

GEOG 204

LECTURE 11

Correlation

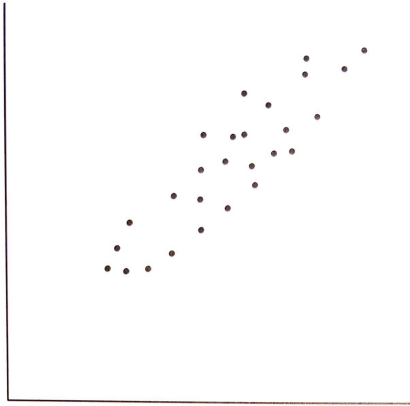
- Often one has to investigate if there is a relationship between two or more variables
 - Relationships total
 - Total household income and total monthly rent
 - Total population and the number of retail stores
 - Crime locations and the distance from police stations
- Correlation is a statistical method used to determine if a relationship between variables



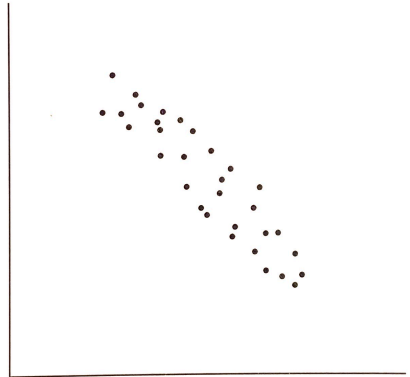
Correlation

- Scatterplots are a common tool used to portray the relationship or association between variables
- Scatterplots provide visual information about
 - Strength of relationship
 - Direction of relationship

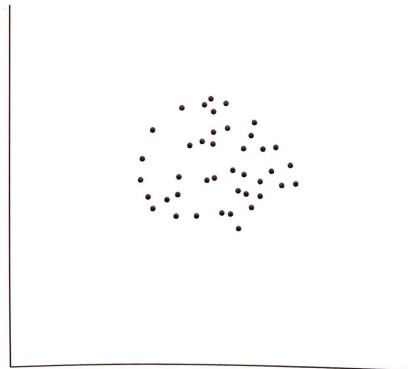
Case 1:
Positive



Case 2:
Negative

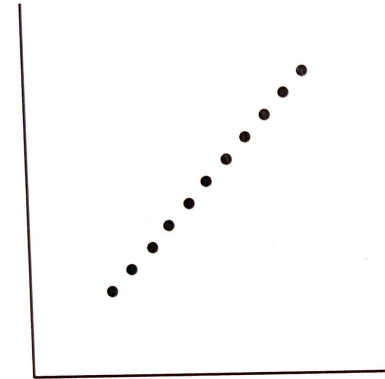


Case 3:
Neutral

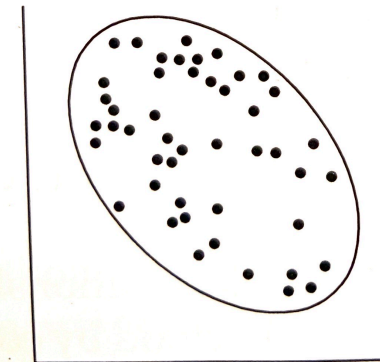


Direction

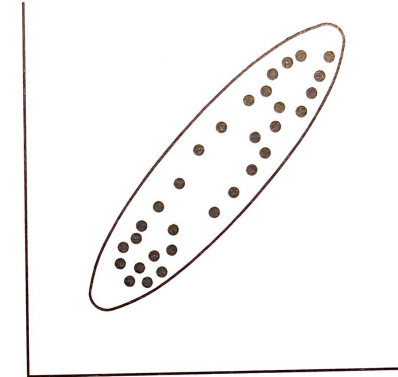
Case 1:
Perfect association



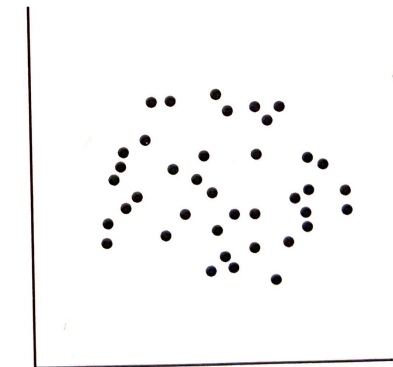
Case 3:
Weak association



Case 2:
Strong association



Case 4:
No association



Strength

Scatterplots

- Direction
 - Positive relationship
 - Increasing values in one variable correspond to increasing values in another variable
 - Decreasing values in one variable correspond to decreasing values in another variable
 - Negative (inverse) relationship
 - Increasing values in one variable correspond to decreasing values in another variable
 - Decreasing values in one variable correspond to increasing values in another variable
- Strength of relationship
 - Determined by the amount of spread in a scatterplot

Covariation

- Covariation:
 - The degree to which variables 'covary' (vary together or jointly)
 - If two variables covary in a similar manner
 - Data have a large covariation
 - Data have a strong correlation
 - If two variables show little consistency in how they covary
 - Then the correlation is weak

Covariance

$$CV_{XY} = \sum (X - \bar{X})(Y - \bar{Y})$$

$$r = \frac{[\sum (X - \bar{X})(Y - \bar{Y})]/N}{S_y S_x}$$

where CV_{xy} = covariation between X and Y

$(X - \bar{X})$ = deviation of X from its mean (\bar{X})

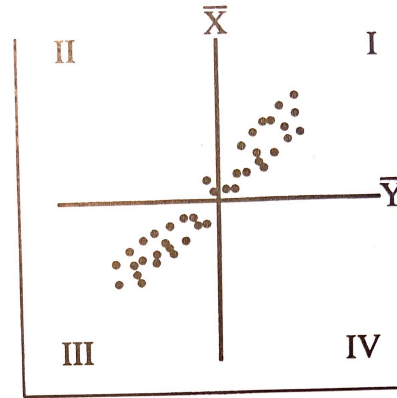
$(Y - \bar{Y})$ = deviation of Y from its mean (\bar{Y})

r = Correlation coefficient

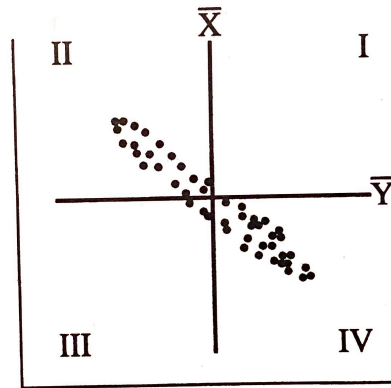
S_y, S_x = Standard deviation of Y and X, respectively

Covariation

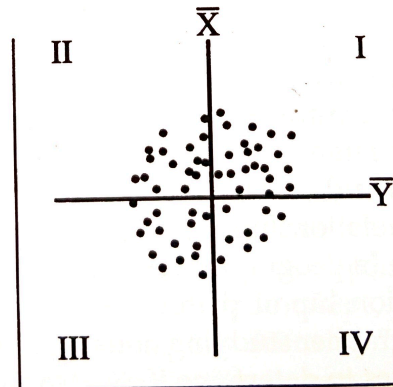
Case 1:
High positive
covariation



Case 2:
High negative
covariation



Case 3:
Low
covariation



Correlation

- Any two variables can be correlated, and the strength and direction of relationship determined.
 - caution must be used when evaluating or interpreting correlations.
 - A relationship does not necessarily imply the existence of a causal relationship.