Regression Analysis

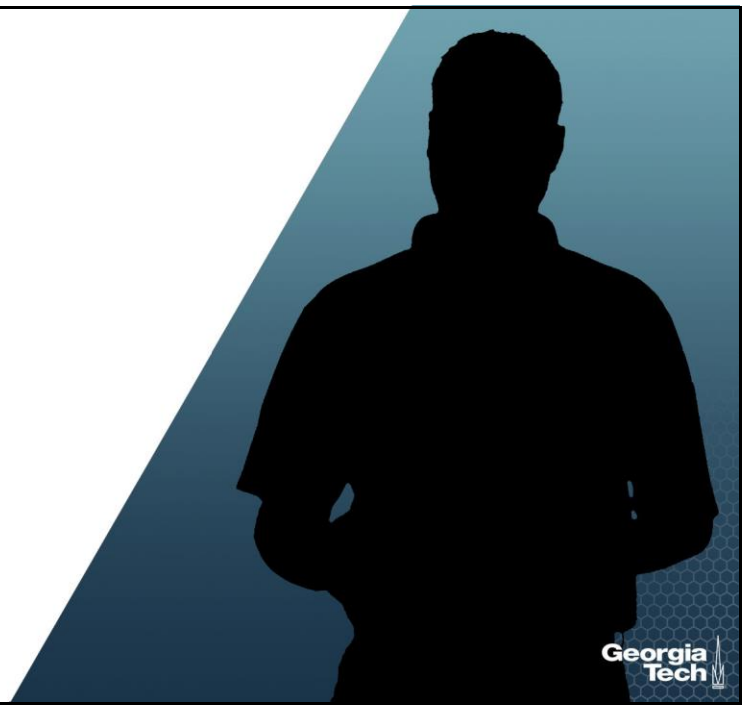
Simple Linear Regression

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

Regression Concepts: Basics

1

About This Lesson



2

Example 1

A company, which sells medical supplies to hospitals, clinics, and doctor's offices, had considered the effectiveness of a new advertising program.

Management wants to know if the **advertisement** is related to **sales**.

This company intends to increase the sales with an effective advertising program.

Georgia Tech

3

Data Example 1

The company observes for **25 offices** the yearly sales **(in thousands)** and the advertisement expenditure for the new program **(in hundreds)**

Sales	ADV
963.50	374.27
893.00	408.50
1057.25	414.31
1183.25	448.42
1419.50	517.88
…	

Georgia Tech

4

Example 2

- The principle of purchasing power parity **(PPP)** states that over long periods of time **exchange rate** changes will tend to offset the differences in **inflation rate** between two countries.
- In an efficient international economy, exchange rates would give each currency the same purchasing power in its own economy. Even if it does not hold exactly, the **PPP** model provides a benchmark to suggest the levels that exchange rates should achieve.

Georgia Tech

5

Data Example 2

The data are recorded for **41** countries, including both developed and developing countries. The data include the following columns.

Country	Inflation.difference	Exchange.rate.change	Developed
Australia	-1.2351	-3.1870	1
Austria	1.5508	1.4781	1
Belgium	1.0371	0.0395	1
Canada	0.0461	-1.6416	1
Chile	-18.4126	-20.6329	0

Georgia Tech

6

Example 3

- In 2000 **Bush** and **Gore** were the main candidates for President in the U.S. Buchanan, a strongly conservative candidate, was also on the ballot. In the **state of Florida**, **Bush** and **Gore** essentially tied, hence the counts were examined carefully county by county.

- **Palm Beach County** exhibited strange results. Even though the people in this county are not conservative, many votes were cast for **Buchanan**. Examination of the voting ballot revealed that it was easy to mistakenly vote for **Buchanan (a conservative candidate)** when intending to vote for **Gore**. We will thus predict whether those who voted for **Buchanan** were indeed going for a conservative candidate.

The data file includes many other variables characterizing the counties. We will focus only on the number of votes in this analysis.

Georgia Tech

7

Variables in Regression

The regression framework is characterized by the following:

1. We have one particular variable that we are interested in understanding or modelling, such as sales of a particular product, or the stock price of a publicly traded firm. This variable is called the **response (dependent) variable**, and is usually represented by Y.

