**EDO STATE UNIVERSITY, IYAMHO**

Department of Mathematics

# STATISTICS

## STA 121

Full Lecture Note with Worked Examples, Past Questions, Assignments and Solutions

Prepared by

**Dr. Michael John**

Academic Lecture Book for Undergraduate Students

April 20, 2026

# Contents

# Preface

These lecture notes are written for undergraduate students taking STA 121: Statistics. The material follows the standard Nigerian university curriculum in introductory statistics and expands it into a self-contained book. Emphasis is placed on explanation, computation, interpretation, graphical presentation and examination-style practice. Every major topic is followed by worked examples, university-style exercises and short review questions. The final chapters contain past-question style practice sets, assignment questions and fully worked solutions.

Chapter 1

# Introduction to Statistics and Data

## 1.1 Meaning of Statistics

Statistics may be understood from three closely related viewpoints.

1. As **facts in numerical form**: for example, the number of students in a department, the mean score in a test, or the percentage of pass in a course.
2. As **numerical summaries**: for example, averages, percentages, indices and measures of variability.
3. As a **discipline**: a branch of science concerned with the collection, organization, analysis and interpretation of data for sound decision making under uncertainty.

A good operational definition is this: statistics is the science of learning from data. The subject provides methods for gathering data, reducing large collections of facts into manageable form, describing the features of a data set and drawing defensible conclusions.

> The four main actions of statistics are: **collect**, **organize**, **analyze** and **interpret**. When students remember these four words, they already understand the heart of elementary statistics.

## 1.2 Why Statistics Matters

Statistics is used in administration, health sciences, agriculture, business, economics, engineering, sports, education and social research. In university work, statistics helps lecturers interpret examination results, helps administrators plan enrolment, helps scientists analyze measurements and helps researchers compare competing explanations.

Examples from Nigerian university life include:

- comparing pass rates between two sessions;

- summarizing CGPA distributions in a department;
- estimating the proportion of students who own laptops;
- analyzing the relationship between attendance and test scores;
- predicting final examination performance from continuous assessment.

## 1.3 Descriptive and Inferential Statistics

**Descriptive statistics** deals with methods of organizing and summarizing data. Tables, bar charts, averages, medians and standard deviations are descriptive tools.

**Inferential statistics** deals with making decisions or generalizations about a population from sample data. Estimation and hypothesis testing belong here, though this book focuses mainly on descriptive techniques and elementary probability.

## 1.4 Variables and Data

A **variable** is a characteristic that takes different values for different individuals or objects. Examples are age, score, income, gender, number of siblings and height.

### 1.4.1 Types of Variables

- **Qualitative variables**: descriptive in nature. Examples are marital status, blood group and state of origin.
- **Quantitative variables**: numerical in nature. Examples are age, salary and test score.

Quantitative variables may be further divided into:

- **Discrete variables**: obtained by counting, e.g. number of books, number of vehicles, number of children.
- **Continuous variables**: obtained by measurement, e.g. weight, height, temperature, time and income measured to decimal accuracy.

### 1.4.2 Primary and Secondary Data

**Primary data** are obtained directly from original sources by observation, interview, questionnaire or experiment. **Secondary data** are obtained from reports, journals, textbooks, internet databases or official publications.

## 1.5 Methods of Data Collection

The choice of data collection method depends on cost, availability of respondents, time and the required level of accuracy.

a. **Observation**: direct watching and recording.

b. **Interview**: asking respondents questions orally.

c. **Questionnaire**: a written instrument completed by respondents.

d. **Experiment**: controlled investigation to observe effects of factors.

e. **Records**: use of existing institutional or governmental data.

**Worked Example 1.1: Classifying variables.** For each of the following, state whether the variable is qualitative or quantitative, and if quantitative state whether it is discrete or continuous.

(i) Number of carry-over courses for a student.

(ii) Religion of a student.

(iii) Height of a lecturer.

(iv) Number of lecture halls in a faculty.

**Solution.**

(i) Quantitative discrete, because it is obtained by counting whole numbers.

(ii) Qualitative, because religion is descriptive.

(iii) Quantitative continuous, because height is measured on a continuum.

(iv) Quantitative discrete, because the halls are counted.

