

Introduction to Research Methods

Prepared by Associate Professor/
Ahmed Rageh Ismail

MGM333

Introduction to Research Methods

Lecture 10



A photograph of a business meeting with several people in blue suits gathered around a table. A semi-transparent light blue rectangle is overlaid in the center, containing the text 'Quantitative Data Analysis'. On the table, there is a laptop on the left, a red folder on the right, and various documents and charts, including a pie chart, in the foreground.

Quantitative Data Analysis

Learning Outcomes

1. Discuss the nature of numerical datasets
2. Explore the appropriateness of statistical analysis
3. Understand the fundamental principles of data analysis procedures

Key Topics

- Techniques for processing and displaying data
- Methods for exploring datasets
- Analysis of relationships between datasets
- Understanding when and why to perform specific analyses

Understanding Variables in Research

Types of Variables

- Nominal Variables
- Ordinal Variables
- Interval Variables

Importance

- Selecting appropriate variables is crucial for meaningful analysis

Nominal Variables

Definition

Categorical, qualitative, or discrete data

Categories with no rank order (e.g., eye color, religion)

Key Analytical Processes

Frequency distribution

Chi-square (χ^2) test

Ordinal Variables

Definition

Data with inherent logical order or rank

Relative magnitude has meaning (e.g., letter grades, top four favorite genres)

Characteristics

Rank-ordered but no meaningful distance between data points

Key Analytical Processes

- Includes nominal data methods
- Mann-Whitney U-test for differences between two groups
- Wilcoxon signed-rank test for comparing the same subjects under different conditions

Interval Variables

Definition

- Strongest level of measurement in statistical analysis
- Categories can be rank-ordered and intervals have meaningful numerical value

Examples

- Age, population size, years in occupation, distance between locations

Characteristics

- Quantifiable and open to arithmetic manipulation
- Provides specific numerical value, allowing for advanced statistical analysis

Level of measure	Analytical procedure
Nominal	Frequency distribution Chi square (χ^2)
Ordinal	Frequency distribution Chi square (χ^2) Wilcoxon Mann-Whitney U test Rank order correlation
Interval	Whole range of parametric statistics

Frequency Distribution

Importance

- First step in analyzing data
- Considered the "bedrock" of data analysis
- Provides an overview of the dataset

Purpose of Frequency Distribution

- Presents the number of responses to each option in your questions
- Essential for giving readers the basic facts before more complex analysis

Misconception

- Often seen as merely data processing, but crucial for understanding data and influencing subsequent analysis

'Book reviews influence people's choice of fiction titles'	Frequency	%	Valid %	Cumulative %
Strongly agree	701	37.0	38.2	38.2
Agree	701	37.0	38.2	76.3
Neither	251	13.2	13.7	90.0
Disagree	163	8.6	8.9	98.9
Strongly disagree	21	1.1	1.1	100.0
Total	1837	96.8	100.0	
Missing	60	3.2		
Total	1897	100.0		

Cross-Tabulation

- A method to explore relationships between two variables
- Expands on frequency distribution by showing potential links between variables
- Allows you to see if and how two variables might be related
- Example: Exploring the relationship between gender and influence of book reviews on fiction title choices

Measures of Central Tendency

Purpose

- Summarize an entire distribution with a single score
- Show the grouping of data around a central point

Types

- Mean
- Median
- Mode

Definition

- Arithmetic average of a dataset

Advantages

- Easily understood and calculated
- Responsive to all scores equally
- Fundamental for further calculations (e.g., standard deviation)
- Most stable measure of central tendency

Disadvantages

- Can be misleading with skewed data
- Applicable only to interval data

The mean = $\frac{\text{total of all values}}{\text{number of cases}}$

or

$$M = \frac{\sum x}{N}$$

Median

- The mid-point or middle value in a ranked dataset
- Calculation involves ordering data and finding the central value

Advantages

- Insensitive to extreme scores in skewed distributions
- Useful for skewed data as it divides the dataset into two equal parts

Disadvantages

- May not be representative of the entire dataset
- Only provides the value of one score

