(4) Describe what a **regression line** is and explain how it is used.

Given a set of points $\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$, a regression line is the line that best approximates the data set. In other words, we call the regression line the best-fit line.

We use regression lines to (1) simplify calculations and

(2) see if the x-values and y-values have relationship that is sufficiently linear.

(5) Given a set of points $\{(x_1, y_1), ..., (x_n, y_n)\}$, the regression line $y = ax + b$ for constants a and b can be calculated using matrices. This problem will walk you through an example calculation with a data set of two points.

(a) Let $(3, 350)$ and $(5, 270)$ represent the population size of an elk herd in a region 3 years after 2015 and 5 years after 2015 respectively. Assume that there exists a line passing through the two points. Express that line as $y = ax + b$ for constants a and b. This is our regression line.

Write down a system of equations, using a and b as variables, using that the two points are on the line $y = ax + b$.

$$\boxed{\begin{cases} 350 = a(3)+b \\ 270 = a(5)+b \end{cases} ;}$$

(b) Rewrite the system using matrices.

$$\begin{pmatrix} 3a+b \\ 5a+b \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}(a) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} b = \boxed{\begin{pmatrix} 3 & 1 \\ 5 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 350 \\ 270 \end{pmatrix};}$$

In general, the matrix equation in linear regression is in the form $\begin{pmatrix} | & 1 \\ x\text{-values} & \vdots \\ | & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} | \\ y\text{-values} \\ | \end{pmatrix}$;

(c) Let $\mathbf{M}$ be the coefficient matrix and $\mathbf{y}$ be the column vector of y-values from part (b). Then, solution can be calculated by

$$\begin{pmatrix} a \\ b \end{pmatrix} = (\mathbf{M}^{tr}\mathbf{M})^{-1}\mathbf{M}^{tr}\mathbf{y}.$$

Find a and b using this method.

$$M = \begin{pmatrix} 3 & 1 \\ 5 & 1 \end{pmatrix}; \quad y = \begin{pmatrix} 350 \\ 270 \end{pmatrix}; \quad M^{tr} = \begin{pmatrix} 3 & 5 \\ 1 & 1 \end{pmatrix};$$

$$M^{tr}M = \begin{pmatrix} 3 & 5 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 5 & 1 \end{pmatrix} = \begin{pmatrix} (3)(3)+(5)(5) & (3)(1)+(5)(1) \\ (1)(3)+(1)(5) & (1)(1)+(1)(1) \end{pmatrix}$$

$$= \begin{pmatrix} 9+25 & 3+5 \\ 3+5 & 1+1 \end{pmatrix} = \begin{pmatrix} 34 & 8 \\ 8 & 2 \end{pmatrix};$$

$$(M^{tr}M)^{-1} = \frac{1}{(34)(2)-(8)(8)}\begin{pmatrix} 2 & -8 \\ -8 & 34 \end{pmatrix} = \frac{1}{68-64}\begin{pmatrix} 2 & -8 \\ -8 & 34 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 2 & -8 \\ -8 & 34 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -4 \\ -4 & 17 \end{pmatrix};$$

$$(M^{tr}M)^{-1}M^{tr} = \frac{1}{2}\begin{pmatrix} 1 & -4 \\ -4 & 17 \end{pmatrix}\begin{pmatrix} 3 & 5 \\ 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} (1)(3)+(-4)(1) & (1)(5)+(-4)(1) \\ (-4)(3)+(17)(1) & (-4)(5)+(17)(1) \end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix} 3-4 & 5-4 \\ -12+17 & -20+17 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 5 & -3 \end{pmatrix};$$

$$\begin{pmatrix} a \\ b \end{pmatrix} = (M^{tr}M)^{-1}M^{tr}y = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 5 & -3 \end{pmatrix}\begin{pmatrix} 350 \\ 270 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} (-1)(350)+(1)(270) \\ (5)(350)+(-3)(270) \end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix} -350+270 \\ 1750-810 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -80 \\ 940 \end{pmatrix} = \begin{pmatrix} -40 \\ 470 \end{pmatrix};$$

Therefore, $\boxed{a = -40 \text{ and } b = 470;}$

(d) Use your answer in the previous part to construct the equation for the regression line.

