

Strategies for Analyzing Quantitative Data

Prof. Peter Fong

ALWAYS LEARNING

1

Exploring and Organizing a Data Set

- Statistics: a group of computational procedures that enable us to find patterns and meaning in numerical data.
- Principles – Exploration of Data:
 1. Where two variables are concerned, one of the variables becomes dominant and governs meaning that emerges from the other.
 2. Whatever the researcher does with the data to prepare it for inspection or interpretation will affect the meaning that the data reveal. Therefore, every researcher should be able to provide a clear, logical rationale for the procedure used to arrange and organize the data.

Leedy & Ormrod
Practical Research: Planning and Design, 10e
Copyright 2016 Pearson Education, Inc. All rights reserved.

11-2

2

Using the Computer to Organize and Analyze Data

- Electronic Spreadsheet: a software program that allows a researcher to manipulate data displayed in a table.
- Uses of electronic spreadsheets:
 - sorting data
 - searching for desired information
 - recoding data
 - graphing from the data
 - calculating formulas
 - employing "trial and error" explorations

Leedy & Ormrod
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-3

3

Considering the Nature of the Data

- Have your data been collected for a single group or, instead, for two or more groups? – it affects a decision on statistical techniques
- Were continuous or discrete variables involved? – indicate if a number of possible answers finite or infinite
- Did data represent nominal, ordinal, interval, or ratio scales? - dictate a strategy for statistical analysis
- Did data reflect a normal or non-normal distribution? – helps to identify patterns in data distribution

11-4

4

Important Terms Related to Distribution

Percentile ranks are often used to report performance on scholastic aptitude and achievement tests.

Raw score—the number of points accumulated on a research instrument (or test). A researcher first determines the raw score to calculate percentile ranks.

Leedy & Ormrod
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-5

5

Functions of Statistics

- Two Major Functions of Statistics:

1. They describe what the data looks like; this is the function of descriptive statistics.

2. They allow us to make *inferences* about large populations by collecting data on relatively small samples; this is the function inferential statistics.

Leedy & Ormrod
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-6

6

Characteristics of Statistics

- Estimates of population parameters
- Different statistics are appropriate for different kinds of data
- Single-group versus multi-group data
- Continuous versus discrete variables
- Scales of measurement (nominal, ordinal, interval, ratio)
- Normal and non-normal distributions
- Parametric versus nonparametric statistics

Leedy & Ormrod
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-7

7

Choosing Between Parametric and Nonparametric Statistics

Parametric statistics are used when

- The data reflect an interval or ratio scale.
- The data fall in a normal distribution

Nonparametric statistics are used when data are ordinal in nature and/or when data distribution is non-normal (i.e., a population is highly skewed in one direction or the other).

Leedy & Ormrod
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-8

8

Measures of Central Tendency

- Mode: the single number or score that occurs most frequently.
- Median: the numerical center of a set of data.
- Mean: the arithmetic average of the scores within the data set.
- Geometric Mean: a measure of central tendency based on a geometric progression, such as growth.

Note: The configuration of the data dictates the measure of central tendency most appropriate for that particular situation.

Leedy & Orm
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-9

9

Measures of Variability-Dispersion and Deviation

- Range: indicates the spread of data from lowest to highest value (highest score – lowest score)
- Average Deviation: the average of differences of each score in a set of scores and the mean score.
- Standard Deviation: the measure of variability most commonly used in statistical procedures; the square of the score-mean differences.
- Norm-Referenced Scores: scores that reflect where each person in the group is relative to other members of the group.
- Standard Score: tells us how far an individual's performance is from the mean with respect to standard deviation units.

11-10

10

Choosing Appropriate Statistics

- Statistics related to central tendency and variability provide a beginning point from which to view data.
- Statistical manipulation of the data is not research.
- Research demands interpretation of the data.

11-11

11

Measures of Association: Correlation

- Correlation: a measure of the relationship between two variables; correlation indicates the strength of the relationship.
- Correlation Coefficient: a number between -1 and +1; most correlation coefficients are decimals (positive or negative) somewhere between these two extremes.
- Positive Correlation: as one variable increases, the other variable also increases.
- Negative Correlation: as one variable increases, the other variable decreases.
- Pearson r : the most widely used statistic for measuring correlation.

Note: *Correlation does not necessarily indicate causation.*

11-12

12

Point Versus Interval Estimates

- Point Estimate: a single statistic that is taken as a reasonable estimate of the corresponding population parameter; will typically not correspond exactly with its true equivalent in the population.
- Interval Estimate: more accurate estimate of the population parameter.
- Confidence Interval: a range within whose limits a population parameter probably lies; provides a certain level of confidence that the estimated range is correct.

11-13

13

Inferential Statistics

- Two main functions of inferential statistics:
 1. To estimate a population parameter from a random sample.
 2. To test statistically based hypotheses.

Leedy & Ormrod
Practical Research: Planning and Design, 10e
Copyright 2016 Pearson Education, Inc. All rights reserved.

