Previews of TDWI course books offer an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews cannot be printed.

TDWI strives to provide course books that are content-rich and that serve as useful reference documents after a class has ended.

This preview shows selected pages that are representative of the entire course book; pages are not consecutive. The page numbers shown at the bottom of each page indicate their actual position in the course book. All table-of-contents pages are included to illustrate all of the topics covered by the course.

TDWI Business Analytics

Exploration, Experimentation, and Discovery

The Data Warehousing Institute takes pride in the educational soundness and technical accuracy of all of our courses. Please send us your comments—we'd like to hear from you. Address your feedback to:

email: info@tdwi.org

Publication Date: April 2011

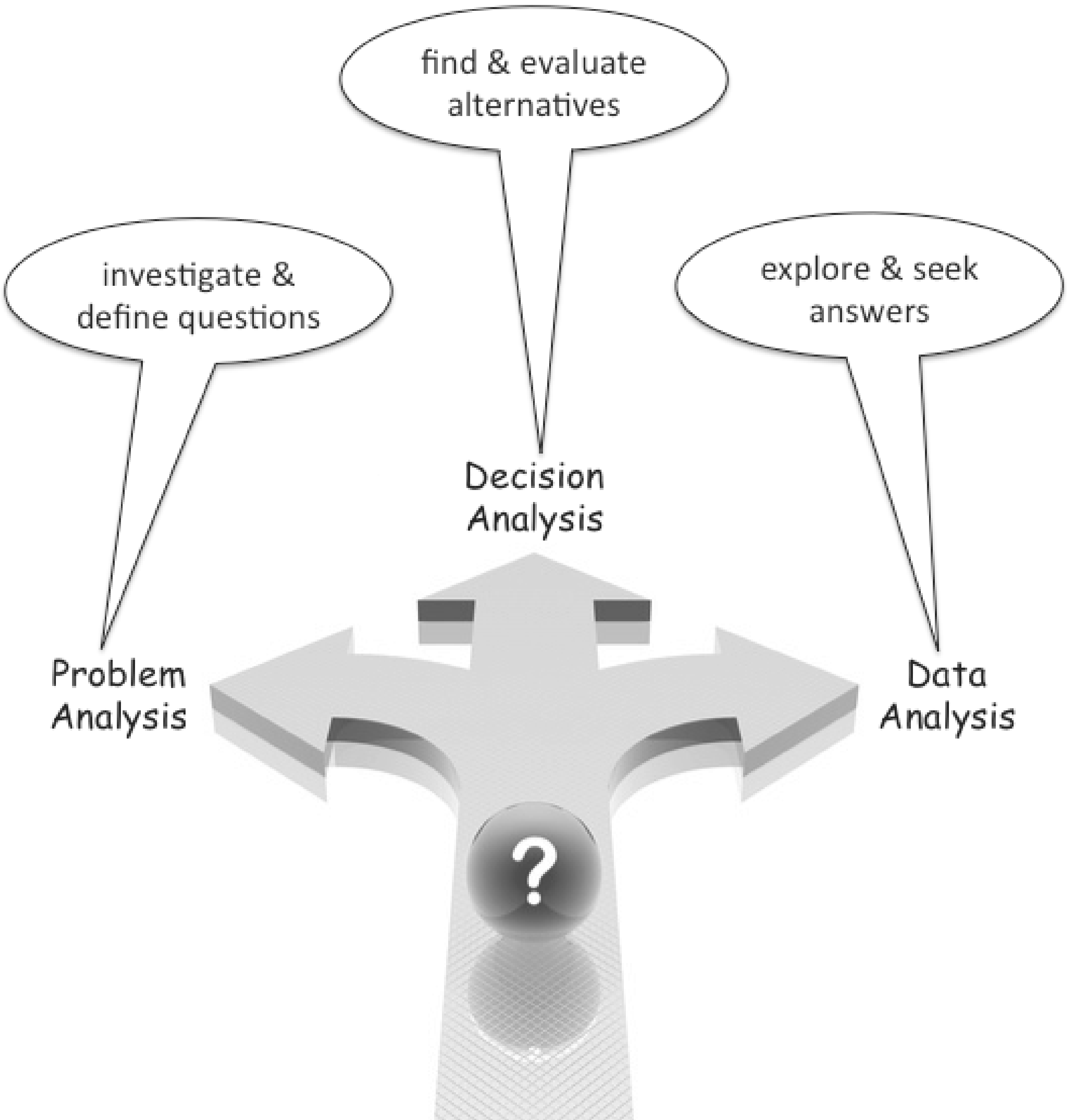
TABLE OF CONTENTS

Module 1

Introduction to Business Analytics

Analytic Processes

Three Kinds of Analysis



Analytic Processes

Three Kinds of Analysis

ANALYTICS WITH PURPOSE

The purpose of business analytics is to support business analysis processes, which typically fall into three categories:

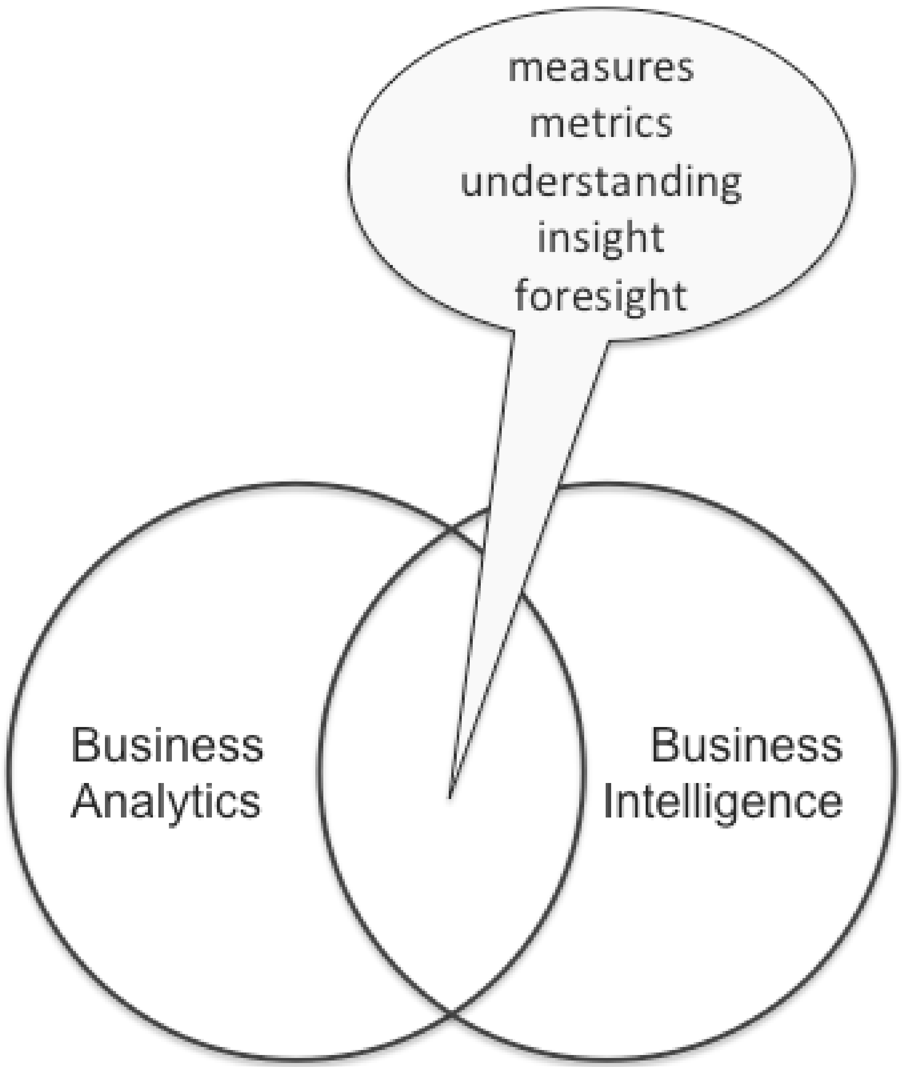
- Decision analysis is used to identify and evaluate alternative courses of action. Decision analysis uses data to develop and confirm understanding of cause and effect.

- Problem analysis is used to understand a poorly defined or ill-structured problem. The purpose of problem analysis is to add definition and structure – to increase understanding of a problem and identify questions that must be answered to resolve the problem. Deeper understanding of the problem is rarely found in data, so don't start problem analysis with the data. Once the problem is well defined, then data analysis and solution seeking can begin.

- Data analysis is the activity of exploring the data to seek answers for a well-defined and well-structured problem. Data analysis is a natural step in the process of decision analysis and in the solution-seeking activities of problem analyses.

