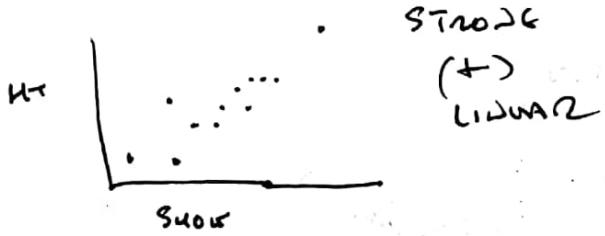
Math One Linear Regression Review

1. A fashion designer thought there might be a relationship between the length of a woman's foot and her height. To test this theory, she collected data from 12 women.

| Shoe Length (in inches) | Height (in inches) |
|---|---|
| 8.9 | 61 |
| 9.6 | 61 |
| 9.8 | 66 |
| 10.0 | 64 |
| 10.2 | 64 |
| 10.4 | 65 |
| 10.6 | 65 |
| 10.6 | 67 |
| 10.5 | 66 |
| 10.8 | 67 |
| 11.0 | 67 |
| 11.8 | 70 |

(a) Sketch a scatterplot of the data.




(b) Write the equation of the line of best fit that models this data. Round to the tenths place.

$\hat{y} = 3.1x + 32.7$

$\hat{y}$ = PRED. HT IN INCHES
$x$ = SHOE LENGTH.

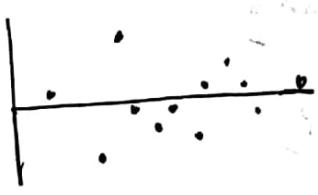
(c) The designer wants to predict the height of a woman whose shoe length is 7.4 inches. What height does the equation predict?

$\hat{y}(7.4) = 55.966$ IN.

(d) What does the equation predict for the shoe length of a woman who is 64 inches tall?

64 → 9 61.002
(10) 64.149

(e) Create a residual plot for the data. What does it tell you about the data?



SCATTERED/RANDOM → GOOD MODEL.

2. The table gives the average hotel rate from 1996 to 2006.

If you plan to stay at a hotel in 2016, what is the rate you would expect to pay?

Equation_____ Rate Expected_____

| Year | Rate ($) |
|---|---|
| 1996 | 70.63 |
| 1997 | 75.31 |
| 1998 | 78.62 |
| 1999 | 81.33 |
| 2000 | 85.89 |
| 2001 | 88.27 |
| 2002 | 83.54 |
| 2003 | 82.52 |
| 2004 | 86.23 |
| 2005 | 90.88 |
| 2006 | 97.78 |

$\hat{y} = 2x - 3934.5$

$113.85

(EXTRAP)

$\hat{y}$ = PREDICTED RATE

$x$ = YR

3. A convenience store manager notices that sales of soft drinks are higher on hotter days, so he assembles the data in the table.

If the high temperature is 91 degrees, what is the predicted number of soft drinks sold?

Equation_____ Expected Cans Sold_____

| High Temp | # of Cans Sold |
|---|---|
| 55 | 340 |
| 58 | 335 |
| 64 | 410 |
| 68 | 460 |
| 70 | 450 |
| 75 | 610 |
| 80 | 735 |
| 84 | 780 |

$\hat{y} = 16.4x - 621.8$

$\hat{y}(91) =$ 872 CANS

EXTRAP

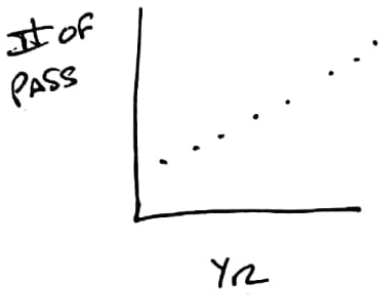
$\hat{y}$ = EXPECTED # OF CANS

$x$ = HIGH TEMP.

4. The accompanying table illustrates the number of passengers (in millions) on ABC Airlines.

| | Year | Number of Passengers |
|---|---|---|
| 0 | 2000 | 30.1 |
| 1 | 2001 | 33.4 |
| 2 | 2002 | 36.2 |
| 3 | 2003 | 39.5 |
| 4 | 2004 | 42.0 |
| 5 | 2005 | 45.8 |
| 6 | 2006 | 48.9 |
| 7 | 2007 | 52.6 |
| 8 | 2008 | 55.3 |
| 9 | 2009 | 58.5 |

(a) Make a scatter plot of the data on your calculator and sketch the graph below. Make sure to label the axes. (Let x=0 represent the year 2000)




(b) Write the linear regression equation for this set of data, rounding values to *five decimal places.*

$\hat{y} = 3.16909x + 29.96909$

$\hat{y}$ = PRED # OF PASS

$x$ = YR.

(c) List the correlation coefficient and explain what it means.

$r = 0.999$ HIGHLY LINEAR DATA.

(d) Using this linear regression equation, find the approximate number of passengers, in millions, in the year 2015. Round to the tenths place.

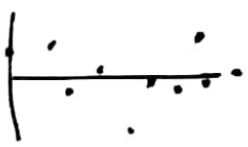
$\hat{y}(15) = 77.5$ PASSENGERS

(e) In what year would the number of passengers exceed 70 million?

70 → 71.167 @ START OF 2013

2012

(f) Create a residual plot and tell what it means about the data.