11-14

14

Estimating Population Parameters

- In conducting research, we use a sample to learn about the larger population from which the sample was drawn.
- Inferential statistics inform how closely sample statistics approximate parameters of the overall population.
- Statistical estimates of population parameters are based on the assumption that *the sample is randomly chosen and representative of the total population*.
- Error: the difference between the population mean and the sample mean.
- Standard Error of the Mean: indicates how much a particular mean is likely to vary from one sample to another *when all samples are the same size and are drawn randomly from the same population*.

11-15

15

Testing Hypotheses

- Null Hypothesis: a statistical hypothesis which postulates that any result observed is the result of chance alone.
- Testing the Null Hypothesis: the process of comparing observed data in a research study with what we would expect from chance alone.
- Significance Level: the probability that researchers use as a cutoff point to decide that a result has not occurred by chance.

11-16

16

Type I Errors, Type II Errors and Statistical Power

- Type I error (α): the probability of rejecting the null hypothesis when it is actually true.
- Type II error (β): the probability of failing to reject the null hypothesis given that the alternative hypothesis is actually true.
- Statistical power ($1 - \beta$): the probability of correctly rejecting the null hypothesis.

17

Making Errors in Hypothesis Testing

- Type I Error: the erroneous conclusion that a result was not due to chance when in fact it was due to chance; incorrectly rejecting the null hypothesis.
- Type II Error: the erroneous conclusion that a result was due to chance when in fact it was not; incorrectly failing to reject a null hypothesis that is actually false; also known as a *beta error*.

11-18

18

Suggestions for Increasing the Power of a Statistical Test

- Use as large a sample as is reasonably possible.
- Maximize the validity and reliability of your measures.
- Use parametric rather than nonparametric statistics whenever possible.

Note: *Whenever we test more than one statistical hypothesis, we increase the probability of making at least one Type I error.*

Leedy & Ormrod
Practical Research: Planning and Design, 10e
© 2011 Pearson Education, Inc. All rights reserved.

11-19

19

Meta-Analysis

- Meta-Analysis: used to analyze and draw conclusions about other researchers' statistical analyses.
- When conducting a meta-analysis, the researcher:
 1. Conducts an extensive search for relevant studies;
 2. Identifies appropriate studies to include in the meta-analysis;
 3. Converts each study's results to a common statistical index.

11-20

20

Advantages of Statistical Software Packages (*SPSS, SAS, SYSTAT, Minitab, StatView*)

- Increased user-friendliness
- Range of available statistics
- Assumption testing
- Speed of completion
- Graphics

11-21

21

Interpretation of the Data

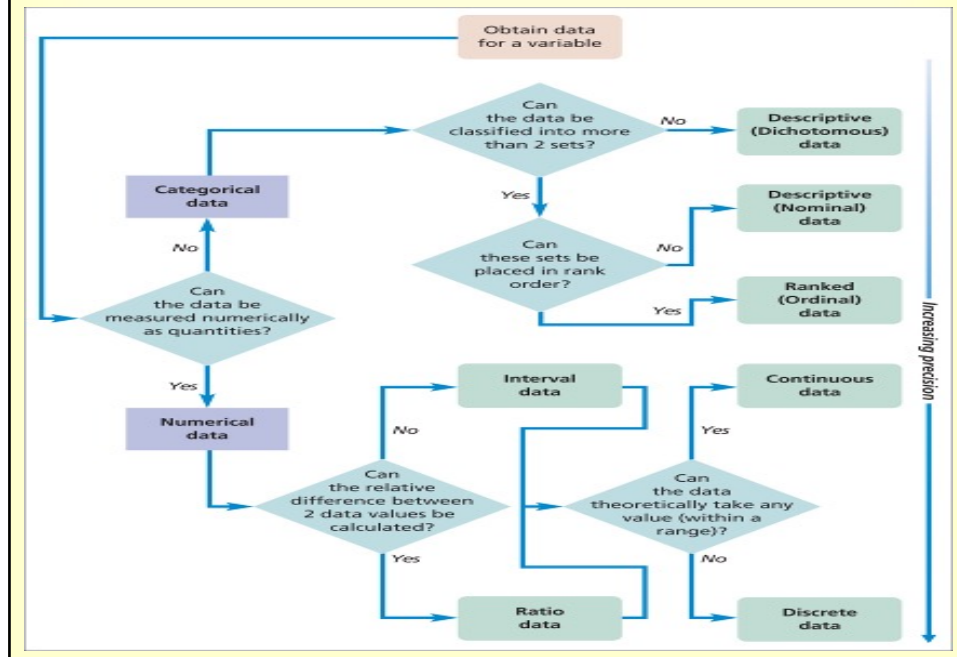
Interpreting the data means several things:

- Relating the findings to the original research problem and to the specific research questions and hypotheses.
- Relating the findings to preexisting literature, concepts, theories, and research studies.
- Determining whether the findings have practical significance as well as statistical significance.
- Identifying limitations of the study.

