Scatter Plots, Best-Fit Lines & Linear Regressions

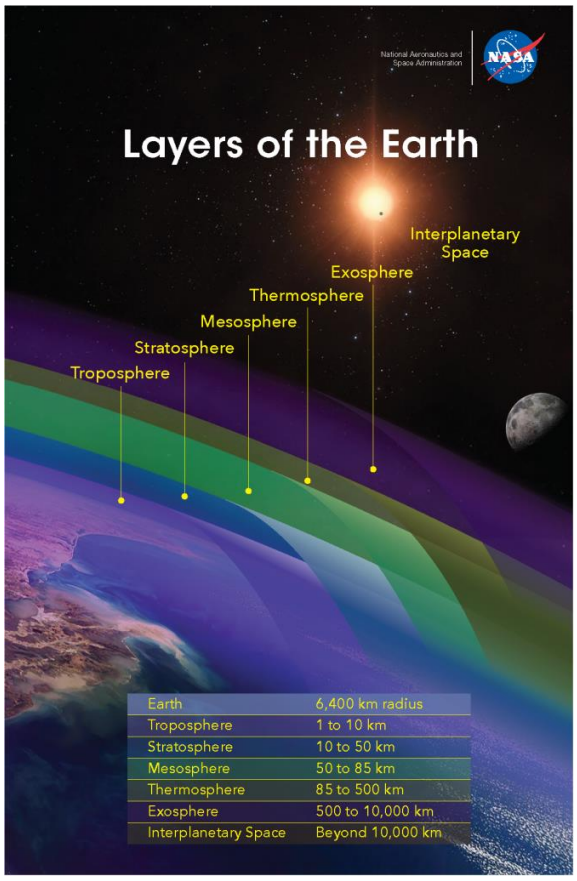
You will create scatter plots, interpret data, and look for correlation with this fun lab! Start by graphing the points on the provided grid, find the equation that best fits the data and then answer the questions about the data. If you'd like to do the linear regression, there are steps included for this as well near the end of this lab.

Math Lab: Temperature & Altitude

Most people think that the higher up in the Earth's atmosphere you go, the colder it gets. Actually, the atmosphere is a really complicated system that we monitor constantly using scientific instruments in order to get accurate information about what's happening now and what we expect to happen in the near future.

The ISA (International Standard Atmosphere) is a model that uses altitude (distance above the Earth's surface) to figure out temperature. Since the Earth changes elevation (we have mountains, valleys, and everything in-between), the gravitational pull is difference depending on where you are.

Temperature increases, decreases and also stays constant as you go up in the atmosphere. At first, that might seem weird – how can temperature increase as you increase your altitude? How can it do all three options?



The lowest part of the Earth's atmosphere is called the **troposphere**, and this layer is heated by the Earth. The further up you go, the farther you are from this heat source, so it makes sense that the temperature decreases by about 6.5°C per km.

The troposphere layer is from the surface to about 12 km high. Most aviation happens here in this layer. It's a dense layer that is compressed by the weight of the rest of the atmospheric layers above it, so most of our weather and almost all the clouds are here.

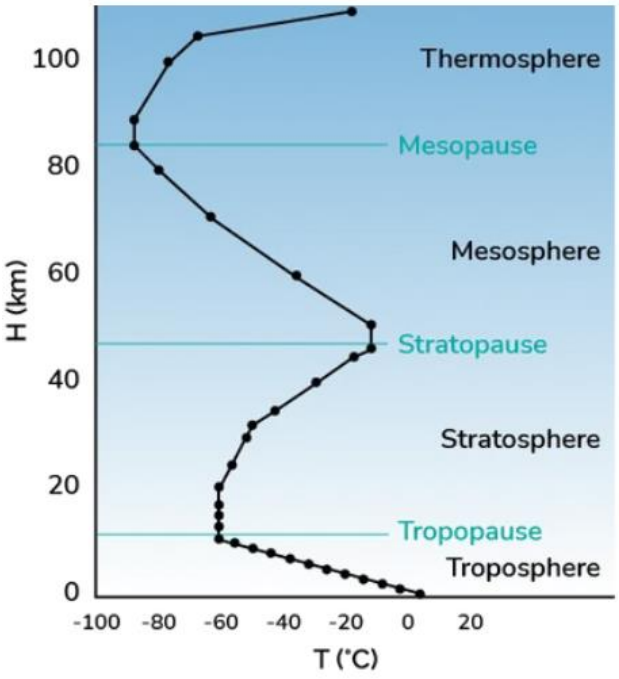
The next layer, the **stratosphere,** from 12km to about 50 km up from the surface. Here the temperature starts increasing by 1-2.8°C per km, because of the UV absorption by the ozone layer. When an ozone molecule (O_3) absorbs harmful UV light, it comes apart into an oxygen molecule (O_2) and a separate oxygen atom (O). Later these two can reform to make ozone. The temperature increases in this zone because the ozone layer acts as a shield and absorbs most of the UV radiation