You have a series of nine scores: 16 6 11 24 17 4 19 9 20

Order these scores in relation to size: 24 20 19 17 16 11 9 6 4

↑
median value

This is simple when you have an odd number of scores. When you have an even number there is an additional step to include in the process:

16 11 24 6 17 4 19 9 20 8

Ranked order shows the following:

24 20 19 17 **16 11** 9 8 6 4

$$\text{Median} = \frac{16 + 11}{2} = 13.5$$

Mode

- Most frequently occurring score in a dataset
- Can be used for nominal data

Advantages

- Easy to obtain
- Unaffected by extreme scores
- Represents a 'typical' value if most values cluster around the mode

Table 24.6 The favourite colour of a sample of pre-school children

Colour	Frequency	Cumulative Frequency
Red	3	3
Blue	16	19
Purple	17	36
Pink	7	43
Orange	3	46
Yellow	2	48

Table 24.6 shows that in this dataset Purple is the mode. The mode is the only measure of central tendency that can be applied to nominal data. There can be more than one modal value in a distribution.

Disadvantages

- Lacks precision compared to other measures
- Dataset may have more than one mode (bimodal or multimodal distributions)
- Less useful in advanced analysis

Measures of Dispersion

- Provide additional information about the spread of data around a central point
- Complement measures of central tendency

Types

- Range
- Variance
- Standard Deviation

The Range

Definition

- Difference between the highest and lowest score in a dataset

Application

- Simple measure but only uses two scores

The Variance

Definition

- Measures how each score varies from the mean

Application

- Helps in understanding the spread of data points around the mean

$$\text{Variance} = \frac{\text{sum of (score - mean score)}^2}{\text{total number of scores}}$$

Table 24.7 illustrates how this is done for dataset 'c' in the example given above, where

Dataset = 1 5 10 15 19

Mean = 10

Table 24.7 The variance for dataset c

Score	score - mean	(score - mean) ²
1	1 - 10 = -9	(-9) ² = 81
5	5 - 10 = -5	(-5) ² = 25
10	10 - 10 = 0	0 ² = 0
15	15 - 10 = 5	5 ² = 25
19	19 - 10 = 9	9 ² = 81

Sum of (score - mean)² = 212

Variance = $\frac{212}{5} = 42.4$

Standard Deviation

Definition

- Most important measure of dispersion
- Indicates the spread of scores around the mean in original units

Application

- Larger standard deviation indicates greater dispersal of scores
- Use caution with skewed data as it may give unrepresentative results

Standard deviation is represented by the Greek lower case sigma = σ .

$$\sigma = \sqrt{\text{Variance}}$$

Correlation

Definition

- Correlation examines the relationship between two sets of data.

Types of Correlation

- Positive Correlation: Both variables increase together.
- Negative Correlation: One variable increases while the other decreases.
- Zero Correlation: No relationship between the variables.

Correlation indicates a general trend but does not establish a fixed rule or causality.

Displaying Data Visually

Importance of Visuals

- Visual representations help communicate patterns in data effectively.
- Commonly used visual tools include bar charts.

Types of Bar Charts

- Simple Bar Charts: Display frequencies or percentages for individual categories.
- Compound Bar Charts: Compare data from multiple groups within the same category.
- Component Bar Charts: Show the breakdown of data within a category.

