Statistical Analysis in Fin Mkts

MSF 502

Li Cai

ILLINOIS INSTITUTE OF TECHNOLOGY

1

Statistics and Data

ILLINOIS INSTITUTE OF TECHNOLOGY

Chapter 1 Learning Objectives (LOs)

LO 1.1: Describe the importance of statistics.

LO 1.2: Differentiate between descriptive statistics and inferential statistics.

LO 1.3: Explain the need for sampling and discuss various data types.

LO 1.4: Describe variables and various types of measurement scales.

Tween Survey

- Survey questions asked to 20 tweens:
  - Q1. Which radio station was playing on your drive to the ski resort?
  - Q2. Rate the quality of the food at the resort on a scale of 1 to 4.
  - Q3. What time should the main dining area close?
  - Q4. How much of your *own* money did you spend at the lodge today?

Tween Survey

- Here are the survey responses from the 20 tweens.

Tween	Q1	Q2	Q3	Q4	Tween	Q1	Q2	Q3	Q4
1	JAMN94.5	4	5:00 pm	20	11	JAMN94.5	3	3:00 pm	0
2	MIX104.1	2	5:00 pm	10	12	JAMN94.5	4	4:00 pm	5
3	KISS108	2	4:30 pm	10	13	KISS108	2	4:30 pm	5
4	JAMN94.5	3	4:00 pm	0	14	KISS108	2	5:00 pm	10
5	KISS108	1	3:30 pm	0	15	KISS108	3	4:00 pm	5
6	JAMN94.5	1	6:00 pm	25	16	JAMN94.5	3	6:00 pm	20
7	KISS108	2	6:00 pm	15	17	KISS108	2	5:00 pm	15
8	KISS108	3	5:00 pm	10	18	MIX104.1	4	6:00 pm	15
9	KISS108	2	4:30 pm	10	19	KISS108	1	5:00 pm	25
10	KISS108	3	4:30 pm	20	20	KISS108	2	4:30 pm	10

Tween Survey

1. Classify the tweens' responses into the appropriate measurement scale.

2. Extract useful information from each measurement scale.

3. Provide management with suggestions for improvement.

1.1 The Relevance of Statistics

LO 1.1 Describe the importance of statistics.

- With knowledge of statistics:

    - Avoid risk of making uninformed decisions and costly mistakes

    - Differentiate between sound statistical conclusions and questionable conclusions.

1.1 The Relevance of Statistics

- **Example 1**. Headline of newspaper states 'What global warming?' after record amounts of snow in 2010.

- **Problem with Conclusion**: Incorrect to draw conclusion based on one data point.

1.1 The Relevance of Statistics

- **Example 2.** A gambler predicts that he will roll a 7 on his next roll of the dice since he was unsuccessful in the last three rolls.

- **Problem with Conclusion.** The probability of rolling a 7 stays constant with each roll of the dice.

1.1 The Relevance of Statistics

- **Example 3**. A *Boston Globe* poll reported a 15-point lead for Martha Coakley in the election for U.S. senator for Massachusetts, implying an easy win for Coakley. Nine days later, Scott Brown wins.

- **Problem with Conclusion**. The *Globe*'s prediction was based on old information and included people that were unlikely to vote.

1.1 The Relevance of Statistics

- **Example 4**. The CFO of Starbucks Corp. claims that business is picking up since sales at stores open at least a year climbed 4% in the quarter ended December 27, 2009.

- **Problem with Conclusion.** The CFO overstated the company's financial position by failing to mention that Starbucks closed more than 800 stores over the past few years.

ILLINOIS INSTITUTE OF TECHNOLOGY

1.1 The Relevance of Statistics

- **Example 5**. Researchers showed that infants who sleep with a nightlight are much more likely to develop myopia.

- **Problem with Conclusion**. This is an example of the correlation-to-causation fallacy. Even if two variables are highly correlated, one does not necessarily cause the other.

