Correleation and Regression

Objectives

Introduce correlation and regression
Review scatter plots

Homework

Do you really need homework to learn scatter plots?

Correlation and Regression

Attempts to answer two questions.

Is there a **relationship** between the **behaviors** of two continuous, quantitative variables?

and

Is that relationship statistically significant?

Correlation

The correlation is one of the most common and useful tests in statistics. A correlation is a single value, denoted r or ρ (rho), that describes the **degree** of **relationship** between two **quantitative, continuous** variables. Correlations are useful because they may indicate a predictive relationship. For example:


An electrical utility may produce less power on a mild day based on the correlation between electricity demand and the weather conditions.


The height of a child might be predicted from the height of the parents.


Success in college may be predicted from SAT scores.

Positive or Negative

In a simple correlation analysis the relationship between the values of two variables can be positive or negative.

A positive relationship indicates a direct variation. In other words both variables increase or decrease at the same time.

id est (i.e.) Education and income.

A negative relationship indicates an inverse variation. In other words one variable increases while the other variable decreases.

i.e. Age and mental acuity.

Objective: Students will will draw a scatterplot for a set of ordered pairs.

Regression

The goal of a **correlation analysis** is to see whether the observations from two variables **co-vary**, and to **quantify** the **strength** of a **linear** relationship between the **variables**. **Regression** expresses the relationship in the form of a **linear equation** for predictive purposes.

For example, for students taking a Math and English test, we could use **correlation** to determine whether students who are good at Math tend also to be good at English, and we could use **regression** to determine whether the grades in English can be **predicted** for given grades in Math.

Variables

In the most simple cases of correlation and regression analyses, there are two variables. The **independent (predictor or explanatory)** variable and the **dependent (response)** variable.

Remember: both variables must be quantitative, continuous and normally distributed.

In an experiment, the **independent** variable is the variable manipulated by the experimenter.

The **dependent** variable, also called the **response** variable, is the variable observed for changes in response to changes in the independent variable.

Objective: Students will will draw a scatterplot for a set of ordered pairs.

Scatter Plot

A scatter plot should be familiar as Algebra students have been experiencing them since Renee DeCartes came up with the Cartesian (coordinate) plane.

Data is displayed as a collection of points, each having the value of the **predictor variable (x)** determining the position on the **horizontal axis** and the value of the **response variable (y)** determining the position on the **vertical axis**. This kind of plot is also called a *scatter chart*, *scattergram*, *scatter diagram* or *scatter graph*.

The variable along the **x-axis is the predictor (independent)** variable. Along the **y-axis is the response (dependent) variable**.

Objective: Students will will draw a scatterplot for a set of ordered pairs.

Scatter Plot

Scatter plots are especially useful when there is a large number of data points. They provide the following information about the relationship between two variables:


Strength


Shape - linear, curved, etc.


Direction - positive or negative


Presence of outliers

Example

Suppose we are interested in discovering if there is any relationship between the time spent in a mall and the amount of money spent shopping. We survey several people and find the following data:

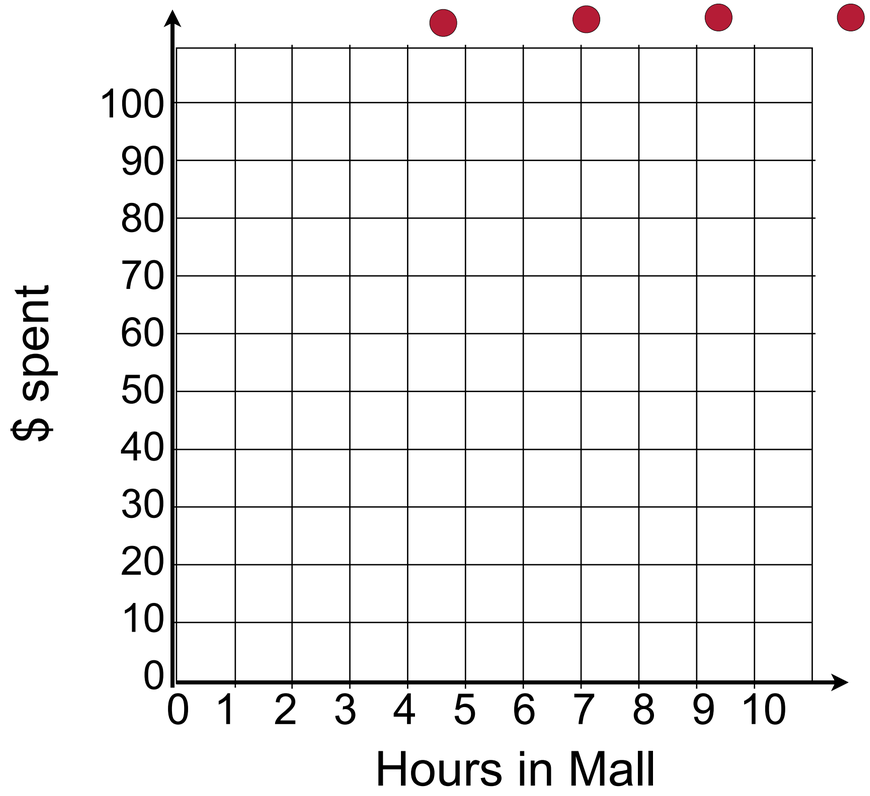
Hours in Mall	10	8	9	3	1	2	5	6	7	8	2	3
Dollars spent	40	15	24	20	10	35	50	70	18	25	100	60

Objective: Students will will draw a scatterplot for a set of ordered pairs.

Statistics 10-2

Example

Hours in Mall	10	8	9	3	1	2	5	6	7	8	2	3
Dollars spent	40	15	24	20	10	35	50	70	18	25	100	60



See any pattern?

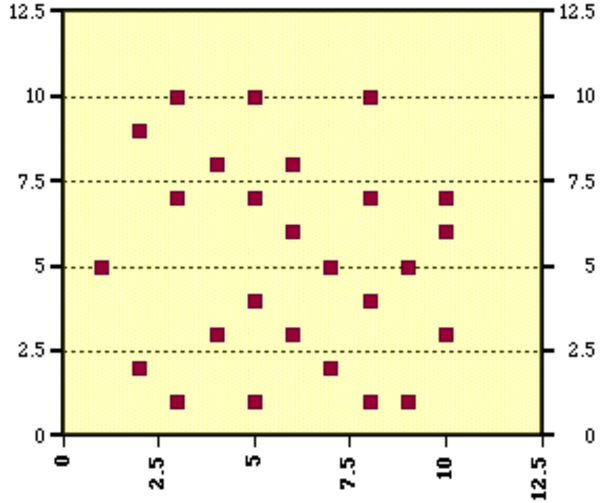
- Strength
- Shape - linear, curved
- Direction - positive or negative
- Presence of outliers

Objective: Students will will draw a scatterplot for a set of ordered pairs.

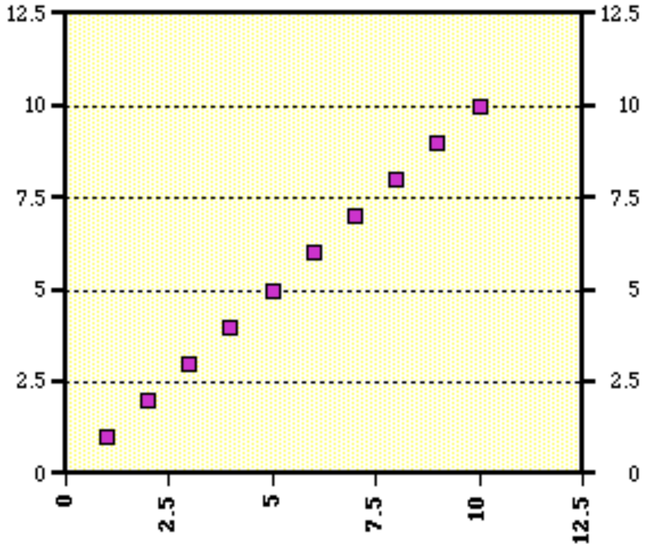
Statistics 10-2

What does it all mean?

