

Statistics: Regressions and Models

Notes, examples, and Practice exercises (with answers)

Topics include linear, quadratic, exponential, tables, graphs, residuals, and more.

Example: Given the exponential model $h(t) = 123(.7)^t$

If $t = 6$ yields a residual of 2.6,

what is the actual value?

$$\text{residual} = \text{"actual"} - \text{"predicted"}$$

$$2.6 = \text{actual} - h(6)$$

$$\text{Actual} = 2.6 + 14.47$$

$$17.07$$

Example: The following chart shows the year and number of Oscar pet stores nationwide.

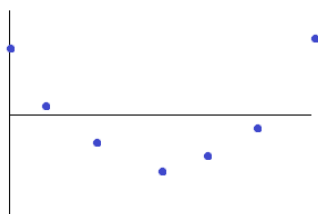
Year	2000	2002	2005	2008	2010	2013	2017
# of pet stores	12	58	351	721	1250	2134	3700

Find the linear, quadratic, and exponential models. Plot their residuals.

Linear

$$y = 210.86x - 481.612$$

residuals: 493.61

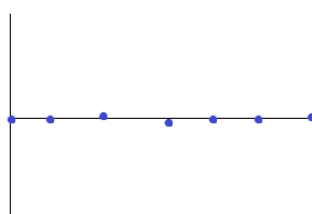


correlation coefficient $r = .9498$

Quadratic

$$y = 13.52x^2 - 14x + 27.8$$

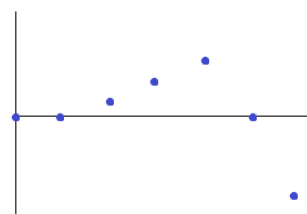
residuals: -15.796



Exponential

$$y = 31.86(1.381)^x$$

residuals: -19.86



Which is the best model? Why?

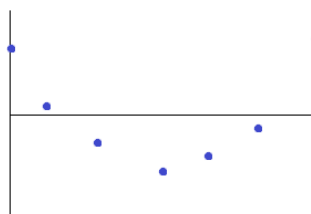
What are the flaws of the other models?

The quadratic model has the lowest residuals, and the values are near the 0.

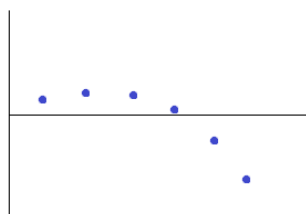
The linear has a pattern where the model underestimates in the beginning and end and overestimates in the middle..

The exponential has a low residual, but the model's predictions will overestimate!

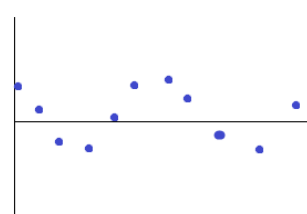
Residual patterns....



