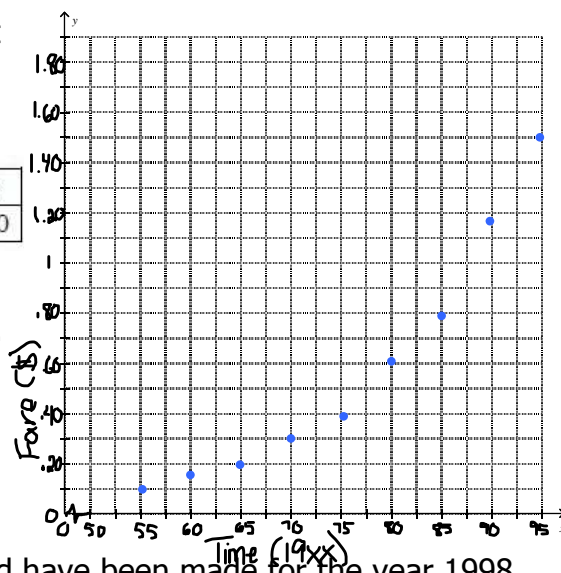Name _____ Class _____

**1** The table below, created in 1996, shows a history of transit fares from 1955 to 1995. On the grid below, construct a scatter plot where the independent variable is years.

| Year (19xx) | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 |
|---|---|---|---|---|---|---|---|---|---|
| Fare ($) | 0.10 | 0.15 | 0.20 | 0.30 | 0.40 | 0.60 | 0.80 | 1.15 | 1.50 |

**a** State the exponential regression equation with the coefficient and base rounded to the *nearest thousandth*.

$$f(x) = 0.002(1.070)^x$$

**b** Using this equation, determine the prediction that should have been made for the year 1998, to the *nearest cent*.

$$f(98) = 0.002(1.070)^{98} \approx \$1.52$$

**c** Calculate the correlation coefficient and explain how accurate this function is in predicting a transit fare.

$r \approx 0.999$

There is a strong, positive, exponential relationship between time and transit fare prices.

**2** A linear regression was done to relate the age of a person, $x$, the number of optimal hours of sleep, $y$. The equation was found to be $y = -0.12x + 11.5$, with a correlation coefficient of $r = -0.86$

**a** Explain what each of the parameters in the linear model represent, i.e. the -0.12 and the 11.5.

$$m = \frac{\Delta y}{\Delta x} = \frac{-0.12 \text{ hrs}}{1 \text{ yr.}}$$

A person starts off by sleeping 11.5 hrs/day and each year he/she gets older, he/she sleeps 0.12 hours less per day.

**b** How do you interpret the fact that the correlation coefficient is negative?

The slope is negative as well. Indicating a negative linear relationship between age and hours of sleep/day.

**3** The table shows the average sale price $p$ of a house in New York City, for various years t since 1960.

| Years since 1960, $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Average sale price (in thousands of dollars), $p$ | 45 | 36 | 29 | 24 | 21 | 20 | 21 |

**a** What type of function most appropriately represents this set of data? Explain your reasoning.

*Quadratic*
*The y-values are symmetrical with a vertex and the second difference is constant (2)*

**b** In what year is the price at the lowest? Explain how you know.

*1965 — The vertex occurs at (5,20) and the time is 1960+t, or in this case 1960+5 = 1965.*

**c** Write a function to represent the data. Show your work.

*Parent function: $p(x)=x^2$*
*You know vertex is (5,20), so it shifted right 5 and up 20. So we know:*
*$f(x) = a(x-5)^2 + 20$*
*Solve for a by plugging in a coordinate (1,36)*
*$36 = a(1-5)^2+20$*
*$36 = a(-4)^2+20$*
*$36 = 16a+20$*
*$16=16a$*
*$1=a$*
*Therefore:*
*$f(x) = (x-5)^2+20$*

**d** Can this function ever be equal to zero? Explain why or why not.

*No, the vertex is (5,20) and since a is positive it opens up and therefore doesn't go below (5,20). 0 is less than 20, so it never reaches zero.*

**e** Mr. Samuels bought his house in New York City in 1970. If the trend continued, how much was he likely to have paid? Explain and provide mathematical evidence to support your answer.

*$f(x) = (x-5)^2+20 \rightarrow X = 10$ (1970-1960=10)*
*$f(10) = (10-5)^2+20$*
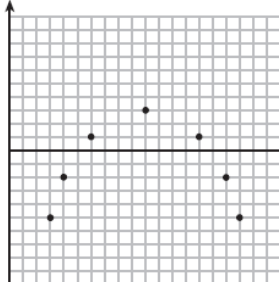*$f(10) = 5^2+20 \rightarrow f(10) = 25+20 = 45$*
*$\$45,000$*
*You can also use the symmetry of the chart by extending to t+6*

**4** Which statistic would indicate that a linear function would *not* be a good fit to model a data set?
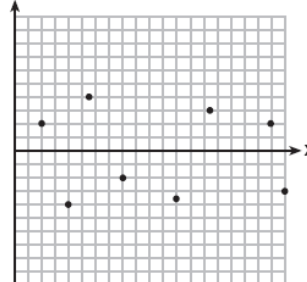
**(1)** $r = -0.93$ *strong, negative correlation*

**(2)** Residual

**(3)** $r = 1$ *perfect, positive correlation*

**(4)** Residual

*You can tell a function is a good fit if the residual plot is scattered. This has a pattern.*