11-22

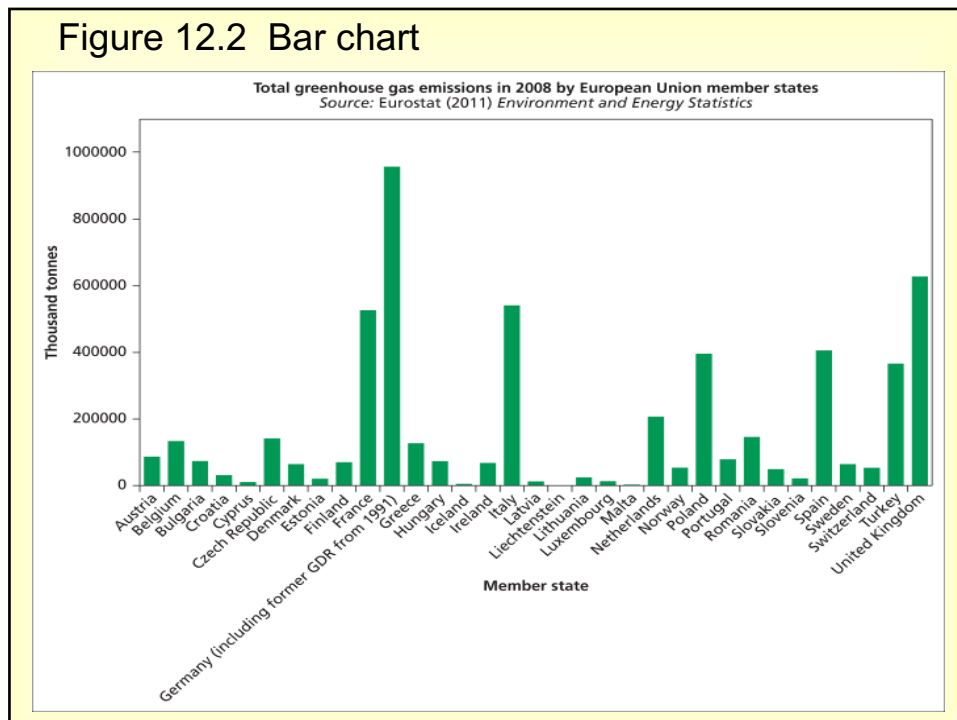
22

Figure 12.1 Defining the data type Source: Mark Saunders (2012), RM for Business students.