The Analytics Environment

Business Analytics and Business Intelligence



The Analytics Environment

Business Analytics and Business Intelligence

BA vs. BI

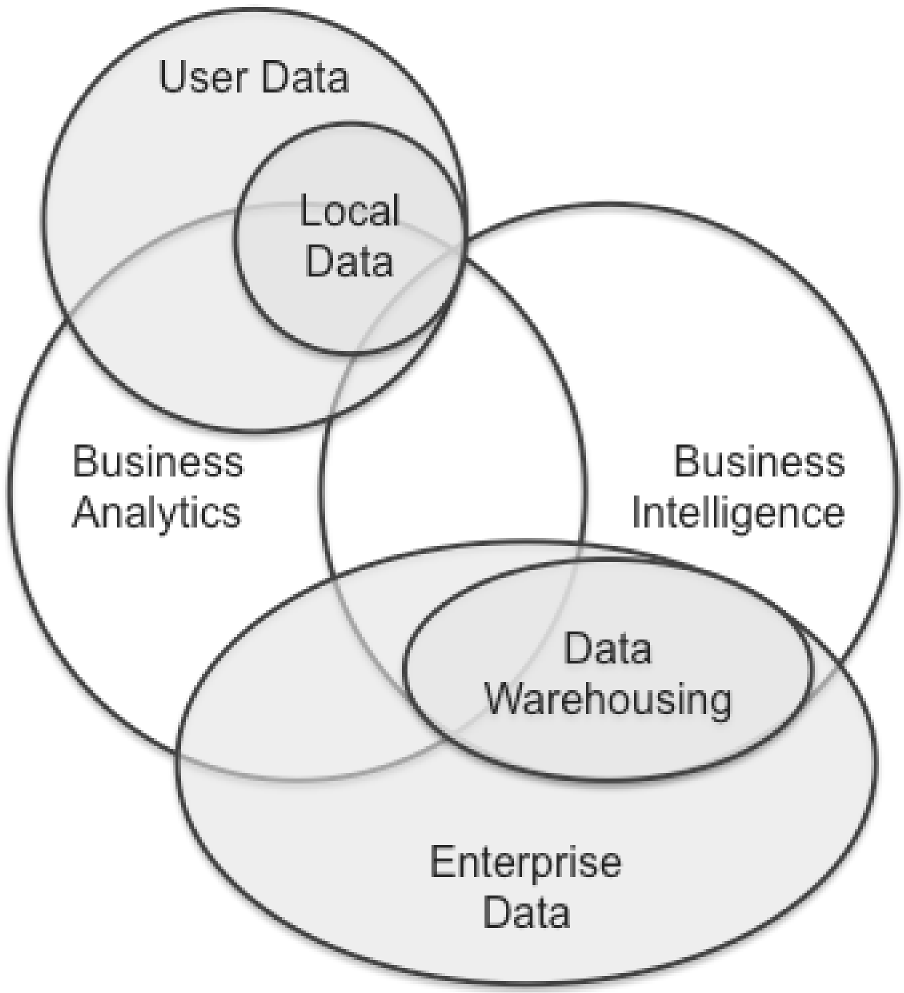
Some define business analytics as a subset of business intelligence, but that definition is misleading. Business intelligence and BA are overlapping fields but they are distinctly different in many ways. The similarities include use of data as measures, use of metrics to evaluate business behaviors, and the shared goals of understanding, insight, and foresight.

But the ways in which data is used and information created differ significantly. Business intelligence is strongly oriented toward querying, reporting, OLAP, and alerts. It answers questions such as what happened, how much, how often, and where the problems occur. Business analytics answers more complex questions about why things happen, the probability of events or conditions occurring, and what is realistic to expect for the future.

Business intelligence reporting typically ends with scorecards and dashboards, which are suited to business monitoring. Business analytics works with the data differently to support business planning. Monitoring can work with relatively static data structures and regularly published reports. Planning is more dynamic. Business planners must work constantly with unanticipated outcomes and other uncertainties. Their job is to mitigate risk and capitalize on opportunity. They interact with data in a different way from what traditional BI supports.

The Analytics Environment

Business Analytics and Data



The Analytics Environment

Business Analytics and Data

THE ROLES AND USES OF DATA

Business analytics makes extensive use of data, both quantitative and quantitative. Quantitative data uses numbers to express business events, behaviors, and trends as measures. Qualitative data segments a set of observations (data instances) by categories. Qualitative data is often referred to as categorical data. Both quantitative and categorical data have roles in statistical analysis. Analytic data comes from many sources.

ENTERPRISE DATA

Enterprise data is widely used across multiple business functions and is defined, managed, and governed from a global or enterprise-wide perspective. Enterprise data typically includes master reference data, metrics that are designated as key performance indicators (KPIs), operational data maintained by ERP systems, and data contained in an enterprise data warehouse. Analytic processes will sometimes, but not always, categorize based on enterprise reference data. KPIs are enterprise quantities, but they are more static and highly aggregated than is desirable for discovery-oriented analytics. The operational data found in ERP databases is typically the enterprise data most useful for analytics.

DATA WAREHOUSING

A data warehouse is a subset of enterprise data, so its utility for business analytics is subject to the same limits as for enterprise data. Many data warehouses use dimensional data architecture that creates additional barriers to analytic use. Dimensional data in OLAP is structured to support drill-down – working from summary to detail. The analytic modeler's needs are better met with direct access to the source data. The modeler works from detail, not from summary, and the key facts needed for iterative and exploratory analysis are often not available in cubes designed to support more predictable analysis patterns.

USER DATA

The data needed for a specific analytic activity is often not available anywhere in enterprise data resources. Business analytics and related planning and decision making are frequently isolated to one or a few business units, processes, or functions. The data found in departmental and end-user databases is valuable and commonly used in analytics.

LOCAL DATA

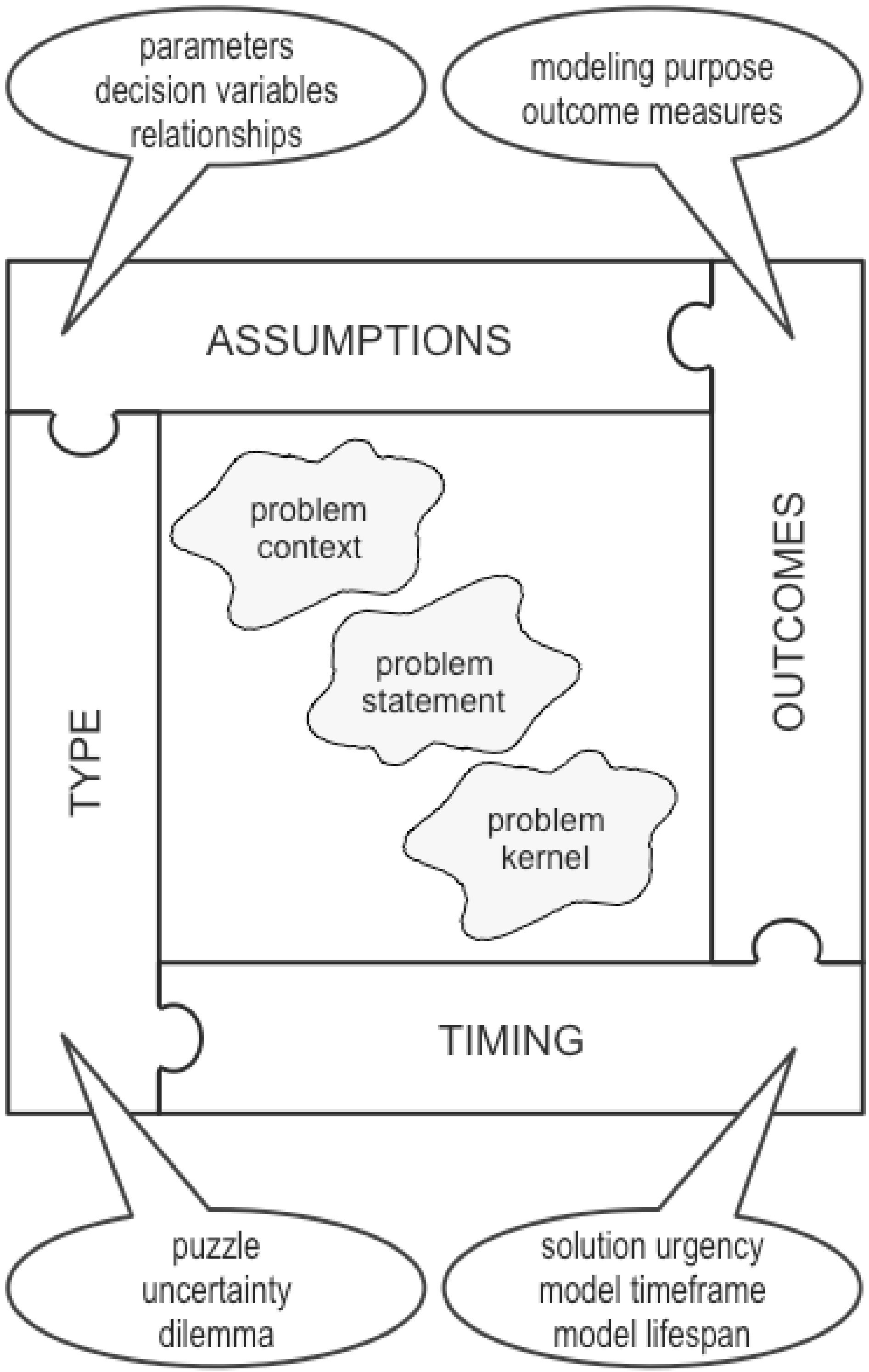
Local data is a distinct subset of user data that is central to analytics processes. This data is typically found in spreadsheets. It may be maintained locally to meet a user's needs, downloaded from data warehouses and then manipulated, created manually to meet a specific need, acquired or derived from external data sources, or generated by earlier analytic processes.

