Warm-up!

Below is data reflecting the amount of solar power generated (in billion kWh) in Costa Rica since 2010.

Use a graphing calculator to obtain a regression equation that models this data. Predict the amount of energy produced in 2023.

2320.349

Exponential Functions Day 6: (2.6) Competing Model Validation

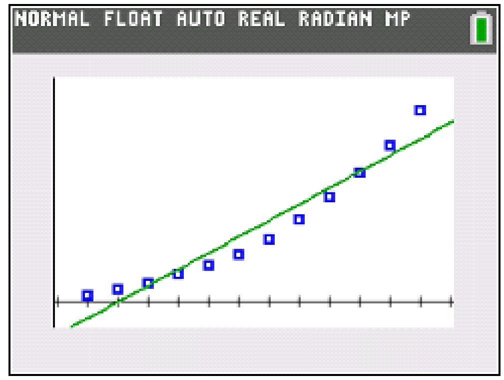
Homework: Competing Model Validation

If you missed the quiz, the smart move is to make it up as soon as possible.

t	10 ~~2010~~	11 ~~2011~~	12 ~~2012~~	13 ~~2013~~	14 ~~2014~~	15 ~~2015~~	16 ~~2016~~	17 ~~2017~~	18 ~~2018~~	19 ~~2019~~	20 ~~2020~~	21 ~~2021~~
Energy (billion kWh)	33.5	66.3	104.2	146.7	202.2	260.2	341.5	445.1	569.8	701	849.1	1036.2

How Good is A Model?

Below is data reflecting the amount of solar power generated (in billion kWh) in Costa Rica since 2010. To the right is a plot reflecting this data along with a linear regression model.



Do you think linear regression was the right decision for this data? Why?

t	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Energy (billion kWh)	33.5	66.3	104.2	146.7	202.2	260.2	341.5	445.1	569.8	701	849.1	1036.2

How Good is a Model?

In the world of statistics, "how appropriate is this model" can be a deep and complex question with many approaches.

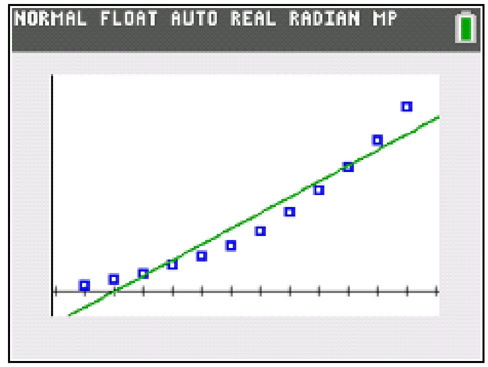
In the world of AP Precalculus focus only on the following three things:

1) Does the context of the data suggest one model may be more appropriate than another?
2) Residuals are without pattern
3) A third thing which I'm going to hold off telling you for now

Residual - Definition

Residual = Actual Minus Predicted

Below is data reflecting the amount of solar power generated (in billion kWh) in Costa Rica since 2010 along with the amount predicted by a linear regression model. To the right is the corresponding plot.