2. We have a set of other variables that we think might be useful in predicting or modelling the response variable (say the price of the product, the competitors' price, and so on; or the profits, revenues, financial position of the firm, and so on). These are called the **predicting *or* explanatory (independent) variables**, and are usually represented by x1, x2, etc.

Georgia Tech

8

Variables in Regression

RESPONSE VARIABLE versus PREDICTING VARIABLE?

Response Variable: It is a Random Variable. It varies with changes in the predictor/s along with other random changes.

Predicting Variable: It is a Fixed Variable. It does not change with the response, but it is set fixed before the response is measured.

Georgia Tech

9

Response vs Predicting Variable

The **effect** of several types of cholesterol medications on LDL levels in humans.

- **Response Variable:** Change in LDL levels
- **Predicting Variable:** Type of Medication

The **relationship** between driving habits and fuel efficiency

- **Response Variable:** Miles Per Gallon **(MPG)** of Fuel
- **Predicting Variable:** Average Driving Speed

The **relationship** between college grade point average **(GPA)** and scores on the SAT

- **Response Variable:** GPA
- **Predicting Variable:** SAT score

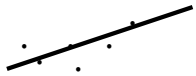
Georgia Tech

10

Linear Regression: General Model

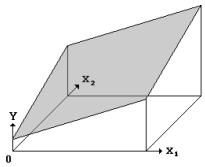
Simple linear regression

$Y = \beta_0 + \beta_1 X + \varepsilon$



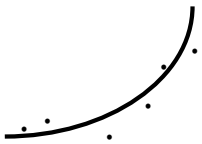
Multiple linear regression

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$



Polynomial Regression

$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$



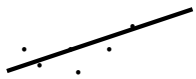
Georgia Tech

11

Linear Regression: General Model

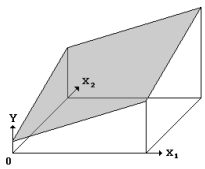
Simple linear regression

$Y = \beta_0 + \beta_1 X + \varepsilon$



Multiple linear regression

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$



Polynomial Regression

$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$

Whether a linear or polynomial model in X, we can estimate the relationship using linear regression.

Georgia Tech

12

Regression: Basics

A regression analysis is used for:

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

Linear Regression: The basis of what we will be talking about 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.

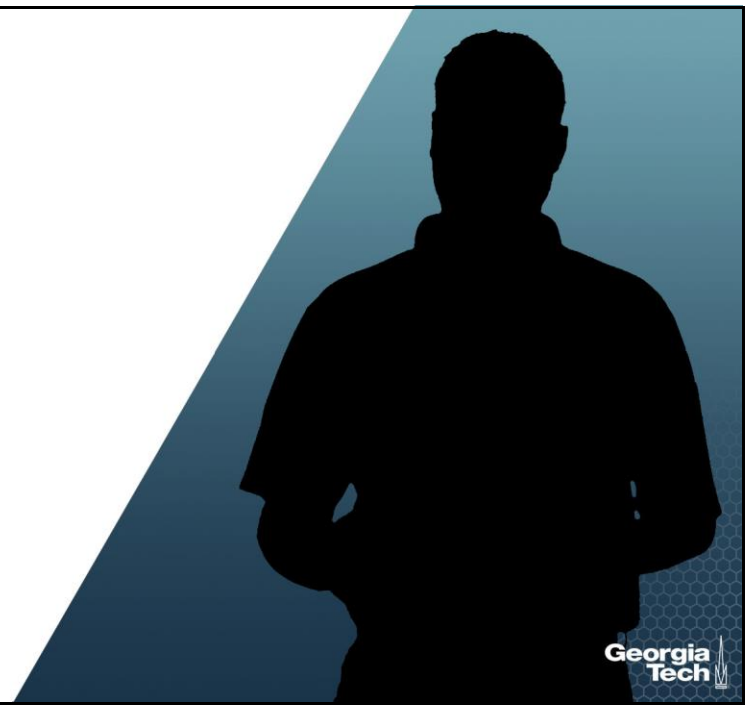
"*All models are wrong, but some are useful*." George Box

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

Georgia Tech

13

Summary



14