Correlation/Scatterplot Activity

Go to: http://stat.istics.net/

A. Choose "Guessing Correlation" * Note: It is not set up for competition mode.

Match the four correlation value to the appropriate scatter plots, and press "Check Answers"

Run the simulation 10 times. If after 5 simulations you are not getting most of them correct, please let me know so I can help.

Record your final score here: _____ /40 _____

B. Scroll to the bottom and choose "Data Program"

For **each of the four set of data** listed below:

1. "Choose a data set" from the drop-down list. It will show a summary of the data. If you want to see the actual data table you can select "Display/Modify Data".
2. Select "Scatter Plots". Choose an appropriate X and Y axis category, paying attention to which variable is the response (y) variable and which is the explanatory (x) variable.
3. On the back, sketch the scatterplot, and write down the equation and the r value.
4. Explain **in context** what the scatterplot and r value are telling you in terms of form (linear?), direction (negative or positive?), and strength (weak, moderate, strong) of the correlation.
5. For any one of the data sets, switch the x and y axis and describe what happens to the r-value.

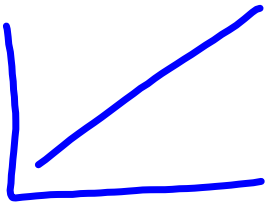
the correlation coeff. (r) stays the same

The data sets are as follows:

Cricket Chirps 2007, IBM Stock Prices, NBA All Stars, Athletes Heights and Weights

1. Cricket Chirp 2007

Sketch:



Equation: $C = -42 + 1.077T$

r-value: .980

Explanation:

Strong (+) linear

As Temp ↑, # of chirps ↑

2. IBM Stock Prices

Sketch:

Equation: $P = 45.45t + .54w$

r-value: .946

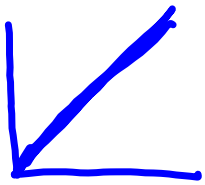
Explanation:

Strong/mod (+) linear

As time ↑, stock $ ↑

3. NBA Allstars

Sketch:



Equation: cm = 2.54(in)

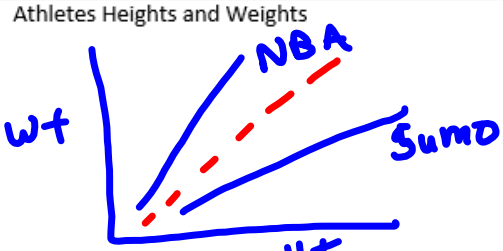
r-value: 1.000

Explanation: Perfect, +, linear

As Int, cm ↑

4. Athletes Heights and Weights

Sketch:



Equation: W = 250.7 − 1.454H

r-value: −.01163

Explanation:

Very weak,

2 different populations gives 2 different lines

If separate, mod-strong linear, (+), corr for both.

C. In "Data Program" again, select the data set "Survey 1". Choose "Modify/Display Data". Play around with the scatterplots by choosing different pairs of variables to plot. Which pair of variables show the strongest correlation and what is that r-value?
* Remember that the variables must both be quantitative to have a correlation.

Age of mother vs. Age of father
r = .766

D. The Monty Hall Problem is a classic probability problem. It doesn't have anything to do with what we are doing in class right now, but there is a fun simulator on this website, so let's try it!

The basic idea is this: In this TV game show, there are three doors for the contestant to choose from. Behind one door is a new car, and behind the other two doors are goats. The contestant chooses one of the three doors. The game show host (Monty Hall) peeks behind the remaining two doors and opens a door to reveal a goat (there will always be at least one door left with a goat). The contestant now has the choice to either stick with their original choice, or switch to the remaining un-opened door. Assuming they want to win a car and not a goat, what should they do?

Go to: http://istics.net/MontyHall/ (or choose the Monty Hall from the home page).

Run the simulation 100 time for each option.

Choose any door to open first and stick with that door 100 times.

Won _____% of the time.

Choose any door to open first and change to the other door 100 times.

Won _____% of the time.

Is that what you expected?????? _____

Monty Hall Solution Explained:

https://betterexplained.com/articles/understanding-the-monty-hall-problem/