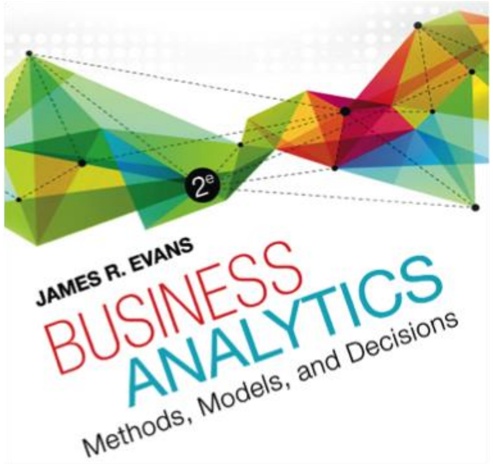



# Chapter 1 Introduction to Business Analytics


Modified and Shortened

# Contents

# Business Analytics

**Analytics** is the use of:

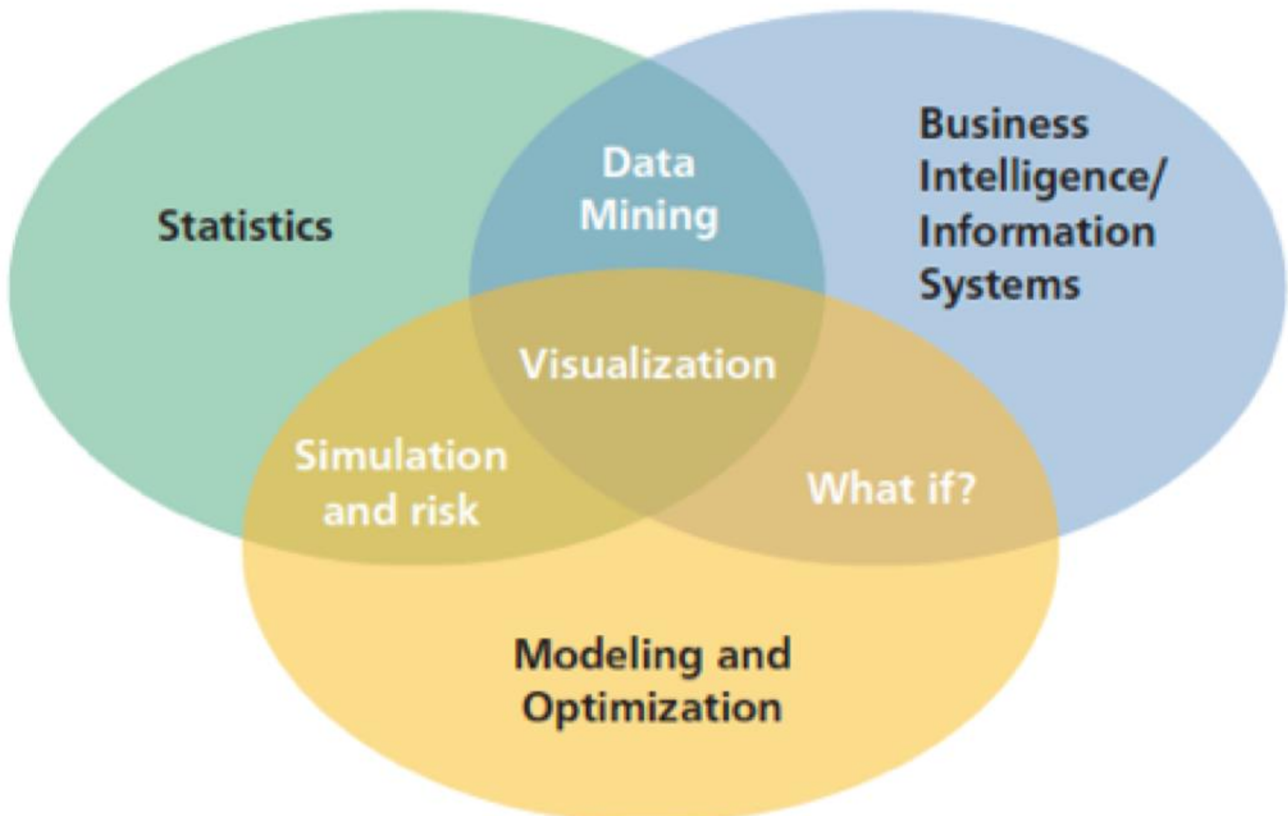
- data,
- information technology,
- statistical analysis,
- quantitative methods, and
- mathematical or computer-based models

to help managers gain improved insight about their business operations and **make better, fact-based decisions.**

