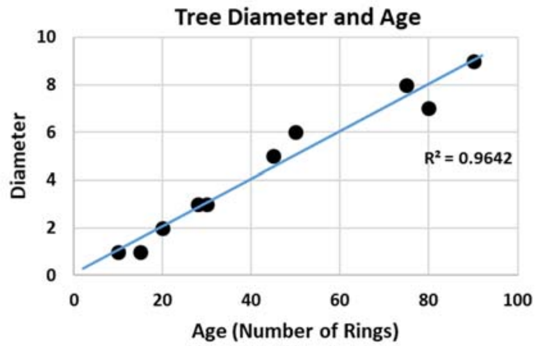
Regression: Tree Rings and Measuring Things

Objectives:

- Measure biological data
- Use biological measurements to calculate means, slope and intercept
- Determine best linear fit of data
- Interpret fit using correlation



Materials:

- Ruler (in millimeters)
- Calculator
- Regression Activity Sheet

Instructional Plan:

Assemble students in groups of 2-3 with a calculator and ruler for each group. Using the sheet 'Example: Regression' have the students examine the scatter plot of growth rings by tree diameter. Have the students consult and draw an approximate best linear fit line through the points. Have them compare the position of the line, the *y*-intercept, and the nearest (*x*, *y*) coordinates to those in the associated 'Calculations' section. How well did they approximate the line?

Next, explain they will be measuring biological data and calculating variables to exactly determine the mean *x* and *y* coordinates, slope, and *y*-intercept. To compare their calculations to their estimation skills, have the students measure the diameter and count the growth rings of the sample trees on the 'Regression Exercise' sheet, entering the data into the table provided. Next have students plot the data. Then have the students calculate the mean of the *x* data ($\bar{x}$) and the mean of the y data ($\bar{y}$). Note that the point ($\bar{x}$, $\bar{y}$) is on the line of best fit for this data. Have the students draw their own line of the best fit

Each student will determine the equation of their line. Also have the students use their calculator to find the equation of best fit.

Also, have students consider what the slope means in terms of the plotted points (rate of change).

Assessments:

Ask the students, "Why is it important to know the best fit line?" The answer is it allows us to predict either DBH (*y*) or the number of growth rings (*x*), using only one coordinate value. As an example, give the students an array of values for either DBH or growth rings or both. Have them use the best fit line to predict what the corresponding value for each. Data could be presented similar to:

Unknown Forest	**Tree 1**	**Tree 2**	**Tree 3**	**Tree 4**	**Tree 5**
Rings (x)	3			6	17
Diameter (y)		30	25		

Following this exercise, discuss what other types of situations would benefit from knowing how to predict values using regression from a smaller data set.

Tennessee Mathematical Standards:

Understand that a function is a rule that assigns to each input exactly one output. The graph of a function is the set of ordered pairs consisting of an input and the corresponding output.

Interpret the equation $y = mx + b$ as defining a linear function, whose graph is a straight line.

Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

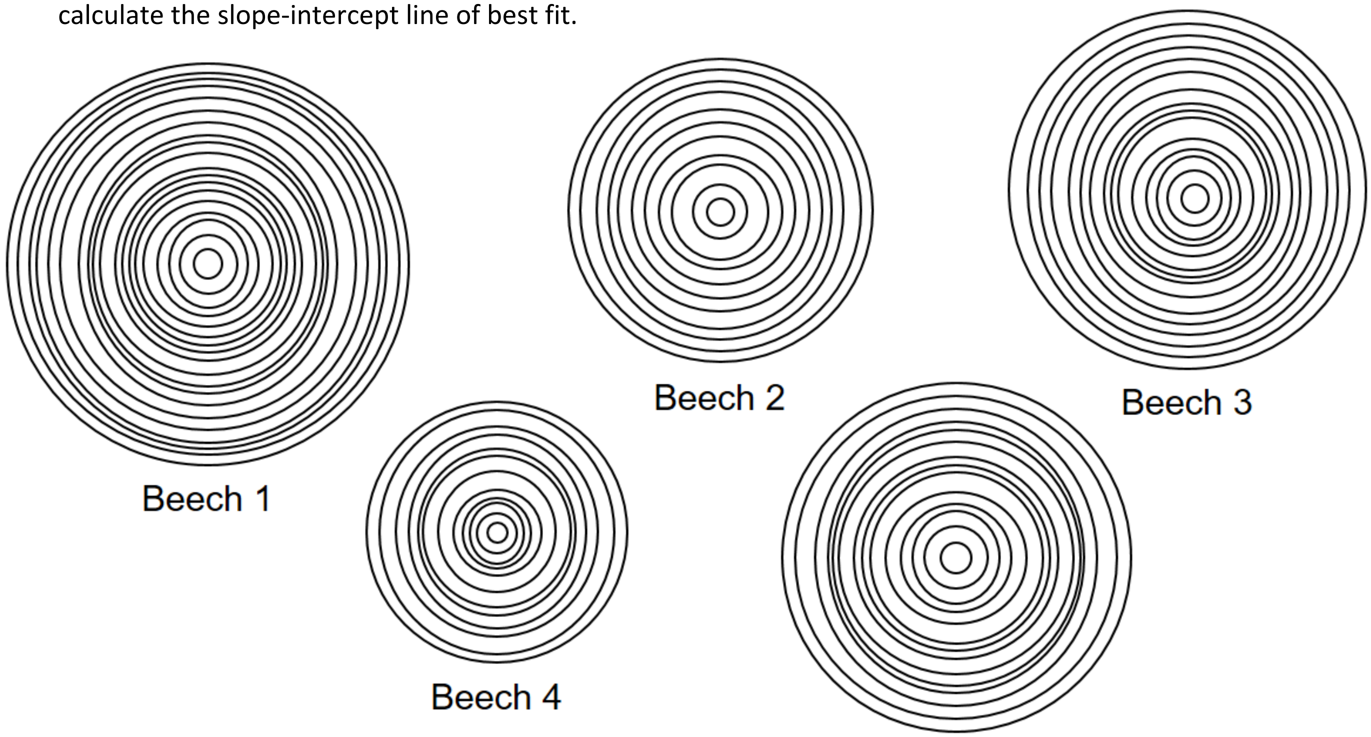
Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.

Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.

Developed by Greg Wiggins, Suzanne Lenhart, and Cameron Cook.

<u>Regression Exercise:</u>

As part of a forest health survey, samples (cross-sections) of American beech were taken from a cove forest in the Great Smoky Mountains National Park. Measure the diameter (in millimeters) and count the corresponding growth rings, record it in the table and, plot the data on the grid. Use the data to calculate the slope-intercept line of best fit.



Forest Samples

Forest Samples	**Beech 1**	**Beech 2**	**Beech 3**	**Beech 4**	**Beech 5**
Growth Rings (*x*)	19	11	15	12	14
Diameter (*y*)	53	40	47	34	46

Plot this data on the grid provided.

Calculate the mean of the x values and the mean of the y values.

$\bar{x}$:

$\bar{y}$:

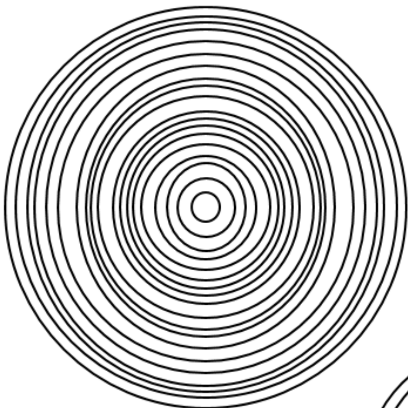
$(\bar{x}, \bar{y})$. This order pair would be on the line of best fit.

Draw your line of best fit for the plotted data. Then find the equation of your line.

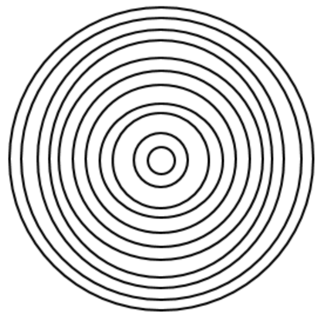
Find the line of best fit using your calculator.

