

Introduction to Data

ECO 230: Business and Economics
Communication and Research

Specific goals

- Appreciate the importance for data analysis in all business roles
- Define different scales of measurement and recognize these in data.

Learning objectives

- LO 1: Develop the ability to define a research or overall business problem.
- LO 3-A: Identify and compare the types of measurement scales used in conducting research.

Specific goals

- Appreciate the importance for data analysis in all business roles
- Define different scales of measurement and recognize these in data.

Learning objectives

- LO 1: Develop the ability to define a research or overall business problem.
- LO 3-A: Identify and compare the types of measurement scales used in conducting research.

Specific goals

- Appreciate the importance for data analysis in all business roles
- Define different scales of measurement and recognize these in data.

Learning objectives

- LO 1: Develop the ability to define a research or overall business problem.
- LO 3-A: Identify and compare the types of measurement scales used in conducting research.

Introduction to Data Analytics

Keeping Up with the Quants, Chapter 1

Davenport and Kim, 2013

Posted on [Canvas](#)

Scale of Measurement

Zikmund et al. (textbook)

Chapter 13, pp. 293-298

Everyone

Basic statistics is expected for all roles.

- Fastenal

We do employ business analysis, but a lot of roles have some amount of analysis and use statistics.

- Organic Valley

Everyone needs to know their numbers. We need to know what is the impact of our decisions.

- Marine Credit Union

The inability to use data will quickly lead to irrelevance.

- Northwestern Mutual

Everyone

Basic statistics is expected for all roles.

- Fastenal

We do employ business analysis, but a lot of roles have some amount of analysis and use statistics.

- Organic Valley

Everyone needs to know their numbers. We need to know what is the impact of our decisions.

- Marine Credit Union

The inability to use data will quickly lead to irrelevance.

- Northwestern Mutual

Everyone

Basic statistics is expected for all roles.

- Fastenal

We do employ business analysis, but a lot of roles have some amount of analysis and use statistics.

- Organic Valley

Everyone needs to know their numbers. We need to know what is the impact of our decisions.

- Marine Credit Union

The inability to use data will quickly lead to irrelevance.

- Northwestern Mutual

Everyone

Basic statistics is expected for all roles.

- Fastenal

We do employ business analysis, but a lot of roles have some amount of analysis and use statistics.

- Organic Valley

Everyone needs to know their numbers. We need to know what is the impact of our decisions.

- Marine Credit Union

The inability to use data will quickly lead to irrelevance.

- Northwestern Mutual

Everyone

Basic statistics is expected for all roles.

- Fastenal

We do employ business analysis, but a lot of roles have some amount of analysis and use statistics.

- Organic Valley

Everyone needs to know their numbers. We need to know what is the impact of our decisions.

- Marine Credit Union

The inability to use data will quickly lead to irrelevance.

- Northwestern Mutual

Good analytical skills are not enough

They get the data, but they have to know how to present it in ways that are useful

Were going to push for the next thing... what is the recommendation?

- Marine Credit Union

You can spend 15 hours doing the best data analysis...

but if you spend 10 minutes presenting it bad, they may not accept it as truth.

- Logistics Health Incorporated

Good analytical skills are not enough

They get the data, but they have to know how to present it in ways that are useful

Were going to push for the next thing... what is the recommendation?

- Marine Credit Union

You can spend 15 hours doing the best data analysis...

but if you spend 10 minutes presenting it bad, they may not accept it as truth.

- Logistics Health Incorporated

Good analytical skills are not enough

They get the data, but they have to know how to present it in ways that are useful

Were going to push for the next thing... what is the recommendation?

- Marine Credit Union

You can spend 15 hours doing the best data analysis...

but if you spend 10 minutes presenting it bad, they may not accept it as truth.

- Logistics Health Incorporated



- Hal Varian
Chief Economist, Google

The ability to take data... to extract value from it, to visualize it, to communicate it is going to be a hugely important skill in the next decades...

Because now we really do have essentially free and ubiquitous data...

So the complimentary scarce factor is the ability to understand that data and extract value from it.

You also want to be able to visualize the data, communicate the data, and utilize it effectively.



- Hal Varian
Chief Economist, Google

The ability to take data... to extract value from it, to visualize it, to communicate it is going to be a hugely important skill in the next decades...

Because now we really do have essentially free and ubiquitous data...

So the complimentary scarce factor is the ability to understand that data and extract value from it.