## 1.6 Population and Sample

A **population** is the complete collection of units of interest. A **sample** is a subset selected from the population. If a lecturer wants to study the reading habits of all 2,400 students in the Faculty of Science, then the population is all 2,400 students. If 120 students are selected and studied, those 120 constitute the sample.

## 1.7 Errors in Data Collection

Common data collection errors include biased questions, inaccurate recording, missing responses, incorrect measurement and poor sampling. A sound statistical study demands careful design before any computation is attempted.

## 1.8 Review Exercises

1. Distinguish between descriptive and inferential statistics.
2. Explain the difference between population and sample using a university example.
3. Classify the following as qualitative, quantitative discrete or quantitative continuous: age, shoe size, department, salary, blood group, number of courses failed.
4. State four methods of collecting primary data and mention one merit of each.

## Chapter 2

# Presentation of Data

## 2.1 Raw Data, Arrays and Frequency Tables

Data collected directly from the field are called **raw data**. To make them meaningful, they are arranged into arrays and frequency tables.

An **array** is an arrangement of the observations in ascending or descending order. A **frequency distribution table** shows each value or class together with the number of times it occurs.

**Worked Example 2.1: Ungrouped frequency distribution.** The scores of 20 students in a test are:

$$8, 3, 2, 7, 4, 5, 4, 5, 8, 5, 5, 2, 1, 0, 8, 1, 9, 6, 3, 9.$$

Arrange the data and construct a frequency table.

**Solution.** First arrange in ascending order:

$$0, 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 5, 5, 6, 7, 8, 8, 8, 9, 9.$$

Hence the frequency table is given below.

| Score | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 1 | 2 | 2 | 2 | 2 | 4 | 1 | 1 | 3 | 2 |

The modal score is 5 because it has the highest frequency, namely 4.

## 2.2 Bar Charts for Categorical Data

A bar chart consists of separated rectangular bars of equal width. The heights are proportional to the observed frequencies. Bar charts are suitable for qualitative data and discrete data.

**Worked Example 2.2: Bar chart of T-shirt colours.** The frequencies of T-shirt colours worn by 20 pupils are as follows:

| Colour | Frequency |
|---|---|
| Black | 4 |
| Blue | 3 |
| Brown | 1 |
| Green | 1 |
| Red | 4 |
| White | 6 |
| Yellow | 1 |

The bar chart is shown in Figure 2.1.

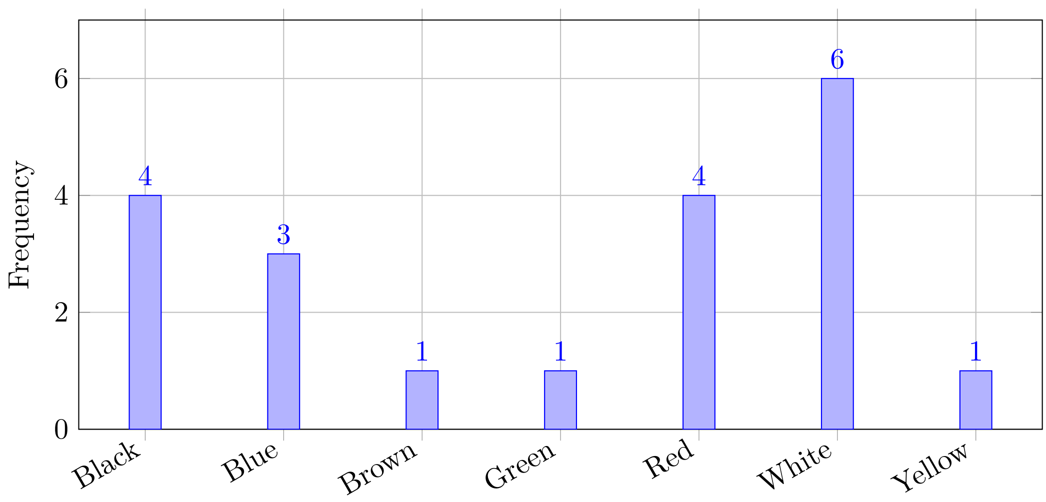


Figure 2.1: Bar chart of T-shirt colour frequencies.