Table 24.8 Frequency distribution for people using park and ride, Mon-Sat

Day	No. of people
Mon	240
Tues	300
Wed	350
Thurs	280
Fri	450
Sat	500

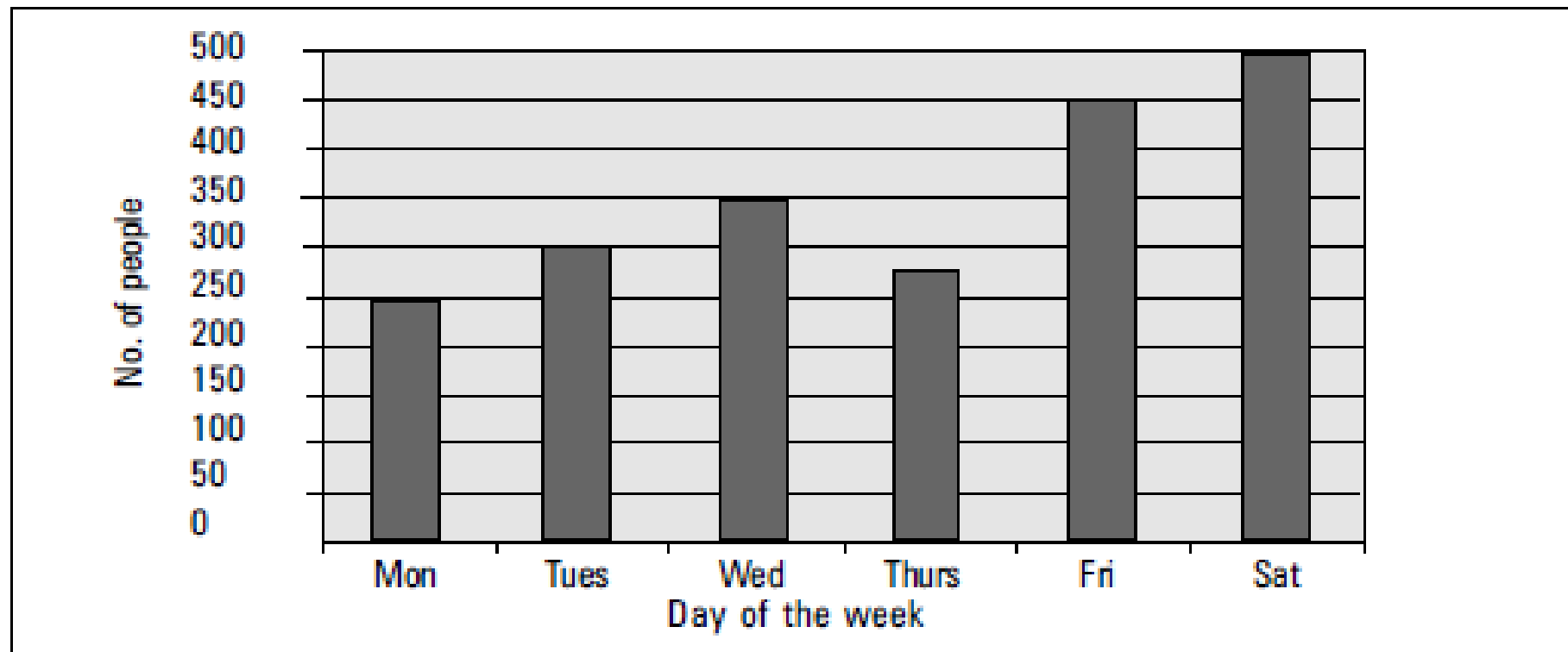


Figure 24.1 Number of people using park and ride, Mon-Sat

Table 24.9 Sales in Britain and Europe 2000–2005 (£ millions)

	Britain	Europe
2000	2.5	1.3
2001	3.0	2.6
2002	1.5	1.2
2003	4.0	2.9
2004	1.5	2.0
2005	2.8	2.1