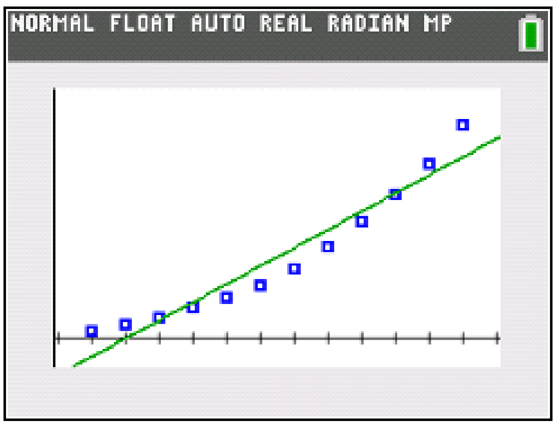


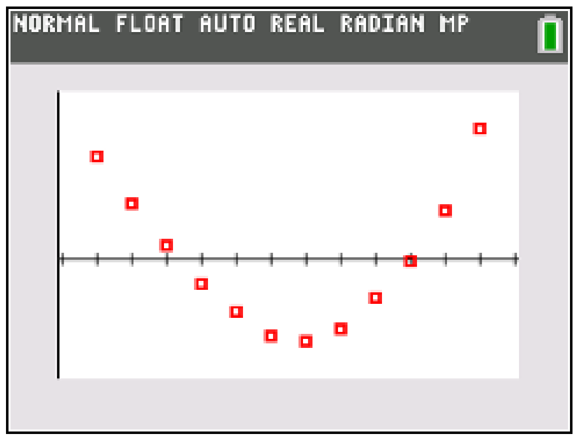
t	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Energy (billion kWh)	33.5	66.3	104.2	146.7	202.2	260.2	341.5	445.1	569.8	701	849.1	1036.2
$f(t)$ Where f is the linear regression model	-87.9	0.2	88.2	176.2	264.3	352.3	440.3	528.4	616.4	704.4	792.5	880.5

Residual Plot

Below is data reflecting the amount of solar power generated (in billion kWh) in Costa Rica since 2010 along with the amount predicted by a linear regression model and the corresponding residual values. The table is followed by a plot of the data and linear regression model and the corresponding residual plot.

t	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Energy (billion kWh)	33.5	66.3	104.2	146.7	202.2	260.2	341.5	445.1	569.8	701	849.1	1036.2
$f(t)$ Where f is the linear regression model	-87.9	0.2	88.2	176.2	264.3	352.3	440.3	528.4	616.4	704.4	792.5	880.5
Residuals	121.4	66.1	16.0	-29.5	-62.1	-92.1	-98.8	-83.3	-46.6	-3.4	56.6	155.7

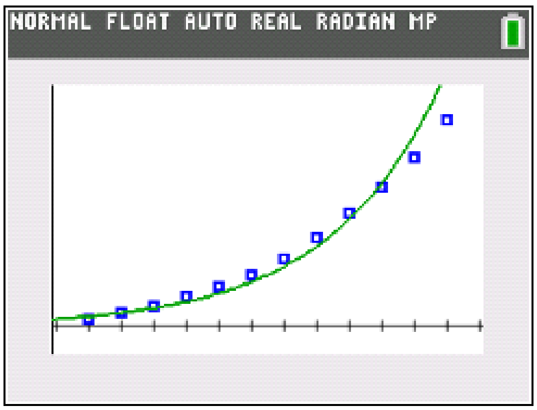


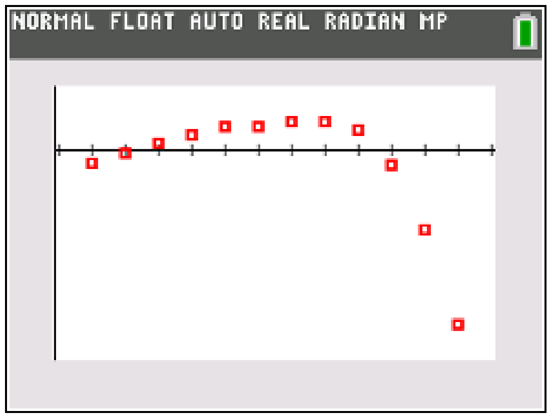


Residual Plot

Below is data reflecting the amount of solar power generated (in billion kWh) in Costa Rica since 2010 along with the amount predicted by an **exponential** regression model and the corresponding residual values. The table is followed by a plot of the data and **exponential** regression model and the corresponding residual plot.

t	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Energy (billion kWh)	33.5	66.3	104.2	146.7	202.2	260.2	341.5	445.1	569.8	701	849.1	1036.2
$f(t)$ Where f is the linear regression model	52.2	69.9	93.6	125.4	167.8	224.7	300.8	402.8	539.3	722.1	966.8	1294.4
Residuals	−18.7	−3.6	10.6	21.4	34.4	35.5	40.7	42.3	30.5	−21.1	−117.7	−258.2





Do we think an exponential model is the most appropriate model choice?

