A quadratic and exponential curve can look similar. How do you tell if a quadratic model or an exponential model would be a better fit for a set of data? For those students familiar with statistics, you know we could look at the value of r and r^2. We are NOT doing that in this course. We are keeping it much simpler.

One way of knowing what model is best, is to examine **contextual clues**. Is something growing at a constant rate (linear)? Is it growing at a linear rate of change (rate is changing at a constant rate = quadratic)? Is it growing at a proportional rate (exponential)?

1. A pizza is removed from the oven and cools down for 5 minutes. During this time interval, the temperature is recorded once every 10 seconds. The recorded temperature with respect to time can be modeled by linear, quadratic, or exponential functions. For each of the three models, they appear to represent the data. Which of the following conclusions is best?

    (A) A linear model is best based on contextual clues.

    (B) A quadratic model is best based on contextual clues.

    (C) An exponential model is best based on contextual clues.

    (D) Contextual clues fail to help in selecting a model for this contextual situation.

Another very useful way to know if a model is appropriate for the data set is to look at the plots of **residuals**. But first….

What is a residual plot?
Recall that regression lines are lines/curves that "fit" a set of data.
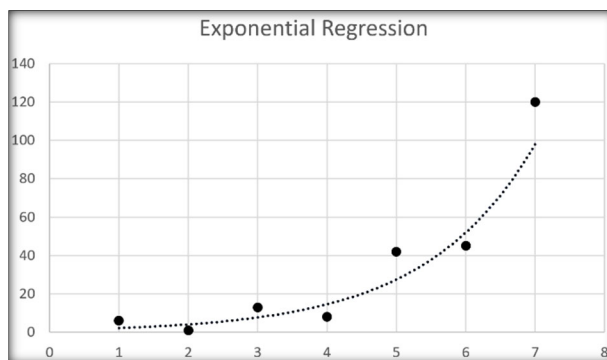
A **residual value** is a measure of how much a regression curve vertically misses a data point. You take the "actual" measured data point and subtract the "predicted" value from the regression line.

$$\text{ACTUAL} - \text{PREDICTED} =$$

This gives us a new set of data that can be graphed as a residual plot.

Exponential Residual Plot
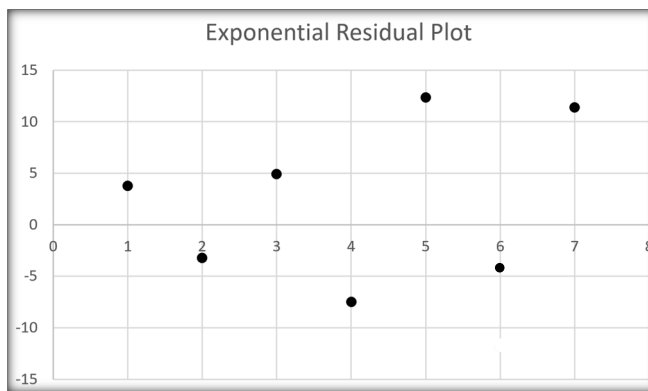
The difference between the predicted and actual values is the _____ in the model. Depending on the data set and context, it may be more appropriate to have an underestimate or overestimate for any given interval.

PREDICTED – ACTUAL =
The line of best fit will give either an overestimate or an underestimate.

*Notice that a "residual" and an "error" are very similar, and often used synonymously. There is a difference though. The sign of the error tells us if the prediction (from the regression model) is an overestimate or an underestimate.

2. A high school student is getting ready for basketball season by getting extra practice before school. Each morning, he records the number of practice shots he took, and records how many he made. The table below represents his recordings.

Shots taken t (input)	100	75	120	200	50	300
Shots made m (output)	35	22	37	60	20	126

a. Find a linear regression curve.

b. Use your model from part a to find the error at $t = 50$.

c. Is the value predicted an overestimate or underestimate of the actual value?