Module 2

Analytic Modeling

Problem Modeling

Framing the Problem



Problem Modeling

Framing the Problem

GETTING SPECIFIC

Most analytic problems begin with some uncertainty about the nature of the problem. Framing is the means to remove vagueness by adding specifics to our understanding of the problem. The goals of problem framing are to:

- Describe the context for the problem.
- Express a clear and non-ambiguous statement of the problem.
- Identify a problem "kernel" – the core element for which an answer or solution is sought.

Ask questions such as: What timeframes should be included? What products, processes, organizations, etc., should be considered? Also remember that the purpose of problem analysis is to build the right analytical models. Ask what decisions the models will support; what uncertainties the models will resolve, and what outcomes the models should produce.

Don't expect to frame a problem perfectly the first time. You will probably reframe it many times – as you begin problem modeling you will learn things that lead you back to reframe the problem repeatedly.

FRAMING DIMENSIONS

Look at each of four dimensions as a guide to framing a problem.

Problem Type classifies a problem as a puzzle, uncertainty, or dilemma. Puzzles are likely to have objective solutions. They are often financial, technical, mechanical, or mathematical. Uncertainties are characterized by unknown future conditions or outcomes, and solutions are contingent on reducing uncertainty through prediction. Dilemmas are the result of simultaneous commitment to incompatible goals. Dilemma solutions seek ways to reduce incompatibility by adjusting goals or removing interdependency.

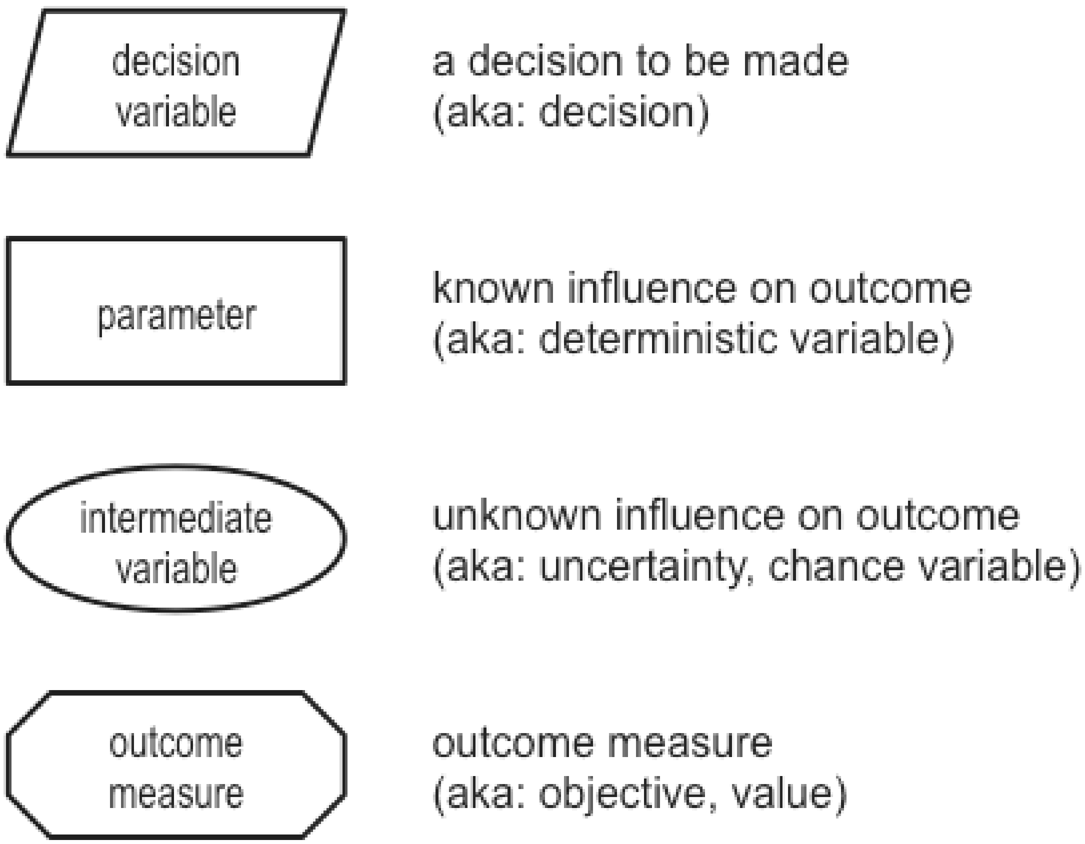
Timing considers the urgency of need for a solution, the timeframe (past and future) of modeling, and the expected lifespan of models.

Outcomes identify the purpose of modeling and the means by which outcomes are quantified. We'll look more at outcome measures later.

Assumptions are the decision variables, parameters, and relationships that are central to the modeling effort. These are described in greater detail later in the course.

Problem Modeling

Influence Diagramming



Problem Modeling

Influence Diagramming

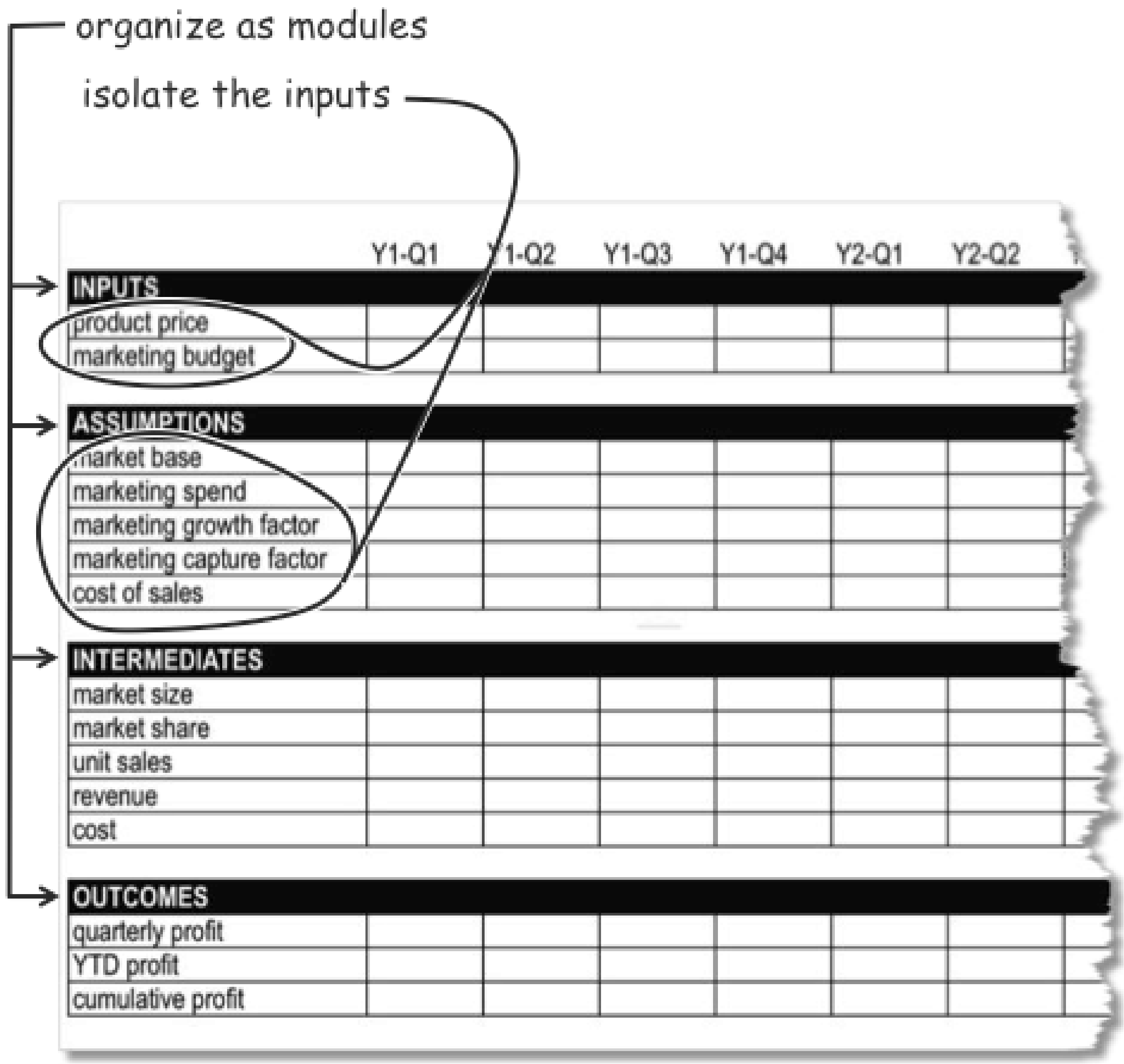
PICTURING THE PROBLEM

Influence diagramming is a graphical technique to understand what is known and unknown relative to decision making about a problem. The diagram represents components of the decision process and the dependencies or influences among those components. The four types of components include:

- Decision variables – the things about which a decision is needed such as pricing, staffing, investment, etc.
- Parameters –influences on outcome for which the values are known, such as budgets and regulations. Parameters are also known as deterministic variables.
- Intermediate variables – influences on outcome for which the values are unknown or uncertain. These are sometimes called chance variables.
- Outcome measures – a quantifiable component that describes the objective or value creation sought through analysis.