from the sun. This layer is nearly cloud and weather-free, and only jet planes can reach this layer.

The **stratopause** is a really thin layer right around 48-51 km where the temperature remains constant at -1°C.

The **mesosphere** is the top layer, about 51-86km high. This is a challenging area to study, because the atmosphere is so thin. Temperature in this layer decreases by 2-2.8°C per km. This layer has the highest clouds. Sounding rockets and rocket-powered aircraft can reach this layer.

In the **mesopause** is the thin layer that sits about 86km high. The temperature here is the coldest on Earth, at a constant -87°C.

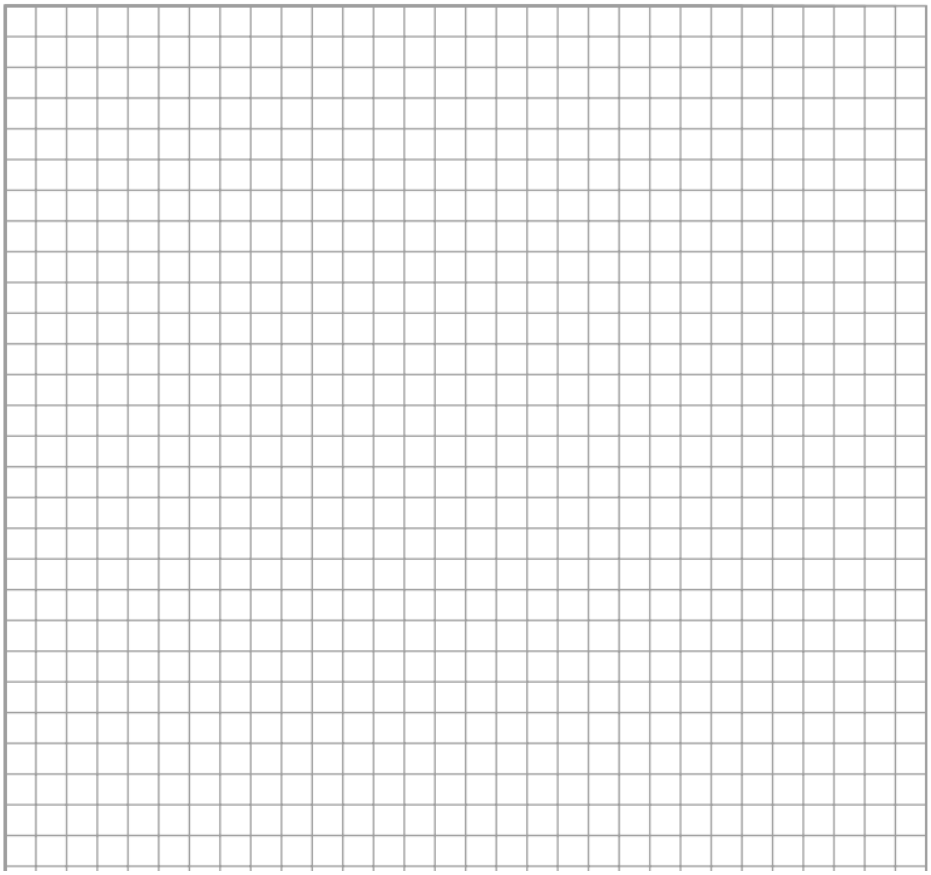


The Weather Balloon Experiment:

We decided to send up a weather balloon with an instrument on board that can measure temperature and altitude. In just two hours, our weather balloon went higher than jets can travel, and it took two measurements every second. I've taken a segment of the and condensed it from thousands of data points to the following:

Altitude (m)	Temperature °C
1	17.99
1000	12.9
2000	7.9
3000	0.2
4000	-7.2
5000	-14.5
6000	-22
7000	-26
8000	-31
9000	-38.5
10,000	-45.6

1. Plot the data on the graph below.
2. Label both axes (temperature on the vertical, altitude on the horizontal).



3. Use your ruler and draw a “best fit” line through the points. You want the same number of data points both above and below your line.

4. Use this formula: $y = mx + b$ to determine the equation of your best fit line:

5. Which atmospheric layer does this equation describe?

6. What do you expect the temperature to be at 4,500 m? 12,000m?

Linear Regression

Sometimes it's hard to figure out what the best fit line should be – there may be more than one line that works, so what can we do?

We can use the *least squares regression* to find the best fit, sometimes called the *regression line.* This method is more accurate than finding it by eye (as we just did).

You'll need a calculator to figure out the "least squares regression". It works like this:

1. For each coordinate (x,y), calculate: xy and x^2

2. Sum (Σ) each: x-values (Σ x), the y-values (Σ y), the xy values (Σ xy) and the x^2 (Σx^2).

3. Use this formula for finding the slope (n = the number of data points):

$$m = \frac{n\Sigma(xy) - \Sigma x\, \Sigma y}{n\Sigma(x^2) - (\Sigma x)^2}$$

4. Use this formula for finding the y-intercept (n = the number of data points):

$$b = \frac{\Sigma y - m\Sigma x}{n}$$

5. Write down the equation: y = mx + b for your best fit line.

6. See example calculation on page 5.

Why does this work? Using this method minimizes the distance between each point and a point on the line with the x-value. We are making the error distances as small as possible for the data set.

Your turn! Go back to page 2. Do the least squares regression for the data provided on page 2. What is your best fit line and how does it compare with the first method?

Example: Write down the coordinates and pull out the x & y values:

(x, y)	x	y	xy	x^2
(1, 2)	1	2		
(2, 5)	2	5		
(3, 5)	3	5		
(4, 8)	4	8		
(6, 13)	6	13		
(8, 15)	8	15		

For each coordinate (x,y), calculate: xy and x^2

(x, y)	x	y	xy	x^2
(1, 2)	1	2	2	1
(2, 5)	2	5	10	4
(3, 5)	3	5	15	9
(4, 8)	4	8	32	16
(6, 13)	6	13	78	36
(8, 15)	8	15	120	64

Now find the sum of each column of values:

(x, y)	x	y	xy	x^2
(1, 2)	1	2	2	1
(2, 5)	2	5	10	4
(3, 5)	3	5	15	9
(4, 8)	4	8	32	16
(6, 13)	6	13	78	36
(8, 15)	8	15	120	64
	Σ x = 24	Σ y = 48	Σ xy = 257	Σ x^2 = 130

Now find the slope (n = 6):