**5** Weather data were recorded for a sample of 25 American cities in one year. Variables measured included January high temperature (in degrees Fahrenheit), January low temperature, annual precipitation (in inches), and annual snow accumulation. The relationships for three pairs of variables are shown in the graphs below (Jan Low Temperature – Graph A; Precipitation – Graph B; Annual Snow Accumulation – Graph C).
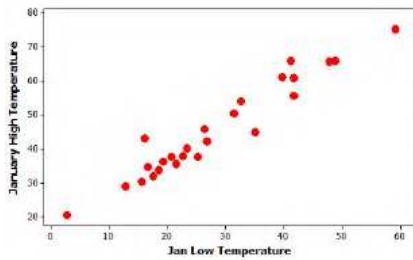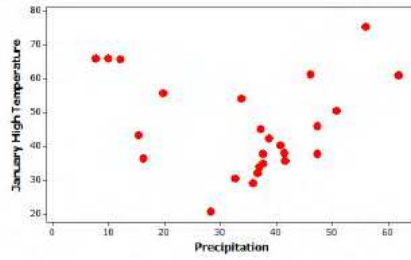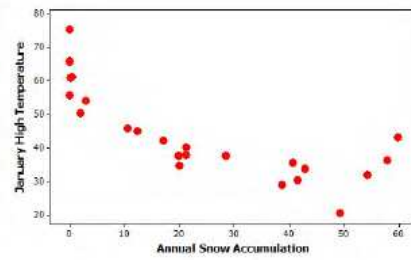


Graph A                    Graph B                    Graph C

**a** Which pair of variables will have a correlation coefficient closest to 0? *(M2:EM#4 – AU5)*

A. Jan high temperature and Jan low temperature

(B.) Jan high temperature and Precipitation

C. Jan high temperature and Snow accumulation

Explain your choice:

*A correlation coefficient of 0 means no relationship between the variables. Choice B has the weakest relationship between the 2 variables*

**b** Which of the above scatterplots would be best described as a strong nonlinear relationship?

Explain your choice:

*Graph C - it shows there is a relationship between the variables but the relationship is curved.*

**6** A local babysitting group takes care of 8 children on Saturday morning. The children's ages are shown below.

$$4, 7, 6, 8, 4, 5, 6, 4$$

Which of the following is the mean age of this group of children?

(1) 5.5          (2) 5.9          (3) 6.4          (4) 6.7

**7** Which of the following is the interquartile range of the data set shown in the box plot below?



$Q_3 - Q_1 =$

$80 - 20 = \boxed{60}$

**8** Which of the following data sets would have a standard deviation closest to zero?

(1) {1, 3, 4, 6, 9}          (3) {6, 6, 6, 7, 7} *least variability*

(2) {15, 18, 20, 25}          (4) {1, 1, 3, 3, 7}

**9** A sample of ten large eggs were weighed with the results shown below, in grams.

58, 64, 61, 56, 60, 57, 60, 62, 59, 68

Which of the following is this sample's standard deviation to the nearest tenth of a gram?

(1) 2.7 ~~(crossed out)~~

(3) 4.8

(2) 3.5 *(circled)*

(4) 5.1

*Handwritten: • Plug into lists & spreadsheets*
*• One-variable statistics*
*• Sx = 3.5*

**10** A survey of 52 graduating seniors was conducted to determine if there was a connection between the gender of the student and whether they were going on to college. Which of the following represents the proportion of those going to college who are male?

(1) 0.53

(3) 0.55 *(circled)*   *Handwritten: $\frac{16}{29} = 0.55$*

(2) 0.31

(4) 0.16

|  | Gender | | |
|---|---|---|---|
|  | Male | Female | Total |
| Going to College | 16 | 13 | 29 |
| Not Going to College | 14 | 9 | 23 |
| Total | 30 | 22 | 52 |

**11** Which of the following correlation coefficient values below indicates the greatest correlation between the two variables?
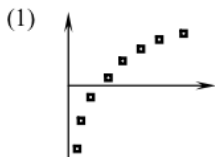
(1) $r = 0.78$

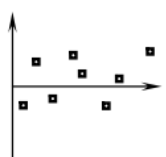(3) $r = 0.12$

(2) $r = -0.54$

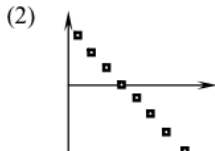(4) $r = -0.97$ *(circled)*   *Handwritten: Closest to 1 or -1*

**12** Four different types of regression were used on a data set. Which of the following four residual plots shows the model that is the most appropriate to use?
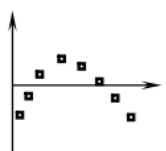
(1)

(3) *(circled)*

*Handwritten: A scattered residual plot indicates a good fit*

(2)

(4)

**13** Economists create a linear regression equation to predict the price of a gallon of gasoline, $y$, based on the price of a barrel of oil, $x$. The equation they find is $y = 0.035x + 0.95$. One of the data points they use is $(60, 3.62)$. What is the residual for this data point?

(1) 0.57 *(circled)*

(3) 0.72   *Handwritten: residual = actual - predicted*

(2) 0.68

(4) 0.84   *Handwritten: predicted: $y = 0.035(60) + 0.95 = 3.05$*
*residual = 3.62 - 3.05 = 0.57*

**14** Which correlation coefficient is most appropriate for the scatterplot shown below?

(1) $r = -1.00$   *Handwritten: There is a strong (not perfect),*
(2) $r = -0.92$ *(circled)*   *negative correlation between*
(3) $r = 0.78$   *the 2 variables.*
(4) $r = 1.00$