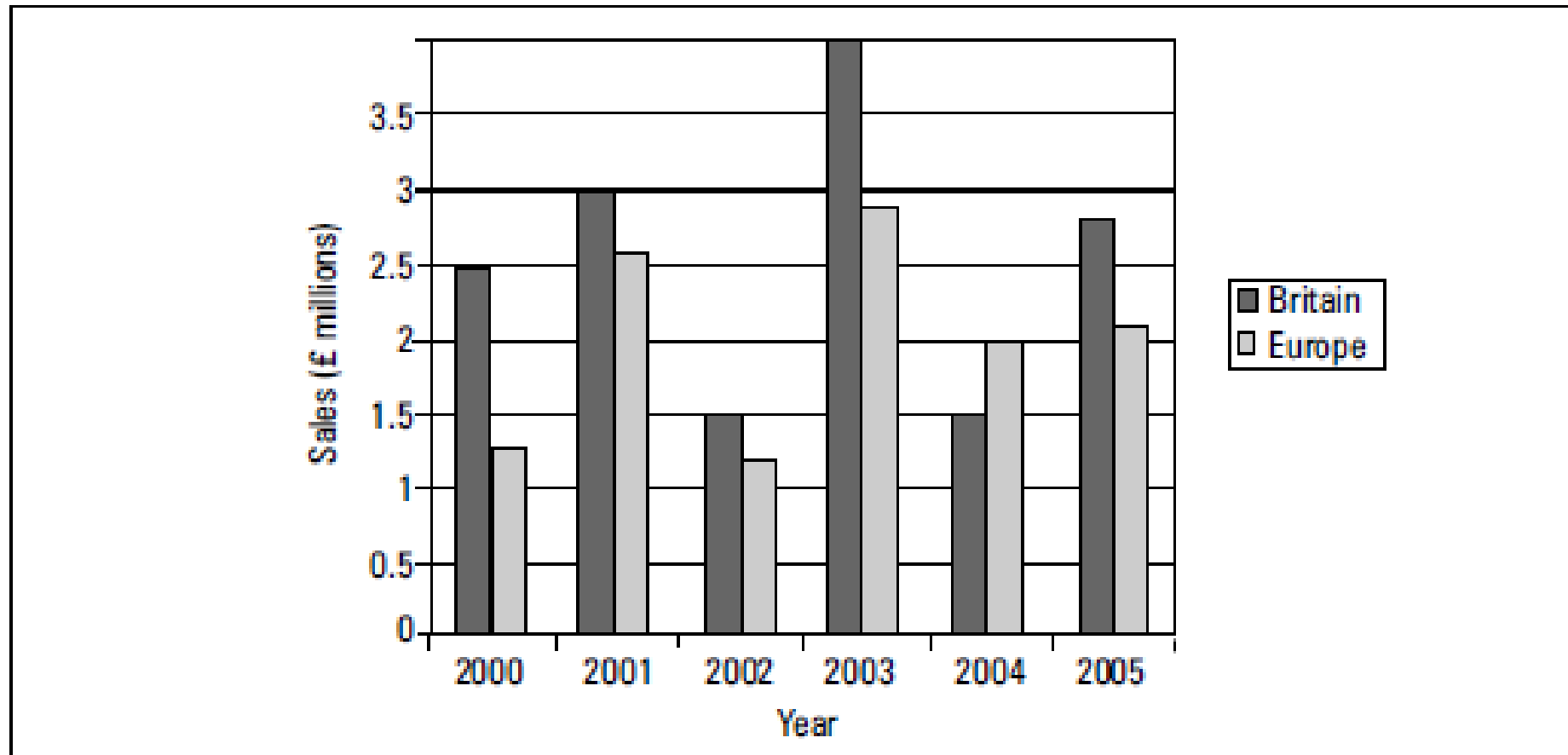


Figure 24.2 Sales in Britain and Europe 2000-2005

Table 24.10 Number of CDs owned by five families by musical category

	Burns	Edwards	French	Stephens	Travis
Blues	18	3	1	10	15
Classical	4	1	14	9	7
Rock	22	2	16	16	1
Techno	1	19	5	2	12