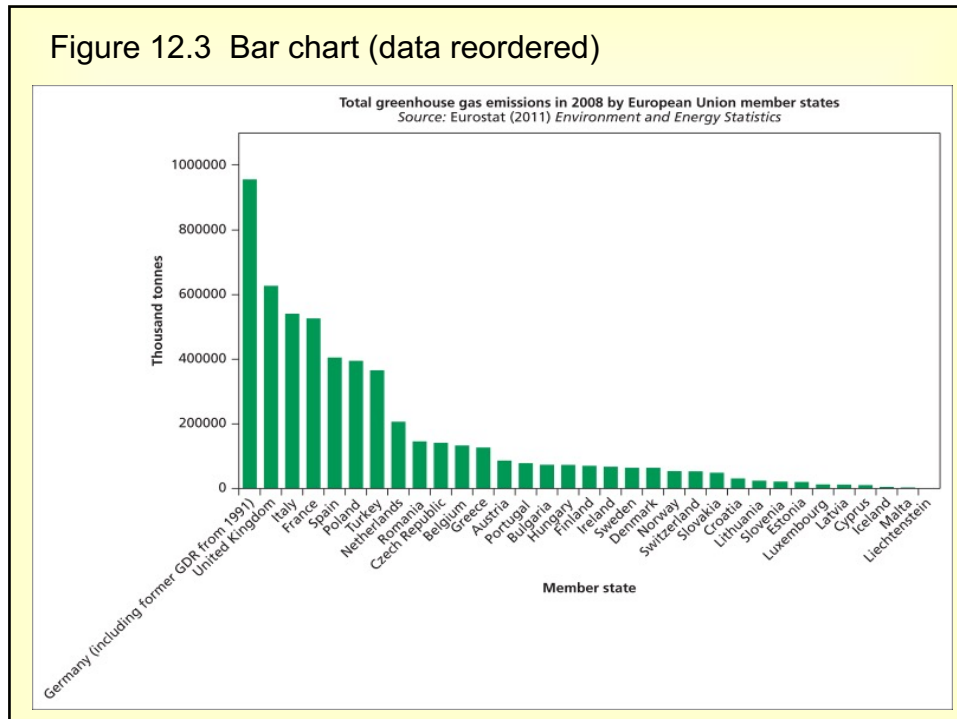
23

Figure 12.2 Bar chart



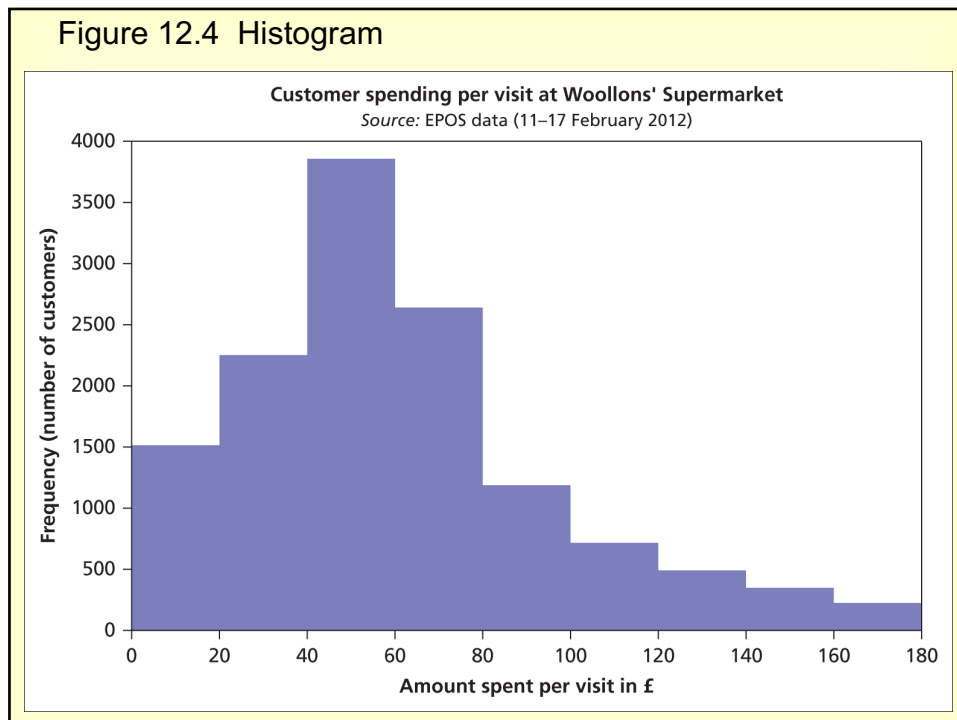
24

Figure 12.3 Bar chart (data reordered)



