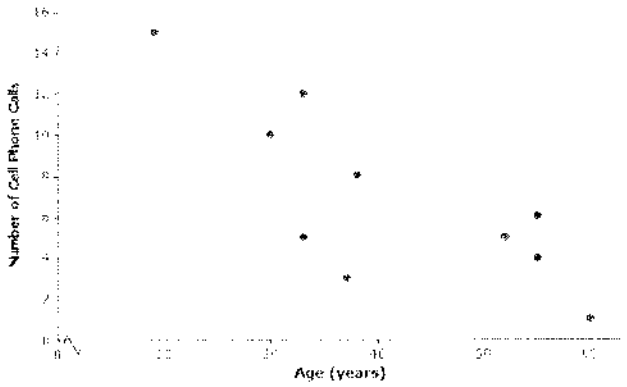
Practice for Quiz 3 Statistics **Name:** Answers

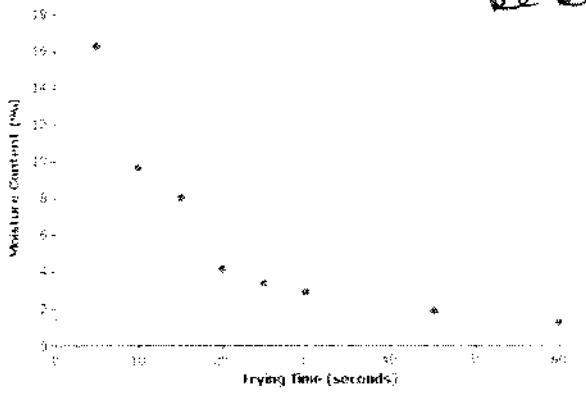
Algebra 1 – Statistics **Quiz next class will be about 25-30 points!**

1. Consider the five scatter plots below for the following questions.
   a) Which plot(s) most likely show a linear relationship? 1 and 5 & perhaps 3 (but weak)
   b) Of these linear relationships, which has a positive correlation? 5 & maybe 3
   c) Estimate the correlation coefficient (r) for scatter plot 1 $r \approx -.75$ or so
   d) Which plot(s) most likely show a quadratic relationship? 4
   e) Which plot(s) most likely show an exponential relationship? 2
   f) If the residual plot for scatter plot 1 had no pattern, what would that tell you? linear is best model
   g) If instead the residual plot for scatter plot 1 had a U-shaped pattern, what would that tell you? A non-linear model would be better.

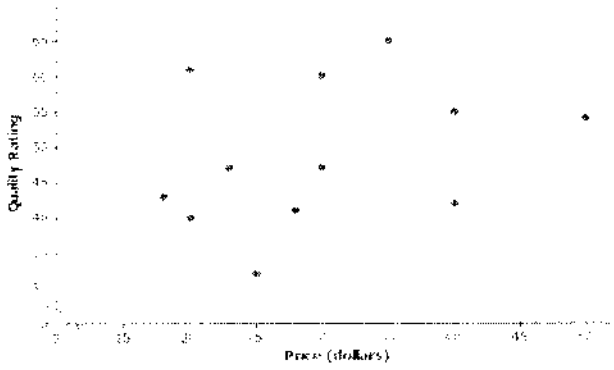
Scatter Plot 1:



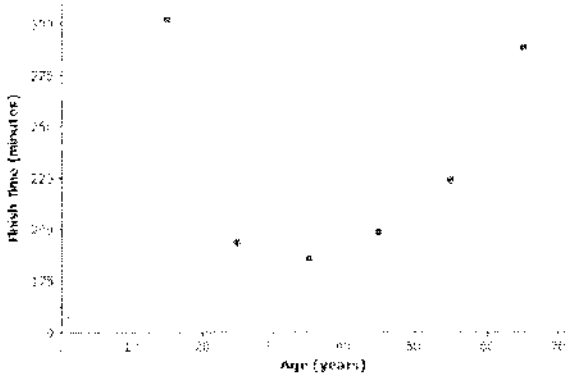
Scatter Plot 2:



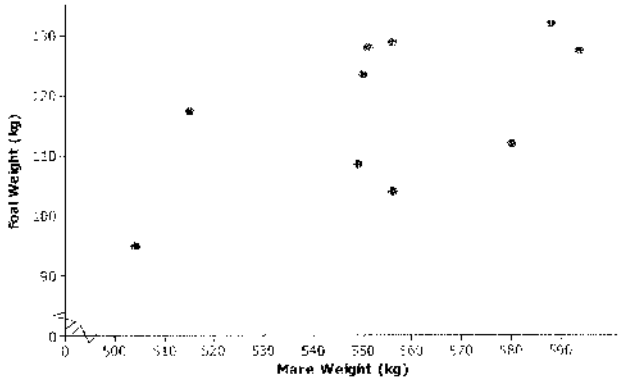
Scatter Plot 3:



Scatter Plot 4:



Scatter Plot 5:



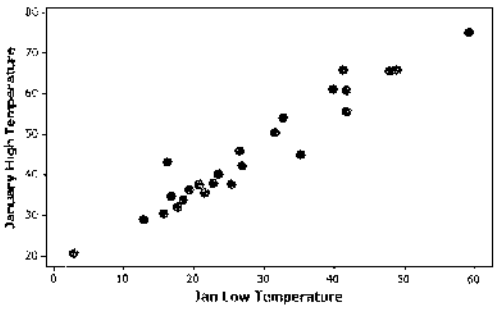
See possible example answers below

2. Choose one relationship shown above. State the variables that are involved and explain what the scatter plot tells us about the relationship between those two variables. Does that relationship seem to make sense given what you know about the world?

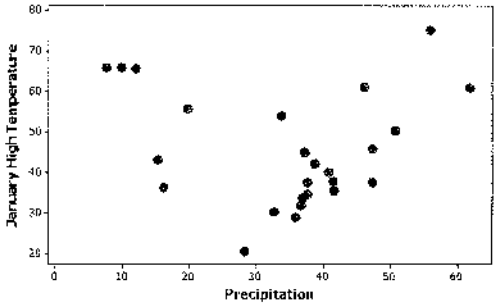
#1 The two variables are age and # cell phone calls. As age increases, fewer cell calls are made. Young people use cells more!

#5 The two variables are foal weight (baby horse) & mare's weight (Mom!). As mare's weight goes up so does foal's. Yes, this makes sense, since bigger horses have larger babies.

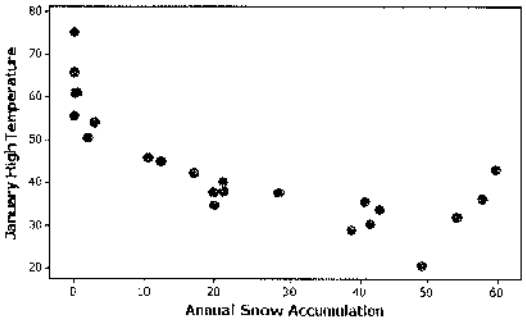
3. Weather data were recorded for a sample of 25 American cities in one year. Variables measured included January high temperature (in degrees Fahrenheit), January low temperature, annual precipitation (in inches), and annual snow accumulation. The relationships for three pairs of variables are shown 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?
1) Jan high temperature and Jan low temperature
2) Jan high temperature and Precipitation
3) Jan high temperature and Snow accumulation

Graph B seems to have no (linear or other) relationship. Graph B involves the variables precipitation (x) and January High Temps (y)

Explain your choice:

b. Which of the above scatterplots would be best described as a strong nonlinear relationship? Explain your choice.

Graph C, which shows the relationship between snow + Jan high temps, shows a strong non-linear relationship. It's curve may be modeled best by a quadratic regression

c. For the city with a January low temperature of 30°F, what do you predict for the annual snow accumulation? Explain how you are estimating this based on the three graphs above.

