Scatterplots & Correlation Coefficient Worksheet (Answer Key)

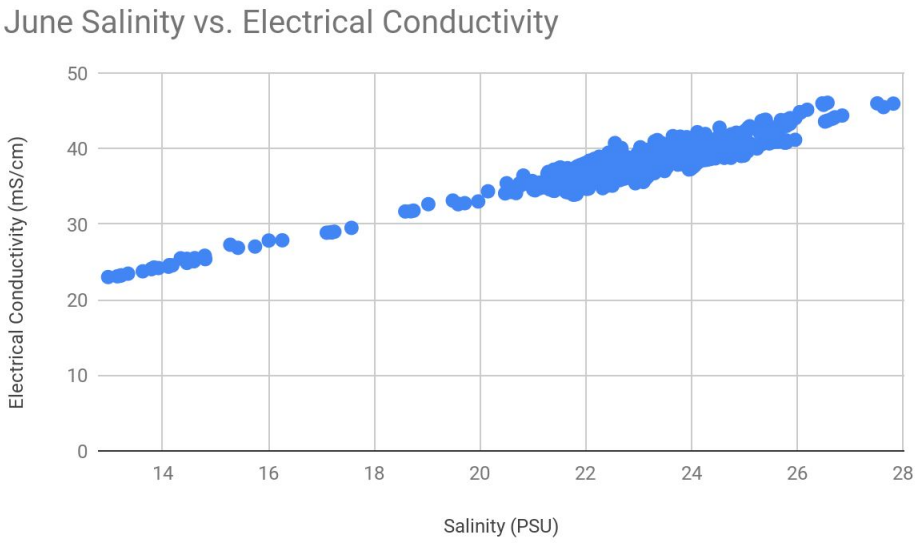
KEY TERMS

- **Scatterplot**: *the primary tool for examining two-variable relationships, when both variables are numerical*
  - In a scatterplot, each point represents one observation and the location of the point depends on the values of the two variables.
- **Trend:** *the general tendency of the scatterplot scanning from left to right*
  - Usually, trends are either increasing (uphill) or decreasing (downhill).
- **Strength:** *the amount of spread the scatterplot has in the vertical direction*
  - Weak associations result in a large amount of scattering in the scatterplot, while strong associations have little vertical variation.
- **Shape:** *the shape that the scatterplot plots form*
  - The simplest shape for a trend is linear.
- **Correlation Coefficient:** *a numerical value that provides a measure of strength and direction of the linear association between the independent variable x and the dependent variable y*
  - $r = (\Sigma ZxZy)/(n-1)$
  - The value of r is always between –1 and +1 and indicates the strength of the linear relationship between x and y. Values of *r* close to –1 or to +1 indicate a stronger linear relationship between x and y.

Illowsky, B., & Dean, S. (2018). Introductory statistics
Gould, R., & Ryan, C. N. (2015). Introductory statistics: Exploring the world through data. Pearson.

EXERCISE 1: Salinity vs. Electrical Conductivity

The below scatterplot contains a sample of 500 randomly selected salinity vs. electrical conductivity data points from the Mokauea loko iʻa data collected in June 2020.



The above data is real data collected from research at Mokauea loko iʻa.
Additional Mokauea loko iʻa data can be explored here: http://grogdata.soest.hawaii.edu/

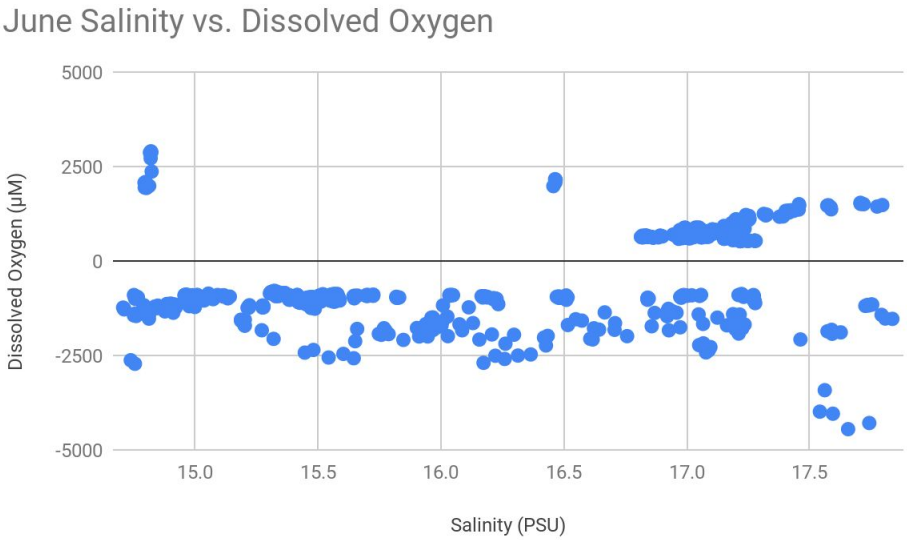
What is a research question that could ask based on variables used in the scatter plot?

What is the trend, strength, and shape of the scatterplot?

What could be determined about the two-variable based on the trend, strength, and shape of the scatterplot?

EXERCISE 2: Salinity vs. Dissolved Oxygen

The below scatterplot contains a sample of 500 randomly selected salinity vs. electrical conductivity data points from the Mokauea loko i‘a data collected in June 2020.



The above data is real data collected from research at Mokauea loko i‘a.
Additional Mokauea loko i`a data can be explored here: http://grogdata.soest.hawaii.edu/

What is a research question that could ask based on variables used in the scatter plot?

What is the trend, strength, and shape of the scatterplot?

What could be determined about the two-variable based on the trend, strength, and shape of the scatterplot?

EXERCISE 3: Correlation Coefficient

Describe what are likely the trend and strength based on the following correlation coefficients found from comparing two Mokauea loko iʻa parameters from March 2020. Assume that the shape is linear.

a) Surface Temperature (°C) vs. Bottom Temperature (°C): r = 0.99799799599

b) Salinity (PSU) vs. Electrical Conductivity (mS/cm): r = 0.97621718894