**Lesson 10.1- Scatter Plots**

**Two-variable data** is a collection of paired variable values, such as a series of measurements of air temperature at different times of day. One method of visualizing two-variable data is called a **scatter plot**: a graph of points with one variable plotted along each axis. A recognizable pattern in the arrangement of points suggests a mathematical relationship between the variables.

**Correlation** is a measure of the strength and direction of the relationship between two variables. The correlation is positive if both variables tend to increase together, negative if one decreases while the other increases, and we say there is “no correlation” if the change in the two variables appears to be unrelated.

One way to quantify the correlation of a data set is with the **correlation coefficient**, denoted by r. The correlation coefficient varies from -1 to 1, with the sign of r corresponding to the type of correlation (positive or negative). Strongly correlated data points look more like points that lie in a straight line, and have values of r closer to 1 or -1. Weakly correlated data will have values closer to 0.

There is a precise mathematical formula that can be used to calculate the correlation coefficient, but it is beyond the scope of this course. It is still useful to learn the qualitative relationship between the appearance of the data and the value of r. The chart below shows examples of strong correlations, with r close.



A **line of fit** is a line through a set of two-variable data that illustrates the correlation. When there is a strong correlation between the two variables in a set of two-variable data, you can use a line of fit as the basis to construct a linear model for the data.

There are many ways to come up with a line of fit. This lesson addresses a visual method: Using a straight edge, draw the line that the data points appear to be clustered around. It is not important that any of the data points actually touch the line; instead the line should be drawn as straight as possible and should go through the middle of the scattered points.

Once a line of fit has been drawn onto the scatter plot, you can choose two points on the line to write an equation for the line.

**Example 1:**

The table below presents two-variable data for seven different cities in the Northern hemisphere.



|  |  |  |
| --- | --- | --- |
| City | Latitude | Average Temp |
| Bangkok | 13.7 | 82.6 |
| Cairo | 30.1 | 71.4 |
| London | 51.5 | 51.8 |
| Moscow | 55.8 | 39.4 |
| New Delhi | 28.6 | 77.0 |
| Tokyo | 35.7 | 58.1 |
| Vancouver | 49.2 | 49.6 |

**Step 1: Graph your coordinate points.**

**Step 2: Draw a line of fit onto your graph. Try to get the same amount of points on both sides of the line.**

**Step 3: Find an equation in the form of y=mx+b, using two coordinate points from your graph.**

**Example 2:** The boiling point of water is lower at higher elevations because of the lower atmospheric pressure. The boiling point of water in some different cities is given in the table.

****

|  |  |  |
| --- | --- | --- |
| **City** | **Altitude** | **Boiling Point** |
| Chicago | 597 | 210 |
| Denver | 5300 | 201 |
| Kathmandu | 4600 | 205 |
| Madrid | 2188 | 207 |
| Miami | 6 | 210 |

**Step 1: Graph your coordinate points.**

**Step 2: Draw a line of fit onto your graph. Try to get the same amount of points on both sides of the line.**

**Step 3: Find an equation in the form of y=mx+b, using two coordinate points from your graph.**

**Homework: Workbook Pages 443-445 (1-12)**