Plug the values back in $y = ax + b$.

ANS: $\boxed{y = -40x + 470;}$

(e) How can you verify your equation? (There are a number of ways to do this. Give at least one.)

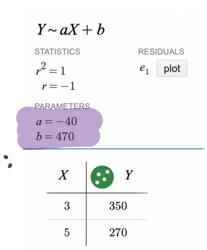
Possible Answer: The points $(3, 350)$ and $(5, 270)$ should be in the line (or close enough).

For $x = 3$: $y = -40(3) + 470 = 350$. ✓

For $x = 5$: $y = -40(5) + 470 = 270$. ✓

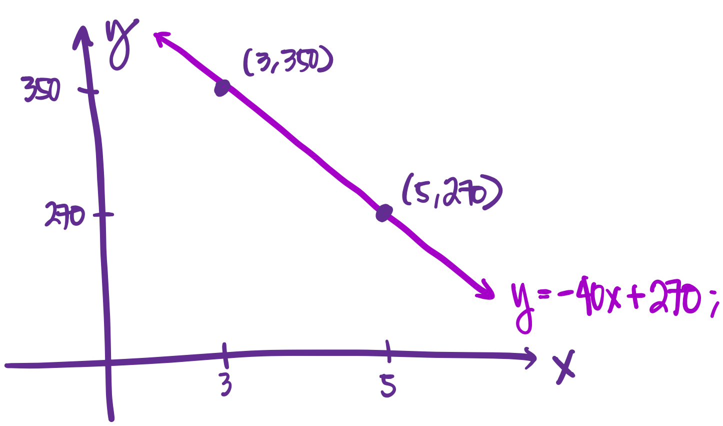
Possible Answer: Use software that does the linear regression for you.

This was the result from Desmos:



(f) Since the data set only involves two points, the regression line will contain those two points. If there are three or more points, the regression line may not pass through all of our points. This is where the term "best fit" comes in.

Provide a sketch of the graph containing the following: (1) the two given points from part (a) and (2) the regression line we've found.



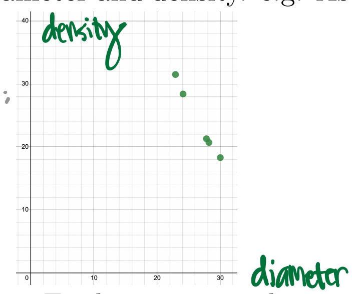
(6) The table below gives the tree diameters measured in inches at breast height and the corresponding density of trees per area in a region.

Diameter (in)	22.9	30.0	27.8	24.1	28.2
Density (trees/acre)	31.5	18.3	21.3	28.4	20.7

(a) Explain, in words, a general relationship between tree diameter and density. e.g. As tree diameter increases, what happens to density?

$\boxed{\text{As tree diameter increases, the density decreases.}}$

The graph looks like this:



(b) Let x represent the tree diameters and y represent density. Find a regression line $y = ax + b$ using the method from Problem (5).

X-values: $\begin{pmatrix} 22.9 \\ 30.0 \\ 27.8 \\ 24.1 \\ 28.2 \end{pmatrix}$; y-values: $\begin{pmatrix} 31.5 \\ 18.3 \\ 21.3 \\ 28.4 \\ 20.7 \end{pmatrix}$; Then, the coordinate matrix $M = \begin{pmatrix} 22.9 & 1 \\ 30.0 & 1 \\ 27.8 & 1 \\ 24.1 & 1 \\ 28.2 & 1 \end{pmatrix}$;

We need to use $\begin{pmatrix} a \\ b \end{pmatrix} = (M^{tr}M)^{-1}M^{tr}y$;

$$M^{tr}M = \begin{pmatrix} 22.9 & 30.0 & 27.8 & 24.1 & 28.2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 22.9 & 1 \\ 30.0 & 1 \\ 27.8 & 1 \\ 24.1 & 1 \\ 28.2 & 1 \end{pmatrix} \overset{\text{Calculator}}{=} \begin{pmatrix} 3573.3 & 133 \\ 133 & 5 \end{pmatrix};$$

$$(M^{tr}M)^{-1} = \frac{1}{(3573.3)(5) - (133)(133)}\begin{pmatrix} 5 & -133 \\ -133 & 3573.3 \end{pmatrix} = \frac{1}{177.5}\begin{pmatrix} 5 & -133 \\ -133 & 3573.3 \end{pmatrix};$$

$$M^{tr}y = \begin{pmatrix} 22.9 & 30.0 & 27.8 & 24.1 & 28.2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 31.5 \\ 18.3 \\ 21.3 \\ 28.4 \\ 20.7 \end{pmatrix} = \begin{pmatrix} 3130.67 \\ 120.2 \end{pmatrix};$$