## 2.3 Multiple Bar Chart

A multiple bar chart compares two or more related series. Consider the data for births recorded in a maternity centre.

| Month | Male | Female | Total |
|---|---|---|---|
| 1 | 20 | 30 | 50 |
| 2 | 30 | 50 | 80 |
| 3 | 55 | 65 | 120 |
| 4 | 60 | 40 | 100 |
| 5 | 40 | 50 | 90 |
| 6 | 70 | 45 | 115 |

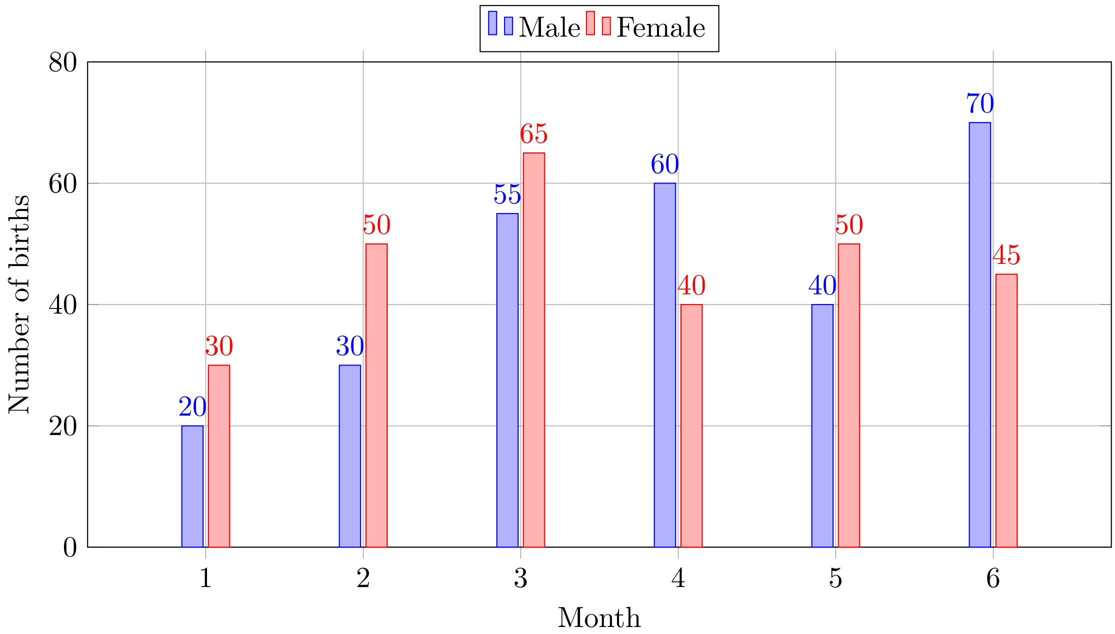


Figure 2.2: Multiple bar chart of male and female births by month.

**Interpretation.** The third month recorded the highest total births. Male births exceeded female births in months 4 and 6, whereas female births exceeded male births in months 1, 2, 3 and 5.

## 2.4 Pie Chart

A pie chart is a circle partitioned into sectors. Each sector angle is computed by

$$\text{sector angle} = \frac{\text{class frequency}}{\text{total frequency}} \times 360^\circ.$$

**Worked Example 2.3: Pie chart of attendance by country.** Attendance at a youth forum is shown below:

| Country | Attendance |
|---|---|
| Liberia | 180 |
| Gambia | 300 |
| Ghana | 240 |
| Nigeria | 360 |
| Sierra Leone | 120 |
| Total | 1200 |

**Solution.** The sector angles are:

$$\begin{aligned}
\text{Liberia} &= \frac{180}{1200} \times 360 = 54^\circ, \\
\text{Gambia} &= \frac{300}{1200} \times 360 = 90^\circ, \\
\text{Ghana} &= \frac{240}{1200} \times 360 = 72^\circ, \\
\text{Nigeria} &= \frac{360}{1200} \times 360 = 108^\circ, \\
\text{Sierra Leone} &= \frac{120}{1200} \times 360 = 36^\circ.
\end{aligned}$$

The angles sum to 360°, so the calculations are correct.