<u>Regression Exercise:</u>

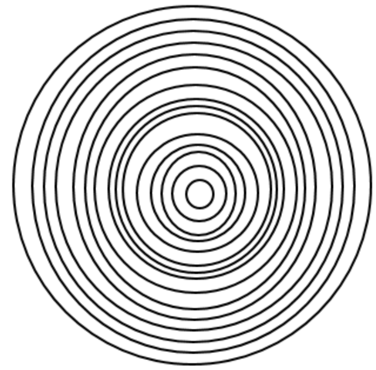
As part of a forest health survey, samples (cross-sections) of American beech were taken from a cove forest in the Great Smoky Mountains National Park. Measure the diameter (in millimeters) and count the corresponding growth rings, record it in the table and, plot the data on the grid. Use the data to calculate the slope-intercept line of best fit.



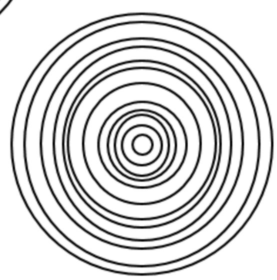
Beech 1



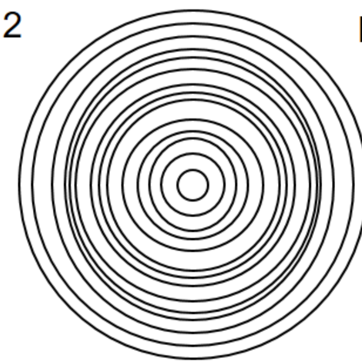
Beech 2



Beech 3



Beech 4



Beech 5

Forest Samples

Forest Samples	**Beech 1**	**Beech 2**	**Beech 3**	**Beech 4**	**Beech 5**
Growth Rings (*x*)					
Diameter (*y*)					

Plot this data on the grid provided.

Calculate the mean of the x values and the mean of the y values.

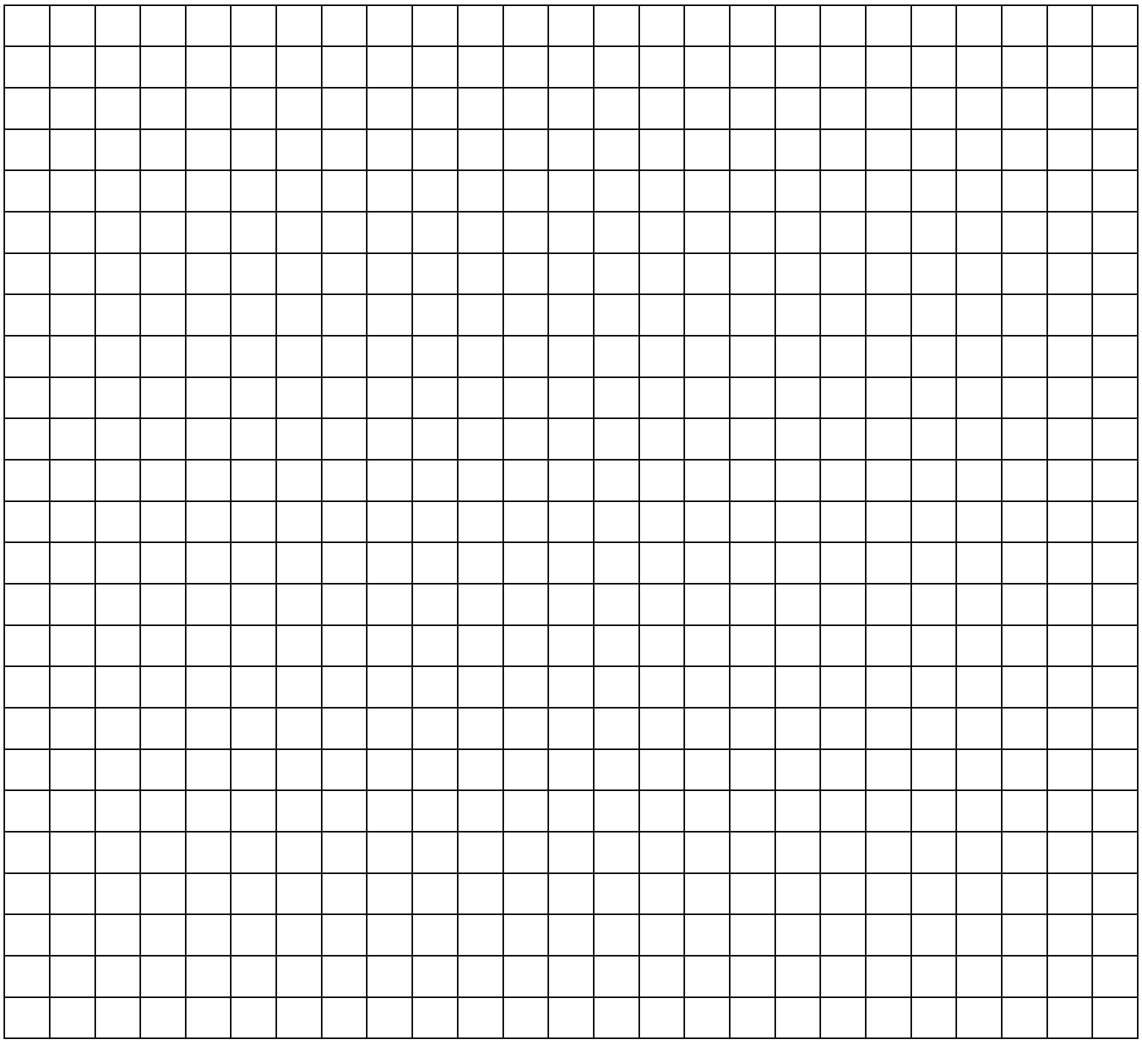
$\bar{x}$:

$\bar{y}$:

$(\bar{x}, \bar{y})$. This order pair would be on the line of best fit.

Draw your line of best fit for the plotted data. Then find the equation of your line.

Find the line of best fit using your calculator.



Regression Exercise Key:

Forest Samples	Beech 1	Beech 2	Beech 3	Beech 4	Beech 5
Growth Rings (*x*)	19	11	15	12	14
Diameter (*y*)	53	40	47	34	46

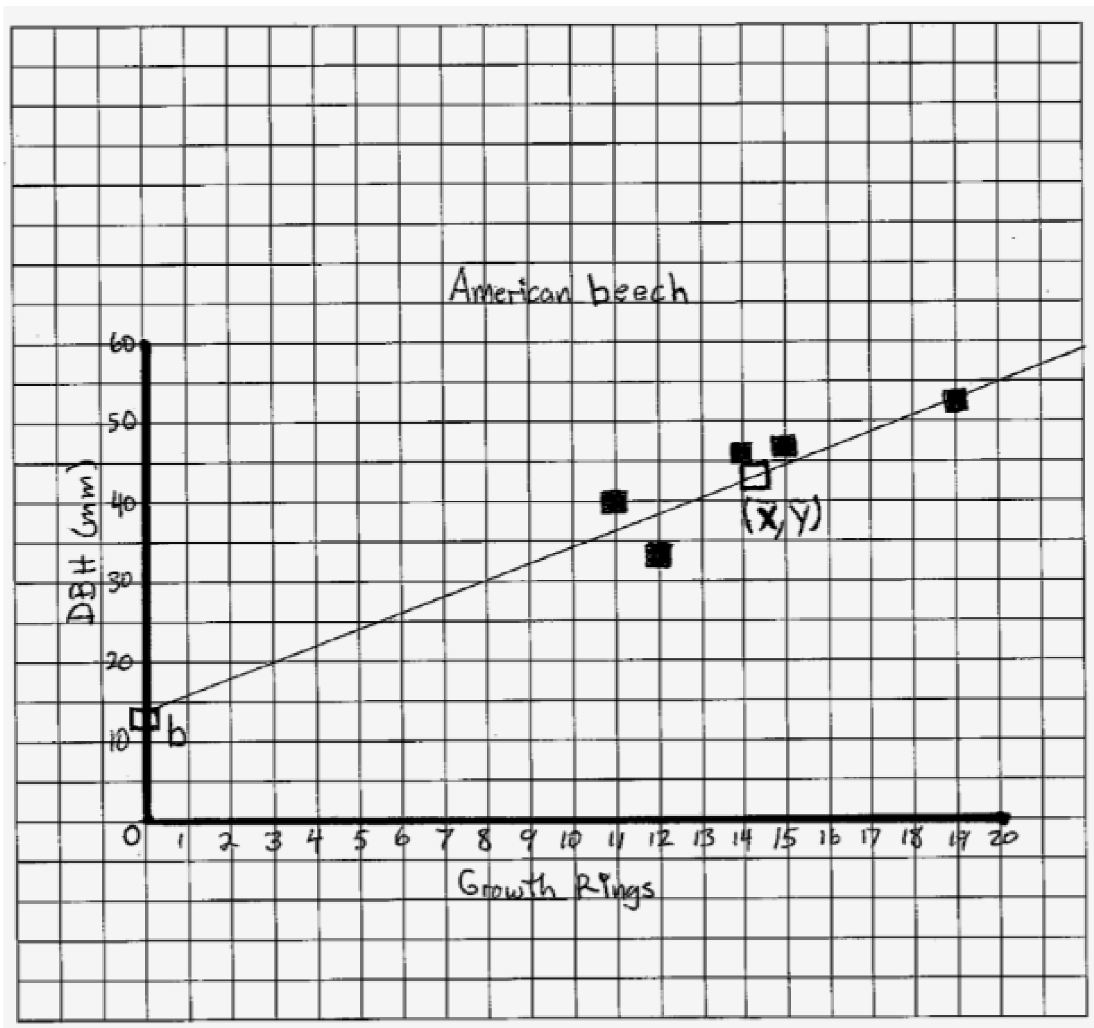
Calculations:

$\bar{x}$: (19+11+15+12+14)/5 = **14.2**

$\bar{y}$: (53+40+47+34+46)/5 = **44**

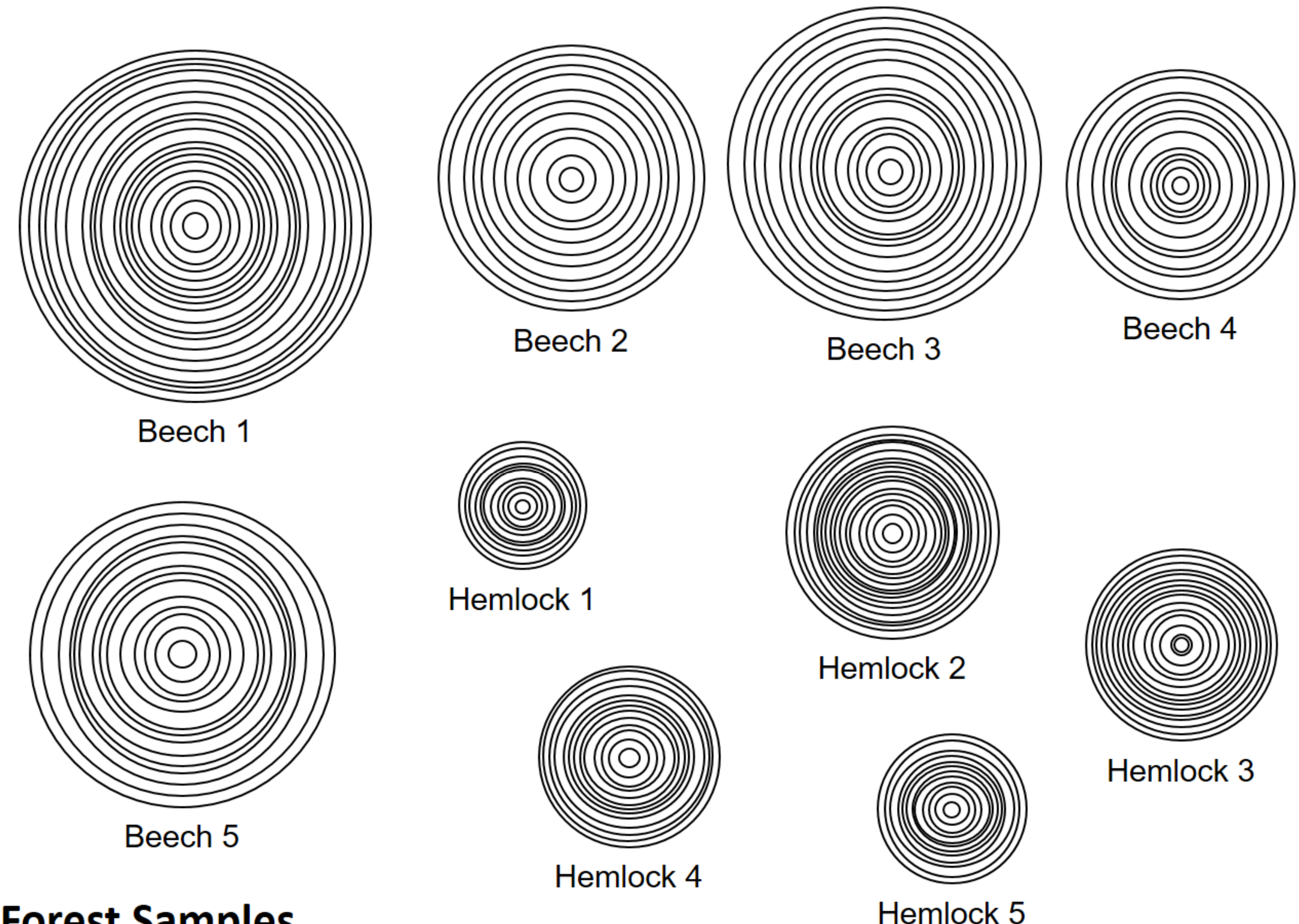
Line of best fit: *y* = **2.05*x* + 14.89**

b: 44-2.05(14.2) = **14.89**



Comparative Regression Exercise:

Samples (cross-sections) of two tree species, American beech and eastern hemlock, were taken from a cove forest in the Great Smoky Mountains National Park to compare growth rate and site suitability (suitability indicated by higher correlation coefficient). Measure the diameter (in millimeters) and count the corresponding growth rings of each sample and record it in the table. Using different color ink or different symbols for each tree species, plot the data on the grid. Use the data to calculate the slope-intercept line of best fit.



Forest Samples

Forest Samples	Beech 1	Beech 2	Beech 3	Beech 4	Beech 5	Hemlock 1	Hemlock 2	Hemlock 3	Hemlock 4	Hemlock 5
Growth Rings (x)										
Diameter (y)										

American Beech Calculations:

$\bar{x}$:

$\bar{y}$:

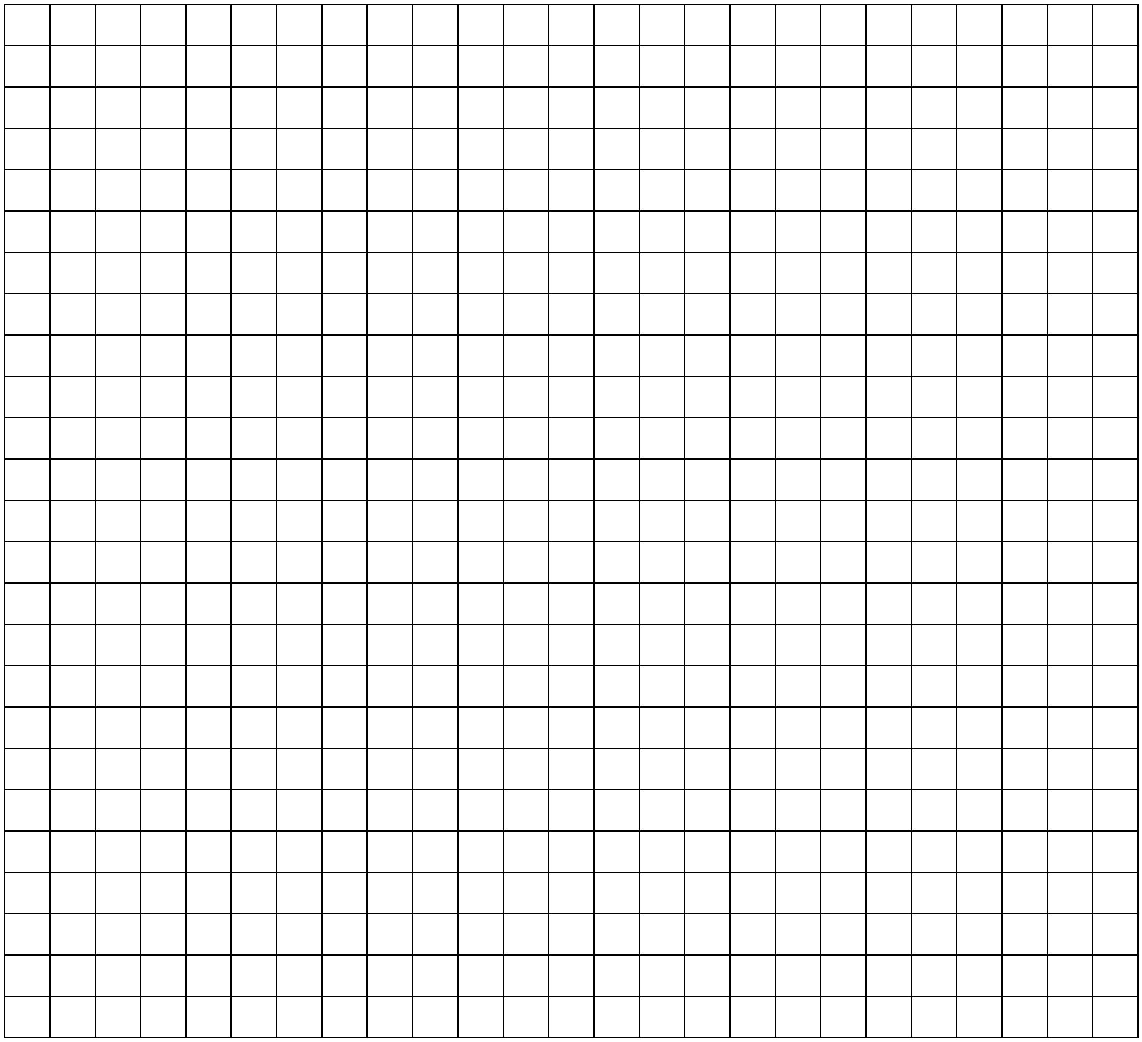
Find line of best fit.

Eastern Hemlock Calculations:

$\bar{x}$:

$\bar{y}$:

***Find line of best fit*:**



Comparative Regression Exercise Key

Forest Samples	Beech 1	Beech 2	Beech 3	Beech 4	Beech 5	Hemlock 1	Hemlock 2	Hemlock 3	Hemlock 4	Hemlock 5
Growth Rings (x)	19	11	15	12	14	11	17	15	13	12
Diameter (y)	53	40	46	34	47	19	32	29	26	22