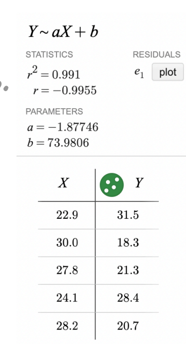
$$\begin{pmatrix} a \\ b \end{pmatrix} = (M^{tr}M)^{-1}M^{tr}y = \frac{1}{177.5}\begin{pmatrix} 5 & -133 \\ -133 & 3573.3 \end{pmatrix}\begin{pmatrix} 3130.67 \\ 120.2 \end{pmatrix} \overset{\text{Calculator}}{\approx} \begin{pmatrix} -1.88 \\ 73.98 \end{pmatrix};$$

Regression line: $\boxed{y = -1.88x + 73.98;}$

Checking: ① Doing all matrix calculations in Symbolab.

$$\left(\begin{pmatrix} 22.9 & 1 \\ 30 & 1 \\ 27.8 & 1 \\ 24.1 & 1 \\ 28.2 & 1 \end{pmatrix}^T \begin{pmatrix} 22.9 & 1 \\ 30 & 1 \\ 27.8 & 1 \\ 24.1 & 1 \\ 28.2 & 1 \end{pmatrix}\right)^{-1} \begin{pmatrix} 22.9 & 1 \\ 30 & 1 \\ 27.8 & 1 \\ 24.1 & 1 \\ 28.2 & 1 \end{pmatrix}^T \begin{pmatrix} 31.5 \\ 18.3 \\ 21.3 \\ 28.4 \\ 20.7 \end{pmatrix} = \begin{pmatrix} -1.87746\ldots \\ 73.98056\ldots \end{pmatrix}$$

② Doing regression on Desmos.



(c) Predict, using the regression line you've found in part (c), the density of trees that have a diameter at breast height of 25 inches.

Find y such that $x = 25$; $y = -1.88(25) + 73.98 \approx 26.98$;

If we use more decimal places for a and b, we should get $y = 27.04$;

ANS. The regression line predicts the density to be $\boxed{27 \text{ trees/acre}}$ when tree height is 25 in.

(d) Provide a (reasonable) range of values for the density of trees that have a diameter of at least 25 inches.

The regression line $y = -1.88x + 73.98$ is decreasing.

ANS. We predict that $\boxed{\text{the density} \le 27 \text{ trees/acre}}$ when tree height ≥ 25 in.

(i.e. between 0 and 27.04)

(e) Predict, using the regression line you've found in part (c), the tree diameter if the tree density is 15 trees per acre.

Find x such that $y = 15$;

$$15 = -1.88x + 73.98;$$

$$1.88x = 73.98 - 15;$$

$$x = \frac{58.98}{1.88} = 31.37;$$

Using more decimal places for a and b, we get $x = 31.415$;

ANS. We predict that $\boxed{\text{tree diameter is 31 in}}$ when density is 15 trees/acre.

---

(7) The table below provides data for the length of a snake (species *Vipera Bertis*) to its weight.

Length (cm)	60	66	64	54	67
Weight (g)	136	194	140	93	172

(a) Let x represent the length data and y represent the weight data. Using the method presented in Problem (5), construct the regression line $y = ax + b$ relating snake length with snake weight.

$$x = \begin{pmatrix} 60 \\ 66 \\ 64 \\ 54 \\ 67 \end{pmatrix}; \quad M = \begin{pmatrix} 60 & 1 \\ 66 & 1 \\ 64 & 1 \\ 54 & 1 \\ 67 & 1 \end{pmatrix}; \quad y = \begin{pmatrix} 136 \\ 194 \\ 140 \\ 93 \\ 172 \end{pmatrix};$$

From earlier, $\begin{pmatrix} a \\ b \end{pmatrix} = (M^{tr}M)^{-1} M^{tr} y$;

We can use Symbolab to do the matrix calculations for us.

