# S.ID.B.6: Regression 1

1 The table below shows the number of grams of carbohydrates, $x$, and the number of Calories, $y$, of six different foods.

| Carbohydrates ($x$) | Calories ($y$) |
|---|---|
| 8 | 120 |
| 9.5 | 138 |
| 10 | 147 |
| 6 | 88 |
| 7 | 108 |
| 4 | 62 |

Which equation best represents the line of best fit for this set of data?

1) $y = 15x$
2) $y = 0.07x$
3) $y = 0.1x - 0.4$
4) $y = 14.1x + 5.8$

2 Emma recently purchased a new car. She decided to keep track of how many gallons of gas she used on five of her business trips. The results are shown in the table below.

| Miles Driven | Number of Gallons Used |
|---|---|
| 150 | 7 |
| 200 | 10 |
| 400 | 19 |
| 600 | 29 |
| 1000 | 51 |

Write the linear regression equation for these data where miles driven is the independent variable. (Round all values to the *nearest hundredth*.)

3 The accompanying table shows the enrollment of a preschool from 1980 through 2000. Write a linear regression equation to model the data in the table.

| Year ($x$) | Enrollment ($y$) |
|---|---|
| 1980 | 14 |
| 1985 | 20 |
| 1990 | 22 |
| 1995 | 28 |
| 2000 | 37 |

4 Omar has a piece of rope. He ties a knot in the rope and measures the new length of the rope. He then repeats this process several times. Some of the data collected are listed in the table below.

| **Number of Knots** | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|
| **Length of Rope** (cm) | 64 | 58 | 49 | 39 | 31 |

State, to the *nearest tenth*, the linear regression equation that approximates the length, $y$, of the rope after tying $x$ knots. Explain what the $y$-intercept means in the context of the problem. Explain what the slope means in the context of the problem.

5 In a mathematics class of ten students, the teacher wanted to determine how a homework grade influenced a student's performance on the subsequent test. The homework grade and subsequent test grade for each student are given in the accompanying table.

| Homework Grade ($x$) | Test Grade ($y$) |
|---|---|
| 94 | 98 |
| 95 | 94 |
| 92 | 95 |
| 87 | 89 |
| 82 | 85 |
| 80 | 78 |
| 75 | 73 |
| 65 | 67 |
| 50 | 45 |
| 20 | 40 |

*a* Give the equation of the linear regression line for this set of data.
*b* A new student comes to the class and earns a homework grade of 78. Based on the equation in part *a*, what grade would the teacher predict the student would receive on the subsequent test, to the *nearest integer*?

6 A real estate agent plans to compare the price of a cottage, $y$, in a town on the seashore to the number of blocks, $x$, the cottage is from the beach. The accompanying table shows a random sample of sales and location data. Write a linear regression equation that relates the price of a cottage to its distance from the beach. Use the equation to predict the price of a cottage, to the *nearest dollar*, located three blocks from the beach.

| Number of Blocks from the Beach ($x$) | Price of a Cottage ($y$) |
|---|---|
| 5 | $132,000 |
| 0 | $310,000 |
| 4 | $204,000 |
| 2 | $238,000 |
| 1 | $275,000 |
| 7 | $60,800 |

7 The accompanying table shows the percent of the adult population that married before age 25 in several different years. Using the year as the independent variable, find the linear regression equation. Round the regression coefficients to the *nearest hundredth*. Using the equation found above, estimate the percent of the adult population in the year 2009 that will marry before age 25, and round to the *nearest tenth of a percent*.

| Year ($x$) | Percent ($y$) |
|---|---|
| 1971 | 42.4 |
| 1976 | 37.4 |
| 1980 | 37.1 |
| 1984 | 34.1 |
| 1989 | 32.1 |
| 1993 | 28.8 |
| 1997 | 25.7 |
| 2000 | 25.5 |

8 The number of newly reported crime cases in a county in New York State is shown in the accompanying table. Write the linear regression equation that represents this set of data. (Let $x = 0$ represent 1999.) Using this equation, find the projected number of new cases for 2009, rounded to the *nearest whole number*.

| Year ($x$) | New Cases ($y$) |
|---|---|
| 1999 | 440 |
| 2000 | 457 |
| 2001 | 369 |
| 2002 | 351 |

9 The data table below shows water temperatures at various depths in an ocean.

| **Water Depth ($x$)** (meters) | **Temperature ($y$)** (°C) |
|---|---|
| 50 | 18 |
| 75 | 15 |
| 100 | 12 |
| 150 | 7 |
| 200 | 1 |

Write the linear regression equation for this set of data, rounding all values to the *nearest thousandth*. Using this equation, predict the temperature (ºC), to the *nearest integer*, at a water depth of 255 meters.

10 The mid-September statewide average gas prices, in dollars per gallon, ($y$), for the years since 2000, ($x$), are given in the table below.

| **Year Since 2000** ($x$) | **Price Per Gallon** ($y$) |
|---|---|
| 1 | 1.345 |
| 2 | 1.408 |
| 3 | 1.537 |
| 4 | 1.58 |

Write a linear regression equation for this set of data. Using this equation, determine how much *more* the actual 2005 gas price was than the predicted gas price if the actual mid-September gas price for the year 2005 was $2.956.