American Beech Calculations:
$\bar{x}$: (19+11+15+12+14)/5 = **14.2**

$\bar{y}$: (53+40+47+34+46)/5 = **44**

Line of best fit: *y* = **2.05*x* + 14.89**

m: 79.5/38.8 = **2.05**

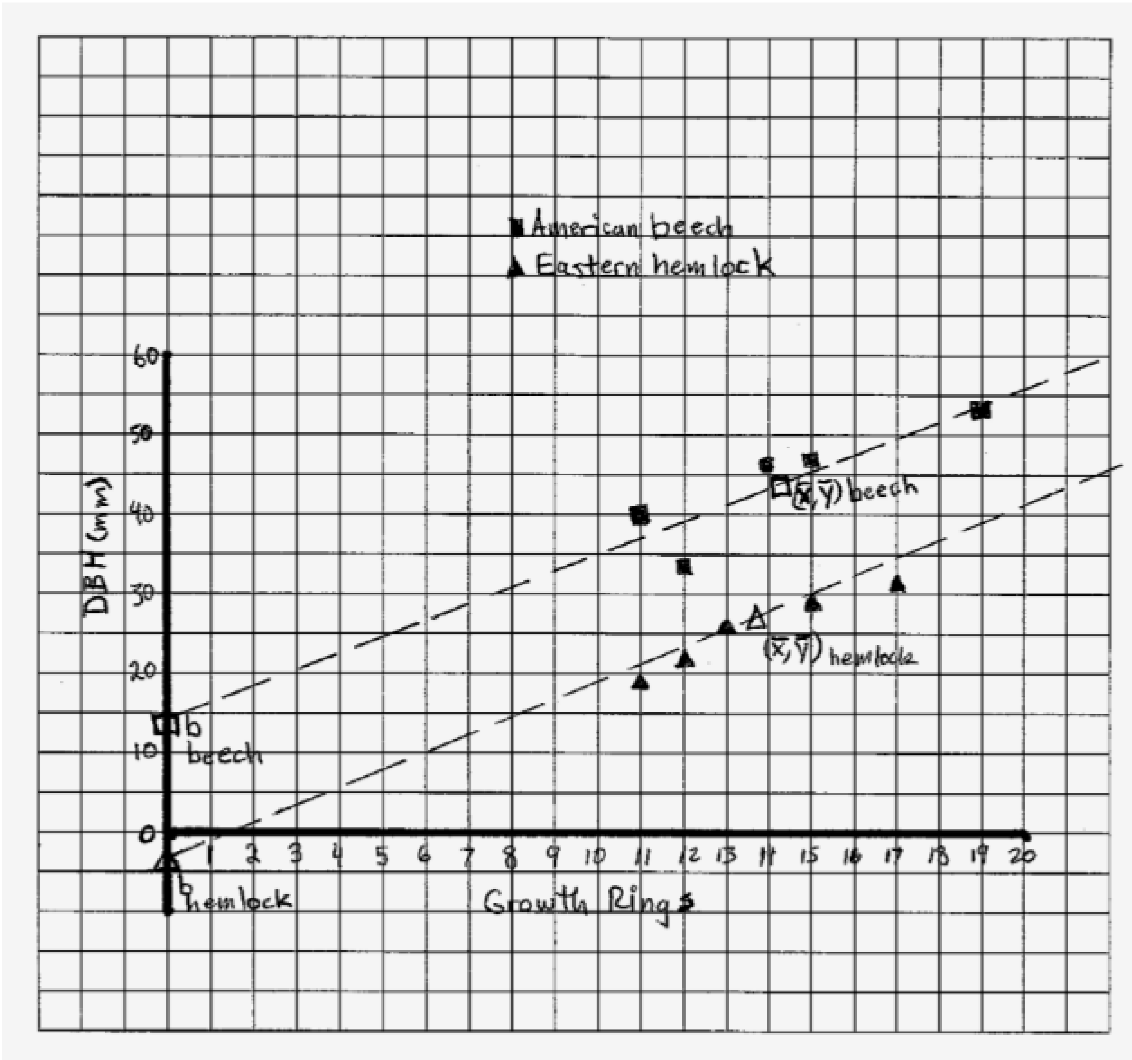
b: 44-2.05(14.2) = **14.89**

Eastern Hemlock Calculations:
$\bar{x}$: (11+17+15+13+12)/5 = **13.6**

$\bar{y}$: (19+32+29+26+22)/5 = **25.6**

Line of best fit: *y* = **2.12*x* – 3.23**

b: 25.6-2.12(13.6) = **-3.23**



American beech
Eastern hemlock
DBH (mm)
Growth Rings
$(\bar{x}, \bar{y})$ beech
$(\bar{x}, \bar{y})$ hemlock
b beech
b hemlock