# Examples of Applications

- **Pricing**
  - setting prices for consumer and industrial goods, government contracts, and maintenance contracts
- **Customer segmentation**
  - identifying and targeting key customer groups in retail, insurance, and credit card industries
- **Merchandising**
  - determining brands to buy, quantities, and allocations
- **Location**
  - finding the best location for bank branches and ATMs, or where to service industrial equipment
- **Social Media**
  - understand trends and customer perceptions; assist marketing managers and product designers

# A Visual Perspective of Business Analytics



# Impacts and Challenges

- **Benefits**
  - …reduced costs, better risk management, faster decisions, better productivity and enhanced bottom-line performance such as profitability and customer satisfaction.

- **Challenges**
  - …lack of understanding of how to use analytics, competing business priorities, insufficient analytical skills, difficulty in getting good data and sharing information, and not understanding the benefits versus perceived costs of analytics studies.

Privacy?

# Scope of Business Analytics

- **Descriptive analytics**: the use of data to understand past and current business performance and make informed decisions

- **Predictive analytics**: predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time.

- **Prescriptive analytics**: identify the best alternatives to minimize or maximize some objective

# Example 1.1: Retail Markdown Decisions

- Most department **stores clear seasonal inventory** by reducing prices.
- *Key question*: When to reduce the **price** and by how much to maximize revenue?

- Potential applications of analytics:
  - Descriptive analytics: examine historical data for similar products (prices, units sold, advertising, …)
  - Predictive analytics: predict sales based on price
  - Prescriptive analytics: find the best sets of pricing and advertising to maximize sales revenue

# Contents

- Introduction to Analytics
- **Tools**
- Data
- Models
- Problem solving with analytics

# Tools

- Database queries and analysis
- Spreadsheets
- Data visualization
- Dashboards to report key performance measures
- Data and Statistical methods
- Data Mining basics (predictive models)

In this course

- Simulation
- Forecasting
- Scenario and “what-if” analyses
- Optimization
- Text Mining
- Social media, web, and text analytics

# Software Support

- **SQL** various databases
- **Excel** Spreadsheets
- Tableau Software Simple drag and drop tools for visualizing data from spreadsheets and other databases.
- IBM Cognos Express An integrated business intelligence and planning solution designed to meet the needs of midsize companies, provides reporting, analysis, dashboard, scorecard, planning, budgeting and forecasting capabilities.
- SAS / SPSS / **Rapid Miner** Predictive modeling and data mining, visualization, forecasting, optimization and model management, statistical analysis, text analytics, and more using visual workflows.
- R / Python Advanced programing-based data preparation, analytics and visualization.

# Contents

# Data for Business Analytics

- **Data**: numerical or textual facts and figures that are collected through some type of measurement process.

- **Information**: result of analyzing data; that is, extracting **meaning** from data to support evaluation and decision making.

# Examples of Data Sources and Uses

- Internal
  - Annual reports
  - Accounting audits
  - Financial profitability analysis
  - Operations management performance
  - Human resource measurements

- External
  - Economic trends
  - Marketing research

- New developments: Web behavior – Social Media – Mobile - IOT
  - page views, visitor’s country, time of view, length of time, origin and destination paths, products they searched for and viewed, products purchased, what reviews they read, and many others.

