Regression Analysis

Multiple Linear Regression

Nicoleta Serban, Ph.D.
Professor
School of Industrial and Systems Engineering

Objectives and Examples

1

About This Lesson

2

Linear Regression: Example 1

Georgia Tech

3

Linear Regression: Data Example 1

The response variable is:

Y = Sales (in thousands of dollars)

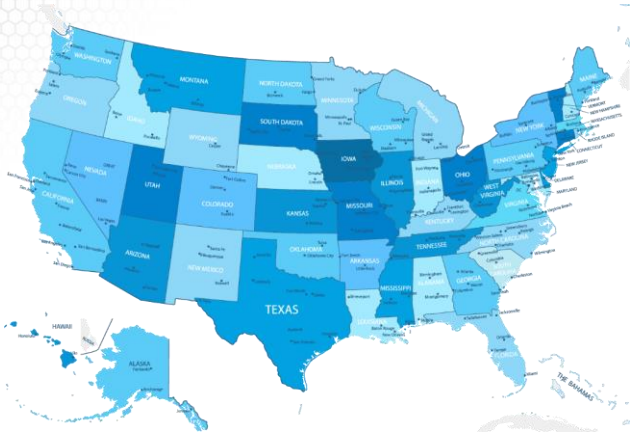
The predicting variables are:

X_1 = Amount (in hundreds of dollars) spent on advertising
X_2 = Total amount of bonuses paid
X_3 = Market share in each territory
X_4 = Largest competitor's sales
X_5 = Region in which territory is located (1 = south, 2 = west, 3 = midwest)

Georgia Tech

4

Linear Regression: Example 2



SAT Mean Score by State – Year 1982

790 (South Carolina) - 1088 (Iowa)

5

Linear Regression: Data Example 2

The response variable is:

Y = State average SAT score (verbal and quantitative combined)

The predicting variables are:

X_1 = % of total eligible high school seniors in the state who took the exam

X_2 = Median income of families of test takers, in hundreds of dollars

X_3 = Average number of years that test takers had in social sciences, natural sciences, and humanities

X_4 = % of test takers who attended public schools

X_5 = State expenditure on secondary schools, in hundreds of dollars per student

X_6 = Median percentile of ranking of test takers within their secondary school classes

6

Linear Regression: Example 3

Bike sharing systems are of great interest due to their important role in traffic management.

Dataset: Historical data for years 2011-2012 for the bike sharing system in Washington D.C.

Data Source: UCI Machine Learning Repository

Georgia Tech

7

Linear Regression: Data Example 2

The response variable is:

Y = Hourly count rentals of bikes

The predicting variables are:

X_1 = Day of the week
X_2 = Month of the year
X_3 = Hour of the day (ranging 0-23)
X_4 = Year (2011, 2012)
X_5 = Holiday Indicator
X_6 = Weather condition (with four levels from good weather for level 1 to severe condition for level 4)
X_7 = Normalized temperature
X_8 = Normalized humidity
X_9 = Wind speed

Georgia Tech

8

Multiple Linear Regression: Objectives

A regression analysis is used for:

1. **Prediction** of the response variable
2. **Modelling** the relationship between the response variable and explanatory variables
3. **Testing** hypotheses of association relationships

Linear Regression: The basis of what we will discussing in most of this course is the linear model. Virtually all other methods for studying dependence among variables are variations on the idea of linear regression.

"*All models are wrong, but some are useful*." – George Box

"*Embrace your data, not your models*." – John Tukey

Georgia Tech

9

Summary

Georgia Tech

10