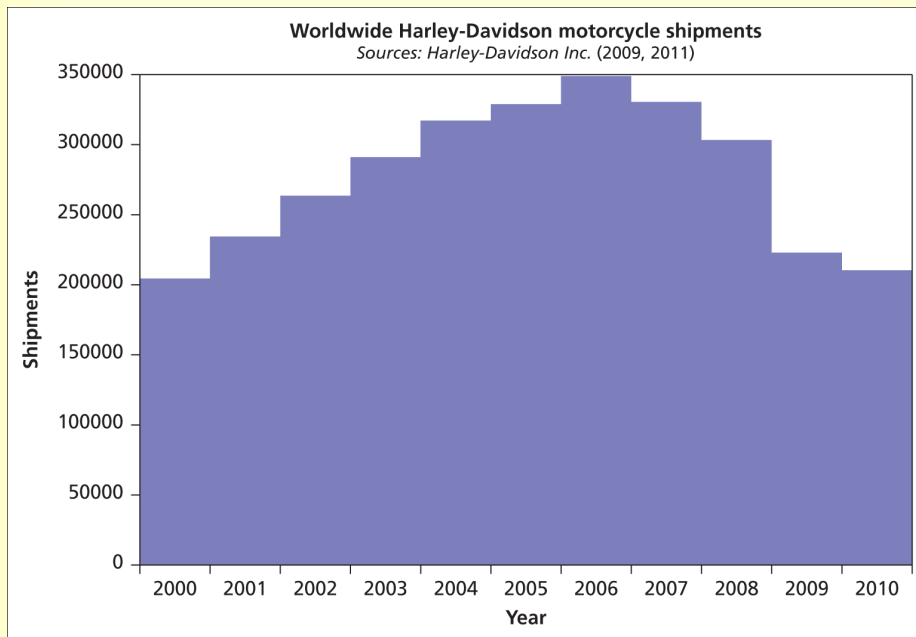
25

Figure 12.4 Histogram



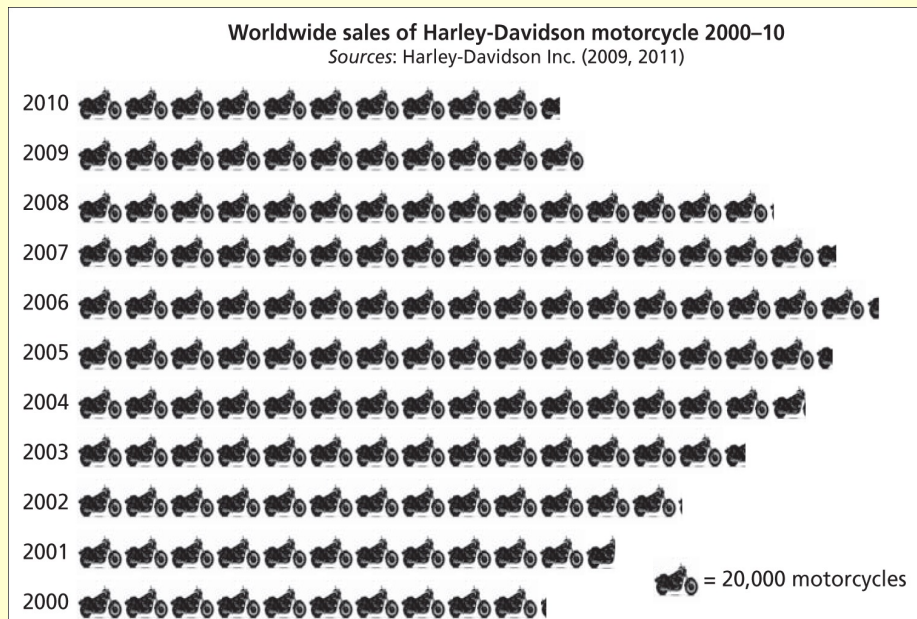
26

Figure 12.5 Histogram



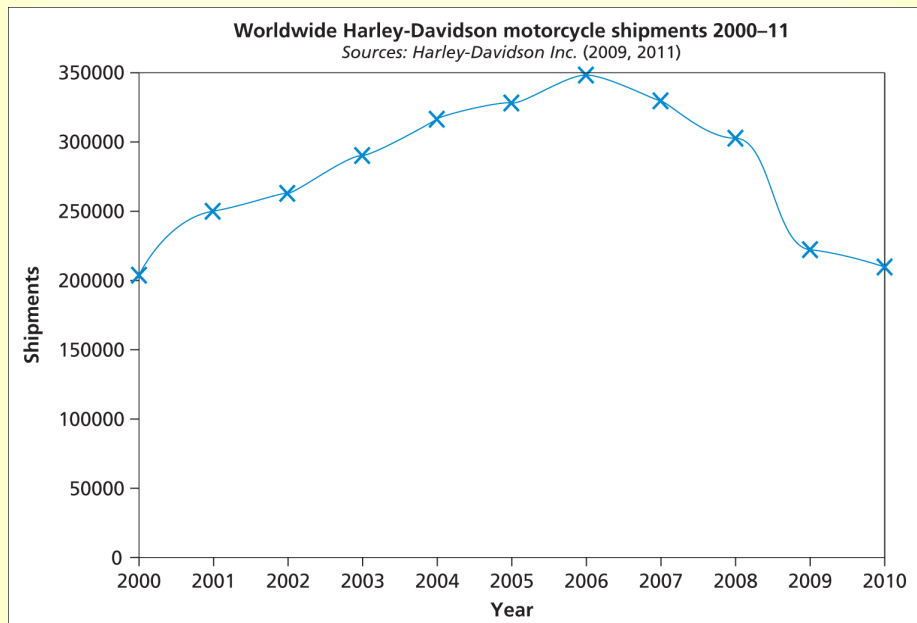