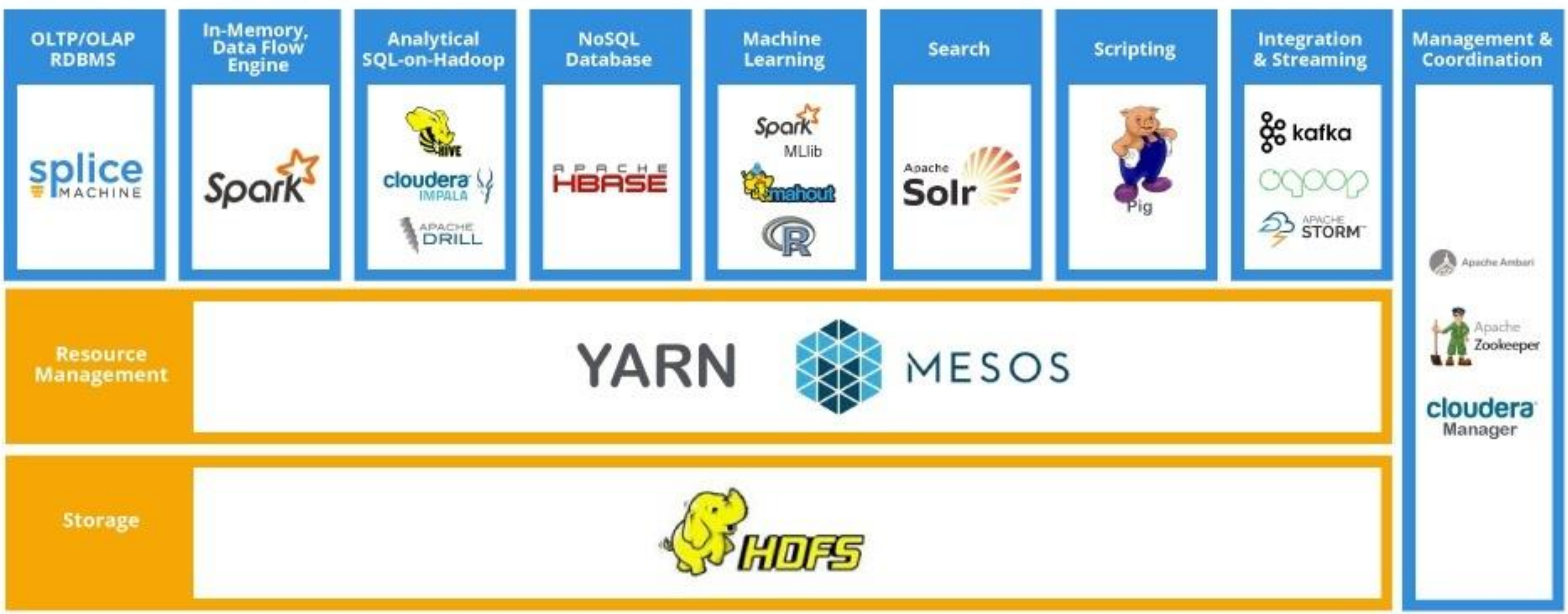
# Big Data

- **Big data** to refer to massive amounts of business data from a wide variety of sources, much of which is available in real time, and much of which is uncertain or unpredictable. IBM calls these characteristics **volume**, **variety**, **velocity,** and **veracity**.

"*The effective use of big data has the potential to transform economies, delivering a new wave of productivity growth and consumer surplus. Using big data will become a key basis of competition for existing companies, and will create new competitors who are able to attract employees that have the critical skills for a big data world.*" - McKinsey Global Institute, 2011

# Big Data

- Apache Hadoop Ecosystem for Big Data



# Data Sets and Databases

- **Database** - a collection of related tables containing records on people, places, or things.
  - In a database table the columns correspond to each individual element of data (called *fields*, or *attributes*), and the rows represent records of related data elements.

Extract (SQL)

- **Data set** - a collection of data (often a single “spread sheet” or data mining table).
  - Examples: Marketing survey responses, a table of historical stock prices, and a collection of measurements of dimensions of a manufactured item.

# Types of Data

- **Discrete** - derived from **counting** something.
  - For example, a delivery is either on time or not; an order is complete or incomplete; or an invoice can have one, two, three, or any number of errors. Some discrete metrics would be the proportion of on-time deliveries; the number of incomplete orders each day, and the number of errors per invoice.
- **Continuous** based on a **continuous scale of measurement.**
  - Any metrics involving dollars, length, time, volume, or weight, for example, are continuous.

# Measurement Scales

- **Categorical (nominal) data** - sorted into categories according to specified characteristics.
- **Ordinal data** - can be ordered or ranked according to some relationship to one another.
- **Interval data** - ordinal but have constant differences between observations and have arbitrary zero points.
- **Ratio data** - continuous and have a natural zero.

Operations have meaning

Equality: Are values the same?

