**Section 3.1 – Scatterplots and Correlation** (pp. 141-164)

Most statistical studies examine data on more than one variable. We will continue to use tools we have already learned as well as adding others to assist us in analysis.

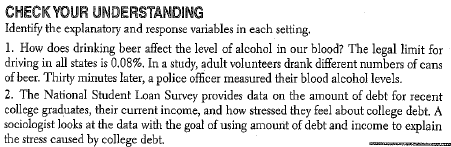
* Plot the data, add numerical summaries
* Look for overall patterns and deviations from those patterns
* If there is a regular pattern, use a simplified model to describe it

**1. Explanatory and Response Variables**

|  |
| --- |
| **Definition**: A **response variable** measures the outcome of a study. An **explanatory variable** *may* help explain or influence changes in a response variable. |

This means that the *explanatory variable* “accounts for” or “predicts” changes in the response variable.

Examples:



**2. Displaying Relationships: Scatterplots**

|  |  |
| --- | --- |
| **Definition:** A **scatterplot** shows the relationship between two quantitative variables measured on the same individuals. The values of one variable appear on the horizontal axis and the values of the other variable appear on the vertical axis. Each individual in the data set appears as a point on the graph. | |
|  | If there is an explanatory variable, it is plotted on the x-axis and the response variable is on the y-axis.  If there is no explanatory-response distinction, either variable can go on the x-axis. |

|  |  |
| --- | --- |
| **How to make a scatterplot:**  1. Decide which variable should go on which axis.  2. Label and scale your axes  3. Plot individual data values  (Common error on AP Exam – failing to label axes.) | **Calculator:** |

**3. Interpreting Scatterplots**

|  |
| --- |
| **How to examine a scatterplot**  Look for *overall pattern* and for striking *departures* from that pattern   * Overall pattern is described by the **direction**, **form**, and **strength** of the relationship. * An important type of departure is an **outlier**, an individual pattern that falls outside the overall pattern of the relationship.   **DOFS** |

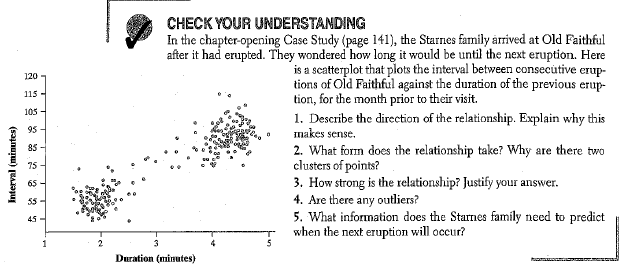
|  |  |
| --- | --- |
|  | D:  F:  S:  O: |

|  |
| --- |
| **Definition**:  Two variables have a **positive association** when the above average values of one tend to accompany above average values of the other and when below average values also tend to occur together.  Two variables have a **negative association** when the above average values of one tend to accompany below average values of the other. |

**\*\*\*\*Causation and Association\*\*\*\***

Association does not imply causation!!!!!

Examples:



**4. Measuring Linear Association: Correlation**

A linear relationship may appear in a scatterplot. The linear relationship is strong if the points lie close to a straight line and weak if they are widely scattered about a line. We are going to use a statistic called **correlation** to measure linearity in a scatterplot. **Correlation** ***r*** measures the *direction* and *strength* of the linear relationship between two quantitative variables.

The correlation r is always a number between -1 and 1. The sign indicates the direction of the association. Values close to 0 indicate a weak linear relationship. As r approaches -1 or 1, the strength of the relationship increases. -1 and 1 only occur if the values lie *exactly* on a straight line.

Refer to figure 3.6 on page 151 for examples of different values of r.

**Team work:** The following data give the weight in pounds and cost in dollars of a sample of 11 stand mixers.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Wt | 23 | 28 | 19 | 17 | 25 | 26 | 21 | 32 | 16 | 17 | 8 |
| Price | 180 | 250 | 300 | 150 | 300 | 370 | 400 | 350 | 200 | 150 | 30 |

1. Scatterplot your data and sketch the scatterplot below. Be sure to scale and label it properly.

2. Calculate the correlation.

3. The last mixer in the table is from Walmart. What happens to the correlation when you remove this point?

4. What happens to the correlation if the Walmart mixer weighs 25 pounds instead of 8 pounds? Add the point (25, 30) and recalculate the correlation.

5. Suppose a new titanium mixer was introduced that weighed 8 points, but the cost was $500. Remove the point (25, 30) and add the point (8, 500). Recalculate the correlation.

6. Summarize what you learned about the effect of a single point on the correlation.

|  |
| --- |
| **How to calculate r** |

What does this mean?

Note: A value of r close to 1 or -1 *does not guarantee a linear relationship between two variables*. A scatterplot with a clear curved form can have a correlation that is near -1 or 1. **Always plot your data!**

**Team work**: Complete Check Your Understanding on pp. 153.

**5. Facts about Correlation**

1. Correlation makes no distinction between explanatory and response variables.

2. Because r uses the standardized values of the observations, r does not change when we change the units of measurement of x, y, or both.

3. The correlation r itself has no unit of measurement.

4. Correlation requires that both variables be quantitative.

5. Correlation measures the strength of only the linear relationship between tow variables. It does not describe curved relationships between variables.

6. The correlation is not *resistant*: it is strongly affected by a few outlying observations.

7. Correlation is not a complete summary of two-variable data. You should always give means and standard deviations of both x and y along with the correlation.

**Team work.** Read and discuss the example on p. 157.

HW: p. 159 – 1, 5, 7, 11, 13, 14-18, 21, 26.