3. Several cars were tested for their braking distance. A regression model was calculated with the car's weight (in pounds) as input values and the braking distance (in feet) as the output values. The given residual plot has a point labeled P at the coordinate $(3,300, -15)$. What does point P indicate in the context?
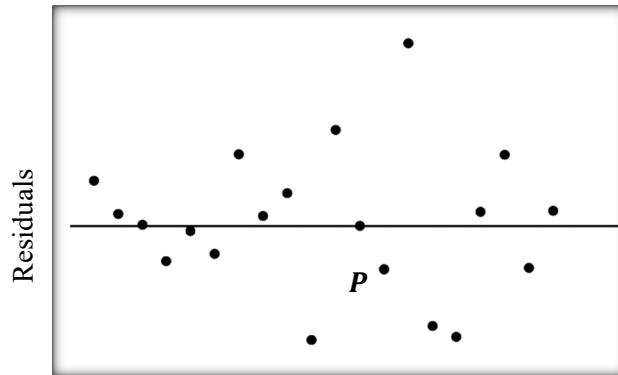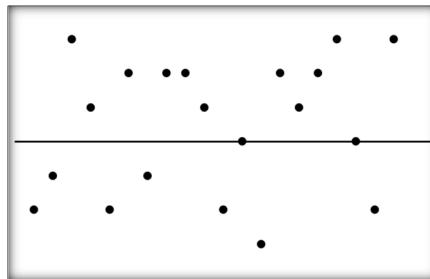

<u>Residual Plot Patterns</u>

How does the shape of a residual plot tell us if the regression curve is an appropriate model for the data set? One way is to ask the question,

Do the residuals exhibit a clear pattern?
1. In a "good" residual plot, the residuals exhibit no clear pattern.
   No pattern = good regression model.

2. In a "bad" residual plot, the residuals exhibit some type of pattern such as a curve or a wave.
   Pattern = bad regression model.
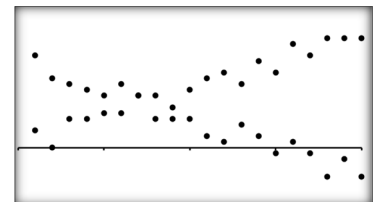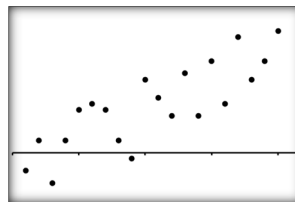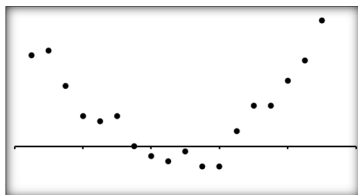
2.6 Competing Function Model Validation
AP Precalculus

1. The salary of a high school teacher typically increases as they have more experience. Below is a list of teacher salaries for one district based on their years of experience.

Years taught t	0	4	10	15	25	30
Annual Salary S	$53,300	$61,500	$68,800	$72,150	$77,600	$79,900

    a. Find a linear regression curve. Round to three decimal places.

    b. Use your model from part a to find the error at $t = 4$.

    c. Is the value predicted an overestimate or underestimate of the actual value?

2. The altitude of the Sun is simply a measurement of how high it appears above the horizon, in terms of an angular measurement. On Halloween, Mr. Bean does his best to measure the sun's altitude throughout the day. The measurements in hours h since sunrise and the angle a in degrees are given below.

| Hours h since sunrise (input)	0.5	3	7	10	11
Altitude a in degrees (output)	6	32	41	12	0

    a. Find a quadratic regression curve. Round to three decimal places.

    b. Use your model from part a to find the error at $t = 10$.

    c. Is the value predicted an overestimate or underestimate of the actual value?