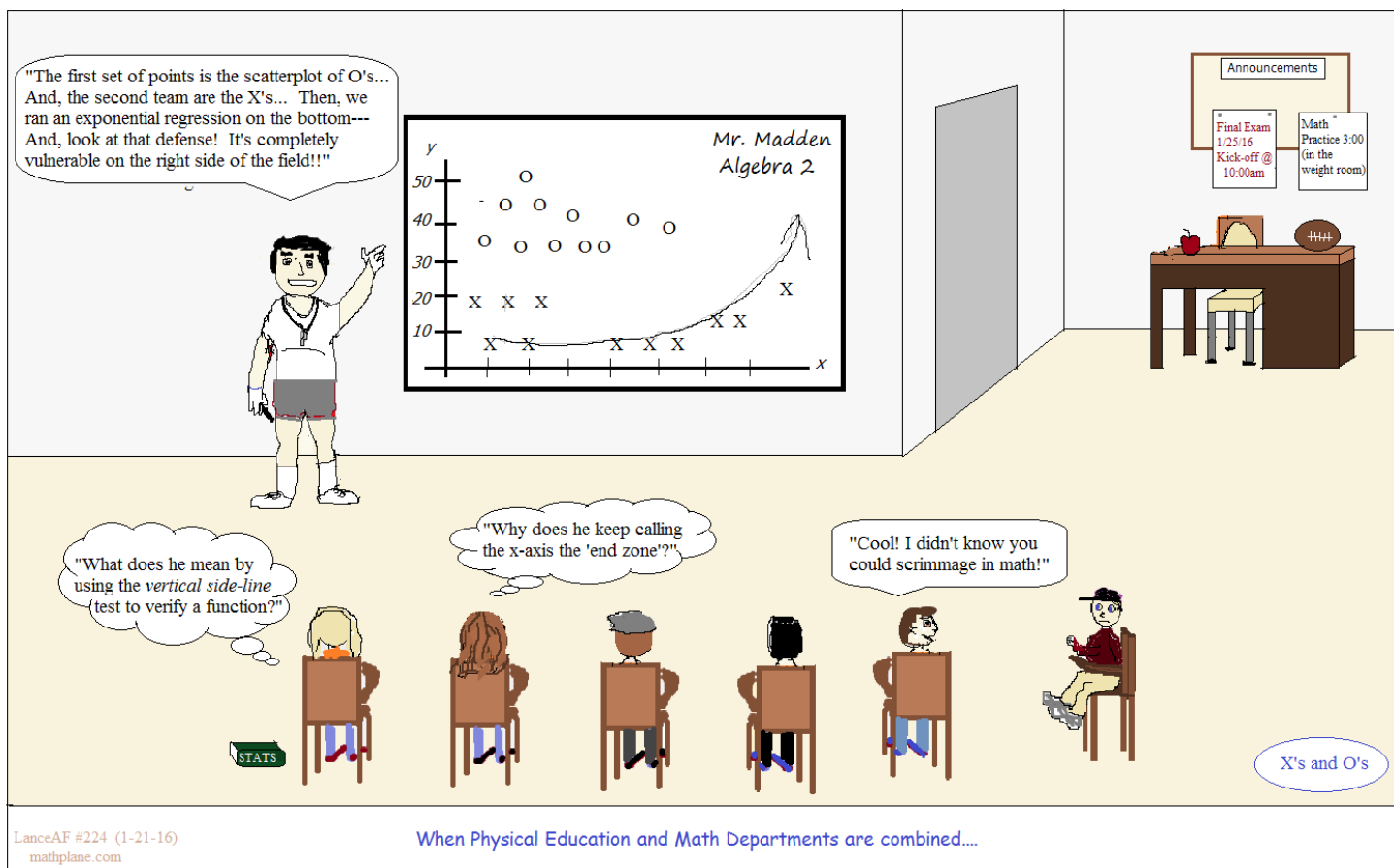
Pattern underestimates the beginning and end. But, overestimates the middle... flawed..



this is likely an exponential model that should be a linear model... The model is increasing more rapidly than the actual values...



It's good... although there is a pattern, the residuals are oscillating over and back over the 0.



Exercises ->

Regression Models

1) The chart on the right shows a set of dog breeds with each average weight and life expectancy.

a) Determine a linear model to find the life expectancy of a dog as it relates to its weight.

b) What is the correlation coefficient?

c) What is the *residual* of the husky?

Dog Breed	Weight (lbs)	Life Expectancy (years)
1) Beagle	24	12
2) Bernese Mountain	100	8
3) Husky	55	13
4) Labrador Retriever	65	12
5) Llasa Apso	15	14
6) Poodle	40	14
7) Shepherd	80	10

d) Using your model, estimate the life expectancy of

i) Great Dane - 140 lbs

ii) Jack Russell Terrier - 15 lbs

iii) Greyhound - 65 pounds

2) Determine which model is best to use, and justify your answer.

(in the open space, sketch graphs showing each model)

Linear, quadratic, or exponential?

1) Car value as a function of time in years.

2) The height of a shot put as it relates to the distance from the olympian.

3) High school student's ACT expected score based on hours of preparation.

4) Expected temperature detemined from the day of the year (1 - 365).

5) Heat use as a function of average temperature outside.

6) Points scored by a basketball player, using minutes played.

- 3) A swimmer steps up to a diving board. When he dives off, his height above the water is
 24 feet @ 1 second,
 18 feet @ 2 seconds, and
 2 feet @ 3 seconds.

Choose the best fit model. Then, answer the following:

- How high is the diving board?
- What is the peak height of the jump? When does it occur?
- When does he hit the water?

Optional: Compare the results using linear vs. quadratic vs. exponential models...

