B – Graphs and Statistics, Lesson 7, Residuals (r. 2018)

GRAPHS AND STATISTICS
Residuals

Common Core Standard	Next Generation Standard
S-ID.B.6b Informally assess the fit of a function by plotting and analyzing residuals. NYSED: Includes creating residual plots using the capabilities of the calculator (not manually).	**STANDARD REMOVED**
S-ID.B.6c Fit a linear function for a scatter plot that suggests a linear association. NYSED: Both correlation coefficient and residuals will be addressed in this standard.	**STANDARD REMOVED**

LEARNING OBJECTIVES

Students will be able to:

1) Understand residuals as the difference between actual and predicted y-values based on a line of best fit.
2) Create residual plots given a table of residual values.
3) Interpret patterns in residual plots as an indication that the regression equation does not fit the data.

Overview of Lesson

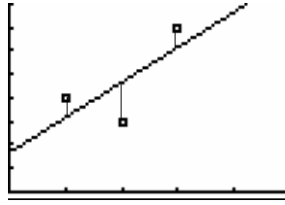| Teacher Centered Introduction	Student Centered Activities
Overview of Lesson - activate students' prior knowledge - vocabulary - learning objective(s) - big ideas: direct instruction - modeling	**guided practice ⬅Teacher: anticipates, monitors, selects, sequences, and connects student work** - developing essential skills - Regents exam questions - formative assessment assignment (exit slip, explain the math, or journal entry)

VOCABULARY

actual y-value   line of best fit
predicted y-value   pattern
residual   fit the data
residual plot

BIG IDEAS

A **residual** is the vertical distance between where a regression equation predicts a point will appear on a graph and the actual location of the point on the graph (scatterplot).   If there is no difference between where a regression equation places a point and the actual position of the point, the **residual** is zero.

A **residual** can also be understood as the difference in predicted and actual y-values (dependent variable values) for a given value of x (the independent variable).

Residual = (actual y-value)-(predicted y-value)

A **residual plot** is a scatter plot that shows the residuals as points on a vertical axis (y-axis) above corresponding (paired) values of the independent variable on the horizontal axis (x-axis).

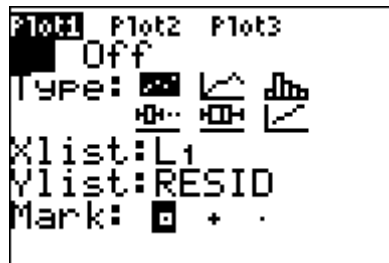Any *pattern* in a residual plot suggests that the regression equation is *not appropriate* for the data.

Patterns in residual plots are bad.

Residual plots with patterns indicate the regression equation is not a good fit.

Residual plots without patterns indicate the regression equation is a good fit.

A **residual plot** without a *pattern* and with a near equal distribution of points above and below the x-axis suggests that the regression equation is a *good fit* for the data.

Residuals are automatically stored in graphing calculators when regression equations are calculated.   To view a residuals scatterplot in the graphing calculator, you must use 2nd LIST to set the Y list variable to RESID, then use Zoom 9 to plot the residuals.
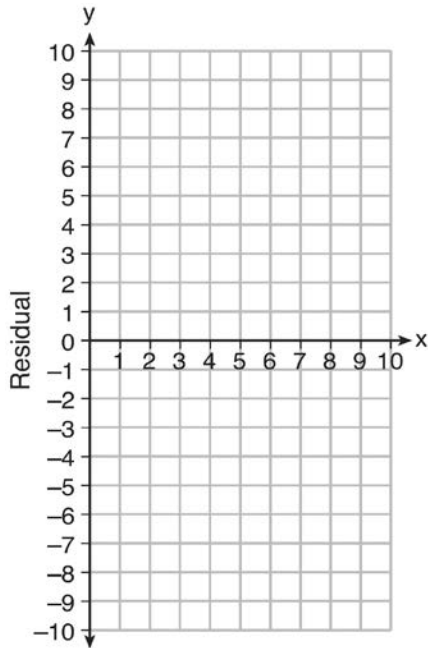

DEVELOPING ESSENTIAL SKILLS

Calculate the residual values:

x	Actual y-value	Predicted y-value	Residual
0	4	-14	10
1	6	-2	8
2	8	2	6
3	10	4	4
4	12	10	2
5	14	14	0

6	16	15	0
7	18	16	2
8	20	15	5
9	22	13	9
10	24	14	10

Plot the residuals and determine if they indicate a good fit or a bad fit.