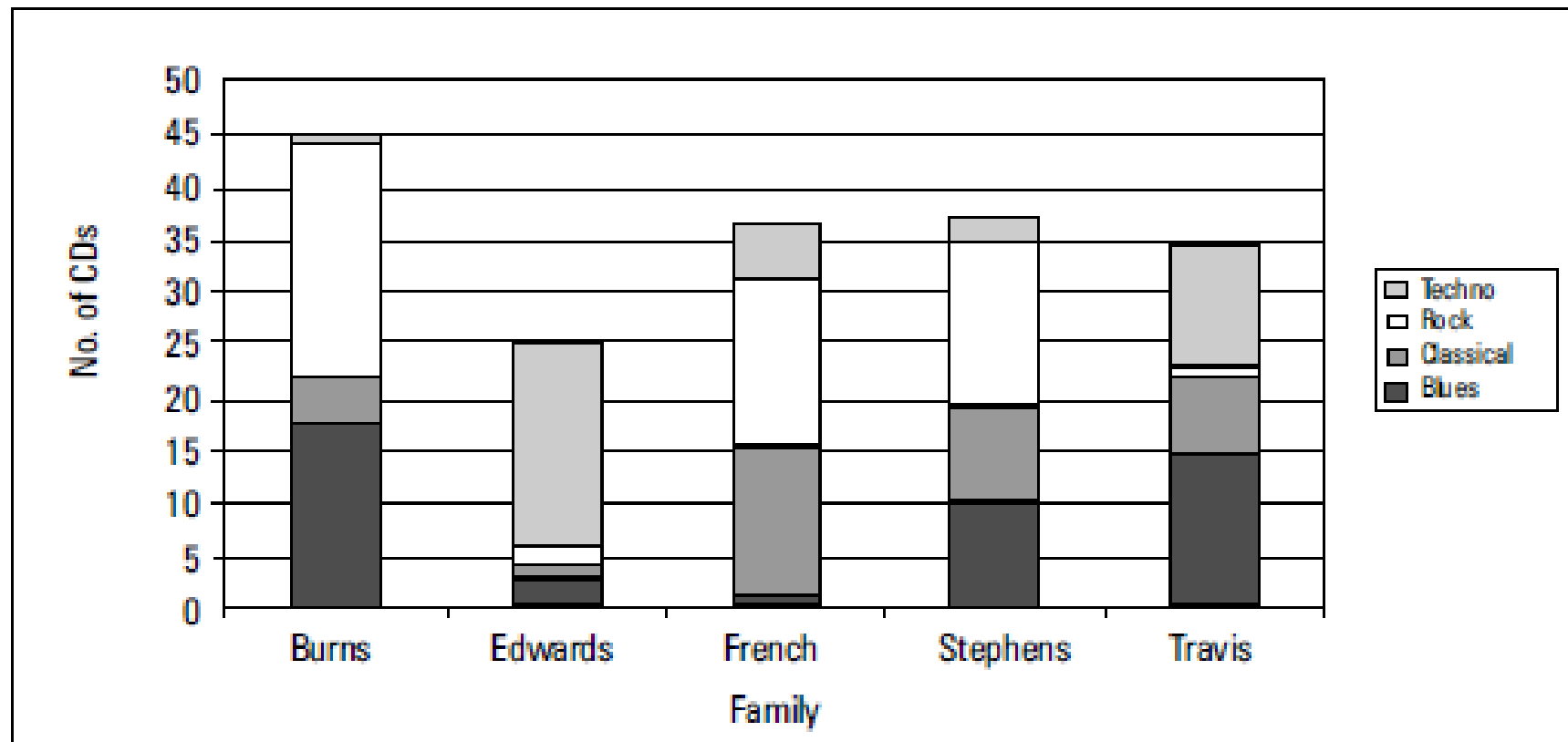


Figure 24.3 Number of CDs owned by five families by musical category

Pie Chart

Purpose

- Show figures in proportion to the whole and in relation to other figures.
- Each segment represents a proportion of the total.
- Clear visualization of data distribution within a whole.