- 4) A series of photo images show a long jumper's position during a jump.
 The following are the distances and height depicted in each image:

0 feet long, 0 feet high (his initial jumping point)
 3' long, 1' 9" high
 15' long, 3' 9" high
 18' long, 3' high

Use the best regression model to find:

- How *far* did he jump?
- How *high* did he peak during his jump?
- At what position(s) was he 2' high?

- 5) An observer claims the population at a baseball game changes exponentially as the weather changes. The table shows statistics from 8 randomly chosen games.

Game Date	Temperature (degrees F)	Attendance
April 21	62	23,567
May 3	68	24,607
June 8	75	29,312
June 27	78	28,778
July 12	67	22,546
July 22	75	27,678
August 3	88	30,133
August 18	91	28,998

- What is the observer's exponential model that would estimate attendance?
- What are potential flaws to this model? (What are its limitations?
 What other factors may influence the attendance?)
- On May 3, the attendance was 24,607.
 What did the exponential model predict attendance would be on that 68 degree day?
- Is a quadratic model better?
- What is the correlation coefficient? Does this suggest that a linear model would be effective?

Identifying Models

Match the description or table with the correct graph...

- 1) Height above ground of a chair on a moving ferris wheel.

2)

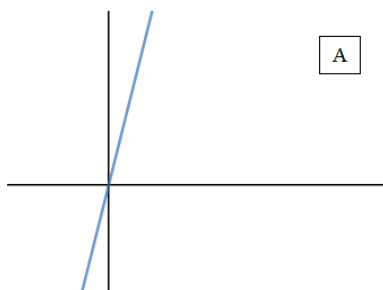
x	-3	-2	-1	0	1
y	-12	-8	-4	0	4

- 3) Your math grade based on the number of points you got on the test...

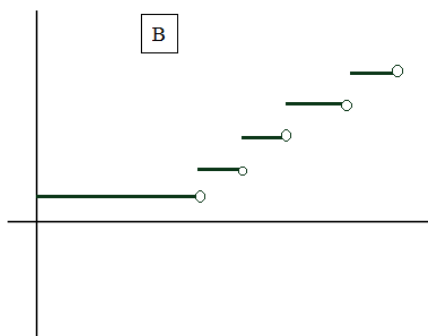
- 4) Population growth of a new town...

- 5) Distance from Chicago when driving from Los Angeles to New York City...

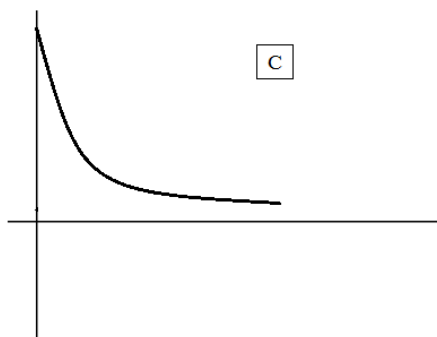
- 6) Value of a car during its life time



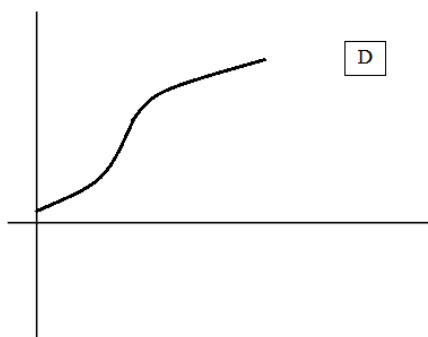
A



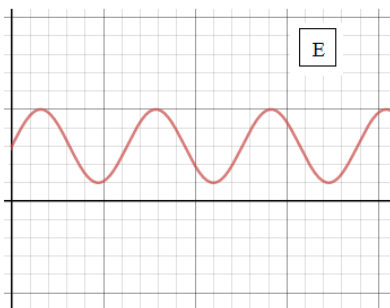
B



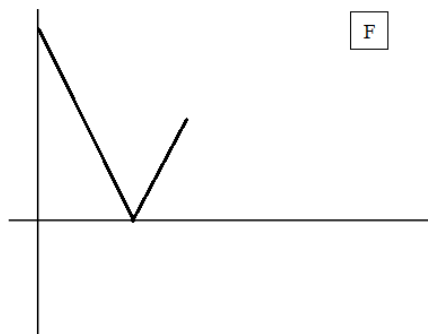
C



D



E



F

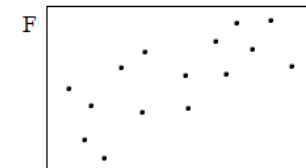
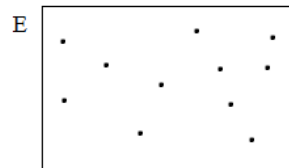
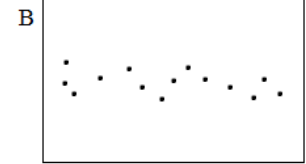
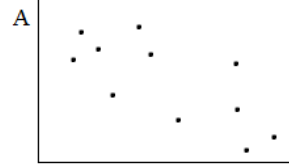
Correlation Coefficient