The residuals form a pattern, so the fit is bad.
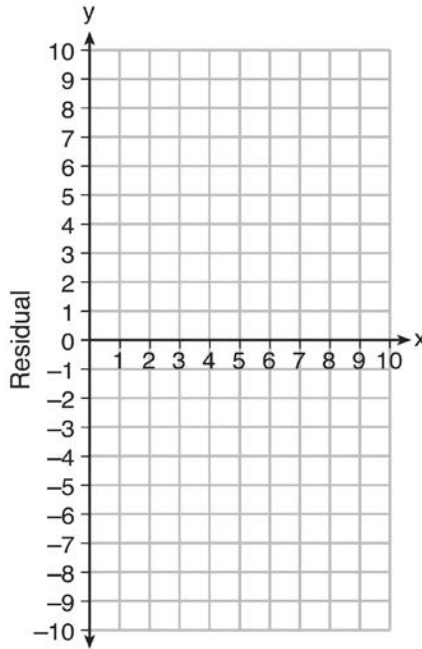
REGENTS EXAM QUESTIONS (through June 2018)

S.ID.B.6b: Residuals

43) Use the data below to write the regression equation ($y = ax + b$) for the raw test score based on the hours tutored. Round all values to the *nearest hundredth*.

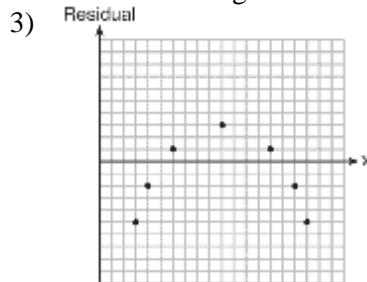| Tutor Hours, x	Raw Test Score	Residual (Actual – Predicted)
1	30	1.3
2	37	1.9
3	35	−6.4
4	47	−0.7
5	56	2.0
6	67	6.6
7	62	−4.7

Equation: _____

Create a residual plot on the axes below, using the residual scores in the table above.
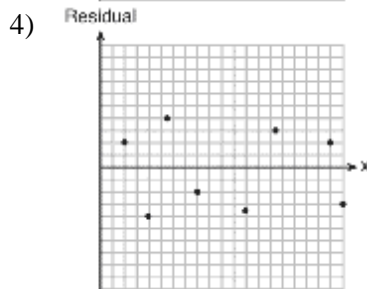

Based on the residual plot, state whether the equation is a good fit for the data.   Justify your answer.
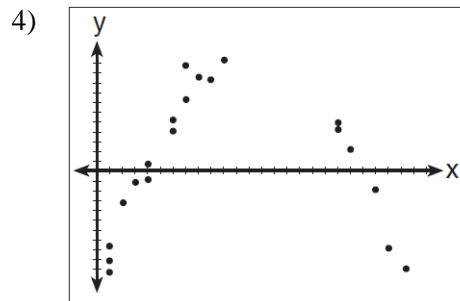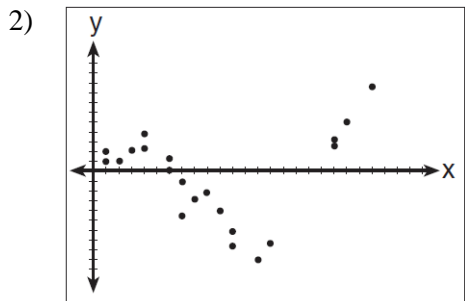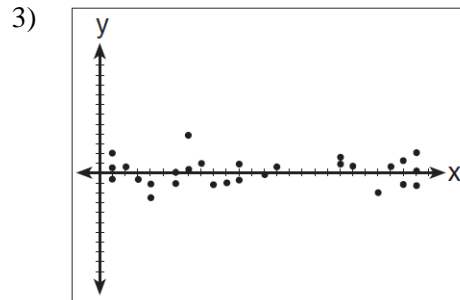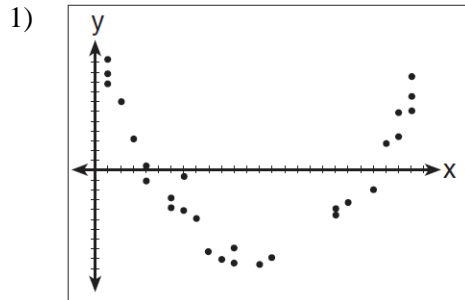
44) Which statistic would indicate that a linear function would *not* be a good fit to model a data set?

1) $r = -0.93$

2) $r = 1$

3)

4)

45) After performing analyses on a set of data, Jackie examined the scatter plot of the residual values for each analysis. Which scatter plot indicates the best linear fit for the data?

1)

3)

2)

4)

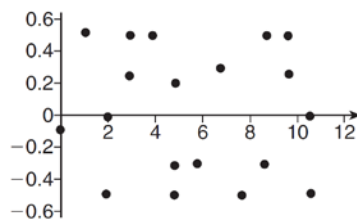

46) The table below represents the residuals for a line of best fit.

x	2	3	3	4	6	7	8	9	9	10
Residual	2	1	−1	−2	−3	−2	−1	2	0	3

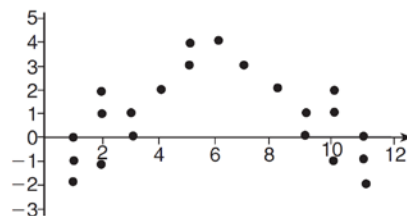Plot these residuals on the set of axes below.

Using the plot, assess the fit of the line for these residuals and justify your answer.