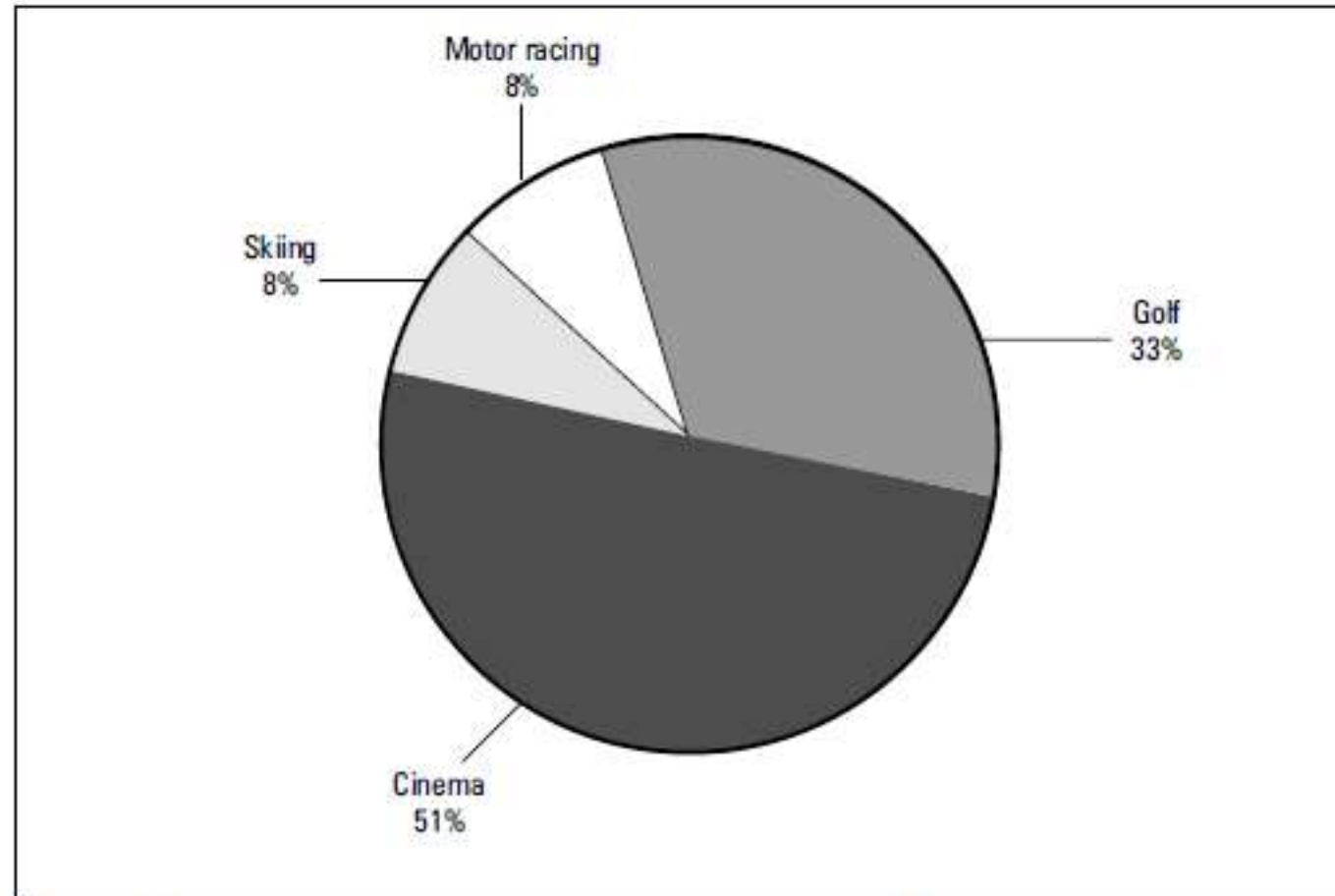


Figure 24.4 Percentage of people who have golf, cinema, skiing or motor racing as hobbies

Software for Quantitative Analysis

- Statistical software has revolutionized data analysis by reducing time and human error.
- Manual analysis is still taught initially to ensure understanding of data handling before using software.
- Software automates analysis, but the researcher must understand which tests to apply and why.
- Incorrect application can lead to misleading results.

SPSS (Statistical Package for Social Sciences)

- Launched in 1989, SPSS is a powerful and flexible statistical tool used widely across disciplines.
- Parametric and Non-Parametric Tests: Includes regression techniques, linear, non-linear, logistic, etc.
- Multivariate Analyses: Cluster, discriminant, principal components, scaling analyses.
- Survival Analysis: Cox proportional hazards with or without time-dependent variables.
- High-Resolution Graphics: Descriptive charts, scatter plots.



Thank You