$$-1 < r < 1$$

0 no correlation

1 direct correlation

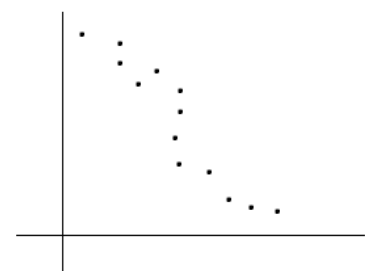
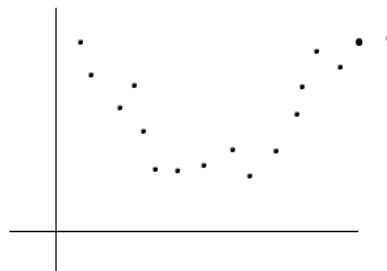
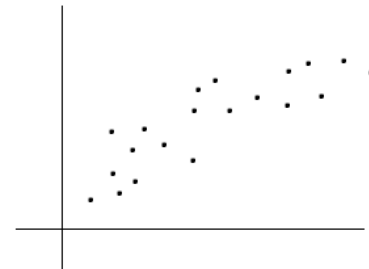
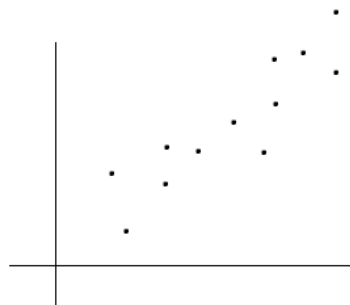
-1 inverse correlation



Place the scatterplots in order from lowest to highest r-value

Identify the best fit model

- a) Linear
- b) Exponential
- c) Logarithmic
- d) Parabolic/Quadratic



Regression Models

1) The chart on the right shows a set of dog breeds with each average weight and life expectancy.

a) Determine a linear model to find the life expectancy of a dog as it relates to its weight.

$$y = -.0616x + 15.194$$

b) What is the correlation coefficient?

$$r = -0.8532 \text{ negatively correlated..}$$

c) What is the *residual* of the husky?

$$\text{residual} = \text{actual value} - \text{model value}$$

$$1.2 = 13 - 11.8$$

$$y = -.0616(55) + 15.194$$

model prediction of
husky: 11.8

d) Using your model, estimate the life expectancy of

i) Great Dane - 140 lbs

6.57 yrs vs. 8 yrs

ii) Jack Russell Terrier - 15 lbs

14.27 vs. 15 yrs

iii) Greyhound - 65 pounds

11.19 vs. 12 yrs

2) Determine which model is best to use, and justify your answer.

(in the open space, sketch graphs showing each model)

Linear, quadratic, or exponential?

1) Car value as a function of time in years. exponential

2) The height of a shot put as it relates to the distance from the olympian.

quadratic

3) High school student's ACT expected score based on hours of preparation.

exponential

4) Expected temperature determined from the day of the year (1 - 365).

quadratic

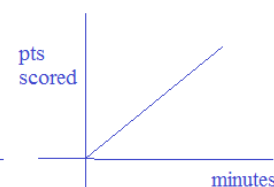
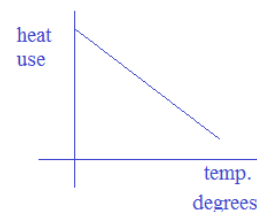
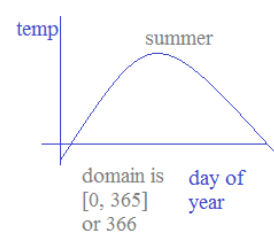
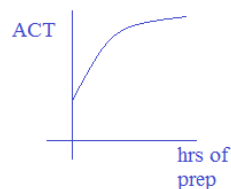
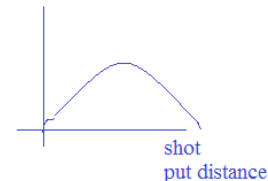
5) Heat use as a function of average temperature outside.

