Quadratic Regression

1. Amery recorded the distance and height of a basketball when shooting a free throw.

Distance (feet), x	Height (feet), f(x)
0	4
2	8.4
6	12.1
9	14.2
12	13.2
13	10.5

a) Find the quadratic equation for the relationship of the horizontal distance and the height of the ball. Round decimals to the hundredths.

$y = -0.13x^2 + 2.26x + 4.06$

b) Using the function what is the approximate maximum height of the ball?

≈ 13.8

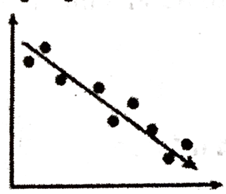
How do we know what is a good fit for our data?

Linear Data

Correlation Coefficient *r* → only tells us what type of correlation we are looking at

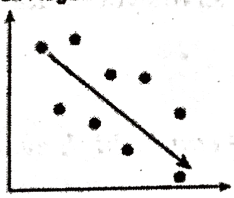
- The correlation coefficient, denoted by *r* varies from -1 to 1. It corresponds to the type of correlation for linear functions.
- Strongly correlated data points have a value of r closer to 1 or -1.
- Weakly correlated data will have values closer to .5

Strong Negative Correlation



r is close to -1

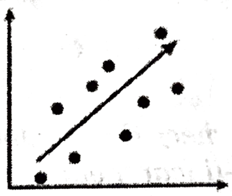
Weak Negative Correlation



r is close to -.5

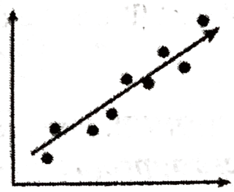
No Corelation
0

Weak Positive Correlation



r is close to .5

Strong Positive Correlation



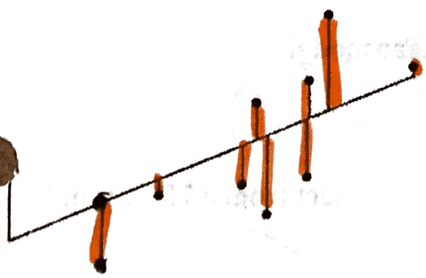
r is close to 1

Residuals

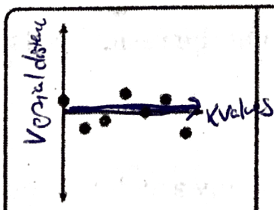
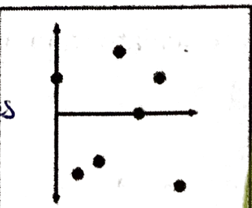
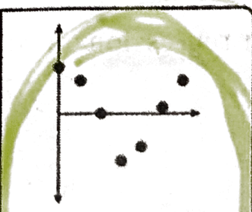
For any set of data, different lines of fit can be created. Some of these lines will fit the data better than others. One way to determine how well the line fits the data is by using residuals. A **residual** is the signed vertical distance between a data point and a line of fit.

After calculating residuals, a residual plot can be drawn. A **residual plot** is a graph of points whose *x*-coordinates are the variables of the independent variable and whose *y*-coordinates are the corresponding residuals.

Looking at the distribution of residuals can help you determine how well a line of fit describes the data. The plots below illustrate how the residuals may be distributed for three different data sets and lines of fit.

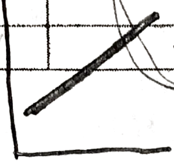


Residuals are the distance between the observed value and the fitted value.

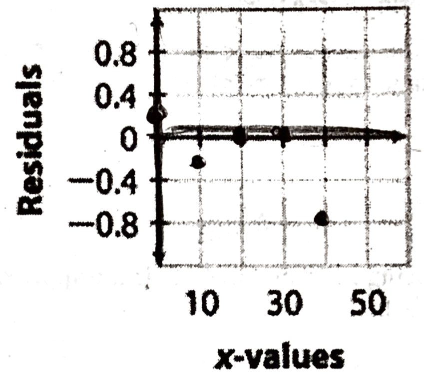
		
Distribution of residuals about the x-axis is random and tight. The line fits the data very well.	Distribution of residuals about the x-axis is random but loose. The line fits the data, but the model is weak.	Distribution of residuals about the x-axis is not random. The line does not fit the data well.

Use residuals to calculate the quality of fit for the line $y = 0.25x + 29$, where y is median age and x is years since 1970.

x	Actual y	Predicted y based on $y = 0.25x + 29$	Residual Subtract Predicted from Actual to Find the Residual.
0	29.2	29	0.2
10	31.3	31.5	-0.2
20	34.0	34	0
30	36.5	36.5	0
40	38.2	39	-0.8



Plot the residuals.