Solution Modeling

Spreadsheet Engineering



Solution Modeling

Spreadsheet Engineering

WHY SPREADSHEETS?

The technique of "spreadsheet engineering" is used in this course to illustrate many techniques of modeling analytic solutions. It is not recommended, nor is it practical, to meet all of your analytic needs with spreadsheets. Yet there are many good reasons to take a spreadsheet view:

- Much of business analysis, especially the analysis performed by business managers, is done with spreadsheets.
- Regardless of the analytic tool that you use, you will work with data that is organized in rows and columns and that has relationships among the cells.
- The kinds of variables illustrated with spreadsheets – inputs, assumptions, intermediates (unknowns), and outcomes – apply for every solution modeling problem and every analysis tool.

WHY ENGINEERING?

It may sound like an ominous term – spreadsheet engineering – but the real goal is to plan and design before building. All too often the initial form of a spreadsheet is determined by the source data that is available. We load the data and that determines the rows and columns. Then we take a circuitous path of fit-and-fix, poke-and-patch until we arrive at something close to a desired solution. A better alternative is to begin at the end – to start with the desired outcome and follow the chain backward to the inputs, carefully managing data relationships and dependencies along the way.

THE BASICS

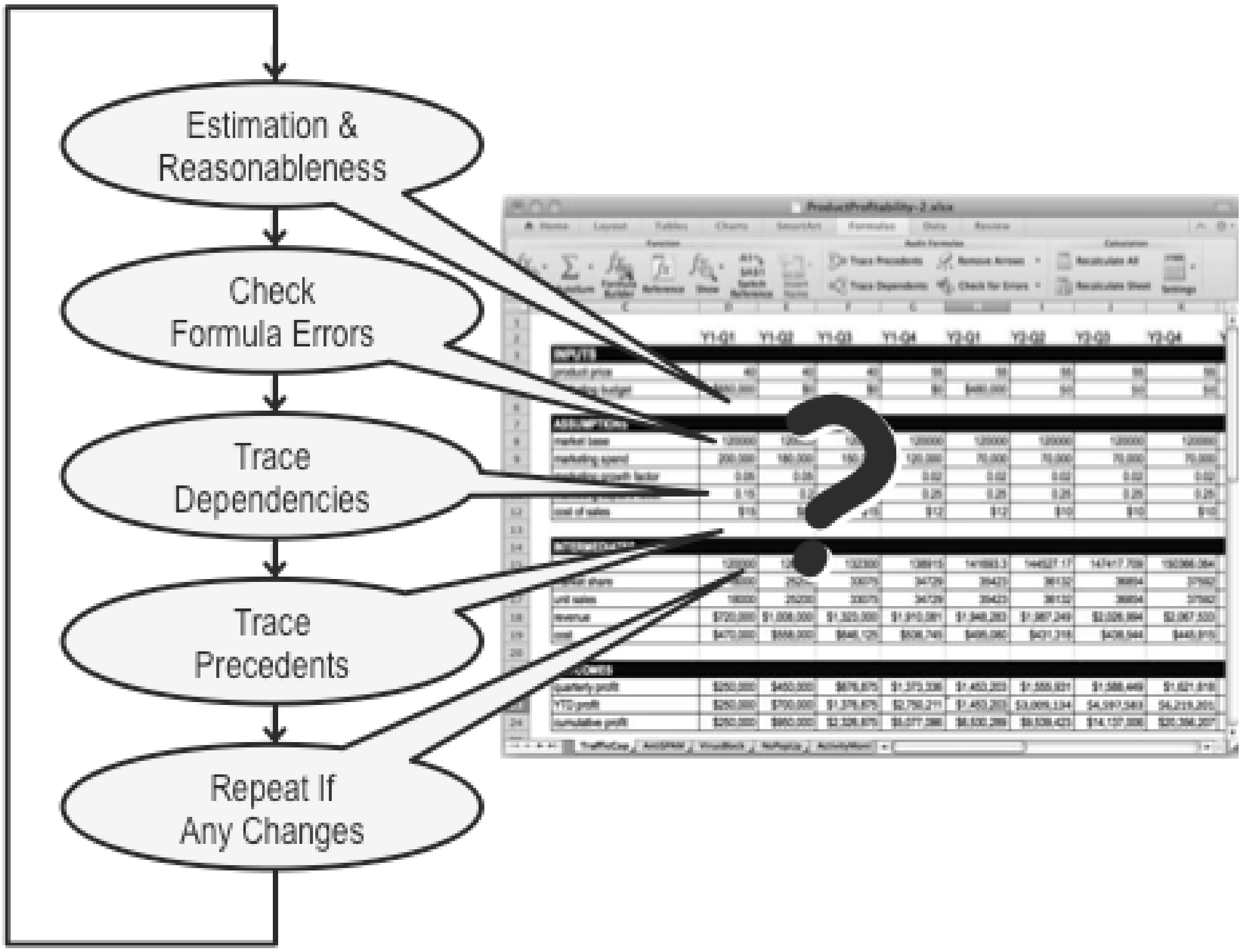
Begin with a quick sketch of the spreadsheet that separates components into modules that are "logically, physically, and visually distinct."[1] The modules may be sections within a worksheet as shown here, or they may be separate worksheets in a workbook for more complex models. The elements of an influence diagram – decision variables, deterministic variables, chance variables, and outcome measures – are a good first cut at modularity.

Use modularity to isolate the input variables. All of the numerical inputs to the model – decision variables, deterministic variables, and assumptions – should be grouped together and modularized. Ideally the inputs are placed at the top of the worksheet and dependencies cascade downward.

[1] *Modeling for Insight,* pp. 37-38, Powell and Batt

Model Refinement

Testing and Iteration



Model Refinement

Testing and Iteration

REVISE AND REPEAT

Testing is an iterative process where you need to re-test everything whenever you make a change to the model. Follow the sequence of:

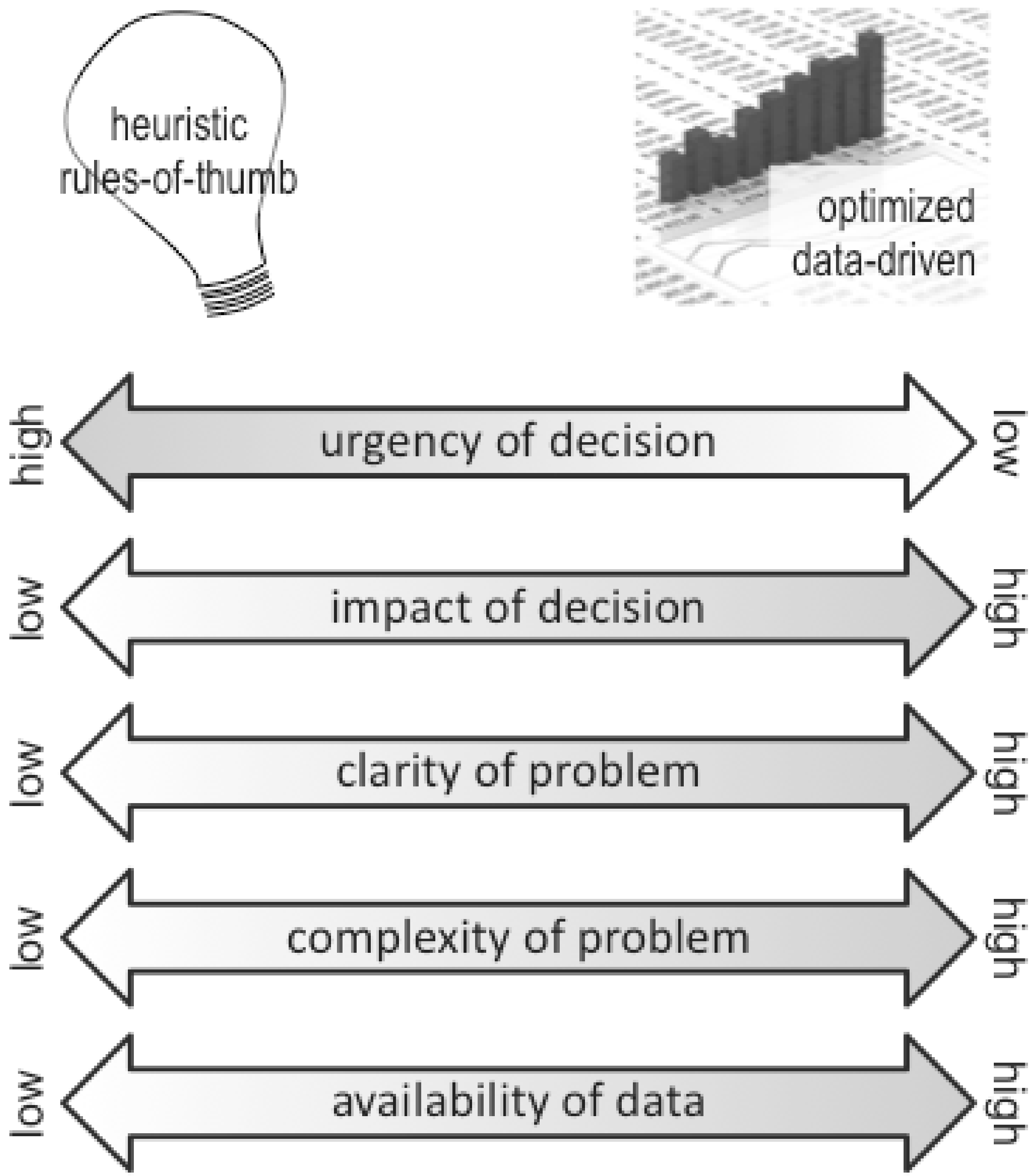
- Estimate and check reasonableness of results.
- Error-check every formula.
- Trace dependents of every input.
- Trace precedents of every outcome.
- When an error is found, correct the error and restart testing from the beginning.