linear (negative correlation)

6) Points scored by a basketball player, using minutes played.

linear

Dog Breed	Weight (lbs)	Life Expectancy (years)
1) Beagle	24	12
2) Bernese Mountain	100	8
3) Husky	55	13
4) Labrador Retriever	65	12
5) Lhasa Apso	15	14
6) Poodle	40	14
7) Shepherd	80	10



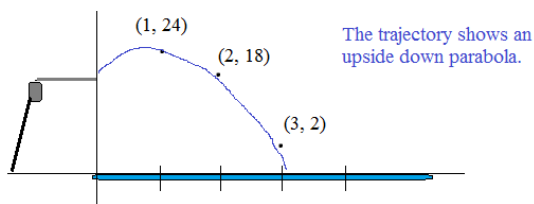
- 3) A swimmer steps up to a diving board. When he dives off, his height above the water is
 24 feet @ 1 second,
 18 feet @ 2 seconds, and
 2 feet @ 3 seconds.

Finding the equation of a quadratic from points

Choose the best fit model. Then, answer the following:

- How high is the diving board?
- What is the peak height of the jump? When does it occur?
- When does he hit the water?

Step 1: Sketch a graph of the dive



Step 2: Determine the equation that models the diver.

Standard form of a quadratic: $y = ax^2 + bx + c$

$$\begin{aligned} (1, 24): 24 &= a(1)^2 + b(1) + c & a + b + c &= 24 \\ (2, 18): 18 &= a(2)^2 + b(2) + c & 4a + 2b + c &= 18 \\ (3, 2): 2 &= a(3)^2 + b(3) + c & 9a + 3b + c &= 2 \end{aligned}$$

Linear system:
3 equations and 3 unknowns

Optional: Compare the results using linear vs. quadratic vs. exponential models...

linear model:

$$h(t) = -11t + 36.67$$

a) $h(0) = 36.67$ feet
too high!!

b) maximum height is 36.67

c) hits water when $h(t) = 0$
 $t = 3.33$

exponential model:

$$h(t) = 114.3(.289)^t$$

$h(0) = 114.3$ feet
definitely too high!!

maximum height is 114.3

$h(t)$ never equals zero!

OR, you can enter the 3 values into a spreadsheet..
Then run a linear regression on the TI nspire CX CAS

Using a TI nspire CX CAS: menu
 3: Algebra
 7: Solve System of Equations
 2: Solve System of Linear Equations
 ('number of equations': enter 3)
 When the *linSolve* function and template appear,
 enter the equations.
 $x + y + z = 24$
 $4x + 2y + z = 18$
 $9x + 3y + z = 2$
 Enter
 Output: $\{-5, 9, 20\}$

The model of the diver is

$$h(t) = -5t^2 + 9t + 20$$

where t is the time of the dive
and $h(t)$ is the height of the diver over the water.

Step 3: Answer the questions

a) height of the diving board: This occurs when the diver starts -- at time $(t) = 0$. $h(0) = 20$

b) Peak of the dive: This would occur at the vertex -- $\frac{-b}{2a} = \frac{-9}{2(-5)} = .9$ seconds

$$h(.9) = -5(.9)^2 + 9(.9) + 20 = 24.05 \text{ feet}$$

c) Diver hits water: This occurs when the height is 0 -- $0 = -5t^2 + 9t + 20$

Since time cannot be negative,

there is only one answer: 3.09 seconds

Using the TI nspire Calculator:

menu
 3: algebra
 1: solve
 Enter the quadratic equation.
 $-5t^2 + 9t + 20 = 0$, t
 control enter
 Output: $t = -1.29$ and $t = 3.09$

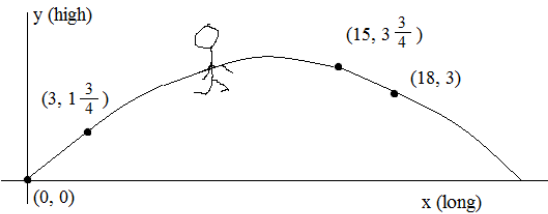
4) A series of photo images show a long jumper's position during a jump.
The following are the distances and height depicted in each image:

0 feet long, 0 feet high (his initial jumping point)
3' long, 1' 9" high
15' long, 3' 9" high
18' long, 3' high

Use the best regression model to find:

- a) How *far* did he jump?
- b) How *high* did he peak during his jump?
- c) At what position(s) was he 2' high?