Sort: Is one value larger/better? Median

Addition/Subtraction: E.g. Average

Multiplication: E.g. % change

# Example 1.3: Classifying Data Elements

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **Purchase Orders** | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | **Supplier** | **Order No.** | **Item No.** | **Item Description** | **Item Cost** | **Quantity** | **Cost per order** | **A/P Terms (Months)** | **Order Date** | **Arrival Date** |
| 4 | Hulkey Fasteners | Aug11001 | 1122 | Airframe fasteners | $ 4.25 | 19,500 | $ 82,875.00 | 30 | 08/05/11 | 08/13/11 |
| 5 | Alum Sheeting | Aug11002 | 1243 | Airframe fasteners | $ 4.25 | 10,000 | $ 42,500.00 | 30 | 08/08/11 | 08/14/11 |
| 6 | Fast-Tie Aerospace | Aug11003 | 5462 | Shielded Cable/ft. | $ 1.05 | 23,000 | $ 24,150.00 | 30 | 08/10/11 | 08/15/11 |
| 7 | Fast-Tie Aerospace | Aug11004 | 5462 | Shielded Cable/ft. | $ 1.05 | 21,500 | $ 22,575.00 | 30 | 08/15/11 | 08/22/11 |
| 8 | Steelpin Inc. | Aug11005 | 5319 | Shielded Cable/ft. | $ 1.10 | 17,500 | $ 19,250.00 | 30 | 08/20/11 | 08/31/11 |
| 9 | Fast-Tie Aerospace | Aug11006 | 5462 | Shielded Cable/ft. | $ 1.05 | 22,500 | $ 23,625.00 | 30 | 08/20/11 | 08/26/11 |
| 10 | Steelpin Inc. | Aug11007 | 4312 | Bolt-nut package | $ 3.75 | 4,250 | $ 15,937.50 | 30 | 08/25/11 | 09/01/11 |
| 11 | Durrable Products | Aug11008 | 7258 | Pressure Gauge | $ 90.00 | 100 | $ 9,000.00 | 45 | 08/25/11 | 08/28/11 |
| 12 | Fast-Tie Aerospace | Aug11009 | 6321 | O-Ring | $ 2.45 | 1,300 | $ 3,185.00 | 30 | 08/25/11 | 09/04/11 |
| 13 | Fast-Tie Aerospace | Aug11010 | 5462 | Shielded Cable/ft. | $ 1.05 | 22,500 | $ 23,625.00 | 30 | 08/25/11 | 09/02/11 |
| 14 | Steelpin Inc. | Aug11011 | 5319 | Shielded Cable/ft. | $ 1.10 | 18,100 | $ 19,910.00 | 30 | 08/25/11 | 09/05/11 |
| 15 | Hulkey Fasteners | Aug11012 | 3166 | Electrical Connector | $ 1.25 | 5,600 | $ 7,000.00 | 30 | 08/25/11 | 08/29/11 |
| | Categorical | Ordinal | Categorical | Categorical | Ratio | Ratio | Ratio | Ratio | Interval | Interval |

# Data Reliability and Validity

- **Reliability** - data are accurate and consistent.
- **Validity** - data measures what it is supposed to measure.

- Examples:
  - A tire pressure gage that consistently reads several pounds of pressure below the true value is **not reliable**, although it is valid because it does measure tire pressure.
  - The number of calls to a customer service desk might be counted correctly each day (and thus is a reliable measure) but **not valid** if it is used to assess customer dissatisfaction, as many calls may be simple queries.
  - A survey question that asks a customer to rate the quality of the food in a restaurant may be **neither reliable** (because different customers may have conflicting perceptions) **nor valid** (if the intent is to measure customer satisfaction, as satisfaction generally includes other elements of service besides food).

# Contents

# Models in Business Analytics

- **Model** - an abstraction or representation of a real system, idea, or object.
  - Often a **simplification** of the real thing.
  - Captures the **most important features.**
  - Can be a written or verbal description, a visual representation, a mathematical formula, or a spreadsheet.