$$m = \frac{n\Sigma(xy) - \Sigma x\,\Sigma y}{n\Sigma(x^2) - (\Sigma x)^2}$$

$m = \frac{(6)(257)-(24)(48)}{(6)(130)-(24)^2}$ = 1.91

$$b = \frac{\Sigma y - m\Sigma x}{n}$$

$b = \frac{(48)-(1.91)(24)}{(6)}$ = 0.36

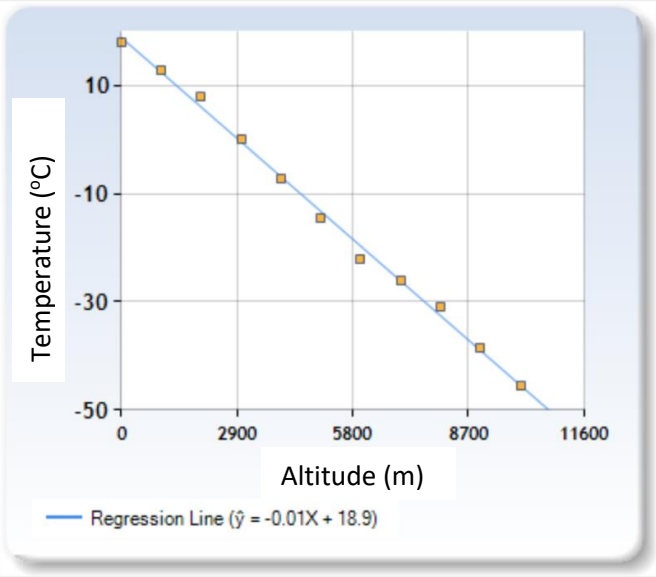
Answer: y = 1.91x + 0.36

Answer Key:

We are in the troposphere (based on altitude).

For 4,500 feet: approximately -10 ^{o}C

For 12,000 feet: approximately -58 ^{o}C (be careful, this is outside of the range of data provided so this is not as accurate as the 4,500 answer).



y = -0.00643 x + 18.90345

Name: _____

LINEAR REGRESSION RIDDLE A

MATCHING Match the data to its corresponding line of best fit and correlation coefficient. Write the letter of the answer in the corresponding numbered blank below to answer the riddle. Round to the nearest hundredth.

x	-20	-11	-7	5
y	66	39	27	-9

1. $y =$ _____
2. $r =$ _____

$(-1, 2), (0.5, 4), (2, 5), (3, 20)$

3. $y =$ _____
4. $r =$ _____

x	4	5	6	7
y	1.88	2.35	2.82	3.29

5. $y =$ _____
6. $r =$ _____

$(0, 6), (2, 3), (3, 4), (5, 3)$

7. $y =$ _____
8. $r =$ _____

y is 6 when x is 2, y is 9 when x is 5, and y is 17 when x is 8

9. $y =$ _____
10. $r =$ _____

x	0	2	4	6
y	5	4	1	2

11. $y =$ _____
12. $r =$ _____

y is 6 when x is 0, y is 1 when x is 2, and y is 5 when x is 4

13. $y =$ _____
14. $r =$ _____

$(1, 1.5), (2, 0.8), (4, 3.5), (5, 2)$

15. $y =$ _____
16. $r =$ _____

G $y = -0.6x + 4.8$

E $r = 1$

I $r = -0.79$

A $y = 3.82x + 3.45$

S $r = 0.59$

N $y = -3x + 6$

U $y = -0.25x + 4.5$

L $r = -0.19$

A $r = 0.81$

R $y = 0.37x + 0.84$

V $y = -0.54x + 5.35$

D $r = -1$

N $r = -0.85$

F $y = 0.47x$

R $r = 0.97$

I $y = 1.83x + 1.5$

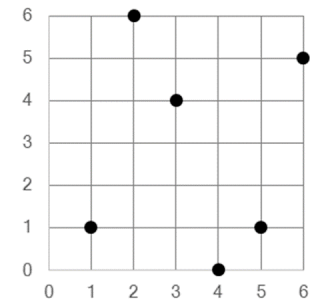
What is a pirate's favorite part of statistics?

___	___	___	___	___	___	___		___	___	___	-	___	___	___	___	___	___
5	9	12	2	8	1	11		3	10	15		7	4	14	13	6	16

Name: _____

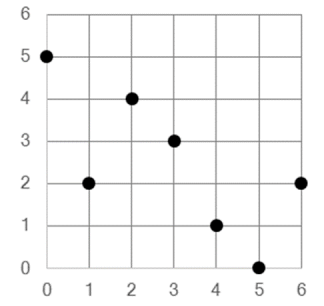
LINEAR REGRESSION RIDDLE B

MATCHING Match the data to its corresponding line of best fit and correlation coefficient. Write the letter of the answer in the corresponding numbered blank below to answer the riddle. Round to the nearest hundredth.



