GEOG 413/613

LECTURE 12

1

Bivariate Analysis

- Multivariate Exploratory Data Analysis
  - Multiple variables
    - Explore patterns
    - Relationships between variable
    - Outliers
  - No functional/causal relationship is assumed between the two variables

2

2

Bivariate Analysis

- Bivariate Analysis
  - Two variables are explored in detail
  - Assumption is that one variable influences/affects the other
  - Example:
    - The relationship between precipitation and population density.
    - The assumption is the amount of moisture at locations influences population density

Bivariate Regression

- Similar to correlation analysis, bivariate regression seeks to examine the influence of one variable on another
- Independent Variable (Explanatory Variable)
  - The variable creating the influence/effect
- Dependent Variable
  - The variable receiving the influence or effect

Relationships in Bivariate Regression

- In correlation, the assignment of the axes for the variables is arbitrary
- In binary regression
  - Independent variable on the X-axis
  - Dependent variable on the Y-axis
- The form of association between the variables can be portrayed using a scatterplot

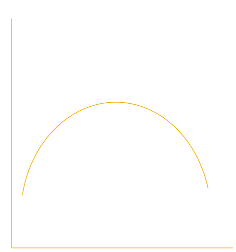
5

5

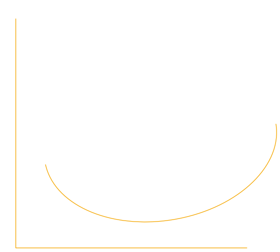
Relationships in Bivariate Regression

- Not Significant (Statistically)

- Linear (Positive, Negative)
- Curve-linear (Concave, Convex)



Concave



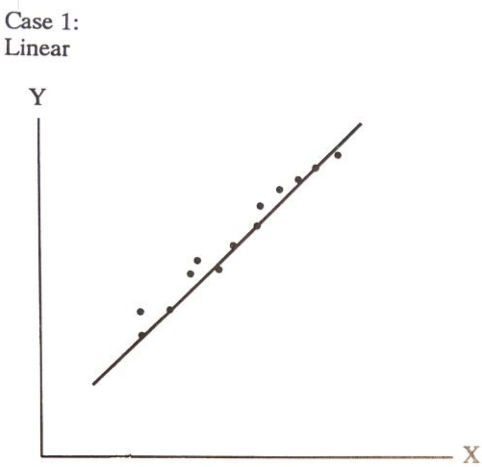
Convex

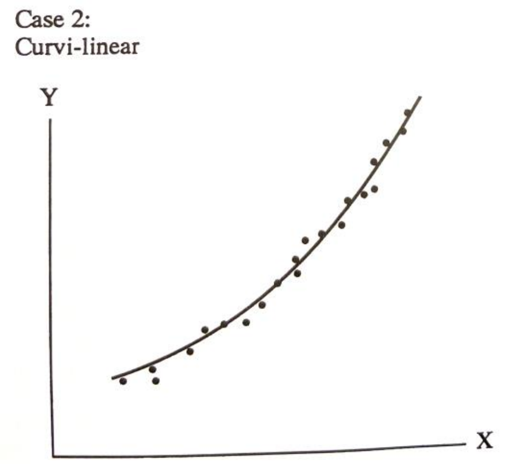
- Undefined complex (statistically significant but relationship cannot be reliably described)

6

6

Relationships





7

7

Bivariate Regression Line

- The point pattern on the scatterplot can can be described with a least-squares regression line
- The least-squares regression line is unique
  - Minimizes the sum of the squared vertical distances between each data point and the line