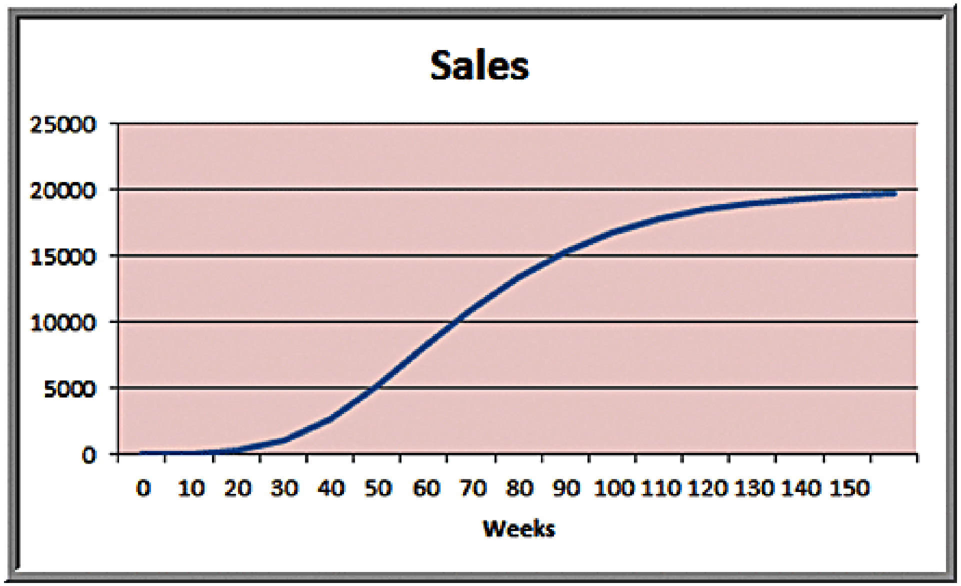
# Example 1.4: Three Forms of a Model

The sales of a new product, such as a first-generation iPad or 3D television, often follow a common pattern.

1. **Verbal description**: The rate of sales starts small as early adopters begin to evaluate a new product and then begins to grow at an increasing rate over time as positive customer feedback spreads. Eventually, the market begins to become saturated and the rate of sales begins to decrease.

# Example 1.4 (continued)

**2. Visual model**: A sketch of sales as an S-shaped curve over time



# Example 1.4 (continued)

**3. Mathematical model**:

$$S = ae^{be^{ct}}$$

where

- $S$ is sales,
- $t$ is time,
- $e$ is the base of natural logarithms, and
- $a$, $b$ and $c$ are constants that need to be estimated.

# From Data to Model

| Week | Price ($) | Coupon (0,1) | Advertising ($) | Store 1 Sales (Units) | Store 2 Sales (Units) | Store 3 Sales (Units) |
|---|---|---|---|---|---|---|
| 1 | $6.99 | 0 | $0 | 501 | 510 | 481 |
| 2 | $6.99 | 0 | $150 | 772 | 748 | 775 |
| 3 | $6,99 | 1 | $0 | 554 | 528 | 506 |
| 4 | $6.99 | 1 | $150 | 838 | 785 | 834 |
| 5 | $6.49 | 0 | $0 | 521 | 519 | 500 |
| 6 | $6.49 | 0 | $150 | 723 | 790 | 723 |
| 7 | $6.49 | 1 | $0 | 510 | 556 | 520 |
| 8 | $6.49 | 1 | $150 | 818 | 773 | 800 |
| 9 | $7.59 | 0 | $0 | 479 | 491 | 486 |
| 10 | $7.59 | 0 | $150 | 825 | 822 | 757 |
| 11 | $7.59 | 1 | $0 | 533 | 513 | 540 |
| 12 | $7.59 | 1 | $150 | 839 | 791 | 832 |
| 13 | $5.49 | 0 | $0 | 484 | 480 | 508 |
| 14 | $5.49 | 0 | $150 | 686 | 683 | 708 |
| 15 | $5.49 | 1 | $0 | 543 | 531 | 530 |
| 16 | $5.49 | 1 | $150 | 767 | 743 | 779 |

Learn model

**Model:** Sales = 500 – 0.05(price) + 30(coupons) + 0.08(advertising) + 0.25(price)(advertising)

If the price is $6.99, no coupons are offered, and no advertising is done (the experiment corresponding to week 1), the model estimates sales as