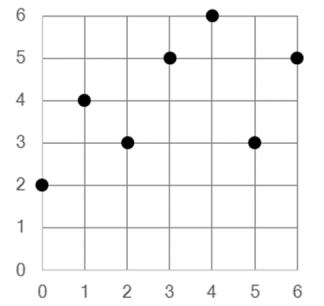
1. $y =$ _____

2. $r =$ _____



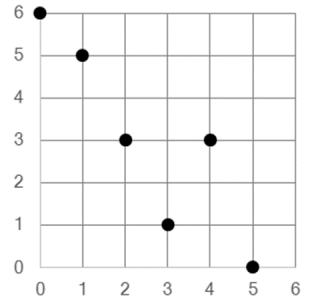
3. $y =$ _____

4. $r =$ _____



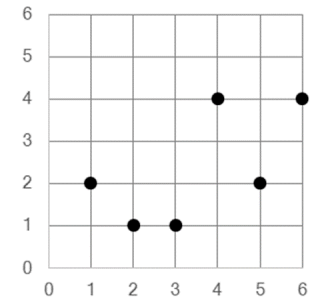
5. $y =$ _____

6. $r =$ _____



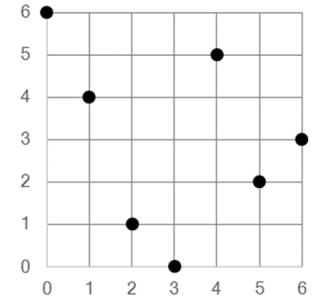
7. $y =$ _____

8. $r =$ _____



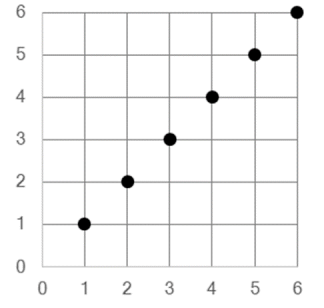
9. $y =$ _____

10. $r =$ _____



11. $y =$ _____

12. $r =$ _____



13. $y =$ _____

14. $r =$ _____

R	$r = -0.32$	L	$y = 0.46x + 0.73$	E	$r = -0.72$
F	$r = 0.63$	T	$y = -0.32x + 3.96$	E	$y = 0.36x + 2.93$
L	$y = -0.57x + 4.14$	T	$r = -0.89$	E	$y = -1.09x + 5.71$
N	$r = 1$	A	$r = 0.02$	O	$r = 0.55$
R	$y = 0.03x + 2.73$	U	$y = x$		

What did the regression equation want to be when it grew up?

__	__	__	__	__	__	__	__	__	__	__	__	__	__
2	10	6	12	8	13	14	5	11	4	9	3	7	1

Name: _____

LINEAR REGRESSION RIDDLE A

Solutions

MATCHING Match the data to its corresponding line of best fit and correlation coefficient. Write the letter of the answer in the corresponding numbered blank below to answer the riddle. Round to the nearest hundredth.

x	-20	-11	-7	5
y	66	39	27	-9

1. $y =$ _____
2. $r =$ _____

$(-1, 2), (0.5, 4), (2, 5), (3, 20)$

3. $y =$ _____
4. $r =$ _____

x	4	5	6	7
y	1.88	2.35	2.82	3.29

5. $y =$ _____
6. $r =$ _____

$(0, 6), (2, 3), (3, 4), (5, 3)$

7. $y =$ _____
8. $r =$ _____

y is 6 when x is 2, y is 9 when x is 5, and y is 17 when x is 8

9. $y =$ _____
10. $r =$ _____

x	0	2	4	6
y	5	4	1	2

11. $y =$ _____
12. $r =$ _____

y is 6 when x is 0, y is 1 when x is 2, and y is 5 when x is 4

13. $y =$ _____
14. $r =$ _____

$(1, 1.5), (2, 0.8), (4, 3.5), (5, 2)$

15. $y =$ _____
16. $r =$ _____

G $y = -0.6x + 4.8$

E $r = 1$

I $r = -0.79$

A $y = 3.82x + 3.45$

S $r = 0.59$

N $y = -3x + 6$

U $y = -0.25x + 4.5$

L $r = -0.19$

A $r = 0.81$

R $y = 0.37x + 0.84$

V $y = -0.54x + 5.35$

D $r = -1$

N $r = -0.85$

F $y = 0.47x$

R $r = 0.97$

I $y = 1.83x + 1.5$

What is a pirate's favorite part of statistics?