47) The residual plots from two different sets of bivariate data are graphed below.
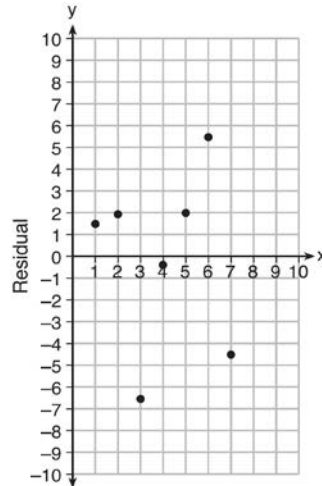

Graph A

Graph B

Explain, using evidence from graph A and graph B, which graph indicates that the model for the data is a good fit.
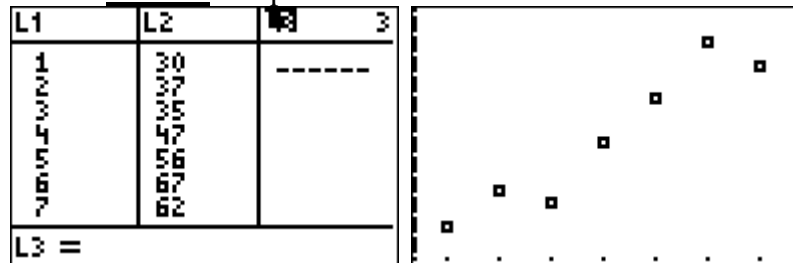
SOLUTIONS

43) ANS:
$y = 6.32x + 22.43$

Based on the residual plot, the equation is a good fit for the data because the residual values are scattered without a pattern and are fairly evenly distributed above and below the *x*-axis.
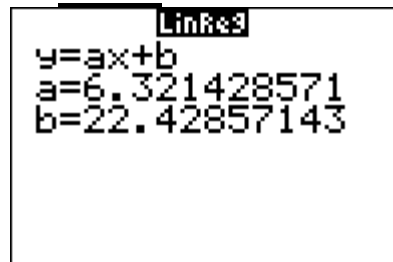
Strategies:
Use linear regression to find a regression equation that fits the first two columns of the table, then create a residuals plot using the first and third columns of the table to see if there is a pattern in the residuals.

- **STEP 1**. Input the data from the first two columns of the table into a graphing calculator.

- **STEP 2**. Determine which regression strategy will best fit the data. The problem states that the regression equation should be in the form $(y = ax + b)$, which means linear reression. The scatterplot produced by the graphing calculator also suggests linear regression.
- **STEP 3**. Execute the lionear regression strategy in the graphing calculator.

Round all values to the nearest hundredth: $y = 6.32x + 22.43$

- **STEP 4**. Plot the residual values on the graph provided using data from the first and third columns of the table. The graph shows a near equal number of points above the line and below the line, and the graph shows no pattern. The regression equation appears to be a good fit.

NOTE: The graphing calculator will also produce a residuals plot.

DIMS: Ask the question, "Does It Make Sense (DIMS)?" Yes. The regression equation produces the same residuals as shown in the table.

PTS: 4   NAT: S.ID.B.6b TOP: Correlation Coefficient and Residuals

44) ANS: 3

Strategy: Use knowledge of correlation coefficients and residual plots to determine which answer choice is **not** a good fit to model a data set.

STEP 1. A correlation coefficient close to –1 or 1 indicates a good fit, so answer choices a and b can be eliminated. Both suggest a good fit.

STEP 2. For a residual plot, there should be no observable pattern and a similar distribution of residuals above and below the *x*-axis. The residual plot in answer choice d shows a good fit, so answer choice d can be eliminated, leaving answer choice c as the correct answer.

DIMS? Does it make sense? Yes. The clear pattern in answer choice c tells us that the linear function is **not** a good fit to model the data set.
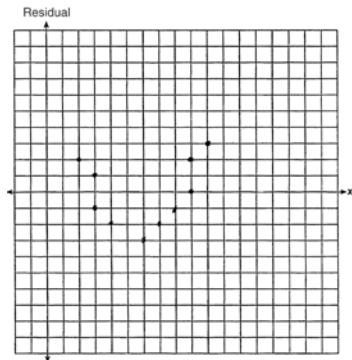
PTS: 2   NAT: S.ID.B.6b TOP: Correlation Coefficient and Residuals

45) ANS: 3

For a residual plot, there should be no observable pattern and about the same number of dots above and below the x axis. Any pattern in a residual plot means that line is **not** a good fit for the data.

PTS: 2   NAT: S.ID.B.6b TOP: Correlation Coefficient and Residuals

46) ANS:

The line is a poor fit because the residuals form a pattern.

PTS: 2   NAT: S.ID.B.6b TOP: Correlation Coefficient and Residuals

47) ANS:
Graph A is a good fit because it does not have a clear pattern, whereas Graph B does have a clear pattern..

PTS: 2   NAT: S.ID.6b   TOP: Correlation Coefficient and Residuals