11 A factory is producing and stockpiling metal sheets to be shipped to an automobile manufacturing plant. The factory ships only when there is a minimum of 2,050 sheets in stock. The accompanying table shows the day, $x$, and the number of sheets in stock, $f(x)$.

| Day $(x)$ | Sheets in Stock $(f(x))$ |
|---|---|
| 1 | 860 |
| 2 | 930 |
| 3 | 1000 |
| 4 | 1150 |
| 5 | 1200 |
| 6 | 1360 |

Write the linear regression equation for this set of data, rounding the coefficients to *four decimal places*. Use this equation to determine the day the sheets will be shipped.

12 The data table below shows the median diameter of grains of sand and the slope of the beach for 9 naturally occurring ocean beaches.

| Median Diameter of Grains of Sand, in Millimeters (x) | 0.17 | 0.19 | 0.22 | 0.235 | 0.235 | 0.3 | 0.35 | 0.42 | 0.85 |
|---|---|---|---|---|---|---|---|---|---|
| Slope of Beach, in Degrees (y) | 0.63 | 0.7 | 0.82 | 0.88 | 1.15 | 1.5 | 4.4 | 7.3 | 11.3 |

Write the linear regression equation for this set of data, rounding all values to the *nearest thousandth*. Using this equation, predict the slope of a beach, to the *nearest tenth of a degree*, on a beach with grains of sand having a median diameter of 0.65 mm.

13 The availability of leaded gasoline in New York State is decreasing, as shown in the accompanying table.

| Year | 1984 | 1988 | 1992 | 1996 | 2000 |
|---|---|---|---|---|---|
| Gallons Available (in thousands) | 150 | 124 | 104 | 76 | 50 |

Determine a linear relationship for $x$ (years) versus $y$ (gallons available), based on the data given. The data should be entered using the year and gallons available (in thousands), such as (1984, 150). If this relationship continues, determine the number of gallons of leaded gasoline available in New York State in the year 2005. If this relationship continues, during what year will leaded gasoline first become unavailable in New York State?

14 The accompanying table illustrates the number of movie theaters showing a popular film and the film's weekly gross earnings, in millions of dollars.

| Number of Theaters (x) | Gross Earnings (y) (millions of dollars) |
|---|---|
| 443 | 2.57 |
| 455 | 2.65 |
| 493 | 3.73 |
| 530 | 4.05 |
| 569 | 4.76 |
| 657 | 4.76 |
| 723 | 5.15 |
| 1,064 | 9.35 |

Write the linear regression equation for this set of data, rounding values to *five decimal places*. Using this linear regression equation, find the approximate gross earnings, in millions of dollars, generated by 610 theaters. Round your answer to *two decimal places*. Find the minimum number of theaters that would generate at least 7.65 million dollars in gross earnings in one week.

15 Since 1990, fireworks usage nationwide has grown, as shown in the accompanying table, where $t$ represents the number of years since 1990, and $p$ represents the fireworks usage per year, in millions of pounds.

| Number of Years Since 1990 (t) | Fireworks Usage per Year, in Millions of Pounds (p) |
|---|---|
| 0 | 67.6 |
| 2 | 88.8 |
| 4 | 119.0 |
| 6 | 120.1 |
| 7 | 132.5 |
| 8 | 118.3 |
| 9 | 159.2 |
| 11 | 161.6 |

Find the equation of the linear regression model for this set of data, where $t$ is the independent variable. Round values to *four decimal places.* Using this equation, determine in what year fireworks usage would have reached 99 million pounds. Based on this linear model, how many millions of pounds of fireworks would be used in the year 2008? Round your answer to the *nearest tenth*.

16 The 1999 win-loss statistics for the American League East baseball teams on a particular date is shown in the accompanying chart.

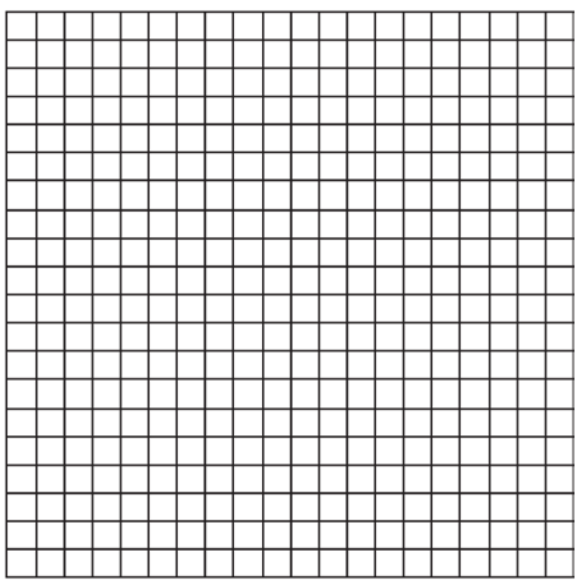
| | W | L |
|---|---|---|
| New York | 52 | 34 |
| Boston | 49 | 39 |
| Toronto | 47 | 43 |
| Tampa Bay | 39 | 49 |
| Baltimore | 36 | 51 |

Find the mean for the number of wins, $\overline{W}$, and the mean for the number of losses, $\overline{L}$, and determine if the point $(\overline{W},\overline{L})$ is a point on the line of best fit. Justify your answer.