Module 3

Discovery Analytics

Heuristic Analysis

Heuristics and Analytics



Heuristic Analysis

Heuristics and Analytics

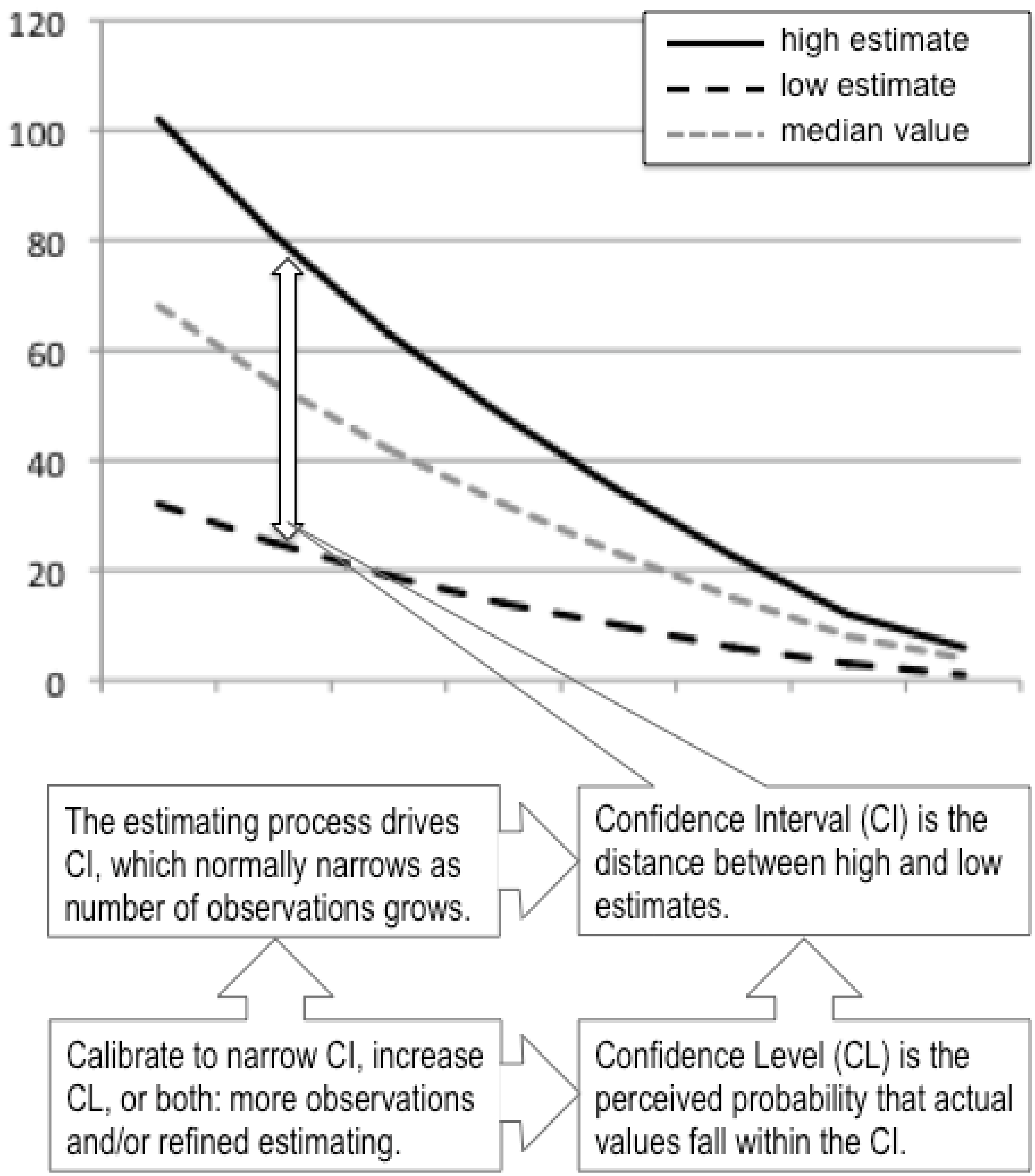
WHEN TO USE HEURISTICS

Most analytics are actually a blend of data-driven and heuristic reasoning. Consider heuristic analysis processes or heuristic elements in data-driven analytic models when:

- A decision is urgent and speed of analysis matters.
- The impact of the decision is sufficiently low that optimized and highly precise answers are not critical.
- The problem under analysis is unclear, vague, or poorly framed.
- The problem is of relatively low complexity, or complexity of a model can be reduced with use of heuristic variables.
- Base data for an important variable is not available and best-guess values must be used in the model.

Subjective Probability

Confidence Intervals and Confidence Levels



Subjective Probability

Confidence Intervals and Confidence Levels

TOPIC

For models with a long lifespan and repeated use it is practical and advisable to fine-tune subjective probabilities based on experience.

CONFIDENCE INTERVALS

One reality of measuring intangibles is the difficulty of providing a single number that expresses an intangible property. Describing intangibles in ranges is common. The size of the range – the difference between high and low estimates – is a confidence interval (CI). Probability of aggressive competition estimated as a range of 10-15% has a confidence interval of 5 percentage points.

CONFIDENCE LEVELS

The degree to which people believe that a true value falls within the range of estimation is known as confidence level (CL). “I am 55% certain that the probability of aggressive competition is between 10 and 15 percent” expresses a confidence level of 85%.

CI & CL TOGETHER

CI is expressed as a range. CL is expressed as a percentage. Together they describe the trust and utility of the numbers. Which of the following is the most useful combination of interval and level? Which is the least useful?

