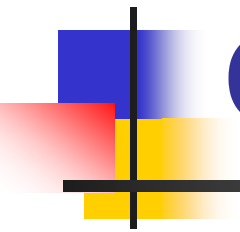


Correlation

Correlation Analysis

- Correlations tell us to the degree that two variables are similar or *associated* with each other. It is a measure of association.

- Correlations are bivariate or multivariate in nature, i.e., relational between variables.

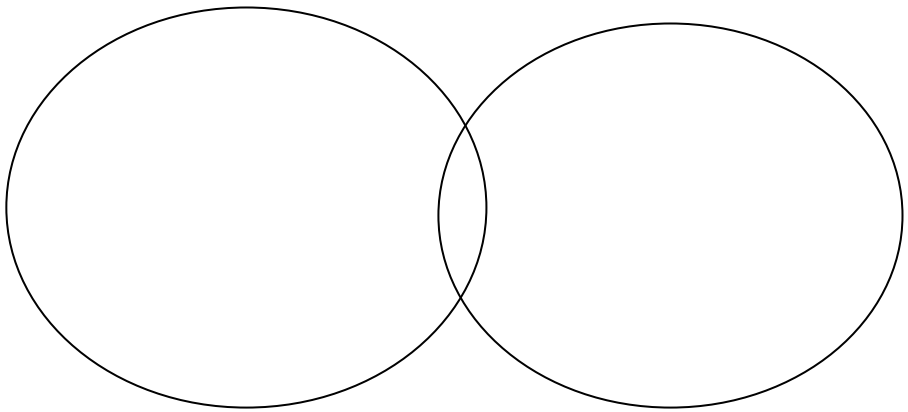
Correlation Analysis

- Correlations measure association by measuring how when one variable changes another changes with it, which means it is measuring dynamic relationships (change).

Conceptualizing Correlation

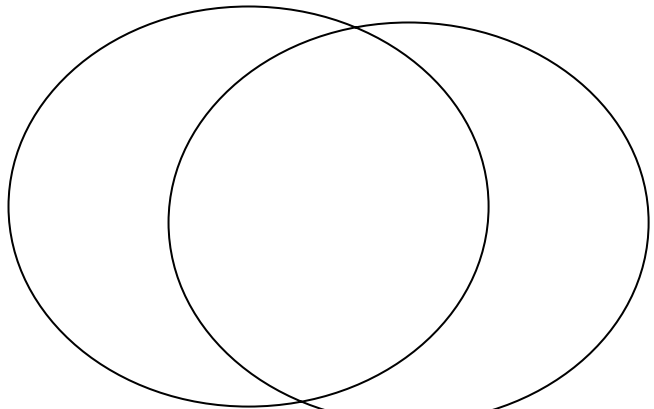
Measuring Development

Weak



GPD POP WEIGHT

Strong



GDP EDUCATION

Correlation will be associated with what type of validity?

Correlation Analysis

- Correlations range from -1 to 1, and what matters in regard to the strength of a relationship between two variables is the absolute value.

- The signs of the correlation coefficient reflect the nature of the relationship: positive (direct), negative (inverse, indirect)

TABLE 5.1 Types of Correlations and the Corresponding Relationship Between Variables

What Happens to Variable X	*What Happens to Variable Y*	*Type of Correlation*	*Value*	*Example*
X increases in value	*Y* increases in value	Direct or positive	Positive, ranging from .00 to +1.00	The more time you spend studying, the higher your test score will be.
X decreases in value	*Y* decreases in value	Direct or positive	Positive, ranging from .00 to +1.00	The less money you put in the bank, the less interest you will earn.
X increases in value	*Y* decreases in value	Indirect or negative	Negative, ranging from –1.00 to .00	The more you exercise, the less you will weigh.
X decreases in value	*Y* increases in value	Indirect or negative	Negative, ranging from –1.00 to .00	The less time you take to complete a test, the more you'll get wrong.

Rules of Thumb

Size of correlation coefficient	General Interpretation
.8 - 1.0	Very Strong
.6 - .8	Strong
.4 - .6	Moderate
.2 - .4	Weak
.0 - .2	Very Weak or no relationship

The strength of the correlation depends on how many data points in the scatter plot are near or far in a pattern.

This is similar to error in the data, which comes from other explanations

The shape of the relationship can be depicted in scatterplots. What type of relationship do you see?