ILLINOIS INSTITUTE OF TECHNOLOGY

1.2 What Is Statistics?

LO 1.2 Differentiate between descriptive statistics and inferential statistics.

- Statistics is the methodology of extracting useful information from a data set.

- To do good statistics, you must
  - Find the right data.
  - Use the appropriate statistical tools.
  - Clearly communicate the numerical information into written language.

ILLINOIS INSTITUTE OF TECHNOLOGY

1.2 What Is Statistics?

- Two branches of statistics
  - Descriptive Statistics
    - collecting, organizing, and presenting the data.
  - Inferential Statistics
    - drawing conclusions about a population based on sample data from that population.

1.2 What Is Statistics?

- Population
  - Consists of all items of interest.
- Sample
  - A subset of the population.
- A sample statistic is calculated from the sample data and is used to make inferences about the population parameter.

The Need for Sampling

- Reasons for sampling from the population
  - Too expensive to gather information on the entire population
  - Often impossible to gather information on the entire population

ILLINOIS INSTITUTE OF TECHNOLOGY

LO 1.3 Types of Data

- Cross-sectional data

  – Data collected by recording a characteristic of many subjects at the same point in time, or without regard to differences in time.

  – Subjects might include individuals, households, firms, industries, regions, and countries.

Types of Data

- Time series data
  - Data collected by recording a characteristic of a subject over several time periods.
  - Data can include daily, weekly, monthly, quarterly, or annual obs
  - This graph plots the U.S GDP growth rate from 1980 to 2010 - it is an example of time series data.
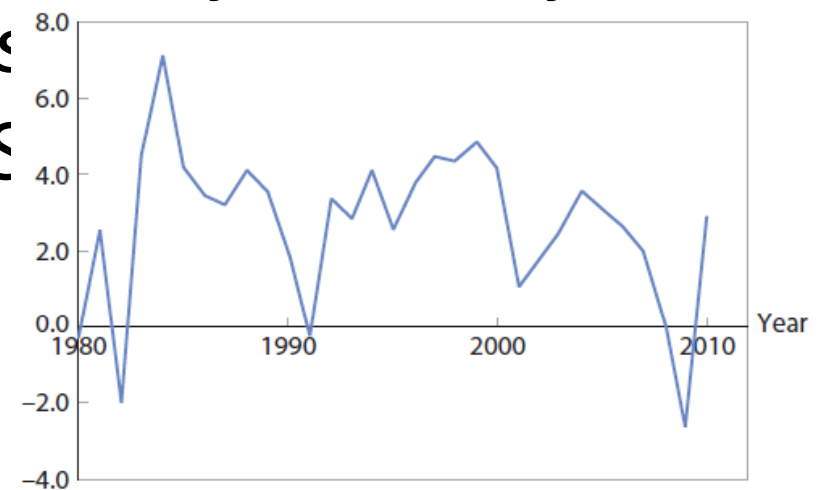
Source: Bureau of Economic Analysis.

ILLINOIS INSTITUTE OF TECHNOLOGY

- There is an abundance of data on the Internet. Here are a few websites for data.

Internet Site	Select Data Availability
Bureau of Economic Analysis (BEA)	National and regional data on gross domestic product (GDP) and personal income, international data on trade in goods and services.
Bureau of Labor Statistics (BLS)	Inflation rates, unemployment rates, employment, pay and benefits, spending and time use, productivity.
Federal Reserve Economic Data (FRED)	Banking, business/fiscal data, exchange rates, reserves, monetary base.
U.S. Census Bureau	Economic indicators, foreign trade, health insurance, housing, sector-specific data.
zillow.com	Recent home sales, home characteristics, monthly rent, mortgage rates.
finance.yahoo.com	Historical stock prices, mutual fund performance, international market data.
The New York Times, USA Today, The Wall Street Journal, The Economist, and *Fortune*	Poverty, crime, obesity, and plenty of business-related data.
espn.com	Professional and college teams' scores, rankings, standings, individual player statistics.