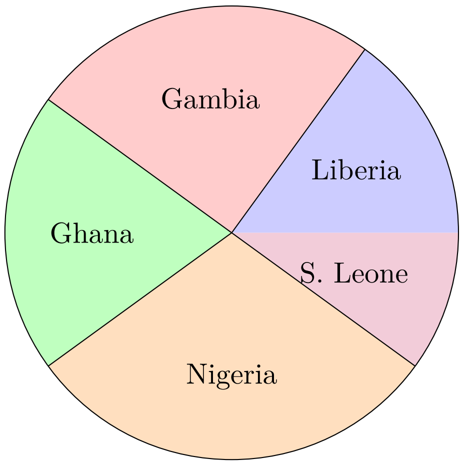


Figure 2.3: Pie chart of forum attendance by country.

## 2.5 Grouped Data, Histograms, Frequency Polygons and Ogives

For large or continuous data sets, raw values are grouped into class intervals. Each interval has class limits, class boundaries, class width and class mark.

Consider the following grouped distribution of scores of 60 students.

| Class | Frequency | Cumulative Frequency | Class Mark | Lower Boundary | Upper Boundary |
|---|---|---|---|---|---|
| 10–19 | 5 | 5 | 14.5 | 9.5 | 19.5 |
| 20–29 | 18 | 23 | 24.5 | 19.5 | 29.5 |
| 30–39 | 11 | 34 | 34.5 | 29.5 | 39.5 |
| 40–49 | 14 | 48 | 44.5 | 39.5 | 49.5 |
| 50–59 | 6 | 54 | 54.5 | 49.5 | 59.5 |
| 60–69 | 4 | 58 | 64.5 | 59.5 | 69.5 |
| 70–79 | 2 | 60 | 74.5 | 69.5 | 79.5 |

### 2.5.1 Histogram

A histogram consists of touching rectangles because the class intervals are continuous.

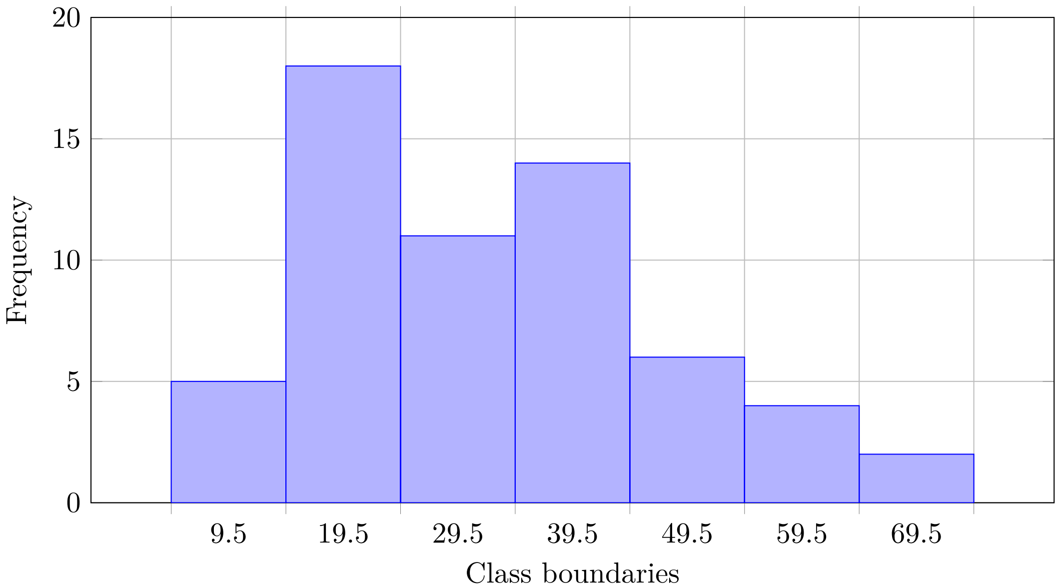


Figure 2.4: Histogram of grouped scores.

### 2.5.2 Frequency Polygon

A frequency polygon joins the class marks against the frequencies.