The data below represent the height and wingspan of the first ten basketball player names that came to my mind. Use this data to construct a **linear** model that predicts wingspan based on height.

	Height (inches)	**Wingspan (inches)**
Michael Jordan	78	83
Kobe Bryant	78	78
LeBron James	81	84
Kareem Abdul-Jabbar	86	89
Shaquille O'Neal	85	91
Kevin Durant	83	89
Russell Westbrook	76	80
Stephen Curry	74	76
Kawhi Leonard	79	87
Jon Scheyer	77	76

A) James Harden is 77 inches tall. Predict his wingspan.

79.891 ← Prediction

B) James Harden actually has a wingspan of 83 ← Actual
inches. Did the model overestimate or underestimate his wingspan?

Residual = Actual − Prediction
Residual = 83 − 79.891

C) What is the residual value associated with using the model to predict Kawhi Leonard's wingspan?

Actual = 87 Prediction = 82.416
Residual = 87 − 82.416 = 4.584

D) Do you believe a linear model is appropriate for predicting wingspan based on height?

E) In which should we have more faith: using this model to predict the wingspan of a 80 inch tall individual or using this model to predict the wingspan of a 90 inch tall individual?

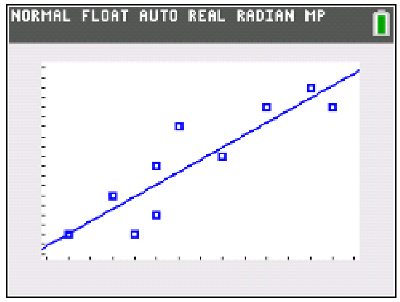
extrapolation interpolation

In general, we trust interpolation more than extrapolation

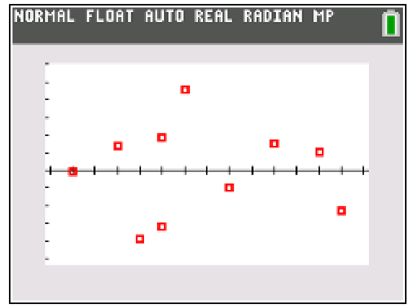
The data below represent the height and wingspan of the first ten basketball player names that came to my mind. Use this data to construct a **linear** model that predicts wingspan based on height.

	Height (inches)	Wingspan (inches)
Michael Jordan	78	83
Kobe Bryant	78	78
LeBron James	81	84
Kareem Abdul-Jabbar	86	89
Shaquille O'Neal	85	91
Kevin Durant	83	89
Russell Westbrook	76	80
Stephen Curry	74	76
Kawhi Leonard	79	87
Jon Scheyer	77	76

D) Do you believe a linear model is appropriate for predicting wingspan based on height?



(data)