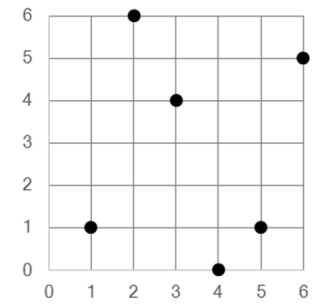
F	I	N	D	I	N	G		A	R	R	-	V	A	L	U	E	S
5	9	12	2	8	1	11		3	10	15		7	4	14	13	6	16

Name: _____

LINEAR REGRESSION RIDDLE B

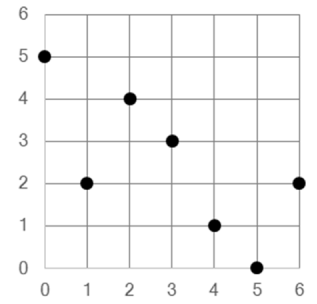
Solutions

MATCHING Match the data to its corresponding line of best fit and correlation coefficient. Write the letter of the answer in the corresponding numbered blank below to answer the riddle. Round to the nearest hundredth.



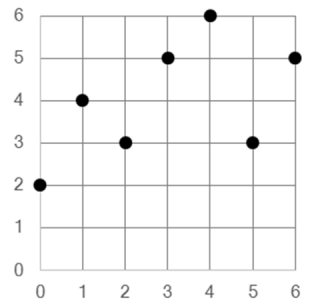
1. $y =$ _____

2. $r =$ _____



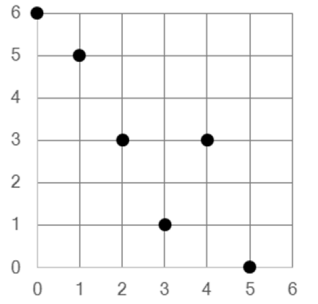
3. $y =$ _____

4. $r =$ _____



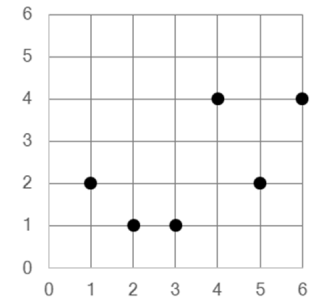
5. $y =$ _____

6. $r =$ _____



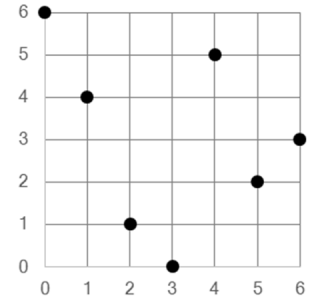
7. $y =$ _____

8. $r =$ _____



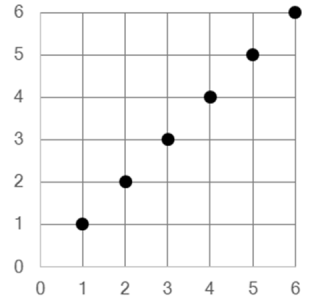
9. $y =$ _____

10. $r =$ _____



11. $y =$ _____

12. $r =$ _____



13. $y =$ _____

14. $r =$ _____

R	$r = -0.32$	L	$y = 0.46x + 0.73$	E	$r = -0.72$
F	$r = 0.63$	T	$y = -0.32x + 3.96$	E	$y = 0.36x + 2.93$
L	$y = -0.57x + 4.14$	T	$r = -0.89$	E	$y = -1.09x + 5.71$
N	$r = 1$	A	$r = 0.02$	O	$r = 0.55$
R	$y = 0.03x + 2.73$	U	$y = x$		

What did the regression equation want to be when it grew up?

A	F	O	R	T	U	N	E		T	E	L	L	E	R
2	10	6	12	8	13	14	5		11	4	9	3	7	1