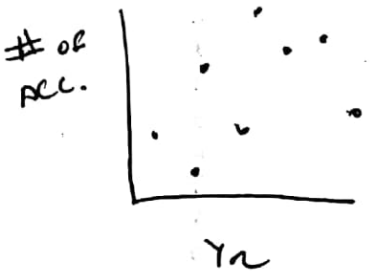
FAIRLY RANDOM + SCATTERED

∴ GOOD MODEL.

5. The table shows the number of accidents in the NASCAR Sprint Cup races from 2001 to 2008.

| | Years | # of Accidents |
|---|---|---|
| 1 | 2001 | 200 |
| 2 | 2002 | 186 |
| 3 | 2003 | 235 |
| 4 | 2004 | 204 |
| 5 | 2005 | 253 |
| 6 | 2006 | 237 |
| 7 | 2007 | 240 |
| 8 | 2008 | 211 |

(a) Sketch a scatterplot of the data. (Let x=0 represent 2000)




WEAK (+) LINEAR

(b) Write the equation for the line of best fit. Round to the tenths place. (remember that x=0 represents 2000)

$\hat{y} = 4.8x + 199.2$ $\hat{y}$ = PRED # OF ACC. x = YR.

(c) How many accidents does the equation predict for the year 2005? How does this compare to the actual number of accidents in 2005?

$\hat{y}(5) = 223.1$
$y(5) = 253$
} UNDERPREDICT

(d) What is the slope of the line? Interpret the meaning of the slope.

SLOPE = 4.8 PREDICT FOR EACH YR, 4.8 ADD'L ACCIDENTS.

(e) What is the y-intercept of the line? Interpret the meaning of the y-intercept.

199.2 IN 2000, PREDICTED 199 ACCIDENTS,

(f) What is the residual value for the year 2003? How did you figure this out?

$\hat{y}(3) = 213.571$

$y(3) = 235$

RESID = A − P = 235 − 213 = 21.4 ACC's.

Name: _____ Date: _____

1. The relationship of a woman's shoe size and length of a woman's foot, in inches, is given in the accompanying table.

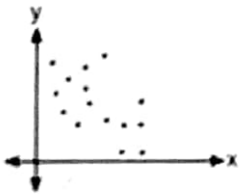
| Women's Shoe Size | 5 | 6 | 7 | 8 |
|---|---|---|---|---|
| Foot Length (in) | 9.00 | 9.25 | 9.50 | 9.75 |

The linear correlation coefficient for this relationship is

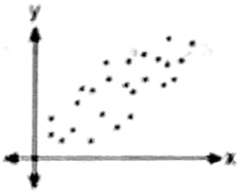
A 1 B −1 C 0.5 D 0

2. Which graph represents data used in a linear regression that produces a correlation coefficient closest to −1?

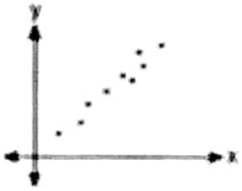
A



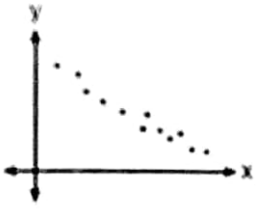

B




C




D






13

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 |

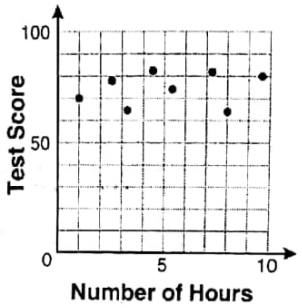
$\hat{y} = 1.08 - 2125$

$\hat{y}$ = PROJ. ENROLLMENT

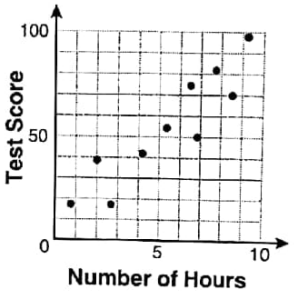
$x$ = YR.

4. There is a negative correlation between the number of hours a student watches television and his or her social studies test score. Which scatter plot below displays this correlation?

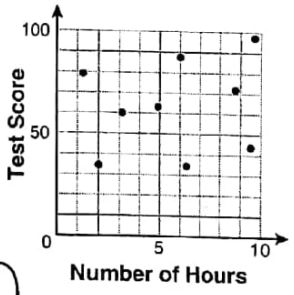
A.



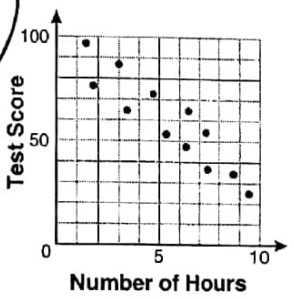

B.




C.

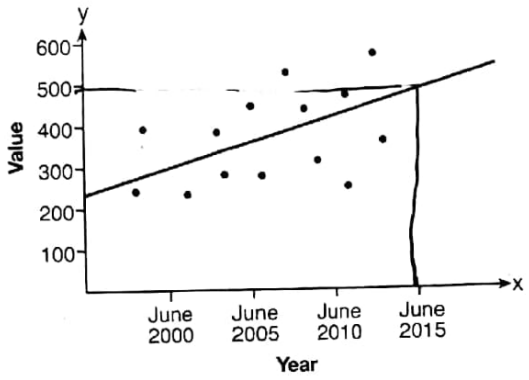



D.




(14)

5. Based on the line of best fit drawn below, which value could be expected for the data in June 2015?




A. 230 B. 310 C. 480 D. 540