ILLINOIS INSTITUTE OF TECHNOLOGY

1.3 Variables and Scales of Measurement

LO 1.4 Describe variables and various types of measurement scales.

- A variable is the general characteristic being observed on an object of interest.

- Types of Variables
  - Qualitative – gender, race, political affiliation
  - Quantitative – test scores, age, weight
    - Discrete
    - Continuous

1.3 Variables and Scales of Measurement

- Types of Quantitative Variables
  - Discrete
    - A discrete variable assumes a countable number of distinct values.
    - Examples: Number of children in a family, number of points scored in a basketball game.

- Types of Quantitative Variables
  - Continuous
    - A continuous variable can assume an infinite number of values within some interval.
    - Examples: Height, investment return.

- Scales of Measure

| Nominal / Ordinal | → | **Qualitative Variables** |
| Interval / Ratio | → | **Quantitative Variables** |

ILLINOIS INSTITUTE OF TECHNOLOGY
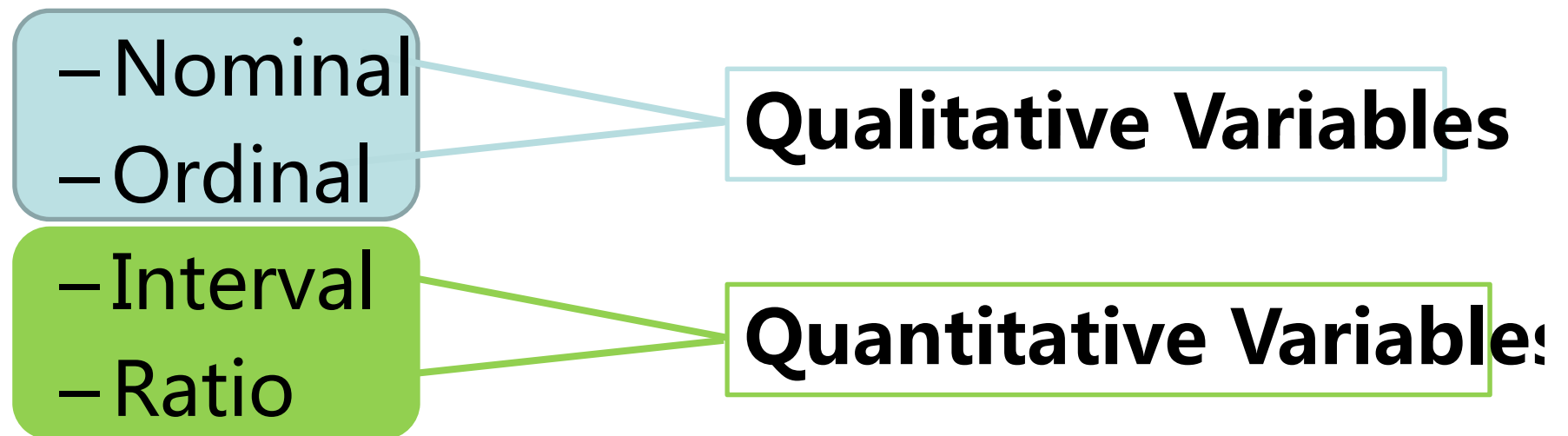
1.3 Variables and Scales of Measurement

- ## The Nominal Scale
  - – The least sophisticated level of measurement.
  - – Data are simply categories for grouping