27

Figure 12.6 Pictogram



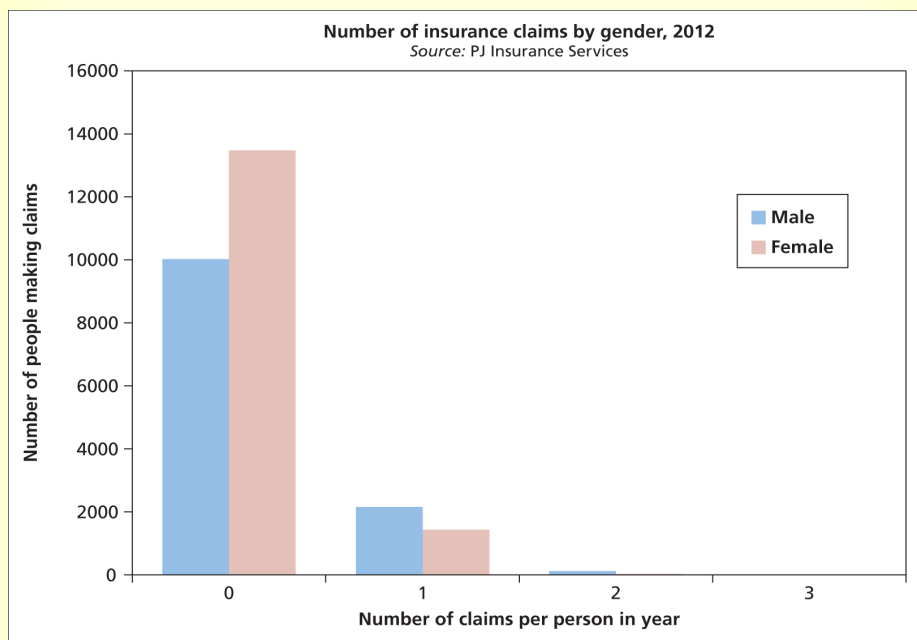
28

Figure 12.7 Line graph



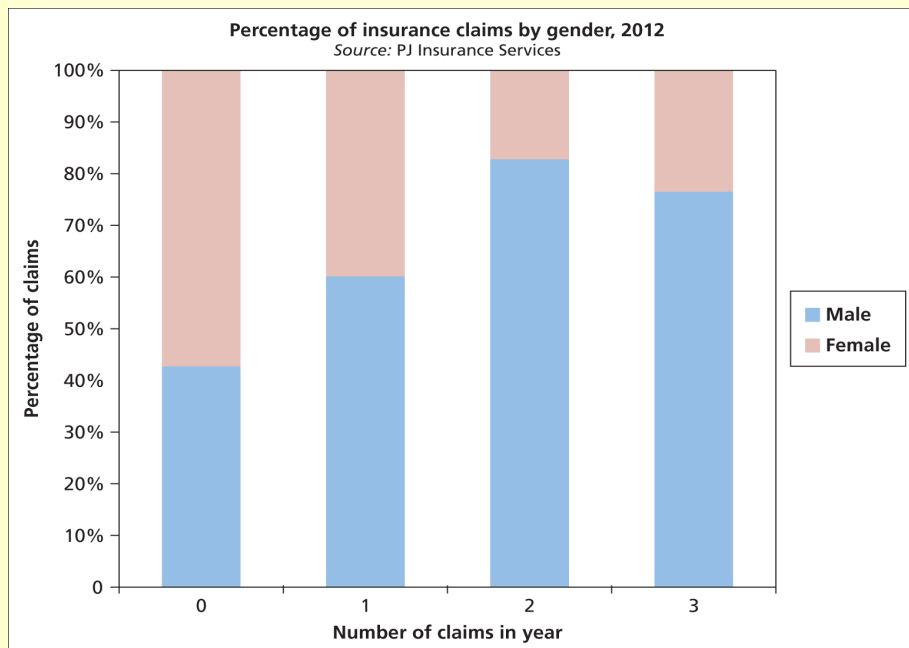
29