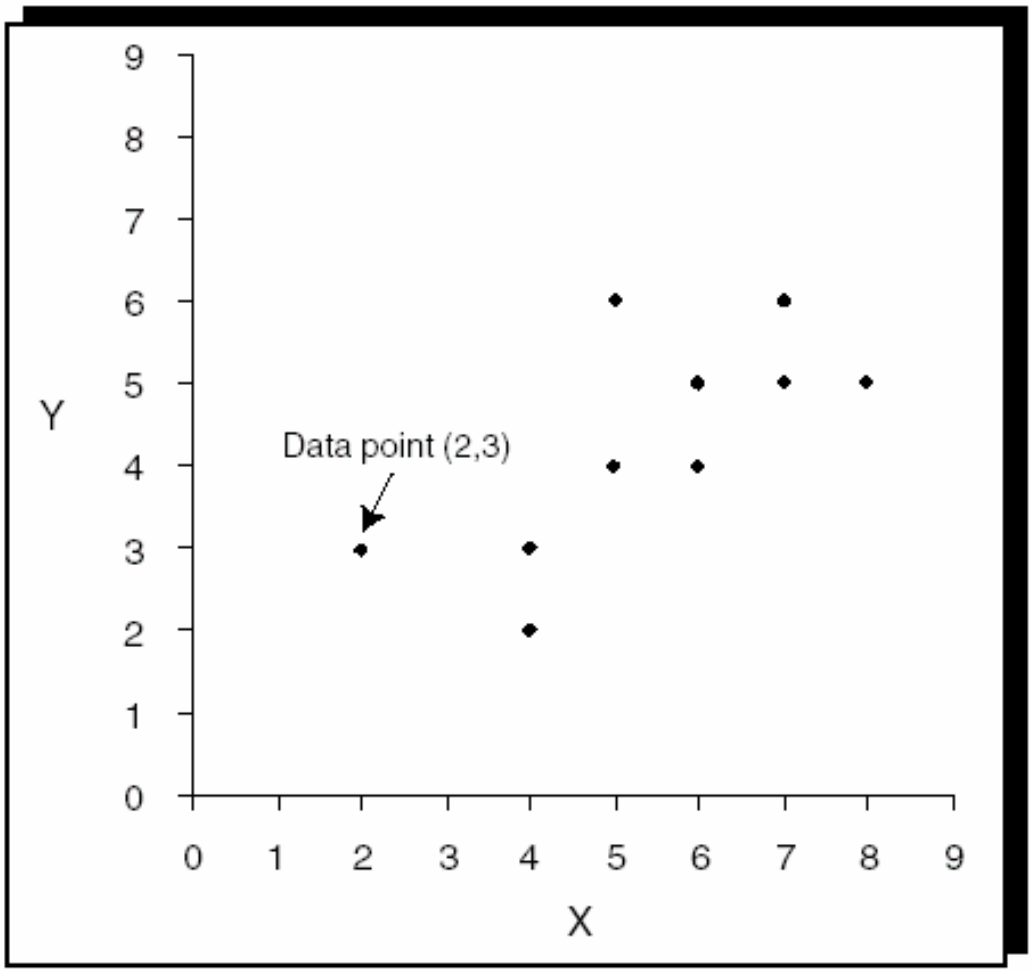


Figure 5.1. A Simple Scatterplot

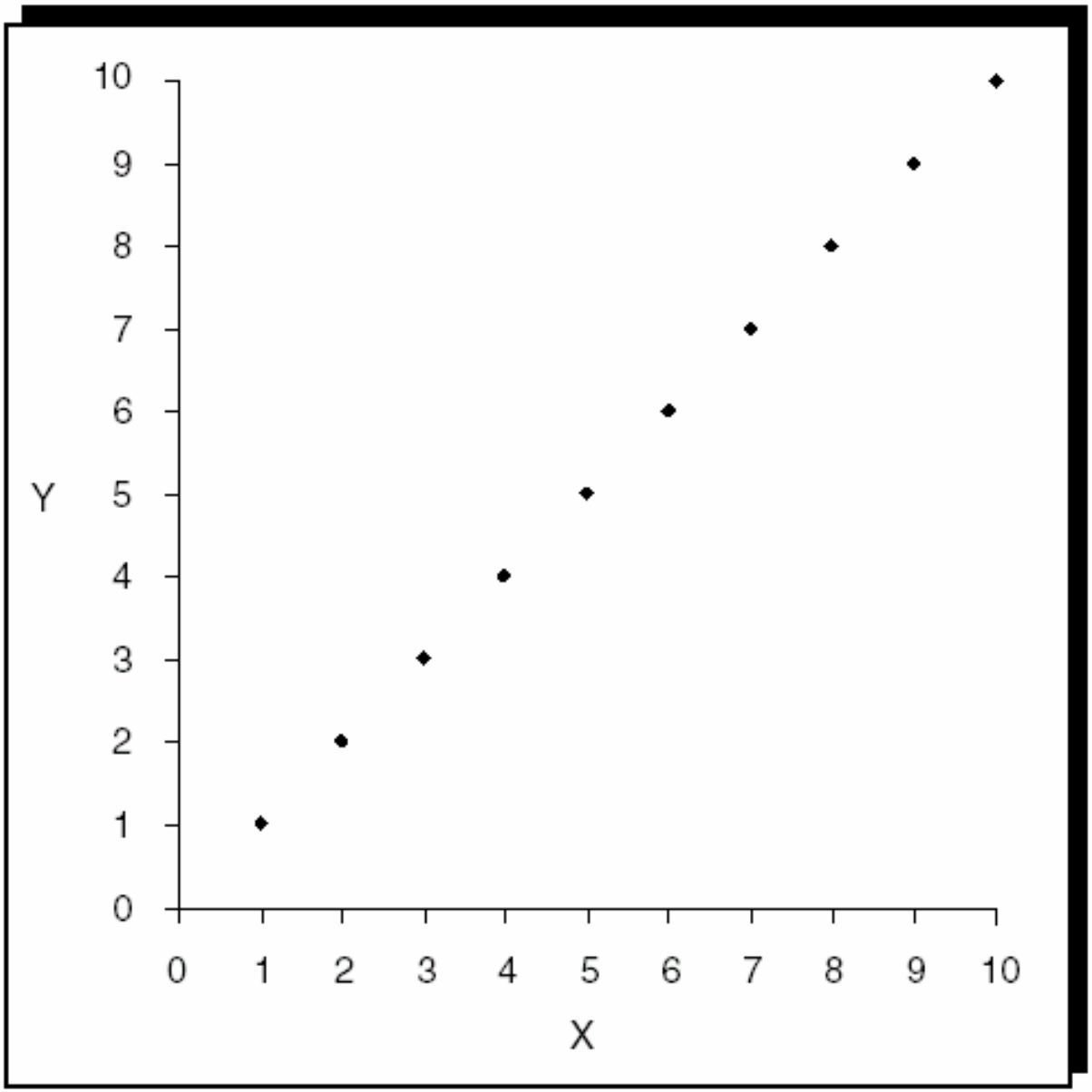


Figure 5.2. A Perfect Direct, or Positive, Correlation

Here is a perfect linear relationship between x and y.

Is it positive or negative?