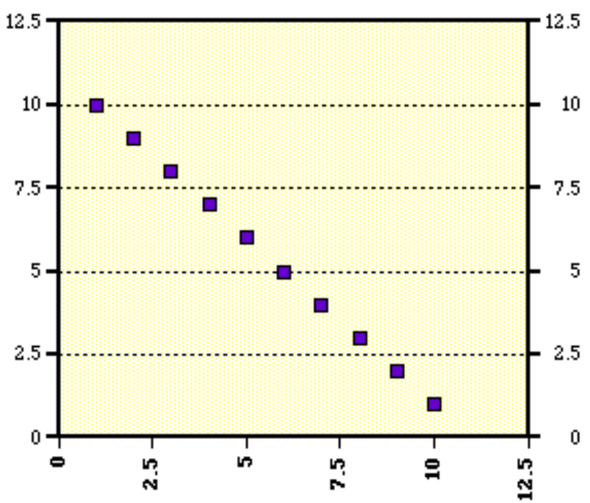
No Correlation



Perfect Positive Correlation



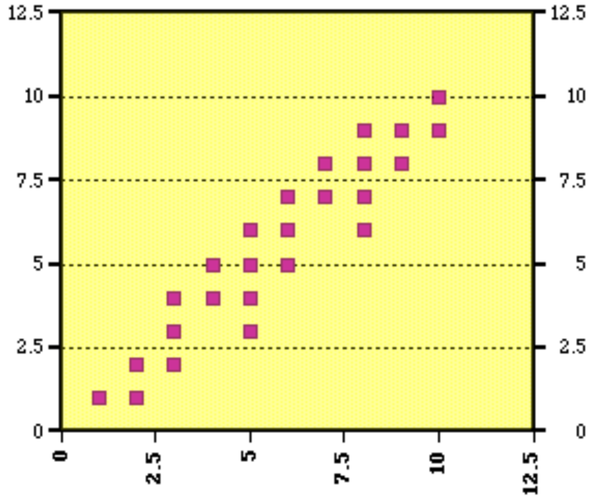
Perfect Negative Correlation



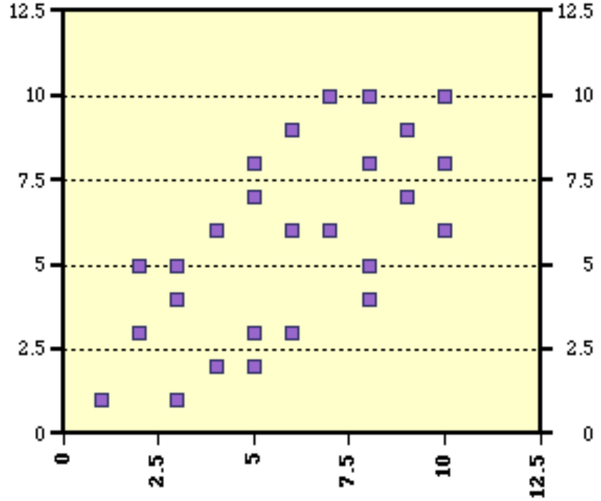
Objective: Students will will draw a scatterplot for a set of ordered pairs.

Statistics 10-2

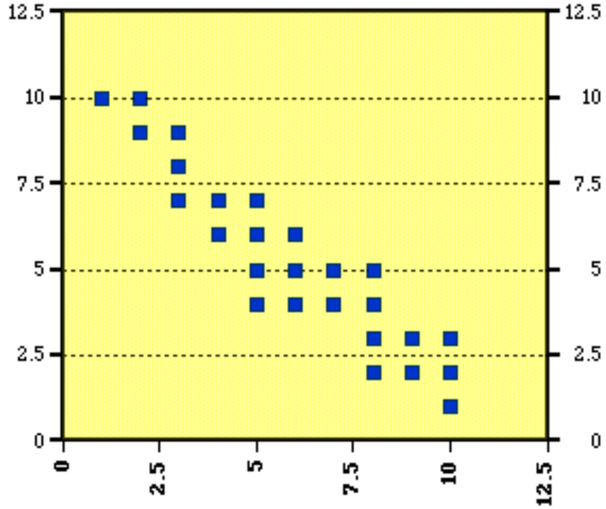
High Positive Correlation



Low Positive Correlation



High Negative Correlation



Low Negative Correlation