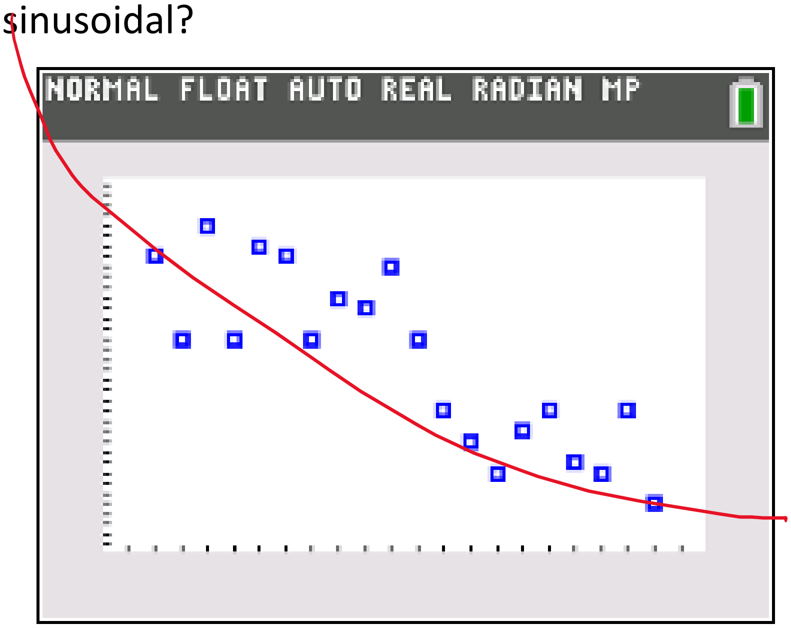


(residual plot)

Below is Data on Air Pollution in Des Moines over the past 20 years. t represents years since 2003. (source: epa.gov/air-data)

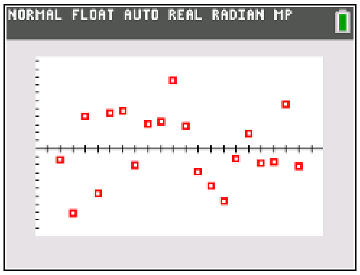
t	Air Pollution in Des Moines (%)
0	37
1	29
2	40
3	29
4	38
5	37
6	29
7	33
8	32
9	36
10	29
11	22
12	19
13	16
14	20
15	22
16	17
17	16
18	22
19	13

Any guess what type of model may be most appropriate? Linear, quadratic, exponential, or sinusoidal?

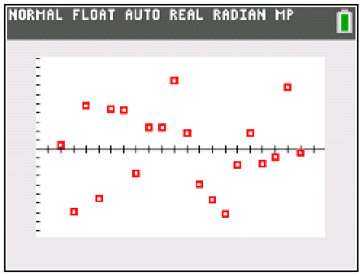


Residual Plots From Various Models of Des Moines Pollution Levels (2003-2022)

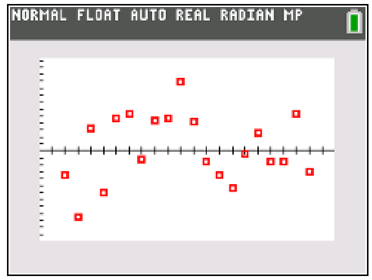
Linear



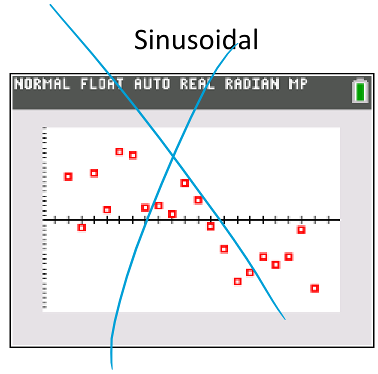
Quadratic



Exponential



Sinusoidal



Which models are/are not appropriate?

Below is Data on Air Pollution in Des Moines over the past 20 years. t represents years since 2003. (source: epa.gov/air-data)

t	Air Pollution in Des Moines (%)
0	37
1	29
2	40
3	29
4	38
5	37
6	29
7	33
8	32
9	36
10	29
11	22
12	19
13	16
14	20
15	22
16	17
17	16
18	22
19	13

Let $P(t)$ be a function that models the level of air pollution in Des Moines t years after 2003.

In 2020, there was a significant reduction in the number of drivers on the roads of Des Moines.

Based on this information, what is a reasonable domain restriction to put on $P(t)$?

Domain: $t < 17$

How Good is a Model?

In the world of AP Precalculus focus only on the following three things:

1) Does the context of the data suggest one model may be more appropriate than another?
2) Residuals are without pattern
3) Depending on context, it may be more appropriate to have an underestimate or overestimate for any given interval.

Taxes, family, and jobs are easy. The hardest part about growing up is learning how much pizza to order for a group. But I've been gathering data! Below is how many large pizzas groups of various sizes ate.