17 The table below shows the results of an experiment that relates the height at which a ball is dropped, $x$, to the height of its first bounce, $y$.

| Drop Height ($x$) (cm) | Bounce Height ($y$) (cm) |
|---|---|
| 100 | 26 |
| 90 | 23 |
| 80 | 21 |
| 70 | 18 |
| 60 | 16 |

Find $\overline{x}$, the mean of the drop heights. Find $\overline{y}$, the mean of the bounce heights. Find the linear regression equation that best fits the data. Show that $(\overline{x},\overline{y})$ is a point on the line of regression. [The use of the grid is optional.]



# S.ID.B.6: Regression 1
# Answer Section

1 ANS: 4 REF: 081421ai

2 ANS:
$y = 0.05x - 0.92$

REF: fall1307ai

3 ANS:
$y = 1.08x - 2125$

REF: 060722b

4 ANS:
$y = -8.5x + 99.2$ The $y$-intercept represents the length of the rope without knots. The slope represents the decrease in the length of the rope for each knot.

REF: 011834ai

5 ANS:
$y = 0.8344648562x + 14.64960064$, 80. $0.8344648562(78) + 14.64960064 \approx 80$

REF: 010328b

6 ANS:
$y = -34739.71292x + 313309.0909$, 209,090. $-34739.71292(3) + 313309.0909 = 209{,}090$.

REF: 010530b

7 ANS:
$y = -0.58x + 1185.09$, 19.9. $y = -0.58(2009) + 1185.09 = 19.9$

REF: 080728b

8 ANS:
$y = -35.5x + 457.5$, 103. $-35.5(10) + 457.5 \approx 103$

REF: 060927b

9 ANS:
$y = -0.112x + 23.448$. $-0.112(255) + 23.448 \approx -5$

REF: 061027b

10 ANS:
$y = 0.0834x + 1.259$, 1.28

REF: 011028b

11 ANS:
$y = 98.8571x + 737.3333$, 14. $2050 = 98.8571x + 737.3333$.

$$x = 14$$

REF: 060631b

12 ANS:
$y = 17.159x - 2.476$. $y = 17.159(.65) - 2.476 \approx 8.7$

REF: 081633ai

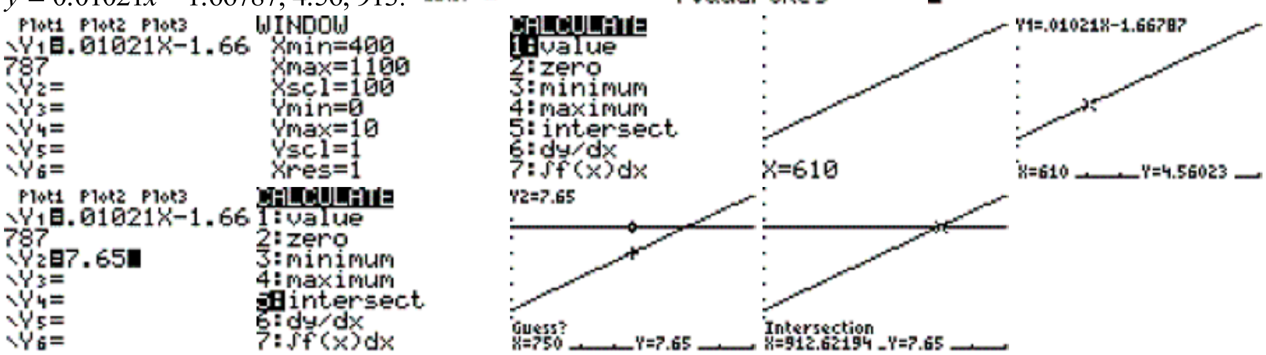
13 ANS:
$y = -6.2x + 12451.2$, 20.2 thousand, 2008. $-6.2(2005) + 12451.2 = 20.2$. $0 = 6.2x + 12451.2$.
$x = 2008$

REF: 080133b

14 ANS:



$y = 0.01021x - 1.66787$, 4.56, 913.




REF: 080533b

15 ANS:
$p = 8.1875t + 72.7860$, 1993, 220.2. $99 = 8.1875t + 72.7860$. $p = 8.1875(18) + 72.7860 = 220.2$
$t = 3.2017$ years since 1990, or 1993

REF: 010633b

16 ANS:
$\overline{W} = 44.6$ and $\overline{L} = 43.2$. If the equation of the line of best fit is $y = -1.007559x + 88.137149$,
$-1.007559(44.6) + 88.137149 = 43.2$, so $(\overline{W}, \overline{L})$ is a point on the line of best fit.

REF: 060134b

17 ANS:
$\overline{x} = 80$, $\overline{y} = 20.8$, and $y = 0.25x + 0.8$, $0.25(80) + 0.8 = 20.8$, so $(\overline{x}, \overline{y})$ is a point on the line of regression.

REF: 080331b