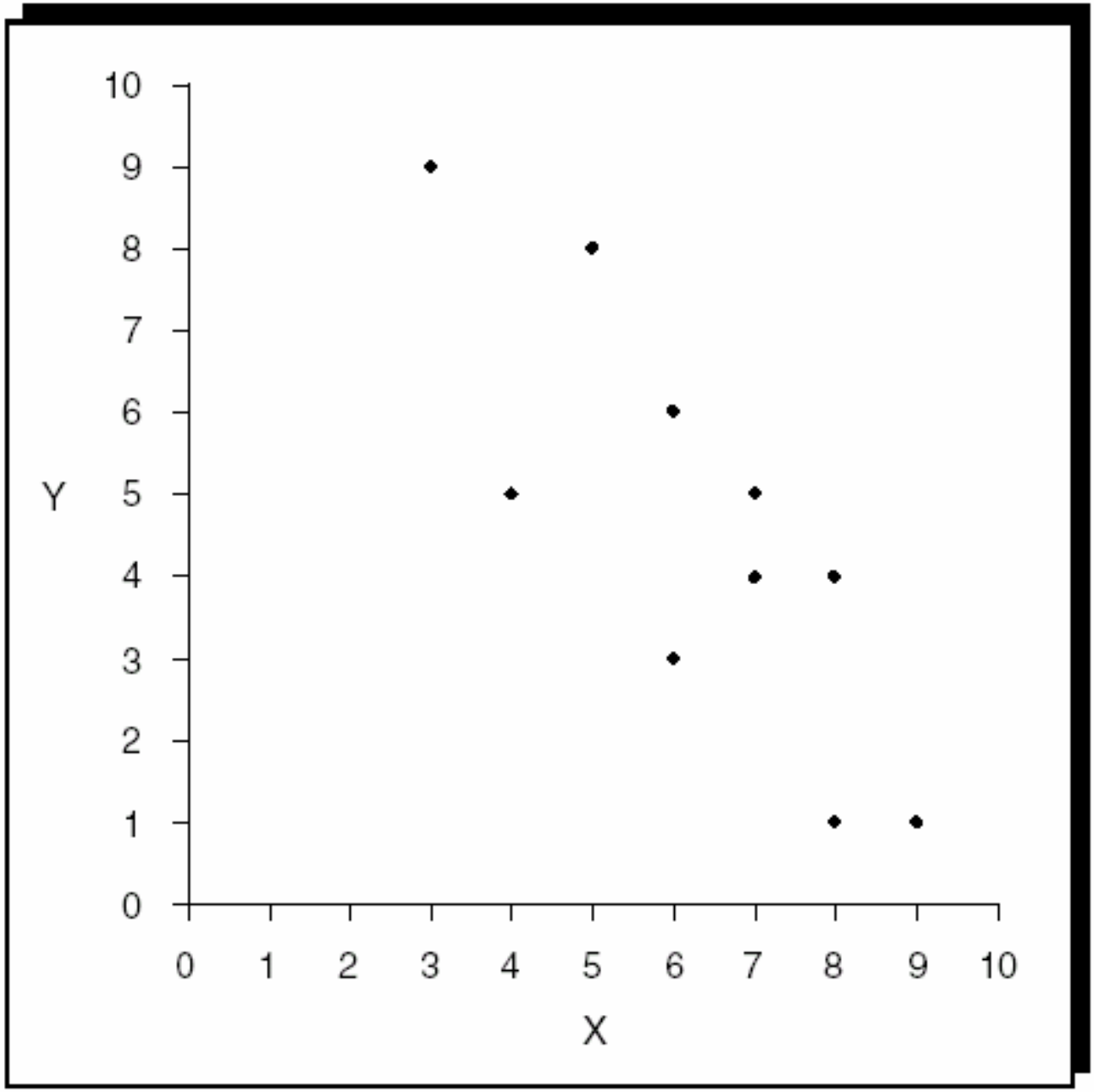


Figure 5.4. A Strong Indirect Relationship

What relationship do you see here?

Correlation Analysis

- We focus here on the Pearson product-moment correlation, which is used for continuous measures (ratio, interval = scalar)
- Spearman's rho Kendall's Tau-b used to compare ordered variables

Correlation Coefficient

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

Correlation and Causation

- We want to know how two variables are related to, overlap with, each other
- If one variable causes another, it will be correlated with it.
- However, some correlations are spurious and association does not automatically mean causation.

Correlating Developmental factors

Correlations

		GNPPERCAP	Urbanpop
GNPPERCAP	Pearson Correlation	1	.643**
	Sig. (2-tailed)	.	.000
	N	202	199
Urbanpop	Pearson Correlation	.643**	1
	Sig. (2-tailed)	.000	.
	N	199	260

**. Correlation is significant at the 0.01 level (2-tailed).

Significance of Variables

- **We can also estimate whether certain variables are important. We do this by ascertaining statistical significance.**

- **<u>Our key question is</u>: What is the probability that an estimate is produced by random chance and there is no relationship between X and Y variables?**

Significance of Variables

- **We measure statistical significance by the probability that what we are observing is wrong (generated by random chance).**
- **A significance level of .05 is conventional. This means that if the significance level is .05, there is a 5 percent chance that our results were generated randomly. A .01 level means there is a 1 percent chance.**

Another Example

Correlations

		Democ	Rebellion
Democ	Pearson Correlation	1	-.182*
	Sig. (2-tailed)	.	.032
	N	262	140
Rebellion	Pearson Correlation	-.182*	1
	Sig. (2-tailed)	.032	.
	N	140	156

*. Correlation is significant at the 0.05 level (2-tailed).

Limitation of correlation coefficients

- They tell us how strongly two variables are related
- May capture causation between variables but cannot differentiate from spurious ones.
- However, r coefficients are limited because they cannot tell anything about:

- Marginal impact of X on Y
- Direction of causation when present
- Forecasting

Because of the above Ordinary Least Square regression analysis (OLS) is most useful