3. A student creates a model for a data set. Which of the following could be an appropriate residual plot for their **linear** regression? More than one answer may apply.
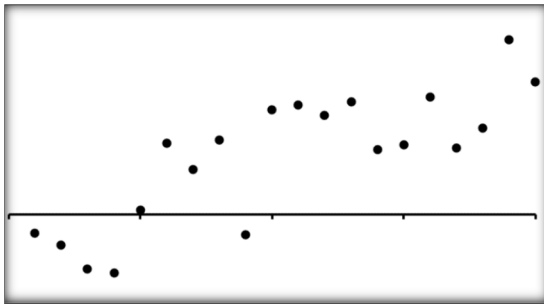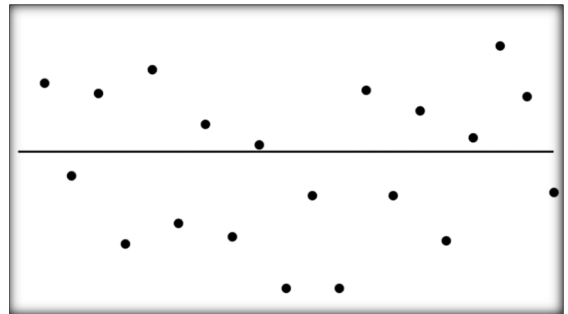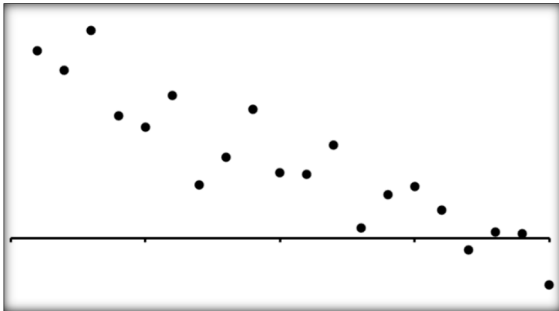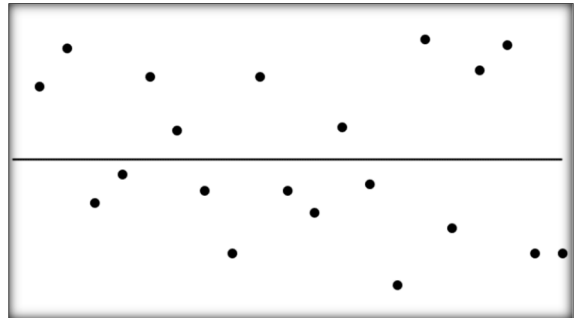
I.
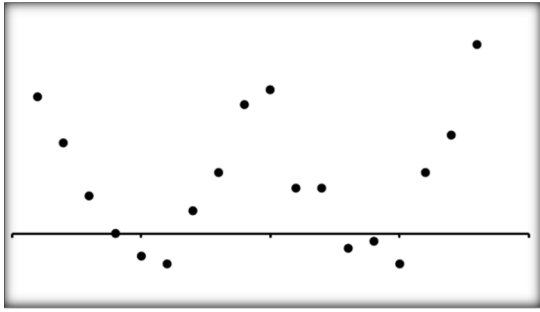
II.
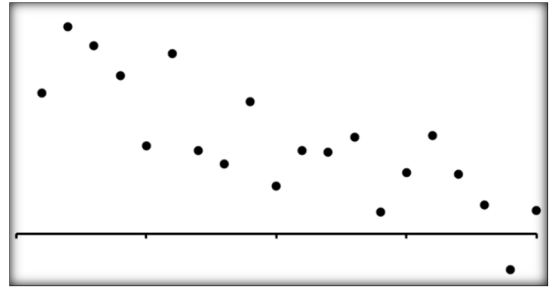
III.

IV.

