**Exact vs. Approximate Linear Models**

**Rationale for selecting/designing this problem/task sequence:**

* In this activity, students will compare exact linear data to approximate linear data. Students will write a model for the exact linear data by hand. Students will evaluate an appropriate linear model using the linear regression features on the graphing calculator. This activity fits well with Sections 3.3 and 3.4 from the Functions and Change text by Evans, Crauder, and Noell.

**Prerequisite Knowledge:**

* Students should be able to calculate average rate of change and/or constant rate of change for given table values.
* Students should be able to write linear equations when given exact linear data.
* Students should be able to use a graphing utility to find a linear regression model for this data.

**Learning objective(s) and alignment with Student Learning Outcomes (SLO From CEP Matrix):**

* Interpret functions using real‐world contexts by translating across multiple representations, including symbols, tables, graphs, and words.
* Identify and analyze families of functions, including linear, polynomial, rational, exponential, and logarithmic functions.
* Determine key characteristics of functions, including global properties and local patterns of change, and interpret their meanings in context, including asymptotes, concavity, end behavior, extrema, increasing/decreasing intervals, and turning points.
* Apply algebraic techniques and digital resources to create, analyze, and interpret appropriate models (either functions or systems of equations) of real‐life phenomena.

**MIP Components of Inquiry:**

**This section outlines how our activity will meet the Mathematical Inquiry Project (MIP) criteria for active learning, meaningful applications, and academic success skills.**

**Active Learning:** Students engage in active learning when they work to resolve a problematic situation whose resolution requires them to select, perform, and evaluate actions whose structures are equivalent to the structures of the concepts to be learned.

* Students will select the appropriate model for the given data by finding AROC and determining if it is a CROC (exactly linear), or if a regression model will be more appropriate.
* Students will evaluate whether the table values support the trends of the model (exact versus approximately linear).
* Students will select calculations such as first and second order differences to determine trends about how the data increases or decreases.
* Students will perform calculations (by hand or with regression) for slope and intercepts and evaluate their meaning in terms of the scenario. This allows students to realize that exact linear models are not appropriate for all scenarios and that approximate models may be a better option.
* Students will find f(c) for a particular c in the real data set and compare the model’s value to the actual value.

**Meaningful Applications:** Applications are meaningfully incorporated in a mathematics class to the extent that they support students in identifying mathematical relationships, making and justifying claims, and generalizing across contexts to extract common mathematical structure.

* Students will use their knowledge of constant rate of change (CROC) with the given data to determine if a linear model is applicable.
* Students will justify their linear model by analyzing key characteristics, such as slope and intercepts.
* Students will distinguish between exact and approximate linear models.

**Academic Success Skills:** Academic success skills foster students’ construction of their identity as learners in ways that enable productive engagement in their education and the associated academic community.

* Students will apply academic success skills by problem solving with application problems that should be of interest to them.
* Students will think critically about the scenarios, how to write the formulas, and the relationship between key characteristics on the equations and graphs.
* Students will use technology to apply mathematical modeling to real-application problems. This allows students to apply and interpret information about modeling to non-exact data, which is more realistic.

**Exact vs. Approximate Linear Models**

Consider the following two scenarios: The first table provides information on the number of hours studied per week and the students’ corresponding grade in their Functions and Modeling course. The second table provides the amount of snow on the ground after the storm has stopped.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hours Studied** | **Grade (%)** |  | **Hours after end of storm** | **Amount of snow on ground (inches)** |
| 4 | 53 |  | 0 | 18 |
| 5 | 56 |  | 2 | 17.25 |
| 5.2 | 65 |  | 4 | 16.5 |
| 5.5 | 58 |  | 6 | 15.75 |
| 6.3 | 70 |  | 8 | 15 |
| 6.5 | 65 |  | 10 | 14.25 |
| 6.7 | 56 |  | 12 | 13.5 |
| 6.8 | 76 |  | 14 | 12.75 |
| 6.9 | 72 |  | 16 | 12 |
| 7 | 60 |  | 18 | 11.25 |
| 7.1 | 68 |  | 20 | 10.5 |
| 7.2 | 69 |  | 22 | 9.75 |
| 7.3 | 74 |  | 24 | 9 |
| 7.4 | 78 |  |  |
| 7.5 | 80 |  |
| 7.6 | 76 |  |
| 7.9 | 79 |  |
| 8 | 92 |  |
| 8.1 | 85 |  |
| 8.2 | 79 |  |
| 8.6 | 84 |  |

1. For both sets of data, determine the AROC for each pair of data points. Based on the AROC, is an exact linear model appropriate? Describe how you know. If an exact linear model is not appropriate, determine whether the data shows concavity (based on the increasing and decreasing trends).

Snow Level vs. Hour Storm AROC: Constant Rate Of Change of $-0.375$ inches per hour

The second order differences on this data are 0, which is consistent with no change in the rate of change (or a constant rate of change).

