Algebra 2 Honors Linear and Quadratic Regression Worksheet 1

Name ____________________________ Date ______________ Hour ________________

Problems

1. The table below lists the total estimated numbers of United States AIDS cases, by year of diagnosis. Find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Predict the number of aids cases for the year 2006.

Year	AIDS Cases
1999	41,356
2000	41,267
2001	40,833
2002	41,289
2003	43,171

2. The table below lists temperatures measured in Fahrenheit and Celsius. Find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Determine the equivalent temperature in Celsius degrees for a body temperature of 98.6 degrees Fahrenheit.

Fahrenheit degrees (ºF)	Celsius degrees (ºC)
32	0
68	20
86	30
122	50
158	70
194	90
212	100

3. According to Roche Pharmaceuticals, a BMI of 30 or greater can create an increased risk of developing medical problems associated with obesity. The chart below shows the height and weight for individuals with a BMI of 30. Find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Determine the weight of a 75-inch tall person who has a BMI = 30.

Height (inches)	Weight (pounds)
61	160
63	170
65	180
67	190
69	200
72	220
73	230

4. The table below lists distances in mega parsecs and velocities for four galaxies moving rapidly away from earth. Find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Determine the velocity of Hydra, a galaxy located 776 mega parsecs from earth.

Galaxy	Distance (Mpc)	Velocity (km/sec)
Virgo	15	1600
Ursa Minor	200	15,000
Corona Borealis	290	24,000
Bootes	520	40,000

Source: Astronomical Methods and Calculations (1994)

5. The following data represents approximate heights for a ball thrown by a shot-putter as it travels x meters horizontally. Find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. What would be the height of the ball if it travels 80 meters?

Distance (m)	Height (m)
7	8
20	15
33	24
47	26
60	24
67	21

6. The concentration (in milligrams per liter) of a medication in a patient's blood as time passes is given by the data in the following table. Find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. What is the concentration of medicine in the blood after 4 hours have passes?

Time (Hours)	Concentration (mg/l)
0	0
0.5	78.1
1	99.8
1.5	84.4
2	50.1
2.5	15.6

7. A ball is rolled down a hallway and its position is recorded at five different times. Use the data given in the table to find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Predict the location of the ball after 12 seconds.

Time (seconds)	Position (meters)
1	9
2	12
4	17
6	21
8	26

8. Suppose you are standing in the observation deck of the CN tower in Toronto. You drop a penny. The distance of the penny from the ground after various times is given the table below. Use the data given in the table to find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Where is the penny located after falling for 10.5 seconds?

Time (seconds)	Distance (feet)
0	1821
2	1757
4	1565
6	1245
8	797
10	221

9. The table below lists the number of Americans (in thousands) expected to be over 100 years old for selected years. Use the data given in the table to find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. How many Americans will be over 100 years old in the year 2008?

Year	Number (thousands)
1994	50
1996	56
1998	65
2000	75
2002	94
2004	110

10. The table below shows the apparent temperature vs. relative humidity in a room whose actual temperature is 72 degrees. Use the data given in the table to find the linear and quadratic regression equations and correlation coefficients. State which model, linear or quadratic, best fits the data. Predict the apparent temperature when the relative humidity reaches 110%.

Relative Humidity (%)	Apparent Temperature (°F)
0	64
10	65
20	67
30	68
40	70
50	71
60	72
70	73
80	74
90	75
100	76

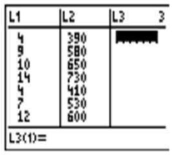
Linear and Quadratic Regression Notes

Find the best fit regression equation.