Company	Exchange	Company	Exchange
3M (MMM)	NYSE	Intel (INTC)	Nasdaq
Alcoa (AA)	NYSE	IBM (IBM)	NYSE
American Express (AXP)	NYSE	Johnson & Johnson (JNJ)	NYSE
AT&T (T)	NYSE	JPMorgan Chase (JPM)	NYSE
Bank of America (BAC)	NYSE	Kraft Foods (KFT)	NYSE
Boeing (BA)	NYSE	McDonald's (MCD)	NYSE
Caterpillar (CAT)	NYSE	Merck (MRK)	NYSE
Chevron Corp. (CVX)	NYSE	Microsoft (MSFT)	Nasdaq
Cisco Systems (CSCO)	Nasdaq	Pfizer (PFE)	NYSE
Coca-Cola (KO)	NYSE	Proctor & Gamble (PG)	NYSE
DuPont (DD)	NYSE	Travelers (TRV)	NYSE
ExxonMobil (XOM)	NYSE	United Tech. Corp. (UTX)	NYSE
General Electric (GE)	NYSE	Verizon Comm. (VZ)	NYSE
Hewlett-Packard (HPQ)	NYSE	Wal-Mart (WMT)	NYSE
Home Depot (HD)	NYSE	Walt Disney (DIS)	NYSE

Source: http://www.finance.yahoo.com.

Qualitative values may be converted to quantitative values for analysis purposes.

Exchange	Number of Companies Trading on Exchange
0	3
1	27

ILLINOIS INSTITUTE OF TECHNOLOGY

- ## The Ordinal Scale
  - Ordinal data may be categorized *and* ranked with respect to some characteristic or trait.
    - For example, instructors are often evaluated on an ordinal scale (excellent, good, fair, poor).

  - Differences between categories are meaningless because the actual numbers used may be arbitrary.
    - There is no objective way to interpret the difference between instructor quality.

Example: Tweens Survey

- What is the scale of measurement of the radio station data?

Tween	Q1	Q2	Q3	Q4	Tween	Q1	Q2	Q3	Q4
1	JAMN94.5	4	5:00 pm	20	11	JAMN94.5	3	3:00 pm	0
2	MIX104.1	2	5:00 pm	10	12	JAMN94.5	4	4:00 pm	5
3	KISS108	2	4:30 pm	10	13	KISS108	2	4:30 pm	5
4	JAMN94.5	3	4:00 pm	0	14	KISS108	2	5:00 pm	10
5	KISS108	1	3:30 pm	0	15	KISS108	3	4:00 pm	5
6	JAMN94.5	1	6:00 pm	25	16	JAMN94.5	3	6:00 pm	20
7	KISS108	2	6:00 pm	15	17	KISS108	2	5:00 pm	15
8	KISS108	3	5:00 pm	10	18	MIX104.1	4	6:00 pm	15
9	KISS108	2	4:30 pm	10	19	KISS108	1	5:00 pm	25
10	KISS108	3	4:30 pm	20	20	KISS108	2	4:30 pm	10

Variables and Scales of Measurement

Example: Tweens Survey

- How are the data based on the ratings of the food quality similar to or different from the radio station data?

Tween	Q1	Q2	Q3	Q4	Tween	Q1	Q2	Q3	Q4
1	JAMN94.5	4	5:00 pm	20	11	JAMN94.5	3	3:00 pm	0
2	MIX104.1	2	5:00 pm	10	12	JAMN94.5	4	4:00 pm	5
3	KISS108	2	4:30 pm	10	13	KISS108	2	4:30 pm	5
4	JAMN94.5	3	4:00 pm	0	14	KISS108	2	5:00 pm	10
5	KISS108	1	3:30 pm	0	15	KISS108	3	4:00 pm	5
6	JAMN94.5	1	6:00 pm	25	16	JAMN94.5	3	6:00 pm	20
7	KISS108	2	6:00 pm	15	17	KISS108	2	5:00 pm	15
8	KISS108	3	5:00 pm	10	18	MIX104.1	4	6:00 pm	15
9	KISS108	2	4:30 pm	10	19	KISS108	1	5:00 pm	25
10	KISS108	3	4:30 pm	20	20	KISS108	2	4:30 pm	10

1.3 Variables and Scales of Measurement

- ## The Interval Scale
  - Data may be categorized *and* ranked with respect to some characteristic or trait.
  - Differences between interval values are equal and meaningful. Thus the arithmetic operations of addition and subtraction are meaningful.
  - No "absolute 0" or starting point defined. Meaningful ratios may not be obtained.