Sales = 500 - 0.05 × $6.99 + 30 × 0 + 0.08 × 0 + 0.25 × $6.99 × 0 = 500 units

**How do we find this model?**

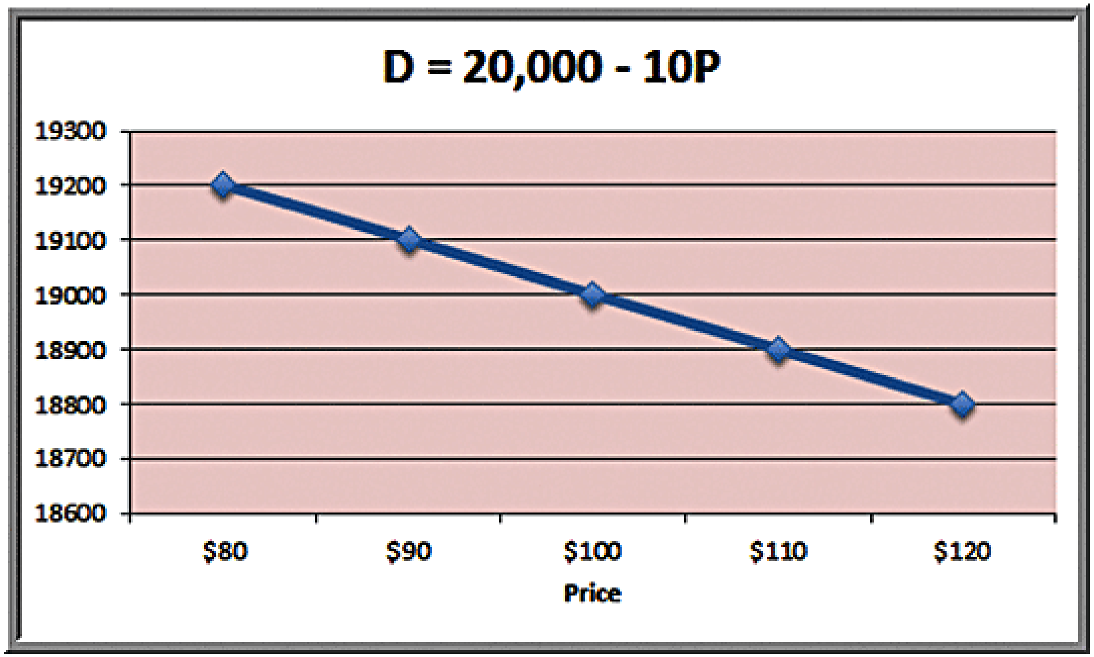
In this case: linear regression

# Model Assumptions

- Assumptions are made to
  - To **simplify** a model and make it more tractable; that is, able to be easily analyzed or solved.
  - To **add prior knowledge** about the relationship between variables.
- The task of the modeler is to select or build an appropriate model that best represents the behavior of the real situation.
- Example: economic theory tells us that demand for a product is negatively related to its price. Thus, as prices increase, demand falls, and vice versa.

# Example 1.9: A Linear Demand Prediction Model

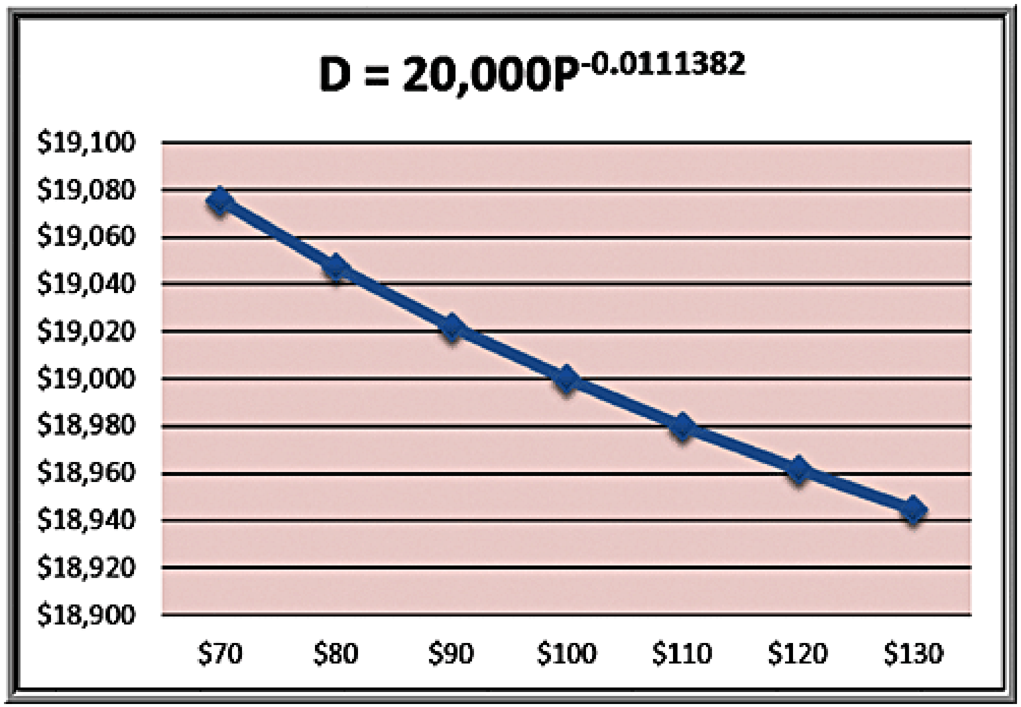
As price increases, demand falls.




