## Scatter Plots and tUne 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 Regressiuns a method used to lind the equation of the best lifting line or curve.
- Extrapolation is the use of the regression curve to make predictions outside the domain of
- Interpolation variable.
is used to make predictions within the domain of values of the independent


## Line of Best Flt 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.

$$
0
$$

| Beverage Can Recycling |  |  |  |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Number of Canned Beverages Sold | 18 | 15 | 19 | 8 | 10 | 13 | 9 | 14 |
| Number of Cans Recycled | 8 | 6 | 10 | 6 | 3 | 7 | 5 | 4 |

a) Plot the points to make a scatter plot.
b) Use a straightedge to approximate the line of best fit by hand.
c) Find an equation of the line of best fit for the data. $(9,5),(13,7)$
$m=\frac{7-5}{13-9}=\frac{2}{4}=\frac{1}{2} \quad 5=\frac{1}{2}(a)+b$

$$
\begin{aligned}
& 5=4.5+b \\
& .5=b
\end{aligned} \quad y=\frac{1}{2} x+\frac{1}{2}
$$

Plot the points moke scatter plot.


## 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. $2^{\text {nd }}$ DATA
4. LinReg $a x+b$ (for linear regression)

ExpReg $a b^{\wedge} x$ (for exponential regression)
11 LL ONE YES CALC
5. $a=$
$b=$
$r=$
6. The equation of the line is $y=a x+b$
7. Correlation Coefficient is r .

| LI: 0 | 5 | 10 | 15 | 18 |
| :--- | :--- | :--- | :--- | :--- |
| $L 2: 585$ | 789 | 1096 | 1693 | 1987 |

2. The table shows the total outstanding consumer debt (excluding home mortgages) in billions of dollars in selected years. (Data is from the Federal Reserve Bulletin.)
et $\mathrm{x}=0$ correspond to 1985 .

| Year, I | $1985(0)$ | $1990(5)$ | $1995(10)$ | $2000(15)$ | 2003 (18) |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Consumer Debt | 585 | 789 | 1096 | 1693 | 1987 |

a) Find the linear regression equation. Also, find the correlation coefficient r. Round to two decimal places.

$$
y=79.86 \quad x+463.35 \quad \ldots 98
$$

b) Find and interpret the slope of the regression line. The consumer debt increases by 79.86 billion dollars per year.
c) Find and interpret the $y$-intercept of the regression line. At year 0, the consumer debt was 463.35 billion dollars. (
d) Find the approximate consumer debt in $1998 . \quad 79.86(13)+463.35=\$ 1501.50$ billia ( 13 years from 1985)
e) Find the approximate consumer debt in 2008. $79.86(23)+463.35=2300.10$ billie (23 years from 1985)
3. The table below shows the number of deaths per 100,000 people from heart disease in selected ears. (Data is from the U.S. National Center for Health Statistics.)

Let $\mathrm{x}=0$ correspond to 1960.

| Year | $1960(0)$ | 1970 | $(16)$ | 1980 | $(20)$ | 1990 | (30) | $2000(40)$ | $2002(42)$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Deaths | 559 | 483 | 412 | 322 | 258 | 240 |  |  |  |

a) ) Find the linear regression equation. Also, find the correlation coefficient $r$. Round to two decimal places.

$$
\begin{aligned}
& \text { Cation coefficient } r \text {. Round to two decimal } \\
& y=-7.62 \quad x+559.25 r=-.999
\end{aligned}
$$

b) Find and interpret the slope of the regression line. The number of deaths per 100,000 people fum heart disease decreased by 7.62 each year
c) Find and interpret the $y$-intercept of the regression line. At year $D$, the number of deaths from heart disease was 559.25 .
d) Find the approximate number of deaths due to heart disease in 1995 using the linear regression equation.

$$
-7.62(35)+559.25=
$$ 293 deaths

e) Find the approximate number of deaths due to heart disease in 2008 using the linear regression equation. 194 deaths

## 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) | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ | $\mathbf{8}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Miles (y) | 35 | 29 | 26 | 20 | 16 | 9 | 6 | 0 |

a. Describe the correlation between the data points on the scatter plot. $r=-.998$. There is a Strong (strong or weak) negative (positive or negative) correlation between hours biking from nome. and \# of miles away
b. Find the linear regression equation for the data. Round answers to 2 decimal places.

$$
Y=-4.94 x+39.86
$$

c. What does the slope represent in the context of the problem? The decrease in miles each hour
d. What does the $y$-intercept represent in the context of the problem? At time $=$ Ohrs, Mike is 39.86 miles away from his hose.

e. Could you use your equation to predict how far Mike would be after 10 hours? Use mathematics to justify your answer.
No, he's already nome.

$$
-4.94(10)+39.86=-9.54
$$

5. Use the table below to answer the questions about the population p (in millions) in Florida. Let $\mathrm{x}=0$ correspond to 2002

| Year, t | $\mathbf{2 0 0 2}$ (0) | $\mathbf{2 0 0 3}$ (1) | $\mathbf{2 0 0 4}$ (2) | 2005 (3) |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Population <br> (millions) | 16.4 | 17.0 | 17.4 | 17.8 |

a. Find the linear regression equation for the data. $y=$ $\qquad$ .46. $x+16.46$
b. Identify and describe the correlation coefficient.
$r=.99$. There is a Strong (strong or weak) posit (positive or negative) correlation between A time and population
c. Using the regression equation, what will be the population in 2020 ?

$$
x=18 \quad y=.46(18)+16.46=24.74 \text { millim }
$$

6. 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.
$t x=0$ 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. $y=19.89 x+1006.16$
b. Identify and describe the correlation coefficient. $r=.96$. There is a sting (strong or weak) positive (positive or negative) correlation between * of years elapsed and carbon dioxide emissim's
c. Using this model, how many residential tons were emitted in 1990 ?
$x=-3 \quad y=19.89(-3)+1006$.
d. Using this model, how many residential tons were emitted in 2010? tons
$x=17 \quad y=19.89(17)+1006.16=1344.29 \mathrm{millm}$ metorctons
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

a. Find the linear regression equation for the data. $y=$ $\qquad$ .67 $x+1.83$
b. Identify and describe the correlation coefficient. strong (strong or weak) positive (positive or negative) correlation between time elapsed and operating ousts of small business
c. Using this model, what will be the operating costs in 2015 ?

$$
x=15 \quad y=.67(15)+1.83=\$ 11.88 \text { thousand }
$$

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.
8. Estimate the number of cards

Theresa can make in 14 hours.
A. 14
B. 10
C. 70
(D.) 41

Card Creation

9. The correlation coefficient of lines of best fit is given below. Which correlation coefficient best represents the graph?
A. 0.095
B. 0.976
C. -0.095
D. -0.976