You also want to be able to visualize the data, communicate the data, and utilize it effectively.



- Hal Varian
Chief Economist, Google

The ability to take data... to extract value from it, to visualize it, to communicate it is going to be a hugely important skill in the next decades...

Because now we really do have essentially free and ubiquitous data...

So the complimentary scarce factor is the ability to understand that data and extract value from it.

You also want to be able to visualize the data, communicate the data, and utilize it effectively.



- Hal Varian
Chief Economist, Google

The ability to take data... to extract value from it, to visualize it, to communicate it is going to be a hugely important skill in the next decades...

Because now we really do have essentially free and ubiquitous data...

So the complimentary scarce factor is the ability to understand that data and extract value from it.

You also want to be able to visualize the data, communicate the data, and utilize it effectively.

Data analytics is defined as...

The use of data, statistical methods, data visualization, and predictive or explanatory models...

To drive **decision making** and **add value**.

Research is defined as...

A systematic inquiry that **investigates** hypotheses...

Answers questions with unknown answers, and...

Produces conclusions and recommendations.

Data analytics is defined as...

The use of data, statistical methods, data visualization, and predictive or explanatory models...

To drive **decision making** and **add value**.

Research is defined as...

A systematic inquiry that **investigates** hypotheses...

Answers questions with unknown answers, and...

Produces conclusions and recommendations.

Structured data

Characteristics

- Easily captures by columns and rows
- Columns correspond to *variables*
- Rows correspond to *observations*
- Readily able to analyze

Examples

- Rectangular spreadsheets
- Relational databases

This class focuses on structured data usually represented in rectangular spreadsheets

Structured data

Characteristics

- Easily captures by columns and rows
- Columns correspond to *variables*
- Rows correspond to *observations*
- Readily able to analyze

Examples

- Rectangular spreadsheets
- Relational databases

This class focuses on structured data usually represented in rectangular spreadsheets

Structured data

Characteristics

- Easily captures by columns and rows
- Columns correspond to *variables*
- Rows correspond to *observations*
- Readily able to analyze

Examples

- Rectangular spreadsheets
- Relational databases

This class focuses on structured data usually represented in rectangular spreadsheets

Structured data

Characteristics

- Easily captures by columns and rows
- Columns correspond to *variables*
- Rows correspond to *observations*
- Readily able to analyze

Examples

- Rectangular spreadsheets
- Relational databases

This class focuses on structured data usually represented in rectangular spreadsheets

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

- Dan Ariely, Professor of Psychology and Behavioral Economics
Duke University

What is it really?

- Volumes of digitally stored unstructured data
- With potential to be structured with automation
- With potential for predictive power, inform decision making

Examples

- 1.2 trillion Google searches in 2017
- 30 billion pieces of content uploaded to Facebook this month
- Click data and purchase data for 300+ million Amazon.com users
- Health sensors for a single cow resulting in 200 MB data per year



Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

- Dan Ariely, Professor of Psychology and Behavioral Economics
Duke University

What is it really?

- Volumes of digitally stored unstructured data
- With potential to be structured with automation
- With potential for predictive power, inform decision making

Examples

- 1.2 trillion Google searches in 2017
- 30 billion pieces of content uploaded to Facebook this month
- Click data and purchase data for 300+ million Amazon.com users
- Health sensors for a single cow resulting in 200 MB data per year



Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

- Dan Ariely, Professor of Psychology and Behavioral Economics
Duke University

What is it really?

- Volumes of digitally stored unstructured data
- With potential to be structured with automation
- With potential for predictive power, inform decision making

Examples

- 1.2 trillion Google searches in 2017
- 30 billion pieces of content uploaded to Facebook this month
- Click data and purchase data for 300+ million Amazon.com users
- Health sensors for a single cow resulting in 200 MB data per year

- **Concept:** Generalized idea of something that represents something interesting to someone (eg: Success of a marketing Facebook post)
- **Variable:** Carefully and specifically defined way to measure a concept (eg: Number of likes in first 48 hours)
- **Observation:** Most basic unit from which multiple variables are measured (eg: Individual post)
- **Observation level:** Description of observation unit

- **Concept:** Generalized idea of something that represents something interesting to someone (eg: Success of a marketing Facebook post)
- **Variable:** Carefully and specifically defined way to measure a concept (eg: Number of likes in first 48 hours)
- **Observation:** Most basic unit from which multiple variables are measured (eg: Individual post)
- **Observation level:** Description of observation unit