4. A student creates a model for a data set. Which of the following could be an appropriate residual plot for their **quadratic** regression? More than one answer may apply.
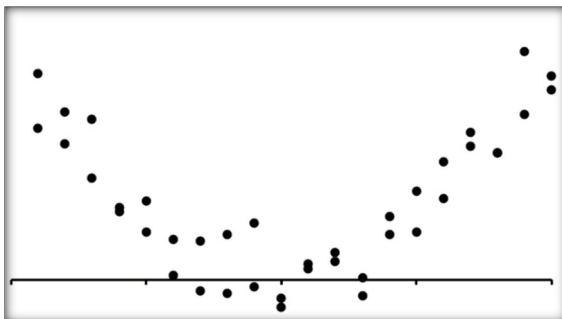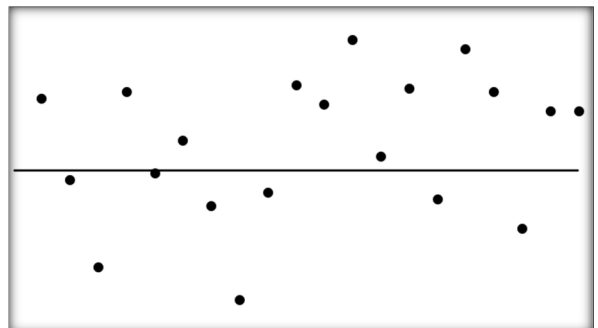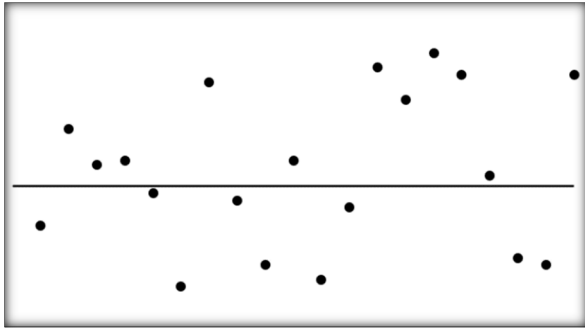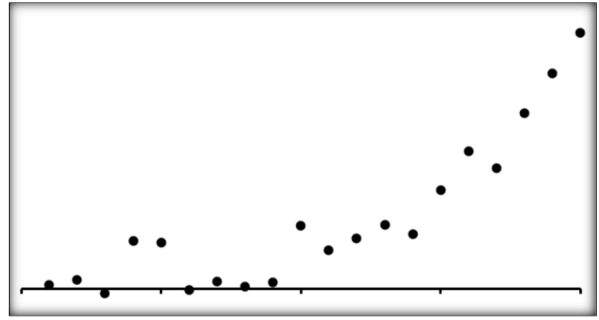
I.

II.

III.

IV.

5. A student creates a model for a data set. Which of the following could be an appropriate residual plot for their **exponential** regression? More than one answer may apply.
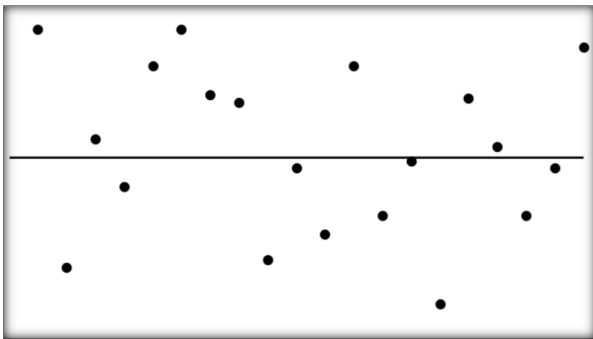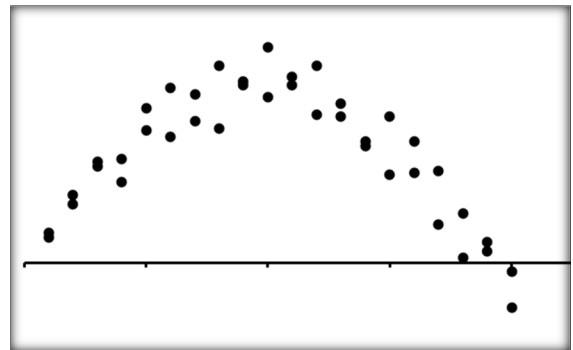
I.

II.

III.

IV.