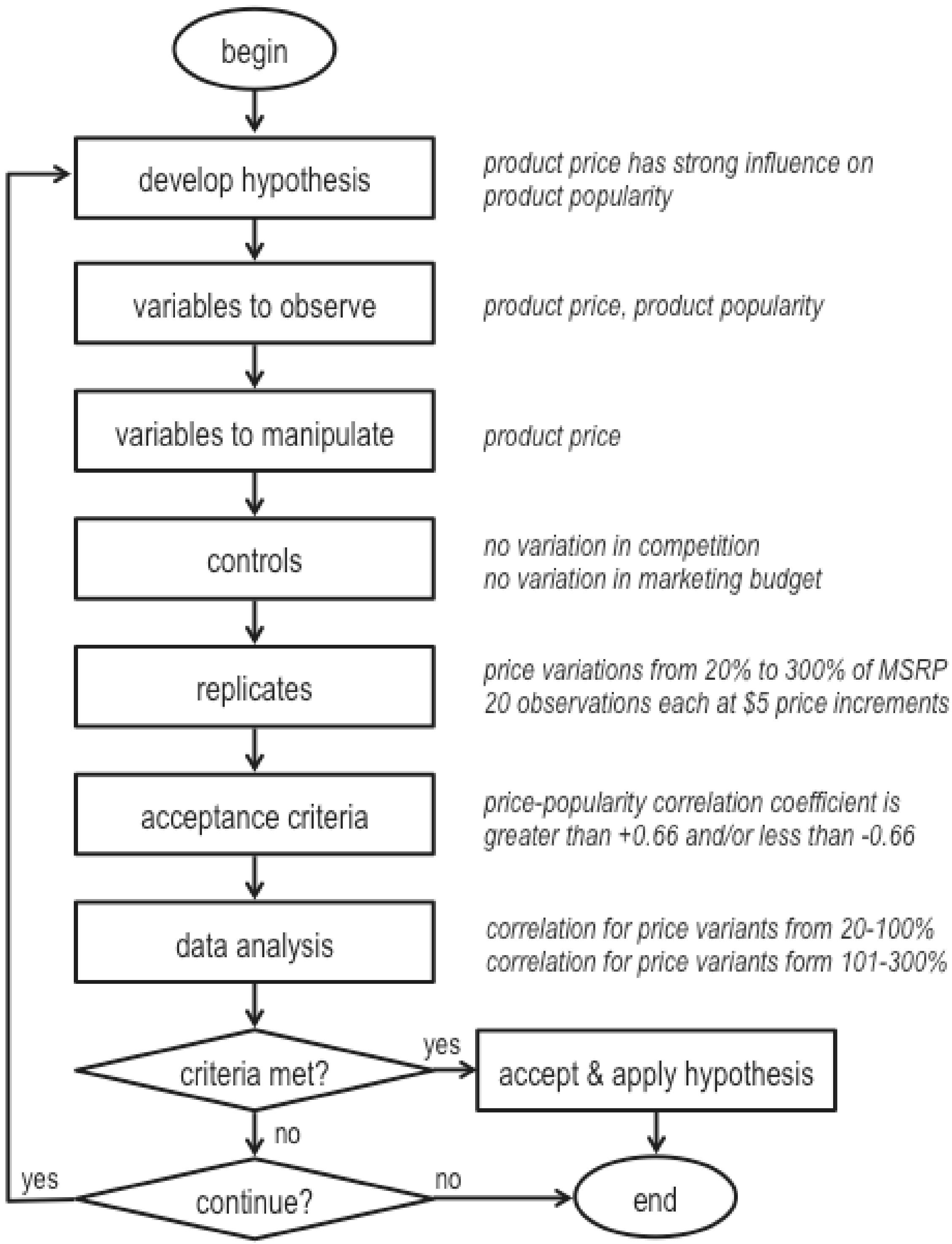
- 100% confident probability of competition is from 10 to 70%.
- 15% confident probability of competition is from 20 to 30%.
- 50% confident probability of competition is from 20 to 50%.
- 85% confident probability of competition is from 10 to 15%.
- 90% confident probability of competition is from 10 to 15%.

CALIBRATION

Calibration is the process of tuning CI and CL. Calibration occurs with time and with use of measures. Observe the contrast of actual values when available, or other indicators to qualitatively judge the accuracy estimates. Use the feedback of observations to tune estimating processes and narrow confidence intervals. Also use the feedback to influence well-informed confidence levels. Time, observation, and feedback are the means to calibrate measures – to achieve smaller confidence intervals and greater confidence levels.

Hypotheses and Experiments

Designing Experiments



Hypotheses and Experiments

Designing Experiments

ANALYTICAL EXPERIMENTS

An experiment is a procedure carried out under controlled conditions that is used to test a hypothesis. The purpose of an **experi**ment is to gain **experi**ence. In analytics, and often in business, it is impractical to test hypotheses with real-world experience. Instead, we use analytic models to experience and observe the results of systematically manipulating variables of interest.

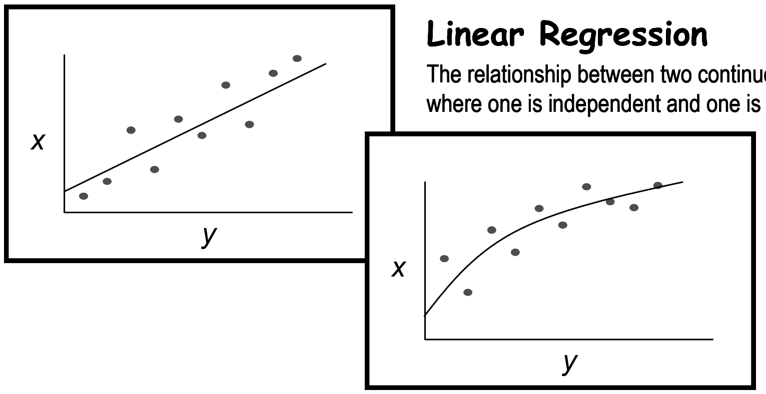
RELIABLE EXPERIMENTATION

Two key concepts appear in the description of experiments: controlled conditions and systematic manipulation of variables. To achieve these goals experiments must be designed carefully and purposefully. The design steps for analytic experiments include:

1. Develop the hypothesis.
   As described on the previous pages, a good testable hypothesis predicts future outcomes. A less predictive assertion leads to exploratory experimentation instead of confirming/disproving experiments. In the example shown here we're working with a subjective prediction of "strong influence." You'll see how the subjectivity and vagueness are overcome in a later design step.
2. Identify the variables to be observed.
   The statement of hypothesis is the source to identify variables. Here we want to observe product price and product popularity.
3. Identify the variables to be manipulated.
   Variables to manipulate must be a subset of variables to observe. In this simple example the variable to be manipulated is product price. Good experimentation manipulates only one variable at a time. When more than one variable is altered, cause and effect become uncertain. When you need to manipulate more than one variable, be prescriptive about sequence: e.g., first establish the minimum threshold, second establish the maximum threshold, and then observe effects of price variations.
4. Establish the controls.
   Here you identify the variables that must remain unchanged for the duration of the experiment. In this example competition must remain fixed because it influences market share, which eventually influences product popularity. Similarly, marketing budget must remain fixed to avoid corrupting the experiment with its influences.

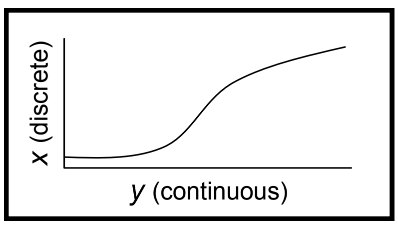
Exploratory Data Mining

Mining and Modeling



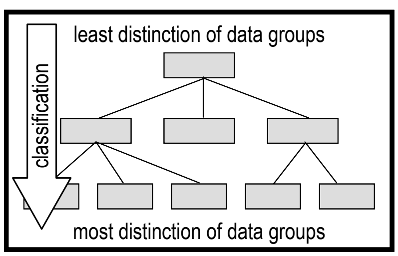
Linear Regression

The relationship between two continuous variables where one is independent and one is dependent.



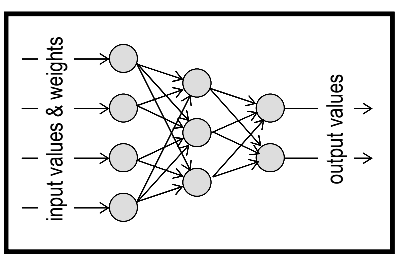
Logistic Regression

The relationship between two variables – a continuous independent variable and a non-continuous (discrete or categorical) dependent variable. Logistic regression is used to predict discrete outcomes such as probability of response to an advertising campaign.



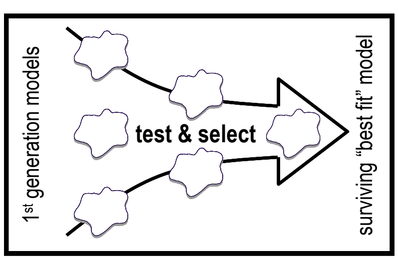
Classification Trees

Classification segments or partitions the set of data on the basis of defined criteria. Hierarchical classification of data helps to better understand the set of data under analysis and to optimize its use in predictive models. Classification is often used to differentiate data groups to create strong separation of dependent variable values when using regression models.



Neural Network

A process of pattern recognition and error minimization constructed as a collection of nodes that are arranged in layers. The first layer is the input layer and the last is the output layer. Through training and testing a neural network is tuned to predict output values from inputs within a prescribed minimum error tolerance.



Genetic Algorithms

An evolutionary process of "survival of the fittest" selection of the best fit model for a task. Many models are compared and adjusted over a sequence of iterative testing. Each model is tested to determine its ability to predict known outcomes. At each iteration models may be combined (mated), adjusted (mutated), or replicated and adapted (cloned).

Exploratory Data Mining

Mining and Modeling

MODEL BUILDING TECHNIQUES

There are many model-building techniques available that allow the modeler to create the structure best suited to the data mining application.

LINEAR REGRESSION

Linear regression describes a linear relationship between two continuous variables. The dependent variable *y* can be estimated by specifying a value for the independent variable *x*. Continuous outcomes can be estimated, such as sales quantities as a function of price.

LOGISTIC REGRESSION

Logistic regression models the relationship of a continuous independent variable *x* with a discrete dependent variable *y*. Discrete outcomes can be estimated, such as probability of responses to a marketing campaign as a function of the funds invested in the campaign.

CLASSIFICATION TREES

Classification trees develop hierarchical groupings of a set of data to determine the groups for each variable likely to produce an expected result. For example, assume a set of data with independent variables for income, number of family members, and education level plus a known dependent variable indicating whether the individual purchased the product. Classification determines the groupings or ranges for each independent variable to predict who is likely to purchase the product.

Decision trees (a variation of classification trees) are described later in this module.

