Chapter 1

Introduction and Data Collection

Chapter Topics

- Why a Manager Needs to Know About Statistics
- The Growth and Development of Modern Statistics
- Some Important Definitions
- Descriptive Versus Inferential Statistics

(continued)

- Why Data Are Needed
- Types of Data and Their Sources
- Design of Survey Research
- Types of Survey Sampling Methods
- Evaluating Survey Worthiness
- Types of Survey Errors
Why a Manager Needs to Know About Statistics

- To Know How to Properly Present Information
- To Know How to Draw Conclusions about Populations Based on Sample Information
- To Know How to Improve Processes
- To Know How to Obtain Reliable Forecasts

The Growth and Development of Modern Statistics

- Needs of government to collect data on its citizenry
- The development of the mathematics of probability theory
- The advent of the computer

Some Important Definitions

- A Population (Universe) is the Whole Collection of Things Under Consideration
- A Sample is a Portion of the Population Selected for Analysis
- A Parameter is a Summary Measure Computed to Describe a Characteristic of the Population
- A Statistic is a Summary Measure Computed to Describe a Characteristic of the Sample
Population and Sample

Population

Sample

Use parameters to summarize features

Use statistics to summarize features

Inference on the population from the sample

Statistical Methods

- Descriptive Statistics
  - Collecting, presenting, and characterizing data
- Inferential Statistics
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Descriptive Statistics

- Collect Data
  - E.g., Survey
- Present Data
  - E.g., Tables and graphs
- Characterize Data
  - E.g., Sample Mean = \( \sum \frac{X_i}{n} \)
Inferential Statistics

Drawing conclusions and/or making decisions concerning a population based on sample results.

- Estimation
  - E.g. Estimate the population mean weight using the sample mean weight
- Hypothesis Testing
  - E.g. Test the claim that the population mean weight value is 120 pounds

Why We Need Data

- To Provide Input to a Survey
- To Provide Input to a Study
- To Measure Performance of Ongoing Service or Production Process
- To Evaluate Conformance to Standards
- To Assist in Formulating Alternative Courses of Action
- To Satisfy Curiosity

Data Sources

- Observation
- Experimentation
- Survey
- Print or Electronic
Design of Survey Research

Choose an Appropriate Mode of Response
- Reliable primary modes
  - Personal interview
  - Telephone interview
  - Mail survey
- Less reliable self-selection modes (not appropriate for making inferences about the population)
  - Television survey
  - Internet survey
  - Printed survey in newspapers and magazines
  - Product or service questionnaires

Identify Broad Categories
- List complete and non-overlapping categories that reflect the theme

Formulate Accurate Questions
- Clear and unambiguous questions use clear operational definitions – universally accepted definitions

Test the Survey
- Pilot test on a small group of participants to assess clarity and length

Write a Cover Letter
- State the goal and purpose of the survey
- Explain the importance of a response
- Provide assurance of respondent anonymity
- Offer incentive gift for respondent participation
Types of Data

- Categorical
  - Qualitative
  - Discrete
  - Continuous
- Numerical
  - Quantitative
    - Discrete
    - Continuous

Type of Data (continued)

- Categorical random variables yield categorical responses
  - E.g. Are you married? Yes or No
- Numerical random variables yield numerical responses
  - Discrete random variables yield numerical responses that arise from a counting process
    - E.g. How many cars do you own? 3 cars
  - Continuous random variables yield numerical responses that arise from a measuring process
    - E.g. What is your weight? 130 pounds

Levels of Measurement and Types of Measurement Scales

- Nominal Scale - distinct categories in which no ordering is implied
  - E.g. Type of stocks invested: growth, income, other and none
- Ordinal Scale - distinct categories in which ordering is implied
  - E.g. Student grades: A, B, C, D or F
- Interval Scale - an ordered scale in which the difference between the measurements does not involve a true zero point
  - E.g. Temperature in degrees Celsius
- Ratio Scale - an ordered scale in which the difference between the measurements involves a true zero point
  - E.g. Weight in pounds
Reasons for Drawing a Sample

- Less Time Consuming Than a Census
- Less Costly to Administer Than a Census
- Less Cumbersome and More Practical to Administer Than a Census of the Population

Types of Sampling Methods

- Non-Probability Samples
  - Quota
  - Judgement
  - Convenience
  - Chunk
- Probability Samples
  - Simple Random
  - Systematic
  - Stratified
  - Cluster
- Convenience

Probability Sampling

- Subjects of the Sample are Chosen Based on Known Probabilities
  - Simple Random
  - Systematic
  - Stratified
  - Cluster
Simple Random Samples

- Every individual or item from the frame has an equal chance of being selected.
- Selection may be with replacement or without replacement.
- One may use table of random numbers or computer random number generators to obtain samples.

Systematic Samples

- Decide on sample size: n.
- Divide frame of N individuals into groups of k individuals: k = N/n.
- Randomly select one individual from the 1st group.
- Select every k-th individual thereafter.

Stratified Samples

- Population divided into 2 or more groups according to some common characteristic.
- Simple random sample selected from each group.
- The two or more samples are combined into one.

N = 64
n = 8
k = 8
First Group
Cluster Samples

- Population Divided into Several “Clusters,” Each Representative of the Population
- A Random Sampling of Clusters is Taken
- All Items in the Selected Clusters are Studied

Advantages and Disadvantages

- Simple Random Sample & Systematic Sample
  - Simple to use
  - May not be a good representation of the population’s underlying characteristics
- Stratified Sample
  - Ensures representation of individuals across the entire population
- Cluster Sample
  - More cost effective
  - Less efficient (need larger sample to acquire the same level of precision)

Evaluating Survey Worthiness

- What is the Purpose of the Survey?
- Is the Survey Based on a Probability Sample?
- Coverage Error – Appropriate Frame
- Nonresponse Error – Follow up
- Measurement Error – Good Questions Elicit Good Responses
- Sampling Error – Always Exists
Types of Survey Errors

- Coverage Error: Excluded from frame
- Nonresponse Error: Follow up on nonresponses
- Sampling Error: Chance differences from sample to sample
- Measurement Error: Bad Question!

Chapter Summary

- Addressed Why a Manager Needs to Know about Statistics
- Discussed the Growth and Development of Modern Statistics
- Addressed the Notion of Descriptive Versus Inferential Statistics
- Discussed the Importance of Data

Chapter Summary (continued)

- Defined and Described the Different Types of Data and Sources
- Discussed the Design of Surveys
- Discussed Types of Survey Sampling Methods
- Evaluated Survey Worthiness
- Described Different Types of Survey Errors