1) Draw a picture and label (map the points)



when is the athlete 2' high?

$$2 = -\frac{1}{36}x^2 + \frac{2}{3}x$$

$$x^2 - 24x + 72 = 0$$

$$x = 3.51 \text{ or } 20.48$$

2) Determine the quadratic equation

Choose 3 points: (0, 0) (3, 1.75) (18, 3)

$$y = ax^2 + bx + c$$

construct 3 equations/3 unknowns

$$a(0)^2 + b(0) + c = 0$$

$$a(3)^2 + b(3) + c = 1.75$$

$$a(18)^2 + b(18) + c = 3$$

$$y = -\frac{1}{36}x^2 + \frac{2}{3}x$$

(note: you can test the points to check the equation)

solve:

$$\begin{bmatrix} 0 & 0 & 1 & | & 0 \\ 9 & 3 & 1 & | & 1.75 \\ 324 & 18 & 1 & | & 3 \end{bmatrix} \rightarrow \begin{bmatrix} -1/36 & a \\ 2/3 & b \\ 0 & c \end{bmatrix}$$

OR, input the table into a spreadsheet and find the quadratic regression

x	long	high y
0		0
3		1.75
15		3.75
18		3

3) Answer the questions

How far? Find where he lands: (x, 0)

$$0 = -\frac{1}{36}x^2 + \frac{2}{3}x \quad (\text{mult. by } -36)$$

$$0 = x^2 - 24x$$

$$x = 0 \text{ or } 24$$

0 is the beginning of his jump;
24 feet is where he lands!

How high? Find the vertex!

axis of symmetry is $x = 12$

$$y = -\frac{1}{36}(12)^2 + \frac{2}{3}(12) = 4$$

so, vertex is (12, 4)

Therefore, he reaches 4 feet high!

5) An observer claims the population at a baseball game changes exponentially as the weather changes. The table shows statistics from 8 randomly chosen games.

a) What is the observer's exponential model that would estimate attendance?

$$y = 13271.3(1.00936)^x$$

where x = temperature
 y = attendance

b) What are potential flaws to this model? (What are its limitations?)
What other factors may influence the attendance?)

As weather gets colder in the fall, the playoffs might offset.

In summer, school is out. More kids can attend.

At some high temperature, attendance will level off or even decline.. Also, attendance is limited by size of the stadium

c) On May 3, the attendance was 24,607.
What did your model predict the attendance would be on a 68 degree day?

$$25,006$$

$$13271.3(1.00936)^{68}$$

d) Is a quadratic model better?

Perhaps... Because extremely low and extremely high temperatures might lead to lower attendance...

e) What is the correlation coefficient? Does this suggest that a linear model would be effective?

$$r = .845881$$

This suggests a 'strong' linear relationship!

Game Date	Temperature (degrees F)	Attendance
April 21	62	23,567
May 3	68	24,607
June 8	75	29,312
June 27	78	28,778
July 12	67	22,546
July 22	75	27,678
August 3	88	30,133
August 18	91	28,998