Number of People	Number of Pizzas
10	2.5
15	4
21	5.5
38	13
44	16.5
55	20
58	20.5

In predicting future pizza orders based on the number of people in the group, the most appropriate model to use would be…

(A) A quadratic model that underestimates the number of pizzas needed

(B) A quadratic model that overestimates the number of pizzas needed

(C) A linear model that underestimates the number of pizzas needed

(D) A linear model that overestimates the number of pizzas needed

Below is data reflecting how many Instagram users there were t years after Instagram debuted in 2010 (i.e. $t = 3$ corresponds with the year 2013). (Data source: https://www.businessofapps.com/data/instagram-statistics/)

Year	3	4	5	6	7	8	9	10	11	12
Users (millions)	110	200	370	500	700	1000	1210	1520	1890	2270

A) Use a calculator to construct a linear model of this data. Use this model to predict the number of Instagram users in 2023.

2292

B) Use a calculator to construct a quadratic model of this data. Use this model to predict the number of Instagram users in 2023.

2678.667

C) Use a calculator to construct an exponential model of this data. Use this model to predict the number of Instagram users in 2023.

4008.621

D) Which of these three models do you believe may be most appropriate to use? Or, do you have reason to believe that none of these models is ideal?

Below is data reflecting how many Instagram users there were t years after Instagram debuted in 2010 (i.e. $t = 3$ corresponds with the year 2013). (Data source: https://www.businessofapps.com/data/instagram-statistics/)

Year	3	4	5	6	7	8	9	10	11	12
Users (millions)	110	200	370	500	700	1000	1210	1520	1890	2270

Using the exponential model…

E) What is the residual value associated with $t = 8$ (year 2018)? Does the model overestimate or underestimate the actual number of Instagram users that year?

$1000 - 803.492 = 196.508$

E) Legend states that when the number of Instagram users reaches 4000 million, the Instagram servers will explode. If this legend is true, what domain restriction should we put on the exponential model?

The model outputs values greater than 4000 when $t > 12.993$

In reality, Instagram would have exploded for that many users.

So we restrict the domain to $t < 12.993$

37.912

F) By what percent does the number of Instagram users increase each year?

The total number of plays, in thousands, for the song since its release can be modeled by the function D given by $D(t) = at^2 + bt + c$, where $D(t)$ is the total number of plays, in thousands, for the song since its release, and t is the number of months after the musician began using the app.

Months after the musician began using the app	0	2	4
Total number of plays for the song since its release (thousands)	25	30	34

C. The quadratic function model D has exactly one absolute minimum or one absolute maximum. That minimum or maximum can be used to determine a domain restriction for D. Based on the context of the problem, explain how that minimum or maximum can be used to determine a boundary for the domain of D.

We need to restrict the domain based on absolute max. Total streams cannot decrease. So the domain for the model should be $t \leq 11$

After $t = 11$, the model decreases. So, the domain must be restricted to exclude this portion. The domain should be $t \leq 11$.

The model decreases after $t = 11$
In reality, total streams cannot decrease
So we restrict the domain to $t \leq 11$

Year	iPads (M)	Basis
2010	7.5	Apple FY unit sales
2011	32.4	Apple FY unit sales
2012	58.3	Apple FY unit sales
2013	71.0	Apple FY unit sales
2014	68.0	Apple FY unit sales
2015	54.9	Apple FY unit sales
2016	45.6	Apple FY unit sales
2017	43.8	Apple FY unit sales
2018	43.5	Apple FY unit sales
2019	49.9	Calendar-year shipments (IDC)
2020	53.2	Calendar-year shipments (IDC)
2021	57.8	Calendar-year shipments (IDC)
2022	61.8	Calendar-year shipments (IDC)
2023	54.1	Calendar-year shipments (Canalys)
2024	56.9	Calendar-year shipments (Canalys)

From <https://chatgpt.com/c/68bca74c-51dc-8328-ac8c-29cb145a1e14>