Then,

$$\begin{pmatrix} a \\ b \end{pmatrix} = \left(\begin{pmatrix} 60 & 1 \\ 66 & 1 \\ 64 & 1 \\ 54 & 1 \\ 67 & 1 \end{pmatrix}^T \begin{pmatrix} 60 & 1 \\ 66 & 1 \\ 64 & 1 \\ 54 & 1 \\ 67 & 1 \end{pmatrix} \right)^{(-1)} \begin{pmatrix} 60 & 1 \\ 66 & 1 \\ 64 & 1 \\ 54 & 1 \\ 67 & 1 \end{pmatrix}^T \begin{pmatrix} 136 \\ 194 \\ 140 \\ 93 \\ 172 \end{pmatrix} = \begin{pmatrix} \frac{1255}{188} \\ -\frac{50425}{188} \end{pmatrix} \approx \begin{pmatrix} 6.68 \\ -268.22 \end{pmatrix};$$

Regression Line: $\boxed{y = \frac{1255}{188}x - \frac{50425}{188}}$ OR $\boxed{y = 6.68x - 268.22;}$

We can confirm this answer using Desmos.

X	Y
60	136
66	194
64	140
54	93
67	172

$Y \sim aX + b$

STATISTICS

$r^2 = 0.8491$

$r = 0.9215$

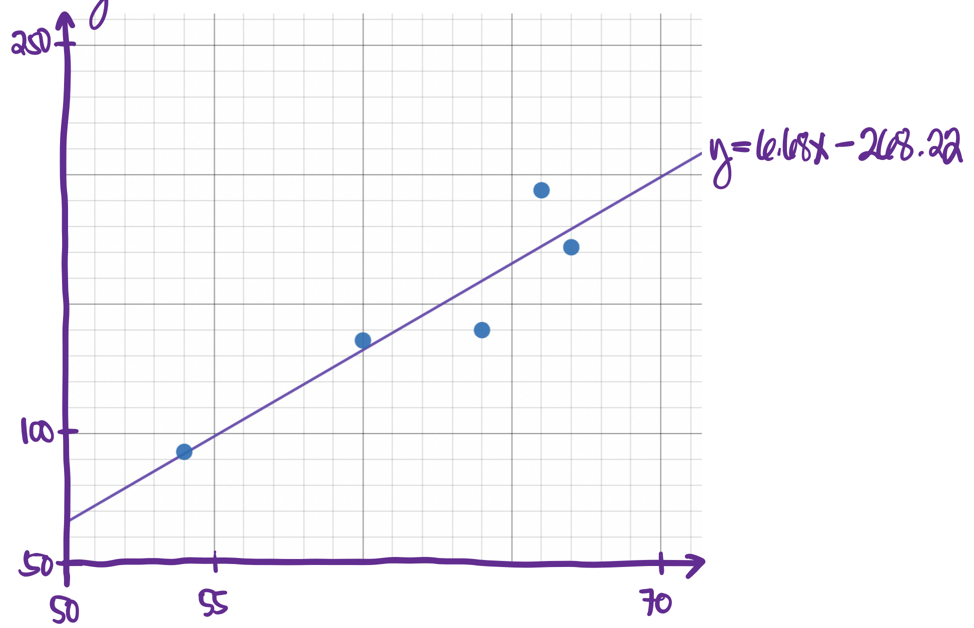
RESIDUALS

e_1 plot

PARAMETERS

$a = 6.67553$

$b = -268.218$



(b) Using the regression line you've found, predict the length of a snake that weighs 50 grams.

Find x such that $y = 50$.

$y = 6.68x - 268.22;$

$50 = 6.68x - 268.22;$

$6.68x = 50 + 268.22;$

$x = \frac{318.22}{6.68} \approx 47.64;$ Using $a = \frac{1255}{188}$ and $b = -\frac{50425}{188}$ yields $x \approx 47.669;$

Ans: We estimate the length of a snake weighing 50 grams to be $\boxed{48 \text{ in.}}$