Correlation Coefficient

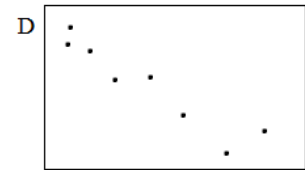
$$-1 < r < 1$$

0 no correlation

1 direct correlation

-1 inverse correlation

SOLUTIONS



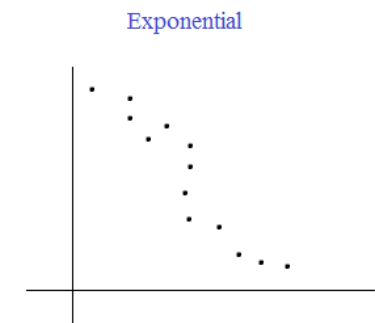
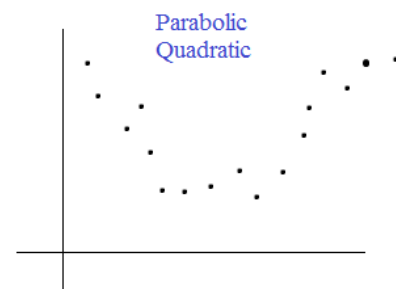
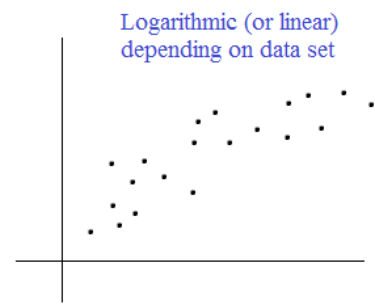
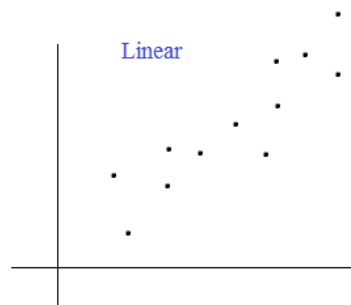
Place the scatterplots in order from lowest to highest r-value

Smallest to largest:

D A B E F C

Identify the best fit model

- a) Linear
- b) Exponential
- c) Logarithmic
- d) Parabolic/Quadratic



Identifying Models

Match the description or table with the correct graph...

Note: it's helpful to identify the independent and dependent variables... (i.e. what are the labels on each axis?)

SOLUTIONS

- 1) Height above ground of a chair on a moving ferris wheel.

E

2)

x	-3	-2	-1	0	1
y	-12	-8	-4	0	4

A

- 3) Your math grade based on the number of points you got on the test...

B

- 4) Population growth of a new town...

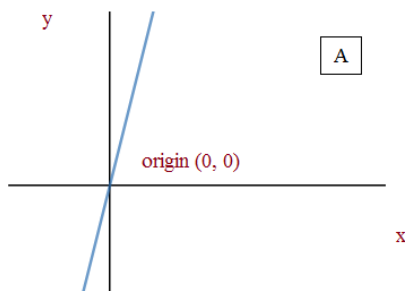
D

- 5) Distance from Chicago when driving from Los Angeles to New York City...