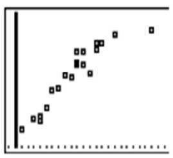
1. Turn on statistical plots. [2ND - STAT PLOT]
2. Turn on Diagnostics to get correlation statistics. [CATALOG (2nd - 0) - DIAGNOSTICS ON]
3. Enter into the calculator list. [STAT – EDIT]
4. Calculate regression statistics. [STAT- CALC - LINREG – L1 (2nd - 1), L2 (2nd - 2), Y1 (VARS - YVARS - FUNCTION - Y1)]
5. Find r, the correlation coefficient.
6. View graph. [GRAPH]
7. Adjust window, if necessary. [ZOOM - ZOOMSTAT]
8. Repeat steps 4 through 7 except calculate regression statistics for best fit quadratic curve (use QUADREG instead for LINREG).
9. The regression that has the best r value is the equation that produces the best fit.

Example of data entry

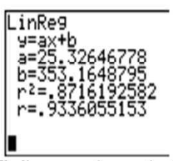
Step 1. Enter the data into the lists.
For basic entry of data, see Basic Commands.



Step 2. Create a scatter plot of the data.
Go to STATPLOT (2nd Y=) and choose the first plot. Turn the plot ON, set the icon to Scatter Plot (the first one), set Xlist to L1 and Ylist to L2 (assuming that is where you stored the data), and select a Mark of your choice.

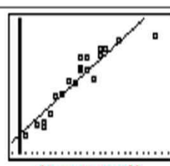


Step 3. Choose Linear Regression Model.
Press STAT, arrow right to CALC, and arrow down to 4: LinReg (ax+b). Hit ENTER. When LinReg appears on the home screen, type the parameters L1, L2, Y1. The Y1 will put the equation into Y= for you.
(Y_1 comes from VARS → YVARS, #Function, Y_1)



The linear regression equation is $y = 25.3x + 353.2$
(answer to part a)

Step 4. Graph the Linear Regression Equation from Y1.
ZOOM #9 ZoomStat to see the graph.

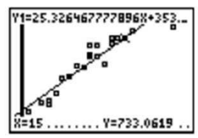


(answer to part b)

Step 5. Is this model a "good fit"?
The correlation coefficient, r, is .9336055153 which places the correlation into the "strong" category. (0.8 or greater is a "strong" correlation)
The coefficient of determination, r^2, is .8716192582 which means that 87% of the total variation in y can be explained by the relationship between x and y. The other 13% remains unexplained.
Yes, it is a "good fit".
(answer to part c)

Step 6. Interpolate: *(within the data set)*
If a student studied for 15 hours, based upon this study, what would be the expected Math SAT score?

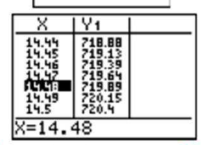
From the graph screen, hit TRACE, arrow up to obtain the linear equation at the top of the screen, type 15, hit ENTER, and the answer will appear at the bottom of the screen.



(answer to part d -- Math SAT score of 733.1)

Step 7. Interpolate: *(within the data set)*
If a student obtained a Math SAT score of 720, based upon this study, how many hours did the student most likely spend studying?

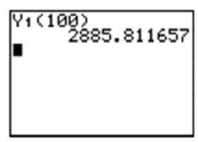
Go to TBLSET (above WINDOW) and set the TblStart to 13 (since 13 hours gives a score of 700). Set the delta Tbl to a decimal setting of your choice. Go to TABLE (above GRAPH) and arrow up or down to find your desired score of 720, in the Y1 column.



(answer to part e -- approx. 14.5 hours)

Step 8. Extrapolate data: *(beyond the data set)*
If a student spent 100 hours studying, what would be the expected Math SAT score? Discuss this answer.

With your linear equation in Y1, go to the home screen and type Y1(100). Press ENTER.



Our equation shows that if a student studies 100 hours, he/she should score 2885.8 on the Math section of the SAT examination. The only problem with this answer is that the highest score that can be obtained is 800. So why is this score so outrageous? ANSWER: When you extrapolate data, the further you move away from the data set, the less accurate your information becomes. In this problem, the largest number of hours in the data set was 22 hours, but the extrapolation tried to jump to 100 hours.
(answer to part f)