Figure 12.9 Multiple bar chart

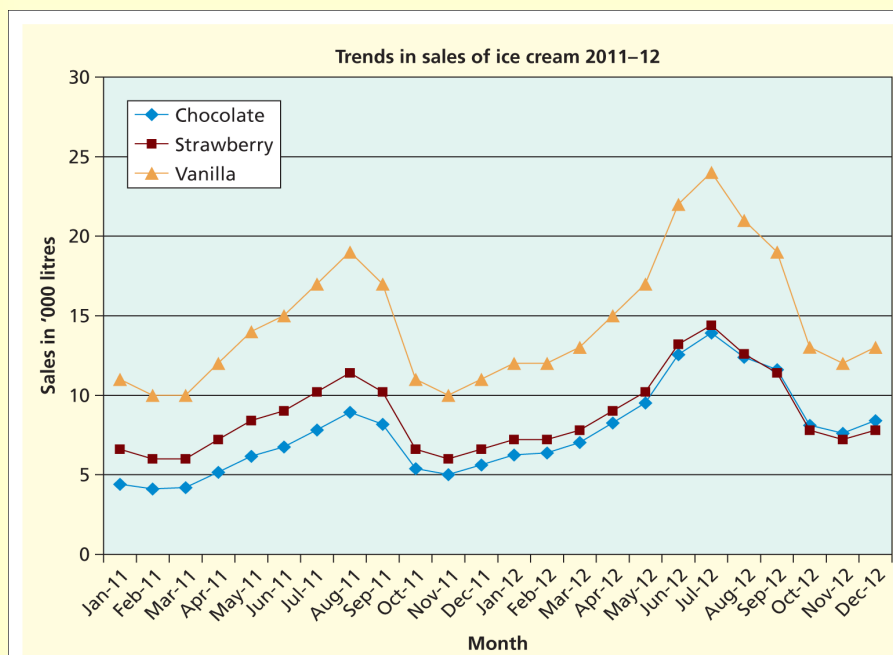


30

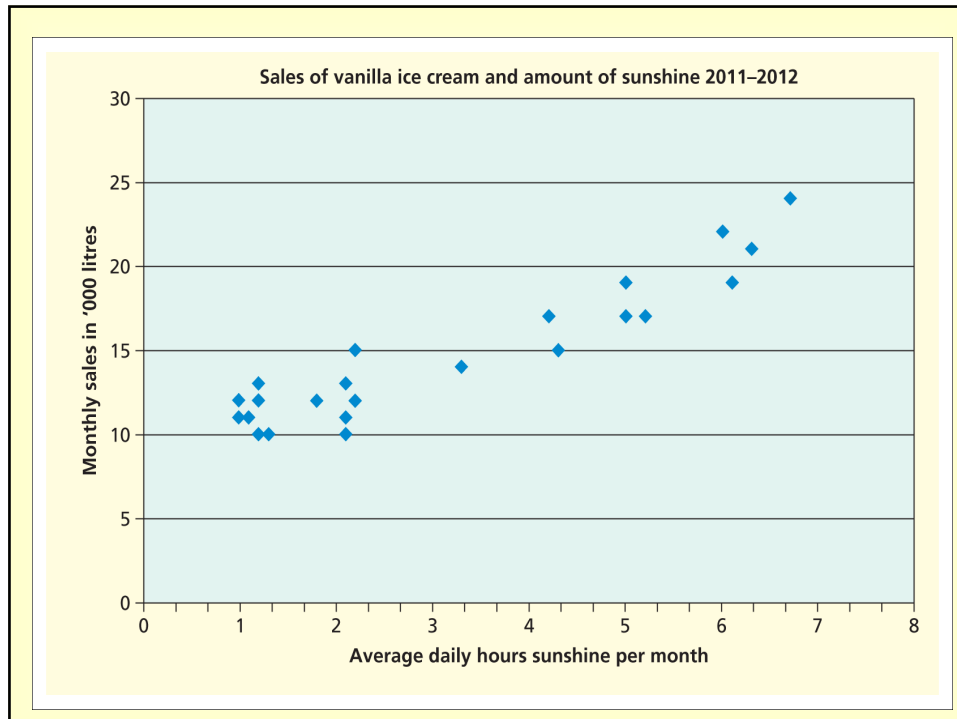
Figure 12.10 Percentage component bar chart



