u>positive</u> (positive or negative) correlation between time elapsed and <u>operating custs of SMall</u> <u>business</u>

c. Using this model, what will be the operating costs in 2015? Y = .67(15) + 1.83 = \$11.88 thousand

For 8 & 9, Theresa started making homemade cards to send to friends and family and to sell at the local craft fair. The scatter plot shows how many cards Theresa made each hour she worked.