6. Students from an AP Precalculus class were brought to the gym to shoot free throws, but only the students with little-to-no basketball experience were allowed to shoot. Each student was given 60 seconds to shoot as many free throws as possible. A regression model was calculated with the student's height (in inches) as input values and the number of shots made as the output values. The residual plot for the regression model and data points is below.
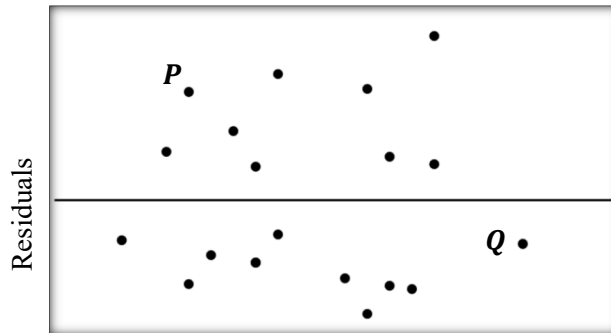
a. The given residual plot has a point labeled P at the coordinate $(62, 4.7)$. What does point P indicate in the context?

b. The point labeled Q is at the coordinate $(74, -2.4)$. What does point Q indicate in the context?

7. After getting a driver's license, the reality of monthly insurance costs (premiums) sets in for many people. Insurance companies charge less for insurance the more years of experience a driver has. A local high school statistics class collects information from several insurance companies regarding their monthly insurance premiums. A regression model was calculated with the driver's experience (in years) as input values and the monthly premiums (in dollars) as the output values. The residual plot for the regression model and data points is below.
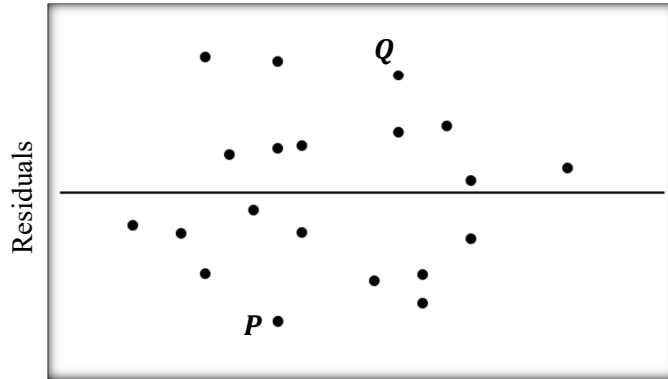

a. The given residual plot has a point labeled P at the coordinate $(6, -102.5)$. What does point P indicate in the context?

b. The point labeled Q is at the coordinate $(10, 97.34)$. What does point Q indicate in the context?

2.6 Competing Function Model Validation

2.6 Test Prep

8. Mr. Brust started a new online business selling Brust-merch. He made estimates about his revenue and used an exponential regression to develop a model for sales over time. The figure shows the graph of the residuals of the exponential regression. Which of the following statements about the exponential regression is true?
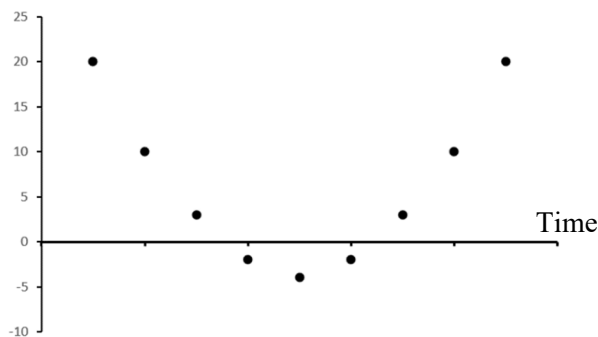
(A) The exponential model is appropriate because there is a clear pattern in the graph of the residuals.

(B) The exponential model is appropriate because there are data points both above and below the x-axis.

(C) The exponential model is not appropriate, because there is a clear pattern in the graph of the residuals.

(D) The exponential model is not appropriate because the residuals are not in an exponential pattern.