AAT Name ______________________________

2-2 Linear Models Date _______________________#3

Goal: Find the line of best fit (line of regression) and use the sum of squared residuals as a measure of that fit.

Notes	Questions

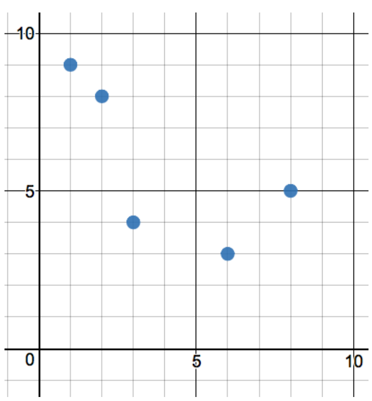
Notes

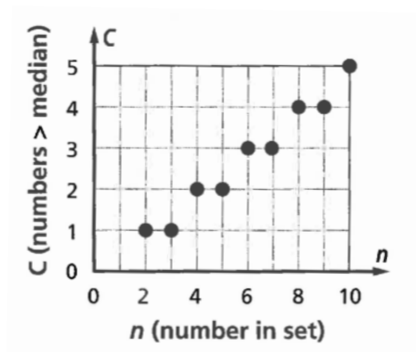
Linear function:

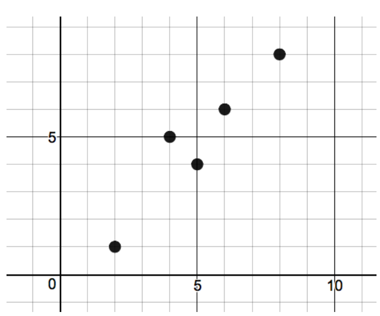
Linear model:

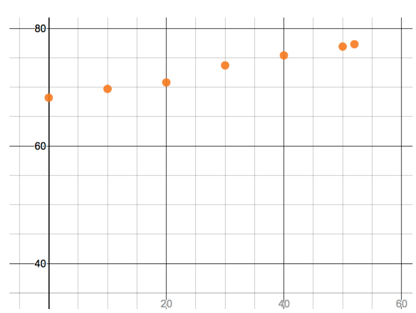
Exact Model Almost Exact Model Impressionistic Model

Fit data "by eye": using a straight edge, draw a trend line of the data:









Questions

Questions | **Notes**

Example 1:

Weight x	Price y (U.S. $)
0.18	702.00
0.17	517.50
0.25	963.00
0.29	1290.00
0.27	1080.00
0.15	484.50
0.20	747.00
0.25	1017.00
0.21	724.50
0.17	529.50
0.35	1629.00
0.33	1417.50
0.26	994.50
0.16	513.00
0.12	334.50
0.18	664.50
0.15	430.50
0.16	507.00
0.16	498.00
0.23	829.50

Source: Journal of Statistics Education

a. Find the equation of the graphed line which relates weight and price.

b. Interpret the equation of the line in context of the problem.

c. Why is the model not good for predicting the cost of a 0.05 carat diamond ring?

d. Why is the set of data not a function?

Interpolation vs extrapolation

Practice: The gold medal winning times for the men's 100-meter dash are listed below for the last 20 Summer Olympic Games.

Questions

City	Year	Winning Time(s)
Beijing	2008	9.69
Athens	2004	9.85
Sydney	2000	9.87
Atlanta	1996	9.84
Barcelona	1992	9.96
Seoul	1988	9.92
Los Angeles	1984	9.99
Moscow	1980	10.25
Montreal	1976	10.06
Munich	1972	10.14
Mexico City	1968	9.95
Tokyo	1964	10.0
Rome	1960	10.2
Melbourne	1956	10.5
Helsinki	1952	10.4
London	1948	10.3
Berlin	1936	10.3
Los Angeles	1932	10.3
Amsterdam	1928	10.8
Paris	1924	10.6

a. The data were graphed and a line fit "by eye" passed through the points (1972, 10.15) and (2004, 9.8). Find the equation of this linear model to relate the year and the winning time.

b. Interpret the slope of your line in the context of the problem.

c. Use the model to predict the winning times for 2012 (London) and 2016 (Rio de Janeiro). Then research and compare your results.

d. Usain Bolt of Jamaica won the 100-m dash at the Beijing 2008 Olympic games in a record of 9.69 seconds. Based on the linear model, when "should" that occur?

Measuring How Well a Lines Models Data

The *average* of a set of data helps us to see what the data tends to do. In other words, what kinds of numbers we expect. Similarly, a linear model gives us an expectation of value and can see how well the *observed data* compares to the *expected data* by calculating the ________________ (by ______________________________), then __________________________________ and ______________________________.

Diamond Ring Prices by Weight of Diamond

Weight	Price ($)	Predicted ($y = 2400x + 400$)	Residuals	Square of Residual
0.15	484.50			
0.16	507.00			
0.18	702.00			
0.25	963.00			
0.27	1080.00			
0.33	1417.50			
0.23	829.50			
			Sum of Squares of Residuals:	

Definition of Sum of Square Residuals

Sum of squared residuals = $\sum_{i=1}^{n}\left(\text{observed } y_i - \text{ predicted } y_i\right)^2$

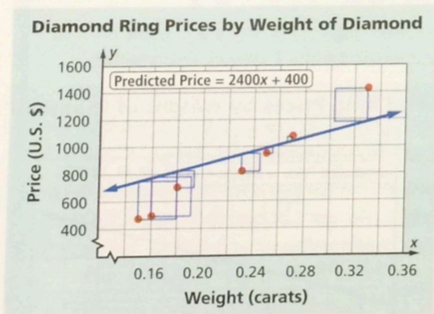
AAT

2-2 Linear Models

#3

Linear Model 1

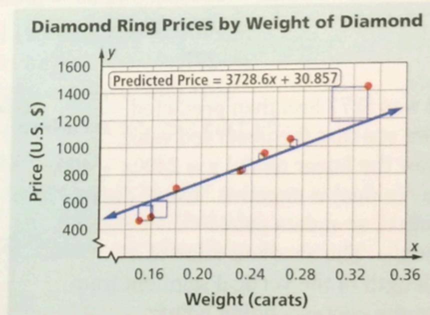
Squares are shown for a line that does not go through any data points.



Total area of the squares ≈ 237,800

Linear Model 2

Squares are shown for a line through two of the data points.



Total area of the squares ≈ 59,870

The second line is a better model of the data because it has a smaller total area of the squares. The total area is the **sum of squared residuals**.

Country	TVs per 100	Unemployed per 100	Predicted $y=-0.3x+17$	Residual	Square of Residual
Argentina	22.3	7.8			
Bulgaria	40	6.3			
India	6.5	6.8			
Israel	29.9	6.1			
Netherlands	51.8	4.5			
New Zealand	52.3	4.0			
Ploan	33.7	9.7			
South Africa	12.3	21.7			
South Korea	34.7	3.2			
			Sum of Squares of Residuals:		

Now, compare with a different model: $y = -0.167x + 13$

Country	TVs per 100	Unemployed per 100	Predicted $y = -0.167x + 13$	Residual	Square of Residual
Argentina	22.3	7.8			
Bulgaria	40	6.3			
India	6.5	6.8			
Israel	29.9	6.1			
Netherlands	51.8	4.5			
New Zealand	52.3	4.0			
Ploan	33.7	9.7			
South Africa	12.3	21.7			
South Korea	34.7	3.2			
			Sum of Squares of Residuals:		

Make a scatterplot of the data and enter the two models in Y1 and Y2 of your calculator.

Which of the two models better describes the data? Explain.