8

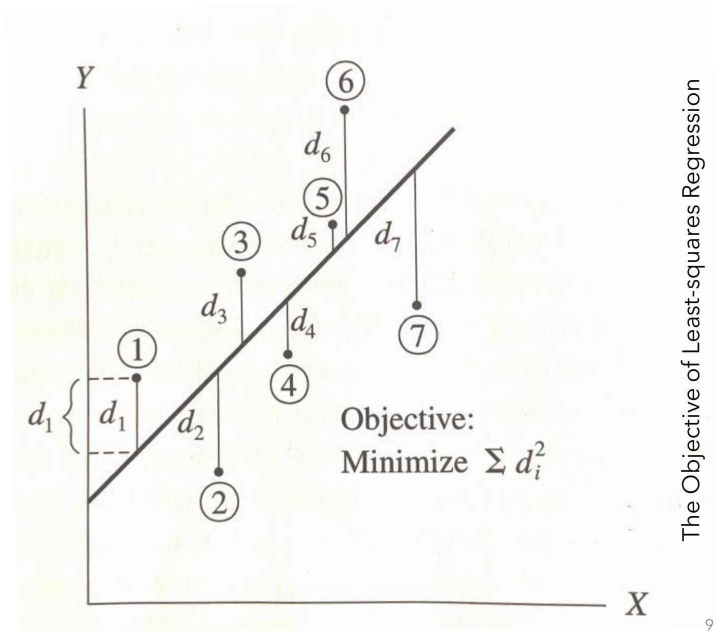
8

Bivariate Regression Line

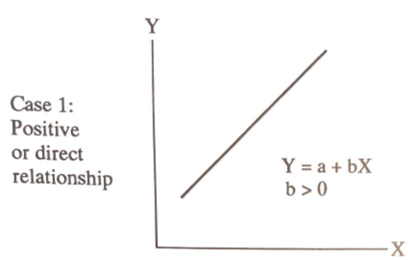
- The regression is given by the straight-line equation

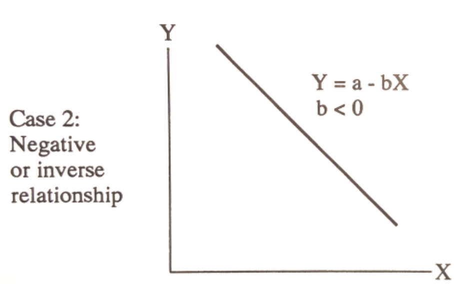
$$Y = a + bX$$

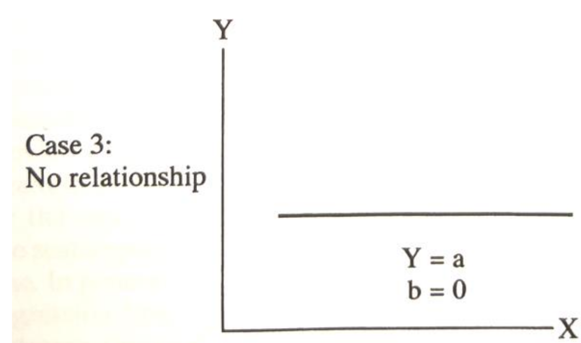
- *a* is the intercept on the Y-Axis
  - Represents the value of Y when X is zero
- *b* represents the slope of the line
  - Also, the correlation coefficient



The Objective of Least-squares Regression





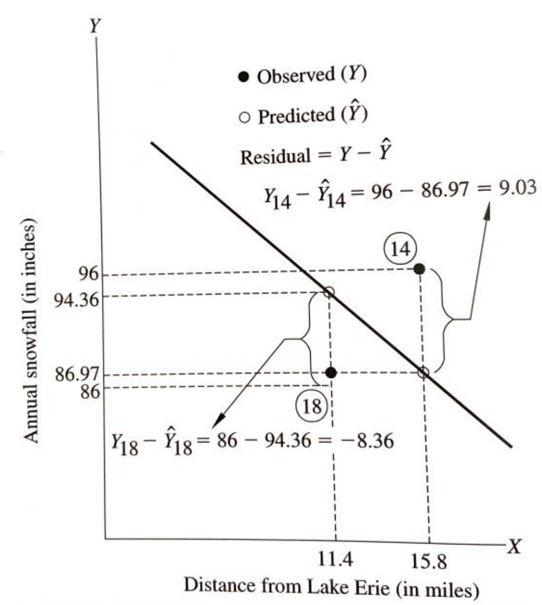


Strength of Relationship

- The ability of the independent to account for variation in Y provides a measure of the strength of the relationship
- The closer the points to the regression line stronger the relation between the variables
- Strength is measured with the coefficient of determination, r^2
- r^2 = ratio of explained variation to total variation

11

11



12

12

Variation in the dependent variable

- Variation in the dependent has two parts
  - Explained Variable
  - Unexplained variable

$$\sum y^2 = \sum y_e^2 + \sum y_u^2$$

$$\sum y^2 \Rightarrow TSS\ (Total\ Sum\ of\ Squares), total\ variation\ in\ Y$$

$$\sum y_e^2 \Rightarrow explained\ variation\ (caused\ by\ independent\ variable)$$

$$\sum y_u^2 \Rightarrow unexplained\ variation$$

13

13

The Coefficient of Determination

$$\sum y_e^2 = \frac{(\sum xy)^2}{\sum x^2}$$

$$\sum x^2 \Rightarrow total\ variation\ of\ X$$

$$(\sum xy)^2 \Rightarrow the\ square\ of\ the\ Covariation\ of\ X\ and\ Y$$

$$r^2 = \frac{\sum y_e^2}{\sum y^2}$$

r^2 = coefficient of determination

14

14

Demo

- Understanding the Variation in the independent variable
  - Create a variable X, where x=x+2
  - Create a variable Y_1, where y_1=x*1.5
  - Create a variable Y_2, where y_2 = y_2 + random number between -2 and +2
  - Y_1 – total variation all explained by X
  - Y_2 – total variation explained in part by X and some unknown variance