- **Concept:** Generalized idea of something that represents something interesting to someone (eg: Success of a marketing Facebook post)
- **Variable:** Carefully and specifically defined way to measure a concept (eg: Number of likes in first 48 hours)
- **Observation:** Most basic unit from which multiple variables are measured (eg: Individual post)
- **Observation level:** Description of observation unit

- **Concept:** Generalized idea of something that represents something interesting to someone (eg: Success of a marketing Facebook post)
- **Variable:** Carefully and specifically defined way to measure a concept (eg: Number of likes in first 48 hours)
- **Observation:** Most basic unit from which multiple variables are measured (eg: Individual post)
- **Observation level:** Description of observation unit

Scale of measurement: Definition for how variables are quantified or categorized

Four Scales of Measurement

- Nominal Data
- Ordinal Data
- Interval Data
- Ratio Data

Scale of measurement: Definition for how variables are quantified or categorized

Four Scales of Measurement

- Nominal Data
- Ordinal Data
- Interval Data
- Ratio Data

Nominal Data: qualitative data that consists of categories that cannot be ordered in a meaningful way

Example: Store location

- Inside mall
- Outdoor shopping complex
- Stand-alone store

Example: Worker classification

- Employed
- Self-employed
- Not in labor force
- Unemployed

Nominal Data: qualitative data that consists of categories that cannot be ordered in a meaningful way

Example: Store location

- Inside mall
- Outdoor shopping complex
- Stand-alone store

Example: Worker classification

- Employed
- Self-employed
- Not in labor force
- Unemployed

Nominal Data: qualitative data that consists of categories that cannot be ordered in a meaningful way

Example: Store location

- Inside mall
- Outdoor shopping complex
- Stand-alone store

Example: Worker classification

- Employed
- Self-employed
- Not in labor force
- Unemployed

Ordinal data: qualitative data, but order is meaningful, but quantitative values assigned to categories are meaningless

Example: Restaurant rating

- Excellent.
- Very good.
- Good.
- Poor.
- Very poor.

It is *not appropriate* to add, subtract, multiply and divide ordinal data (and therefore cannot take averages).

Ordinal data: qualitative data, but order is meaningful, but quantitative values assigned to categories are meaningless

Example: Restaurant rating

- Excellent.
- Very good.
- Good.
- Poor.
- Very poor.

It is *not appropriate* to add, subtract, multiply and divide ordinal data (and therefore cannot take averages).

Ordinal data: qualitative data, but order is meaningful, but quantitative values assigned to categories are meaningless

Example: Restaurant rating

- Excellent.
- Very good.
- Good.
- Poor.
- Very poor.

It is *not appropriate* to add, subtract, multiply and divide ordinal data (and therefore cannot take averages).

Interval data

- Order is meaningful, *and* distances are meaningful
- However, there is *no natural zero*

Example

Temperature

Appropriate Methods

- Can add and subtract, take means
- Can conduct most parametric statistics (estimate means, correlations, regression equations)
- **Cannot** divide values, take ratios

Interval data

- Order is meaningful, *and* distances are meaningful
- However, there is *no natural zero*

Example

Temperature

Appropriate Methods

- Can add and subtract, take means
- Can conduct most parametric statistics (estimate means, correlations, regression equations)
- **Cannot** divide values, take ratios

Interval data

- Order is meaningful, *and* distances are meaningful
- However, there is *no natural zero*

Example

Temperature

Appropriate Methods

- Can add and subtract, take means
- Can conduct most parametric statistics (estimate means, correlations, regression equations)
- **Cannot** divide values, take ratios

Ratio data

- Typical numeric data
- Order, differences, distances, and zero are all meaningful

Appropriate Methods

- Can add and subtract, take means
- Can conduct most parametric statistics (estimate means, correlations, regression equations)
- **Can** divide values, take ratios

Ratio data

- Typical numeric data
- Order, differences, distances, and zero are all meaningful

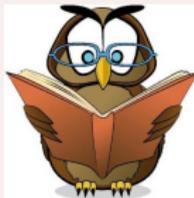
Appropriate Methods

- Can add and subtract, take means
- Can conduct most parametric statistics (estimate means, correlations, regression equations)
- **Can** divide values, take ratios

Step 1:
Identify a problem



Step 2: Review
existing knowledge



Step 3:
Organize Data



Step 4:
Analyze Data



Step 5:
Communicate



Step 6:
Recommend