Grade (%) vs. Hours studied AROC: Not constant AROC

In this data set, the trend is increasing data. To be concave up, the data would be increasing at an increasing rate. Therefore, the second order differences would be positive. To be concave down, the data would be increasing at a decreasing rate. Therefore, the second order differences would be negative. For the given grade data, the second order differences vary greatly. This shows that there is no definitive concavity with the data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Hours after end of storm** | **Amount of snow on ground (inches)** | AROC | Second Order Differences |
| 0 | 18 | $$\frac{17.25-18}{2-0}=\frac{-0.75}{2}=-0.375$$ |  |
| 2 | 17.25 |
| 4 | 16.5 | $$\frac{16.5-17.25}{4-2}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 6 | 15.75 | $$\frac{15.75-16.5}{6-4}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 8 | 15 | $$\frac{15-15.75}{8-6}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 10 | 14.25 | $$\frac{14.25-15}{10-8}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 12 | 13.5 | $$\frac{13.5-14.25}{12-10}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 14 | 12.75 | $$\frac{12.75-13.5}{14-12}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 16 | 12 | $$\frac{12-12.75}{16-14}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 18 | 11.25 | $$\frac{11.25-12}{18-16}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 20 | 10.5 | $$\frac{10.5-11.25}{20-18}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 22 | 9.75 | $$\frac{9.75-10.5}{22-20}=\frac{-0.75}{2}=-0.375$$ | 0 |
| 24 | 9 | $$\frac{9-9.75}{24-22}=\frac{-0.75}{2}=-0.375$$ | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Hours Studied** | **Grade (%)** | AROC | Second Order Differences |
| 4 | 53 | $$\frac{56-53}{5-4}=\frac{3}{1}=3$$ | 42 |
| 5 | 56 |
| 5.2 | 65 | $$\frac{65-56}{5.2-5}=\frac{9}{0.2}=45$$ |
| 5.5 | 58 | $$\frac{58-65}{5.5-5.2}=\frac{-7}{0.3}=-23.33$$ | -68.33 |
| 6.3 | 70 | $$\frac{70-58}{6.3-5.5}=\frac{12}{0.8}=15$$ | 38.33 |
| 6.5 | 65 | $$\frac{65-70}{6.5-6.3}=\frac{-5}{0.2}=-25$$ | -40 |
| 6.7 | 56 | $$\frac{56-65}{6.7-6.5}=\frac{-9}{0.2}=-45$$ | -20 |
| 6.8 | 76 | $$\frac{76-56}{6.8-6.7}=\frac{20}{0.1}=200$$ | 245 |
| 6.9 | 72 | $$\frac{72-76}{6.9-6.8}=\frac{-4}{0.1}=-40$$ | -240 |
| 7 | 60 | $$\frac{60-72}{7-6.9}=\frac{-12}{0.1}=-120$$ | -80 |
| 7.1 | 68 | $$\frac{68-60}{7.1-7}=\frac{8}{0.1}=80$$ | 200 |
| 7.2 | 69 | $$\frac{69-68}{7.2-7.1}=\frac{1}{0.1}=10$$ | -70 |
| 7.3 | 74 | $$\frac{74-69}{7.3-7.2}=\frac{5}{0.1}=50$$ | 40 |
| 7.4 | 78 | $$\frac{78-74}{7.4-7.3}=\frac{4}{0.1}=40$$ | -10 |
| 7.5 | 80 | $$\frac{80-78}{7.5-7.4}=\frac{2}{0.1}=20$$ | -20 |
| 7.6 | 76 | $$\frac{76-80}{7.6-7.5}=\frac{-4}{0.1}=-40$$ | -60 |
| 7.9 | 79 | $$\frac{79-76}{7.9-7.6}=\frac{3}{0.3}=10$$ | 50 |
| 8 | 92 | $$\frac{92-79}{8-7.9}=\frac{13}{0.1}=130$$ | 120 |
| 8.1 | 88 | $$\frac{88-92}{8.1-8}=\frac{-4}{0.1}=15$$ | -115 |
| 8.2 | 90 | $$\frac{90-88}{8.1-8}=\frac{2}{0.1}=20$$ | 5 |
| 8.6 | 94 | $$\frac{94-90}{8.6-8.2}=\frac{4}{0.4}=10$$ | -10 |

1. Create a function to model both sets of data. If one, or both, of the data sets has a CROC, find this function by hand. If one, or both, of the data sets did not have a CROC, use a graphing utility to find a linear regression model for this data.
	1. Desmos regression: Enter your data by clicking the `+` button and selecting `Table'. Input your x- and y-values, then find a linear regression model by typing 'y\_1~mx\_1+b' in a new input field.
	2. Calculator regression: Enter your data by first pressing STAT. Select Edit by pressing enter. Use L1 for your x-values and L2 for your y-values. Exit table by pressing 2nd, then mode. Enter STAT window, again, and scroll right to CALC. Select LinReg for your linear model, ensure that your x-value list and y-value list are appropriately selected. Enter three times to calculate your regression model.