15

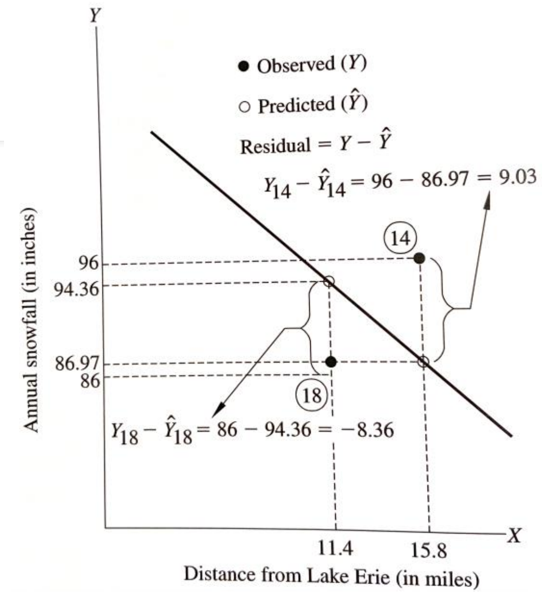
15

X	Y1	Y2
2	3	5
4	6	8
6	9	8
8	12	11
10	15	13
12	18	19
14	21	23
16	24	23
18	27	26
20	30	32
22	33	31
24	36	38
26	39	39
28	42	40
30	45	46
32	48	48
34	51	50
36	54	54
38	57	58
40	60	60
42	63	61
44	66	64
46	69	71
48	72	72
50	75	75
52	78	79
54	81	83
56	84	85
58	87	87

Var X	280		
Var Y1	630		
Var Y2	634.256837		
Covariance (X&Y1)	420		
Covariance (X&Y2)	420.758621		
Exp Var Y1	(420*420)/280 = 630		
Exp Var Y2	??		

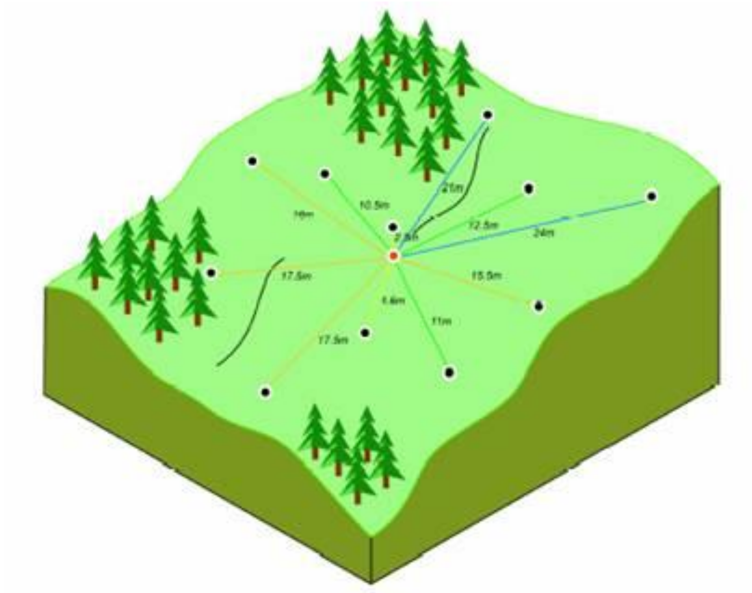
16

16



Semivariance

- In the context of spatial points on a surface there is a need to understand the degree of association between points
- Semivariance used in the process of Kriging
  - Equal to half the variance of the differences between all possible points spaced a constant distance apart
  - At d=0 semivariance is Zero, as the points further are considered, the semivariance increases until it reaches the variance of the whole surface.
    - This is is the maximum distance at which two points are related
    - This maximum distance is called the range
    - The range defines the size of the neighborhood over which control points should be selected to predict other points



19

19