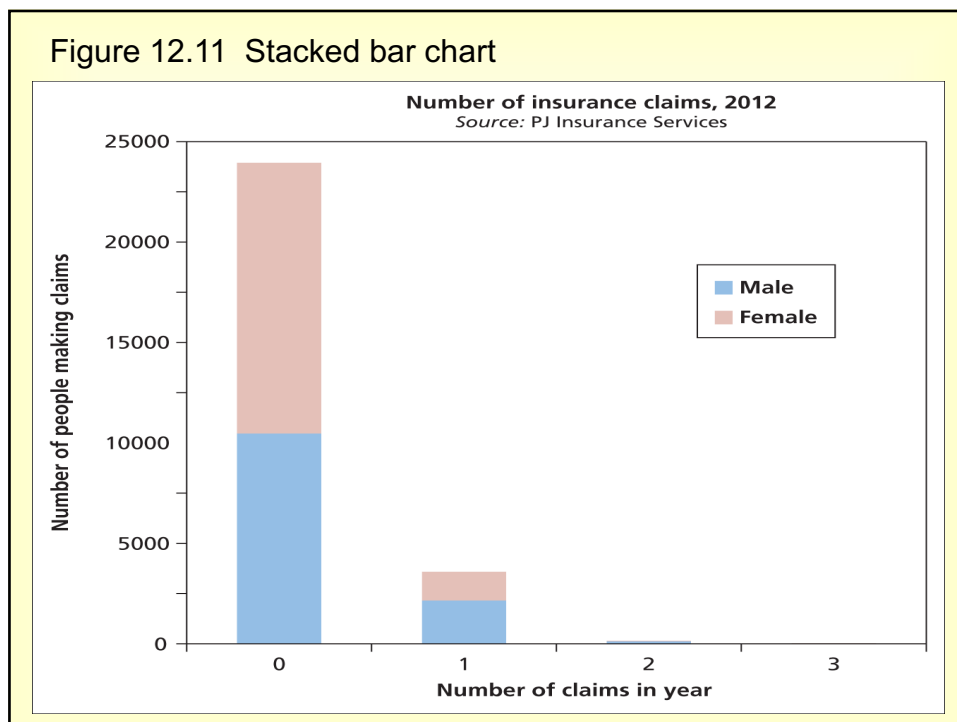
31



32

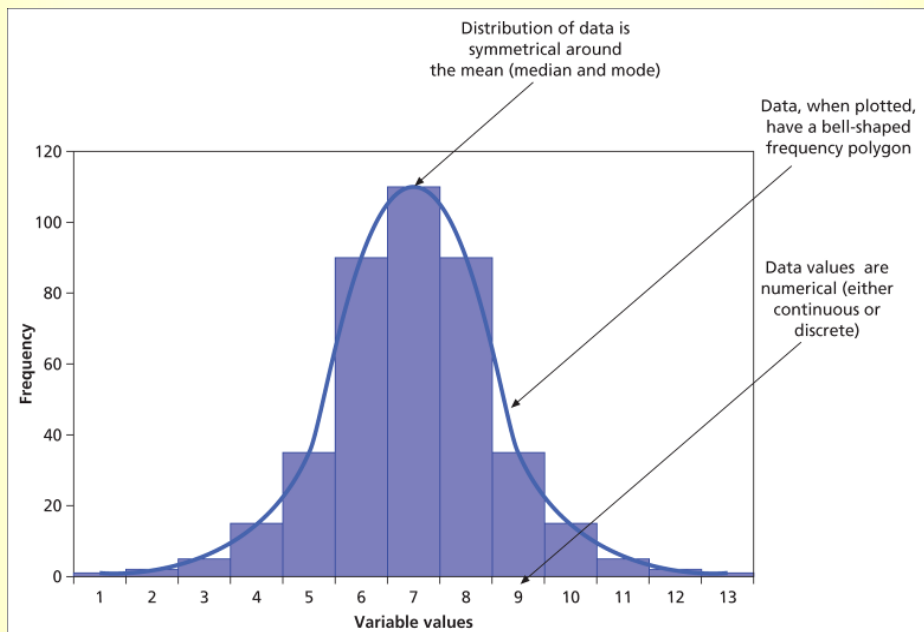


33



34

Figure 12.13 Annotated frequency polygon showing a normal distribution



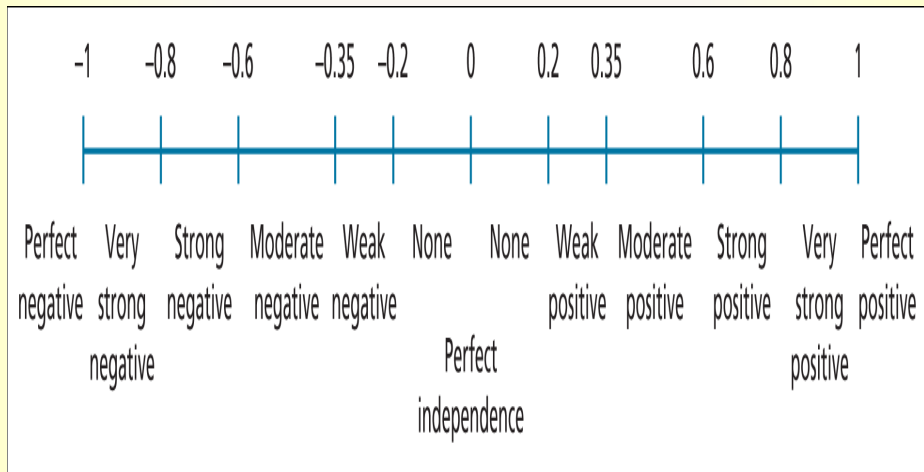
35

Figure 12.14 Type I and Type II errors

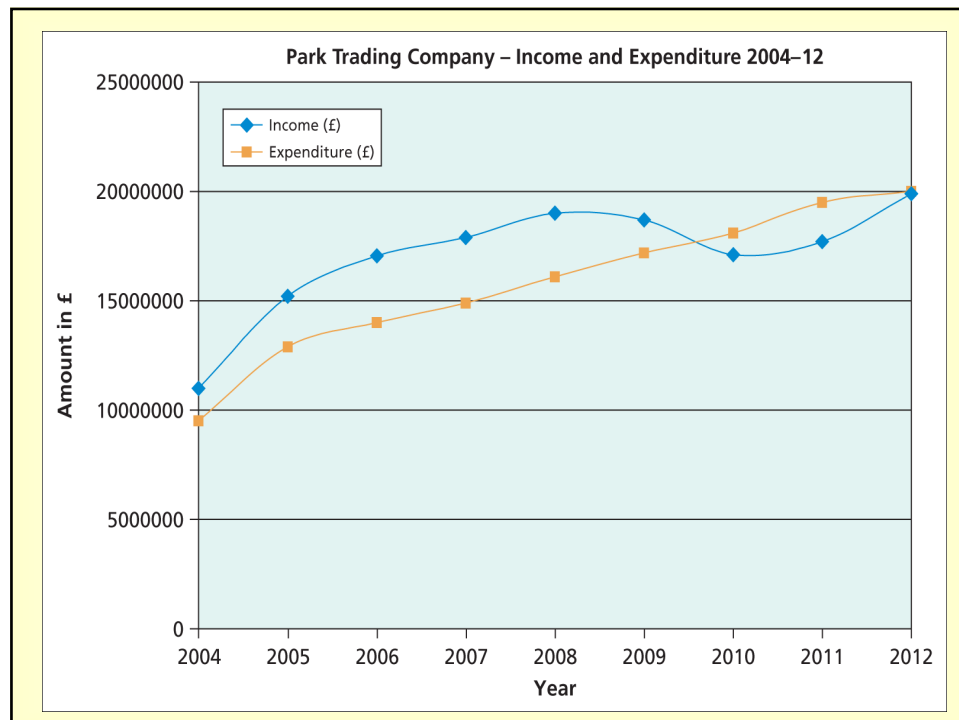
		Likelihood of making a	
		Type I error	Type II error
Significance level at	0.05	Increased	Decreased
	0.01	Decreased	Increased

36

Figure 12.15 Values of the correlation coefficient



37



38