**Issues:** Demand can become negative + empirical data has a poor fit.

# Example 1.10 A Nonlinear Demand Prediction Model

Assumes price elasticity is constant (constant ratio of % change in demand to % change in price)



# Uncertainty and Risk

- **Uncertainty** is **imperfect knowledge** (of what will happen in the future).
- **Risk** is the potential of (gaining or) losing something of value. It is the **consequence of actions** taken under uncertainty.

Often measured using standard deviation of variables. (=Deviation risk measure)

*"To try to eliminate risk in business enterprise is futile. Risk is inherent in the commitment of present resources to future expectations. Indeed, economic progress can be defined as the ability to take greater risks. The attempt to eliminate risks, even the attempt to minimize them, can only make them irrational and unbearable. It can only result in the greatest risk of all: rigidity."*

– Peter Drucker

# Prescriptive Decision Models

- **Prescriptive decision models** help decision makers identify the best solution.

- **Optimization** - finding values of decision variables that minimize (or maximize) something such as cost (or profit).
  - **Objective function** - the equation that minimizes (or maximizes) the quantity of interest.
  - **Constraints** - limitations or restrictions.
  - **Optimal solution** - values of the decision variables at the minimum (or maximum) point.

# Example 1.11: A Prescriptive Pricing Model

- A firm wishes to determine the best pricing for one of its products in order to maximize profit.
- Analysts determined the following **predictive model**:
  Sales = -2.9485(price) + 3240.9
  Total revenue = (price)(sales)
  Cost = 10(Sales) + 5000
- Identify the price that maximizes profit, subject to any constraints that might exist.

**max. Profit**
**s.t. Sales >= 0**
**Sales is integer**

# Contents

# Problem Solving With Analytics

1. Recognize a problem
2. Define the problem
3. Structure the problem
4. Analyze the problem
5. Interpret results and make a decision
6. Implement the solution

Focus of the remainder of this course

# Recognize a Problem

Problems exist when there is a gap between what is happening and **what we think should be happening.**

- For example, costs are too high compared with competitors.

# Define the Problem

- Clearly defining the problem is not a trivial task.

- Complexity increases when the following occur:
  - large number of courses of action
  - the problem belongs to a group and not an individual
  - competing objectives
  - external groups are affected
  - problem owner and problem solver are not the same person
  - time limitations exist

- **What is part of the problem? What not?**

# Structure the Problem

- Stating **goals** and objectives
- Characterizing the possible decisions
- Identifying any **constraints** or restrictions

# Analyze the Problem

- Analytics plays a major role.

- Analysis involves some sort of experimentation or solution process, such as evaluating different scenarios, analyzing risks associated with various decision alternatives, finding a solution that meets certain goals, or **determining an optimal solution**.

# Interpret Results and Make a Decision

- What do the results found by the model mean for the application?

- Models cannot capture every detail of the real problem. Managers must understand the **limitations of models** and their underlying assumptions and often incorporate judgment into making a decision.

# Implement the Solution

- Translate the results of the model back to the real world.

- Requires providing adequate resources, motivating employees, eliminating resistance to change, modifying organizational policies, and developing trust.

# How to do an analytics project? CRISP-DM Reference Model

- **C**ross **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining
- De facto standard for conducting data mining and knowledge discovery projects.
- Defines tasks and outputs.
- Now developed by IBM as the Analytics Solutions Unified Method for Data Mining/Predictive Analytics (ASUM-DM).
- SAS has SEMMA and most consulting companies use their own process.