1.3 Variables and Scales of Measurement

- ## The Interval Scale
  - – For example, consider the Fahrenheit scale of temperature.
  - – This scale is interval because the data are ranked and differences (+ or –) may be obtained.
  - – But there is no "absolute 0" (What does 0°F mean?)

    What does $\dfrac{80°F}{40°F}$ mean?

1.3 Variables and Scales of Measurement

- ## The Ratio Scale
  - The strongest level of measurement.
  - Ratio data may be categorized *and* ranked with respect to some characteristic or trait.
  - Differences between interval values are equal and meaningful.
  - There *is* an "absolute 0" or defined starting point. "0" *does* mean "the absence of ..." Thus, meaningful ratios may be obtained.

1.3 Variables and Scales of Measurement

- ## The Ratio Scale
  - ### The following variables are measured on a ratio scale:
    - General Examples: Weight, Time, and Distance
    - Business Examples: Sales, Profits, and Inventory Levels

Example: Tweens Survey

- How are the time data classified? In what ways do the time data differ from ordinal data? What is a potential weakness of this measurement scale?

Tween	Q1	Q2	Q3	Q4	Tween	Q1	Q2	Q3	Q4
1	JAMN94.5	4	5:00 pm	20	11	JAMN94.5	3	3:00 pm	0
2	MIX104.1	2	5:00 pm	10	12	JAMN94.5	4	4:00 pm	5
3	KISS108	2	4:30 pm	10	13	KISS108	2	4:30 pm	5
4	JAMN94.5	3	4:00 pm	0	14	KISS108	2	5:00 pm	10
5	KISS108	1	3:30 pm	0	15	KISS108	3	4:00 pm	5
6	JAMN94.5	1	6:00 pm	25	16	JAMN94.5	3	6:00 pm	20
7	KISS108	2	6:00 pm	15	17	KISS108	2	5:00 pm	15
8	KISS108	3	5:00 pm	10	18	MIX104.1	4	6:00 pm	15
9	KISS108	2	4:30 pm	10	19	KISS108	1	5:00 pm	25
10	KISS108	3	4:30 pm	20	20	KISS108	2	4:30 pm	10

ILLINOIS INSTITUTE OF TECHNOLOGY

1.3 Variables and Scales of Measurement

Example: Tweens Survey

- What is the measurement scale of the money data?

Tween	Q1	Q2	Q3	Q4	Tween	Q1	Q2	Q3	Q4
1	JAMN94.5	4	5:00 pm	20	11	JAMN94.5	3	3:00 pm	0
2	MIX104.1	2	5:00 pm	10	12	JAMN94.5	4	4:00 pm	5
3	KISS108	2	4:30 pm	10	13	KISS108	2	4:30 pm	5
4	JAMN94.5	3	4:00 pm	0	14	KISS108	2	5:00 pm	10
5	KISS108	1	3:30 pm	0	15	KISS108	3	4:00 pm	5
6	JAMN94.5	1	6:00 pm	25	16	JAMN94.5	3	6:00 pm	20
7	KISS108	2	6:00 pm	15	17	KISS108	2	5:00 pm	15
8	KISS108	3	5:00 pm	10	18	MIX104.1	4	6:00 pm	15
9	KISS108	2	4:30 pm	10	19	KISS108	1	5:00 pm	25
10	KISS108	3	4:30 pm	20	20	KISS108	2	4:30 pm	10

ILLINOIS INSTITUTE OF TECHNOLOGY

Synopsis of Tween Survey

- 60% of the tweens listened to KISS108. The resort may want to direct its advertising dollars to this station.
- 55% of the tweens felt that the food was, at best, fair.
- 95% of the tweens would like the dining area to remain open later.
- 85% of the tweens spent their own money at the lodge.

End of Chapter