See graph A: A city with low temp (x) of 30° has a high temp (y) of about 40°. Now look at graph C. Find 40° on the y + see what x matches it... About 18 inches of snow seems right.

4. Suppose times (in minutes) to run one mile were recorded for a sample of 100 runners. ages 16-66 years and the following least squares regression line was found:
Predicted time in minutes to run one mile = 5.35 + 0.25 x (*age*)

a. Using this least squares line, predict the time it would take for a 50 year old to run one mile.

time = 5.35 + .25(50) = 17.85 min.

b. Let's say that one of the runners in the study was 50 years old and had a time of 18.5 minutes. Use your findings from part a to find this runner's residual.

residual = actual − predicted
= 18.5 − 17.85 = .65

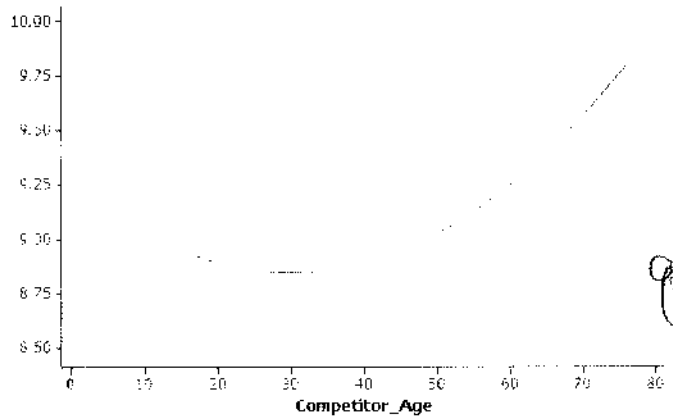
c. Explain what it would mean in the context of this study for a runner to have a negative residual.

If a runner has a negative residual, that means that their actual time was less than the predicted time. actual − predicted = neg. #

d. When you find a least squares line and find the residuals, **the sum of the residuals is always equal to zero**. Explain what this means and why this makes sense.

The residuals are the differences between all the predicted times & actual times. The line of best fit is the line that's closest to all the data points... Some points are above it & some below it, so some residuals are + and some − when they add together the "cancel each other out" to equal 0.

e. Suppose instead someone suggests using the following curve to predict time to run one mile. Explain what this model implies about the relationship between running time and age, and why that relationship might make sense in this context.



This model would show a quadratic relationship i.e. a quadratic model would work best. this might make sense. A young person around 8 years would take more time to run a mile than a 25 year old. And then a 50 year old would again take more time than a 25 year old. I

5. Blast from the past problems will be on this quiz! **Solve each equation for x.**

a) $2x+5=7x-30$

$-2x \quad -2x$

$5 = 5x - 30$

$+30 \quad +30$

$35 = 5x$

$\frac{35}{5} = x$

$\boxed{7 = x}$

b) $3(x+2)-2x=9-5x+3-2x$

$3x+6-2x = 12-7x$

$x+6 = 12-7x$

$+7x \quad +7x$

$8x+6 = 12$

$-6 \quad -6$

$8x = 6$

$x = \frac{6}{8}$

$\boxed{x = \frac{3}{4}}$

c) $\frac{2}{3}x+3=-11$

$-3 \quad -3$

$\frac{3}{2} \cdot \frac{2}{3}x = -14 \cdot \frac{3}{2}$

$\boxed{x = -21}$

d) $\frac{x}{3}+2=4x-1$

$-2 \quad -2$

$3 \cdot \frac{x}{3} = (4x-3) \cdot 3$

$x = 12x - 9$

$-12x \quad -12x$

$-11x = -9$

$x = \frac{-9}{-11}$

$x = \frac{9}{11}$

Make sure you look back at the practice sheets for quiz 1 and quiz 2 in statistics because there just may be a box plot or dot plot or histogram on this quiz as well!