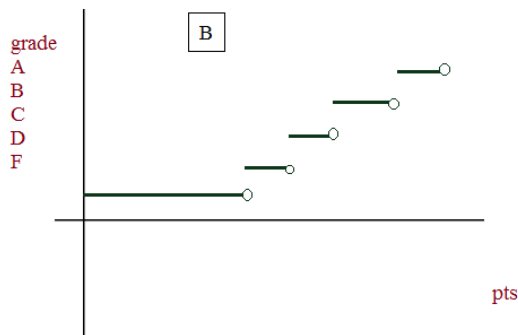
F

- 6) Value of a car during its life time

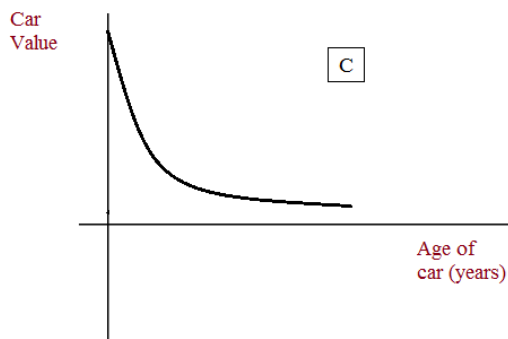
C



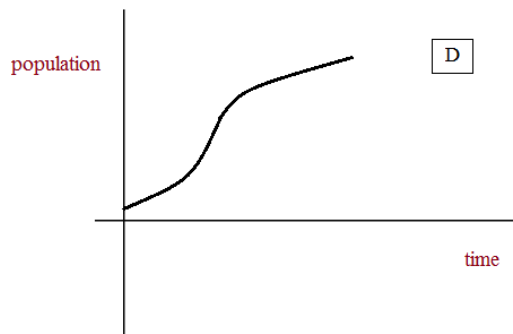
A



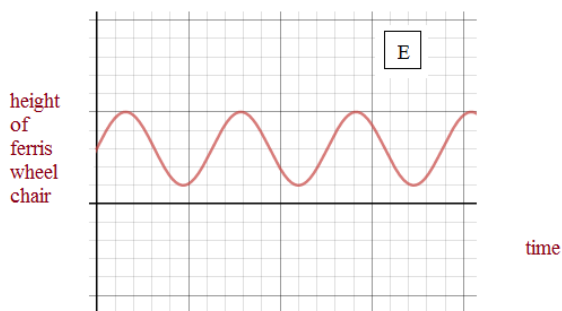
B



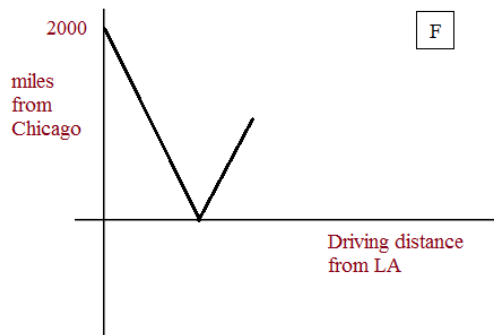
C



D



E

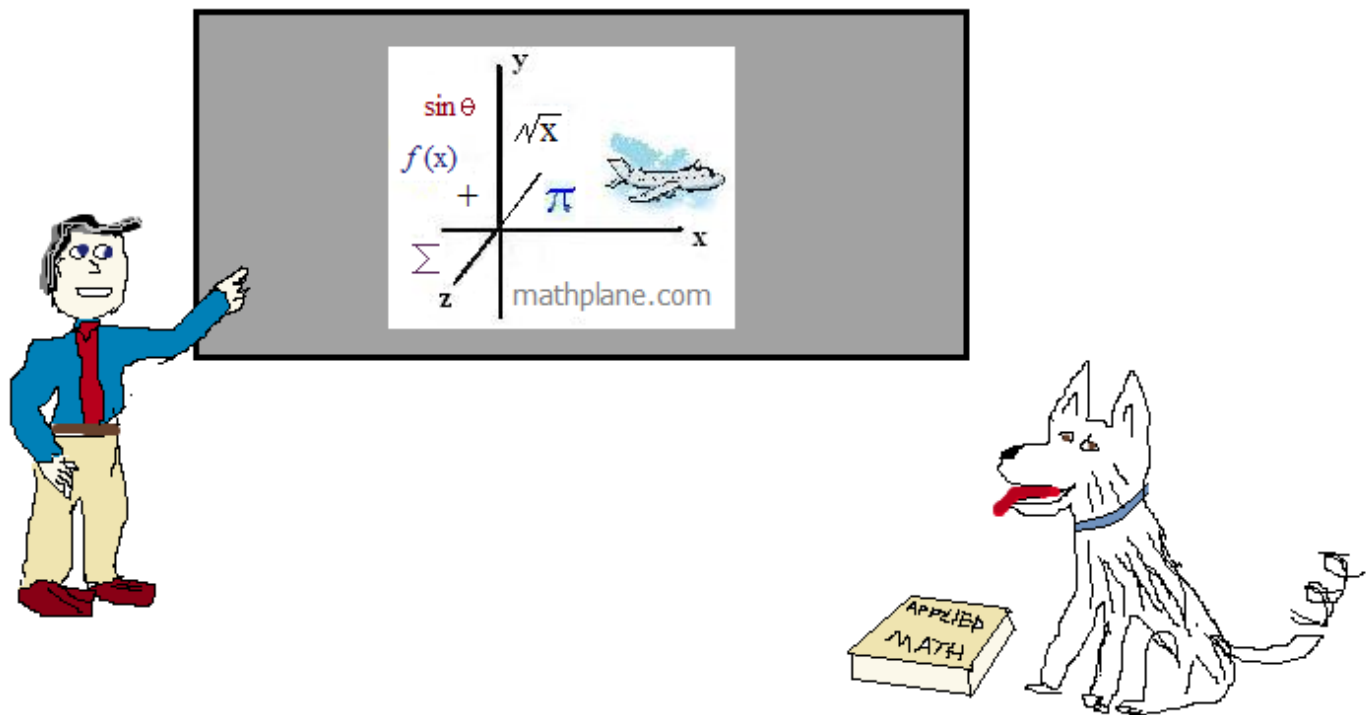


F

Thanks for visiting! (Hope it helped a bit.)

If you have questions, suggestions, or requests, let us know.

Cheers.



Find our store at [TeachersPayTeachers.com](https://www.teacherspayteachers.com)

And, more content at [Mathplane.ORG](https://www.mathplane.org)