NEURAL NETWORKS

Neural networks are non-linear, non-statistical tools used to model complex relationships between inputs and outputs or to find patterns in data. We'll clarify neural networks with an example.

GENETIC ALGORITHMS

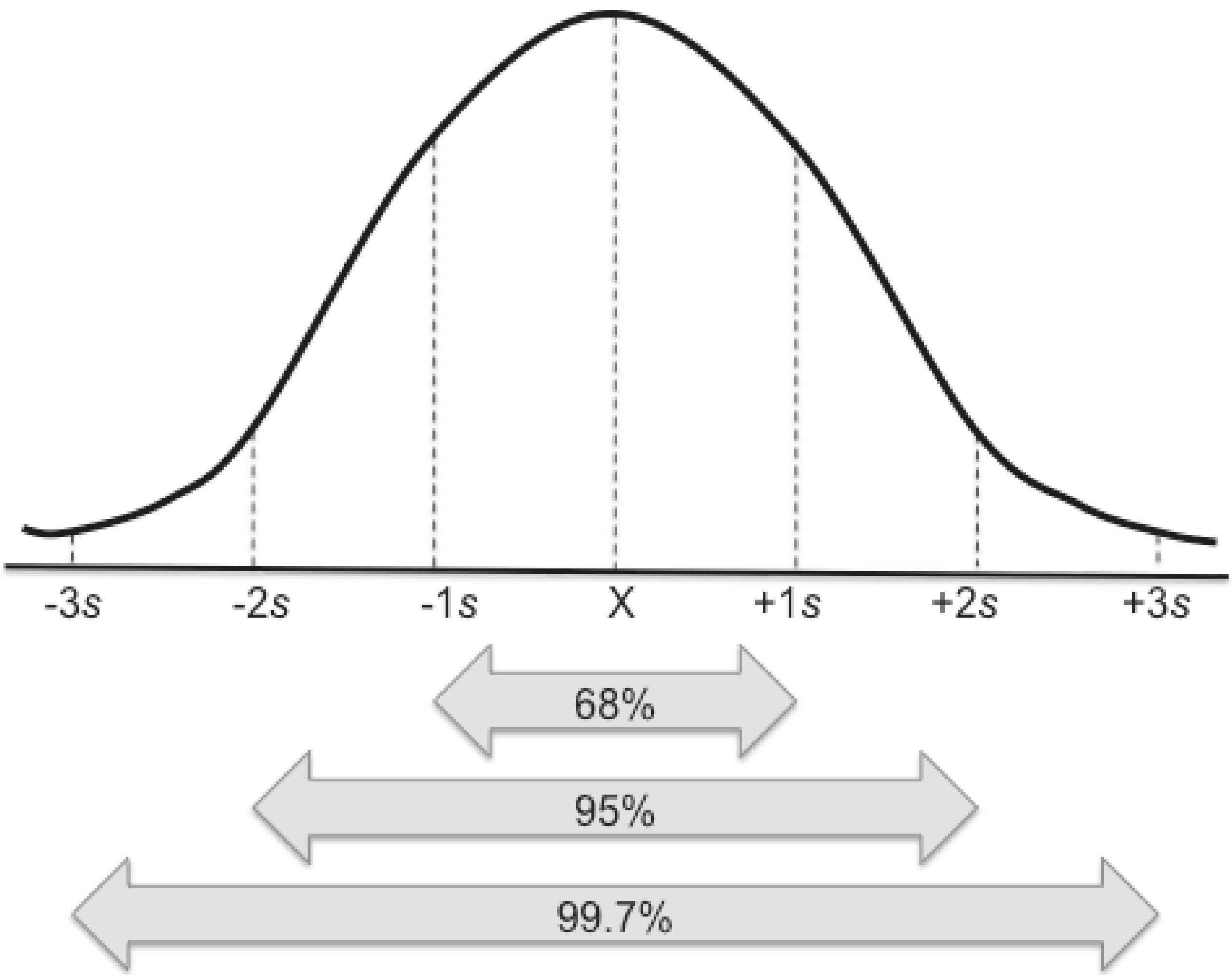
A genetic algorithm (GA) is a search technique used in computing to find true or approximate solutions to optimization and search problems. Genetic algorithms are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, selection, and crossover.

Module 4

Statistics and Data

Statistics and Analytics

Variance and Standard Deviation



Statistics and Analytics

Variance and Standard Deviation

RANGE

The range is the difference between the largest value of a data set and the smallest value. The major disadvantage of the range is that it does not include all of the values. Only the two most extreme values are included. They may be outliers and non-typical observations.

VARIANCE

Variance offers a stronger measure of location in a data set than range. Variance is calculated as the average of the squared deviations of all observations from the mean value. To calculate variance:

- Find the mean.
- Find the difference between each distinct value and the mean.
- Square each difference.
- Sum the squared differences.
- Divide the sum by the number of values minus one. The n-1 rule is used to correct for bias in a sample data set.

STANDARD DEVIATION

Variance is an intermediate value calculated to derive standard deviation. Both variance and standard deviation provide the same information; one can always be obtained from the other. Calculating standard deviation always involves calculating variance. Standard deviation is the square root of variance.

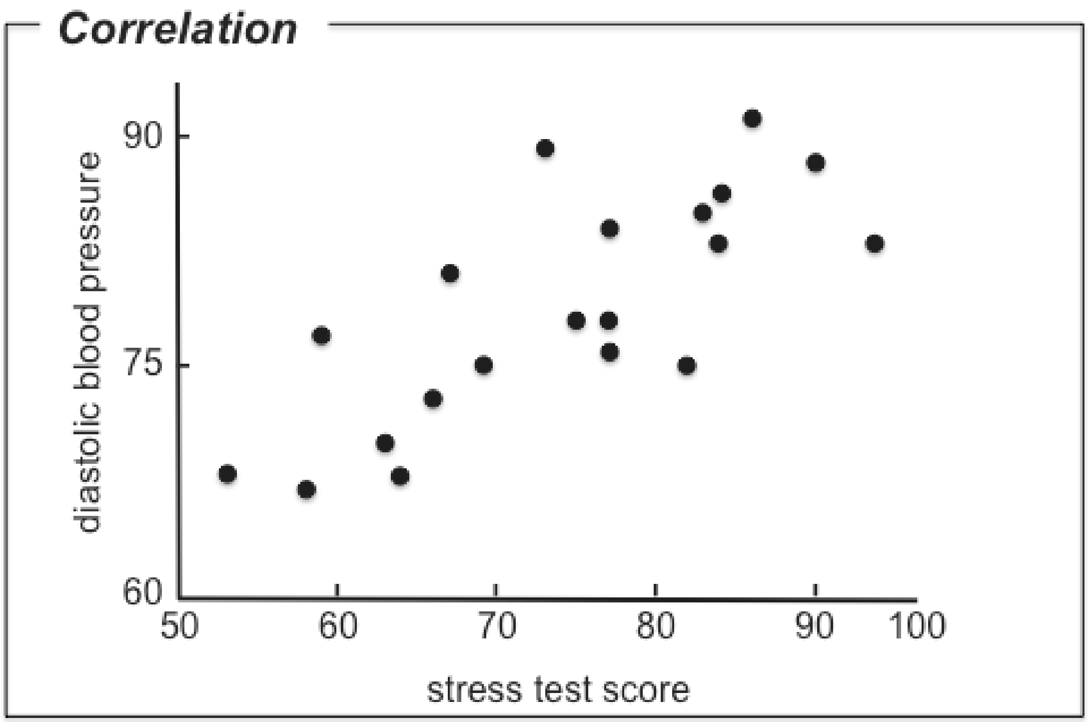
EMPIRICAL RULE

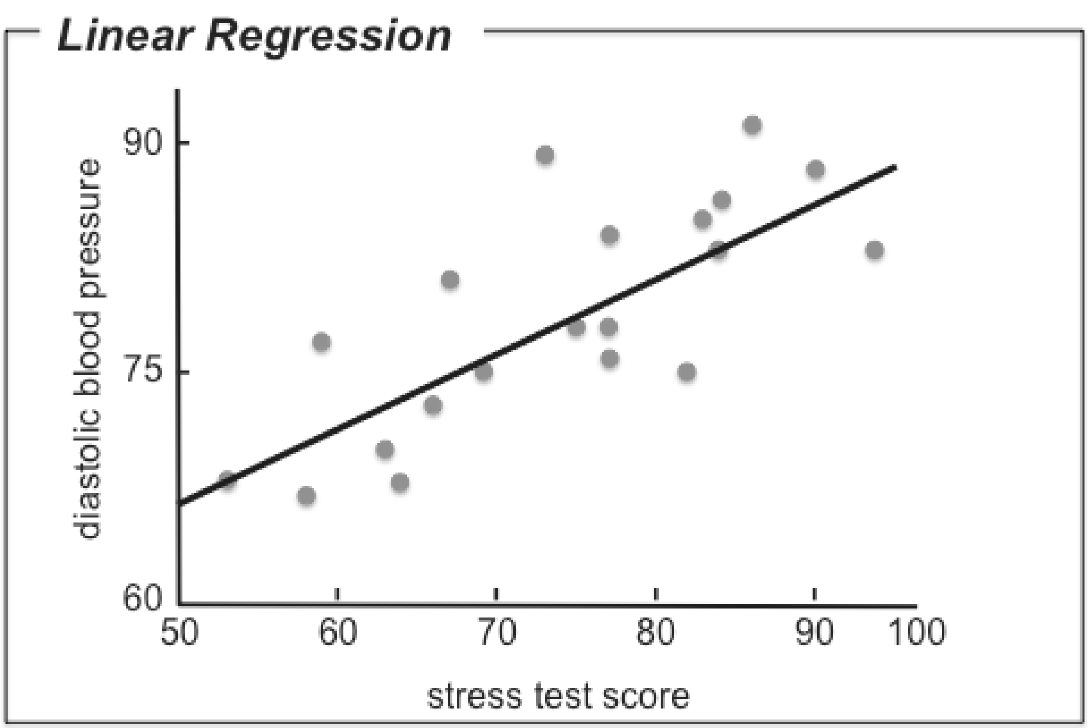
For a normal frequency distribution the following will be true:

- About 68% (0.6826895) of the observations are within 1 (+/-) standard deviation of the mean.
- About 95% (0.9544997) of the observations are within 2 (+/-) standard deviations of the mean.
- More than 99% (0.9973002) of all observations are within 3 (+/-) standard deviations of the mean.

Statistics and Analytics

Correlation and Regression





Statistics and Analytics

Correlation and Regression

CO-RELATION AND CO-VARIANCE

Correlation and linear regression are the most commonly used techniques to investigate the relationship between two quantitative variables. The goal of correlation analysis is to determine the degree to which two measurement variables co-vary, and to quantify the strength of the linear relationship between the variables. It is important to note that correlation does not necessarily imply causation. Two variables may illustrate strong correlation for a variety of reasons:

- The relationship is purely coincidental – a function of chance.
- The correlation results from a third causal or confounding variable.
- There is, in fact, a causal relationship between the variables.

Correlation is typically visualized with scatter graphs. The facing page illustrates the correlation of blood pressure with stress.

REGRESSION

In regression analysis, we are interested in the nature of the relationship between a dependent variable and an independent variable. The dependent variable is also known as a response variable – it responds to change in the independent variable.

Regression analysis fits a linear model to help estimate the probable response of the dependent variable to change in the value of the independent variable. The regression line is plotted using *the method of least squares*.

In the least squares method the difference between a data point and the regression line is calculated for each data point. The differences are squared and the sum of the squares calculated. The line that yields the lowest sum is the best-fit regression model.

Visual Design

Graphing Choices

	Whole and Parts	Simple Comparison	Multiple Comparison	Trends	Frequencies	Correlation	Spatial Relationships
Line Graph	no	maybe	maybe	**yes**	**yes**	**yes**	no
Column Graph	maybe	**yes**	**yes**	**yes**	**yes**	no	no
Bar Graph	maybe	**yes**	**yes**	**yes**	**yes**	no	no
Pictograph	maybe	**yes**	maybe	**yes**	maybe	no	maybe
Pie Chart	**yes**	**yes**	no	no	no	no	maybe
Donut Chart	**yes**	**yes**	maybe	no	no	no	maybe
Cosmograph	maybe	maybe	no	no	maybe	maybe	**yes**
Scatter Graph	no	maybe	no	maybe	**yes**	**yes**	no
Area Graph	maybe	maybe	no	maybe	maybe	maybe	maybe
Surface Graph	no	no	no	maybe	**yes**	maybe	maybe
Bubble Graph	no	maybe	yes	maybe	no	no	maybe

Visual Design

Graphing Choices

BACKGROUND

A graph's primary purpose is to describe and communicate the shapes that represent relationships among quantitative variables. Each relationship type has different communication needs. Understanding the key elements for different types of data relationships helps to make informed choices about selecting suitable graphing methods.

TYPES OF DATA RELATIONSHIPS

The following types of data relationships are commonly found in business analytics applications.

- Whole and Part
  - Show component breakdown of a quantitative variable (whole) into contribution levels at a detail level (parts)
- Simple Comparison
  - Show quantitative values of a single variable for visual comparison
- Multiple Comparison
  - Show quantitative values of multiple variables for visual comparison
- Trends
  - Show how a quantitative variable fluctuates or changes compared to a continuous categorical variable. The most common categorical variable for comparison is time.
- Frequencies
  - Present how a set of quantitative values is distributed across its full range from lowest to highest and how often each range occurs in the data set
- Correlation
  - Shows if the quantitative values from a paired set vary in relation to each other including the direction (positive or negative) and strength (strong or weak)
- Spatial Relationships
  - Shows how quantitative values are related in terms of location specific attributes.

USAGE

The table on the facing page provides a summary perspective about the suitability for different graph types. The graph types were described in an earlier section of this material.

Module 5

People and Technology

The Human Side of Analytics

Collaboration

The Human Side of Analytics

Collaboration

COLLECTIVE EFFORT

We discussed collaboration at the start of the course, but it makes sense to revisit it here. With multiple stakeholders – managers, analysts, modelers, and data experts – it is clear that business analytics is not an individual or solitary process. But the most effective analytics go well beyond teamwork to embrace collaborative analytics. Collaboration is the act of working jointly – two or more people combining their efforts toward achieving shared or intersecting goals. Collaborative analytics, then, is a set of analytic processes where the analysts work jointly and cooperatively to achieve shared or intersecting goals. Collaborative analytics includes data sharing, collective analysis, and coordinated decisions and actions. The goals of conventional analytics are to find answers and make decisions. Collaborative analytics encompasses these goals but seeks to achieve more – to increase visibility of important business facts and to improve alignment of decisions and actions across the entire business.

The Technology Landscape

Next-Generation Analytics Technology

FILL THE MIDDLE & BUILD THE BRIDGES

easy access to all types of local data
self-service connection to enterprise data
adaptable to changing requirements
columnar, in-memory data management
data in view all of the time
designed to facilitate discovery and exploration
intuitive interface – easy to learn and use
embedded statistical functions – analytics for non-statisticians
desktop, server-based, and cloud options
individual, departmental, and enterprise scalability
customization, collaboration, and personalization
managed metadata and traceability

The Technology Landscape

Next Generation Analytics Technology

BEST OF BOTH WORLDS

The next generation of analytic technologies will embrace the advantages of both centralized and decentralized approaches. These emerging tools fill the middle of the analytic technology continuum and build bridges between the two extremes. Modern analytic technologies should offer:

- Easy access to all types of local data, including end-user spreadsheets and databases.
- Self-service connection to enterprise-managed data.
- Easy adaptability to changing analysis requirements.
- High performance even with very large data sets through columnar, in-memory data management.
- Experimentation and discovery with all of the data visible all of the time. This is important because analysts need to work with data, not with abstractions of the data.
- Modeling and analysis features designed to facilitate discovery and exploration processes.
- Intuitive and familiar user interface in tools that are easy to learn and use. Much business analysis is performed by business managers who are comfortable with Excel. They're unlikely to change tools unless the ease-of-use and comfort level are similar.
- Embedded statistical functions making statistical analysis accessible to non-statisticians.
- Desktop, server-based, and cloud deployment options that support individual, departmental, and enterprise scaling.
- Customization and personalization with built-in features to support collaborative analytics.
- Managed metadata and traceability for those analyses that have legal, regulatory, or mission-critical impacts.

Module 6

Summary and Conclusion

Summary of Key Points

A Quick Review

- Analytics and performance management are related but different.
- Performance management is retrospective and published. Analytics is prospective, dynamic, and frequently ad hoc.
- Analytic processes include problem analysis, decision analysis, and data analysis.
- Business analytics intersects with business intelligence but is in many ways a separate discipline.
- Analytics is not data warehouse dependent. In fact much of the data used for analytics is local and user data.
- Influence diagramming is an effective problem modeling technique. Causal loop models and GQMM are also useful techniques.
- Spreadsheet engineering illustrates the fundamental modeling concepts that apply whether using spreadsheets or more advanced analytic tools.
- Modularity and parameterization are core techniques of solution modeling.
- Formulas are a critical component. They are expressions of data dependencies and are the most error-prone of solution model components.
- Testing, analysis, and refinement are important steps of the solution modeling process.
- Common model-driven analyses include base case, what-if, break-even, and optimization.
- Discovery analysis includes heuristics, subjective probability, hypothesis testing, experimentation, and exploratory data mining.
- Heuristics are useful to include intangible and uncertain variables in analytic models.
- Design of analytic experiments focuses on systematic manipulation of variables under controlled conditions.
- Statistics are the core of analytics. A basic understanding of statistics is essential.
- Data visualization is important and it is a learnable skill.

Summary of Key Points

A Quick Review

SUMMARY

The facing page summarizes many of the key points from this course. It can be a useful quick-reference as you find your place in the world of business analytics.