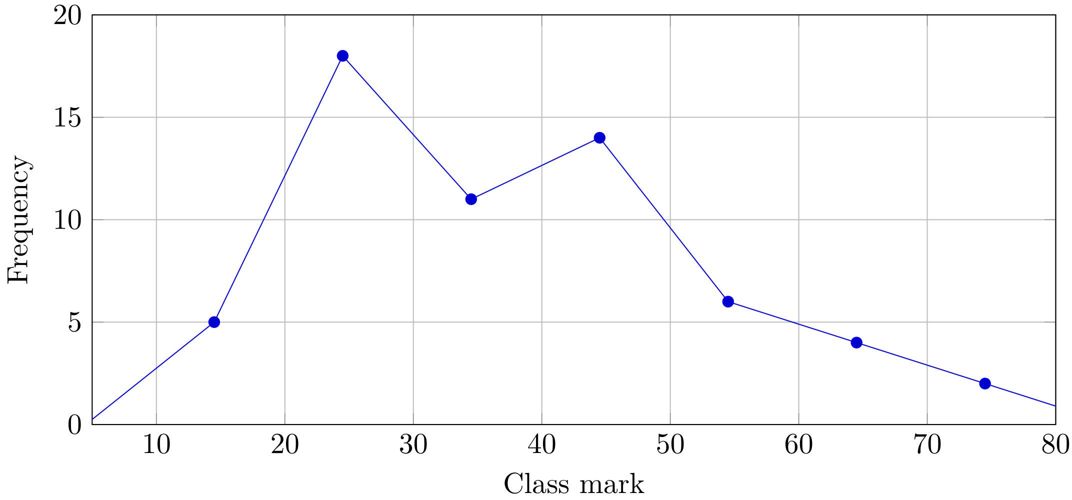


Figure 2.5: Frequency polygon for the grouped scores.

### 2.5.3 Less-than Ogive

The less-than ogive is obtained by plotting upper class boundaries against cumulative frequencies.

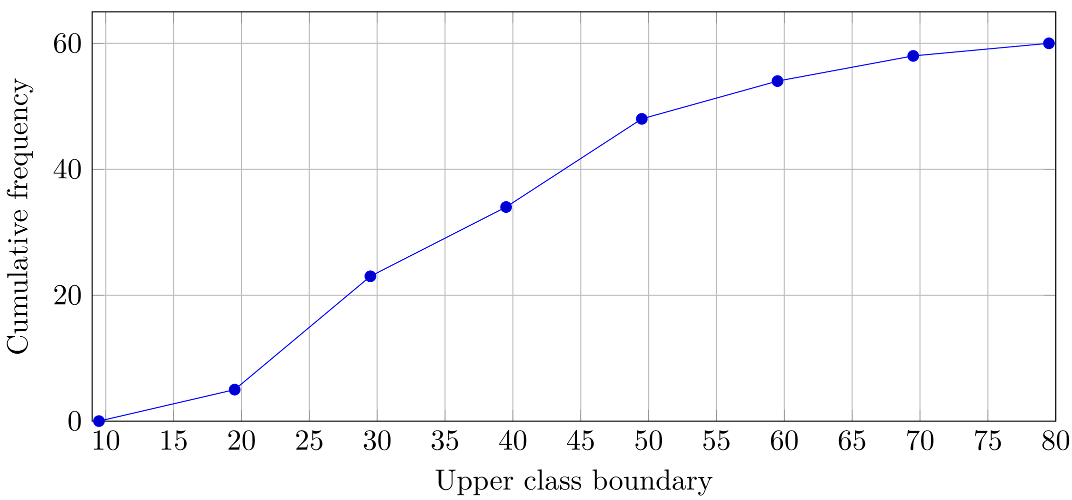


Figure 2.6: Less-than ogive for the grouped scores.

**Worked Example 2.4: Reading information from graphs.** From the ogive above, the median corresponds approximately to the point where the cumulative frequency is $N/2 = 30$. The value read on the horizontal axis lies in the class 30–39, which agrees with the formal median class obtained algebraically.

## 2.6 Sturges' Rule

If the number of classes is not given, a useful rule is

$$K = 1 + 3.322 \log_{10}(n),$$

where $n$ is the number of observations.

**Worked Example 2.5: Number of classes and class width.** For a data set of size $n = 60$, we have

$$K = 1 + 3.322 \log_{10}(60) = 1 + 3.322(1.7782) \approx 6.91 \approx 7 \text{ classes.}$$

If the minimum and maximum values are 12 and 79 respectively, then the range is $R = 79 - 12 = 67$. The class width is approximately

$$W = \frac{R}{K} = \frac{67}{7} \approx 9.57,$$

so a convenient width of 10 may be used.