Grade percentage (regression): $G(h)=7.62182h+18.6362$

Snow level (exact): $m=\frac{17.25-18}{2-0}=-\frac{3}{8}$

$$s-18= -\frac{3}{8}(h-0)$$

$$s-18= -\frac{3}{8}h$$

$s= -\frac{3}{8}h+18$ or $s=-0.375h+18$

1. Verify your findings by plotting the data for both data sets using a graphing utility. Make a scatter plot with each data set along with plotting the found model for each data set.

The first graph below is grade percentage (y-axis) vs. hours studied (x-axis). The second graph below is snow level (y-axis) vs. hours after the storm has ended (x-axis). Students’ graphs should look similar to the following:

|  |  |
| --- | --- |
| Hours studied vs. Grade percentage | Hours after storm ends vs. Snow level |
| Chart | Chart |

1. What type of function(s) worked best for each data set?

Both data sets show that a linear function will be appropriate. When making a scatter-plot for grades vs hours studied, the data takes a shape more close to a line compared to a nonlinear model. Students can be encouraged to try other models and asked if this new regression encompasses more of the data points compared to that of the linear model. Visually, students should note that, for the grade data, there is a general increasing trend. Students can determine that the increasing trend is not strictly increasing at an increasing or decreasing rate by checking the second order differences. This shows that the amount of increase varies and that the model does not have consistent concavity. This would support a linear model (compared to a nonlinear model). The hours studied versus grade percentage is better modeled by linear regression, while snow level is exactly linear.

1. Determine if the functions have any x- or y-intercepts, slope, vertices, local maximum, or local minimums. Interpret the meaning of each that you evaluate in terms of the data given.

Grade percentage

Slope: $7.62182$; for every hour increase of study time, grade percentage will increase $7.62182$ percentage points.

y-intercept: 18.6362; at zero hours of study time, grade percentage is 18.8382%.

x-intercept: -2.445; at a grade percentage of 0%, the student studied -2.445 hours. Students should comment on how this intercept makes no logical sense per the scenario.

Snow level

Slope: $-\frac{3}{8}$; after every hour since the storm stopped, the snow level decreases by $\frac{3}{8}$ inches. This matches our AROC calculations from part (a).

y-intercept: 18; when the storm stops, initially, there is 18 inches of snow on the ground

x-intercept: After 48 hours, the snow is completely gone.

1. In terms of the grades in the Functions and Modeling scenario, what does the calculation

$7.62182(10)+18.6362$

mean in terms of study time and grade percentage?

This calculation will give the theoretical grade, as a percentage, if a student studies for 10 hours.

1. In terms of the grades in the Functions and Modeling scenario, what are you solving for in

$0=7.62182h+18.6362$?

Find this value and interpret it within the context of the scenario. Determine an appropriate interval for the domain of this function, to the nearest tenth of an hour. Explain your reasoning.

This calculation gives the theoretical hours studied in order to have a 0 percent grade in Functions and Modeling.

$$0=7.62182h+18.6362$$

$$-18.6362=7.62182h$$

$$h=-\frac{18.6362}{7.62182}$$

$$h≈-2.445$$

A student would need to study negative 2.445 hours a week in order to obtain 0% in Functions and Modeling according to the regression model.

$I=[0,10.7)$; Students cannot study negative hours in order to receive a 0% based off this model. Similarly, students cannot receive more than 100% in their course, theoretically.

1. If either data set has a function obtained by using regression, plug in three independent values from that given data and compare the answer to the corresponding dependent values from the data.

Three points chosen to test may vary. An example would be:

$G(6.3)= 7.62182(6.3)+18.6362=66.65\%$; underestimate, but close to the actual value given by the data $G(6) = 70\%$.

$G(7)= 7.62182(7)+18.6362=71.99\%$; overestimate, which is much larger than the previously tested data value.

$G(8)= 7.62182(8)+18.6362=79.61\%$; underestimate, which is even larger compared to the two previously tested values.

1. Does the model's theoretical outputs closely relate to the exact value? If not, what could account for the differences?

It depends on the tested value. Some are close to the regression line, while others are further away from the regression line. This could be due to individual students enjoying the subject more than others or having a better memory if the grade is determined by assignments which rely heavily on memorization.

A regression equation is an approximate equation for data that is close to, but not exactly linear.

1. Create your own real-world scenario where the data is exactly linear. Give at least 5 data value points, but you do not need to do any calculations.

Answers will vary, but a possible example can be the amount of money earned vs hours worked for a retail job.

1. Create your own real-world scenario where the data is near, but not exactly, linear. Give at least 5 data value points, and create your regression model for the data.

Answers will vary, but a possible example can be the heights of newborn to 5-year-old boys.

1. Give an example of where a linear model would not be appropriate.

Answers will vary, but a possible example would be stock performance for a company, or S&P 500.