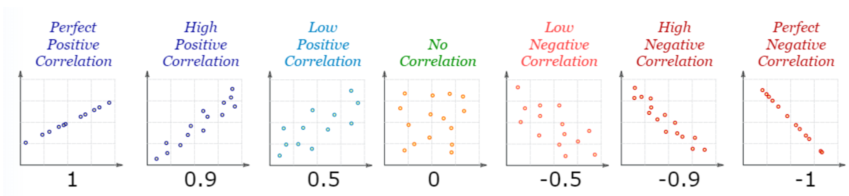
Pier Data Activity

How does salinity relate to depth?

1. Produce a scatterplot of salinity (y-axis) vs. depth (x-axis). Describe the direction (positive or negative), strength (weak, moderate, strong) and type (linear, quadratic, log, exponential, etc.) of the relationship you observe between salinity and depth.

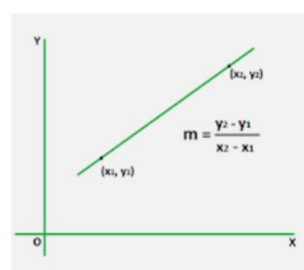
2. Determine the correlation between salinity and depth and interpret this number.

Definition*: The* ***correlation coefficient****, r, is a measure of the strength of the linear relationship between two variables, x and y. The correlation can be any value between -1 and 1, inclusive,* $-1 \leq r \leq 1$*. The closer the correlation is to -1 or 1, the stronger the correlation. The sign on the correlation indicates the direction of the relationship. Below are some examples of scatterplots and their respective correlation coefficients.*

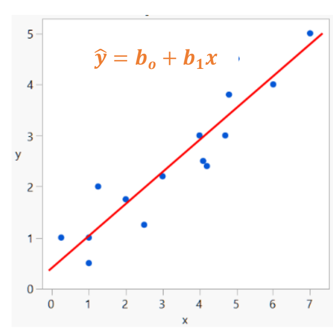


3. Fit a simple linear regression model to the salinity (y) vs. depth (x) data. Provide the equation of the least squares regression line predicting salinity from depth using good statistical notation.

Recall: In algebra, you learned about fitting a line to 2 points (x_1, y_1) and (x_2, y_2). The equation of this line can be represented by the equation y = mx + b, where m = slope and b = y-intercept.



In statistics, we never have only 2 points. Instead we measure many pairs of (x_i, y_i) on observational units that we call a sample. As you saw in Question #1, when we plot these points in a scatterplot, they rarely, if ever, fall perfectly on a line. The relationship between the variables x and y though, may be linear. In which case, it is appropriate to model the behavior of the (x,y) pairs with a line. We call this line a **simple linear regression model**, or **the least-squares regression line**. In general, the equation of the least-squares regression line may be represented as



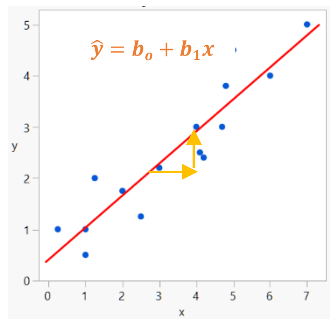
$$\hat{y} = b_o + b_1x$$

Where $\hat{y}$= predicted value of the response or dependent variable, b_o= y-intercept, and b_1= slope.

4. Interpret the slope and intercept of your least squares regression line from question #3.

The slope, b_1, represents the predicted increase (or decrease) in the y-variable for a 1-unit increase in the x-variable. We say, "for each 1-unit increase in the x-variable, the predicted value of the y-variable increases (or decreases) by , b_1."

The y-intercept, b_0, represents the predicted value of the y-variable when the x-variable value is 0. We say "When the x-variable is 0, the predicted y-variable value is b_0." The intercept may only be meaningful if it makes sense for the x-variable to take a value of 0.



5. Is there a *statistically significant* linear association between salinity and depth? Explain, citing an appropriate test statistic and p-value from your analysis to support your answer.

Recall: For the linear relationship between x and y to be statistically significant, we carry out a statistical hypothesis test for the slope of the regression line. If the observed slope, b_1, differs significantly from 0, we say the linear association between x and y is statistically significant. To decide whether the observed slope, b_1, is significantly different from 0, we use the p-value for the slope. A p-value < 0.05 indicates the observed slope (or something more extreme) is unlikely to have happened by random chance alone if there is no true linear association between x and y. Thus, when the p-value < 0.05, we say there is a statistically significant linear association between the two variables x and y.

6. Use the least-squares regression equation to predict the salinity at a depth of 4.5 meters. Show your work.

7. Use the least-squares regression equation to predict the salinity at a depth of 45 meters. What are you assuming about the relationship between salinity and depth when using the least-squares regression line to extrapolate to this depth?

8. Repeat questions 1-7 for temperature vs. depth.