Evaluate the quality of fit to the data for the line $y = 0.25x + 29$.

yes

R^2 is called the coefficient of determination. It is the statistical measure of how close the data is to the fitted regression **line.**

R^2 is always between 0 and 100% (or 0 and 1.0).

0% means that the model explains none of the variability of the response data around its mean.
100% means the model explains all of the variability of the response data around its mean.

In other words, the higher the value of r-squared the better the model fits your data. Use it to determine if your line of best fit fits the data well.

CAREFUL THOUGH! You must always check the residuals first. R^2 can be high even when a linear model is less appropriate than a non-linear model.

Example: **CAR ACCIDENTS**

This table shows the number of car accidents (in 1000's) based on driver age in 2009:

Age (years)	16	20	24	28	32	36	40	44	48	52	56
# accidents	2020	2189	1978	1685	1335	964	598	412	301	198	225

Perform a linear regression on your calculator and plot the residuals. Is a linear model a good fit?

No - residuals form a pattern

Find the quadratic regression. Use it to predict how many accidents you would expect from a 75 year old.

$y = 0.44x^2 + 89.38x + 3653.12$ -575

Identify each table as linear, quadratic, or exponential.

1.

-4	-3	-2	-1	0	1
3	12	48	192	768	3072

2.

0	1	2	3	4	5
25	20	15	10	5	0

3.

3	4	5	6	7	8
10	17	26	37	50	65

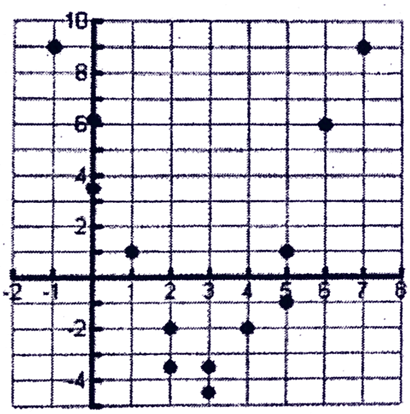
4. Which is the quadratic equation that best fits th scatterplot? **Explain why.**

a) $f(x) = (x - 3)^2 - 4$

b) $f(x) = (x + 3)^2 + 4$

c) $f(x) = (x - 4)^2 - 3$

d) $f(x) = (x + 4)^2 + 3$



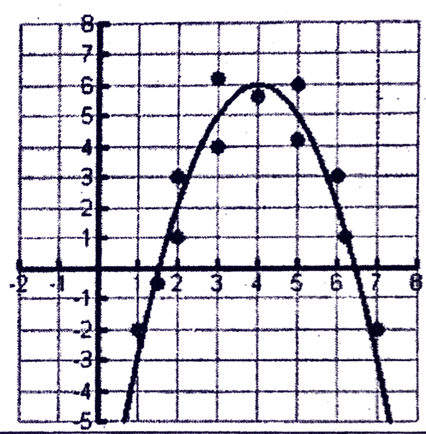
5. Which is the quadratic equation that best fits the scatterplot? **Explain why.**

a) $f(x) = x^2 - 8x + 22$

b) $f(x) = -x^2 - 8x - 10$

c) $f(x) = -x^2 + 8x - 32$

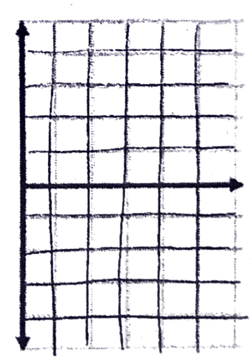
d) $f(x) = -x^2 + 8x - 10$



Residuals

1. Linear regression equation: $y = 0.5x$

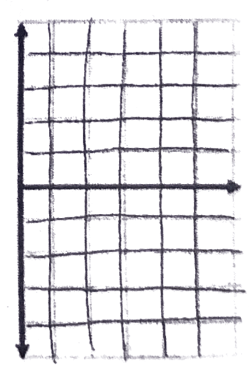
x	y	Predicted Value	Residual Value
5	3		
10	4		
15	9		
20	7		
25	13		
30	15		



Does the residual plot suggest a linear relationship? Explain.

2. Linear regression equation: $y = -0.4x + 16.3$

x	y	Predicted Value	Residual Value
2	5		
4	15		
6	26		
8	23		
10	11		
12	3		



Does the residual plot suggest a linear relationship? Explain.