## 2.7 Presentation of Data: Past-question Style Problems

**Problem 1.** The following data give the preferred means of transport of 80 university students: Bus 24, Taxi 10, Bike 12, Tricycle 18, Trekking 16. Represent the data using (i) a table, (ii) a bar chart and (iii) a pie chart.

**Problem 2.** Construct a grouped frequency table for the data range 21 to 78 using class width 10, then draw a histogram, frequency polygon and less-than ogive.

**Problem 3.** A library recorded the numbers of books borrowed on seven consecutive days as 45, 56, 62, 40, 51, 49, 60. Present the data with a line graph and comment on the trend.

Chapter 3

# Measures of Location

## 3.1 Arithmetic Mean

The arithmetic mean is the most common average. For ungrouped data $x_1, x_2, \ldots, x_n$,

$$\bar{x} = \frac{\sum x}{n}.$$

For grouped data with frequencies $f$ and class marks $x$,

$$\bar{x} = \frac{\sum fx}{\sum f}.$$

**Worked Example 3.1: Mean of ungrouped data.** Using the data set

$$5, 3, 6, 4, 8, 2, 7, 8, 3, 1, 9, 6, 4, 2, 4,$$

compute the mean.
**Solution.** The sum is

$$5 + 3 + 6 + 4 + 8 + 2 + 7 + 8 + 3 + 1 + 9 + 6 + 4 + 2 + 4 = 72.$$

Since $n = 15$,

$$\bar{x} = \frac{72}{15} = 4.8.$$

### 3.1.1 Weighted Mean

If observations carry different weights, then the weighted mean is

$$\bar{x}_w = \frac{\sum wx}{\sum w}.$$

This is useful for course grading where credits differ.

**Worked Example 3.2: Weighted mean.** A student scores 70, 62 and 80 in three courses of credit units 3, 2 and 4 respectively. Find the weighted average.
**Solution.**

$$\bar{x}_w = \frac{3(70) + 2(62) + 4(80)}{3 + 2 + 4} = \frac{210 + 124 + 320}{9} = \frac{654}{9} = 72.67.$$

## 3.2 Median

The median is the middle observation after arranging the data in order.

- If $n$ is odd, the median is the observation in position $(n + 1)/2$.
- If $n$ is even, the median is the mean of the two central observations.

**Worked Example 3.3: Median of ungrouped data.** Arrange

$$5, 3, 6, 4, 8, 2, 7, 8, 3, 1, 9, 6, 4, 2, 4$$

in ascending order:

$$1, 2, 2, 3, 3, 4, 4, 4, 5, 6, 6, 7, 8, 8, 9.$$

Since $n = 15$, the median position is

$$\frac{15 + 1}{2} = 8.$$

The 8th observation is 4, so the median is $\boxed{4}$.

### 3.2.1 Median of Grouped Data

For grouped data,

$$\text{Median} = L + \left( \frac{\frac{N}{2} - c_f}{f_m} \right) h,$$

where $L$ is the lower class boundary of the median class, $N$ is the total frequency, $c_f$ is the cumulative frequency before the median class, $f_m$ is the frequency of the median class, and $h$ is class width.

**Worked Example 3.4: Median of grouped data.** Consider the distribution:

| Class | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70–79 | 80–89 | 90–99 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 2 | 5 | 6 | 10 | 11 | 8 | 5 | 3 |

**Solution.** First compute cumulative frequencies:

$$2, 7, 13, 23, 34, 42, 47, 50.$$

Hence $N = 50$ and $N/2 = 25$. The median class is 60–69 because the first cumulative frequency that exceeds 25 is 34. Now

$$L = 59.5, \quad c_f = 23, \quad f_m = 11, \quad h = 10.$$

Thus

$$\text{Median} = 59.5 + \left(\frac{25 - 23}{11}\right) 10 = 59.5 + 1.818 = 61.32.$$

So the median is approximately $\boxed{61.32}$.

## 3.3 Mode

The mode is the most frequent observation. For grouped data,

$$\text{Mode} = L + \left(\frac{d_1}{d_1 + d_2}\right) h,$$

where $L$ is the lower boundary of the modal class, $d_1 = f_1 - f_0$, $d_2 = f_1 - f_2$, and $h$ is class width.

**Worked Example 3.5: Mode of grouped data.**

| Class interval | Frequency |
|---|---|
| 0–9 | 11 |
| 10–19 | 6 |
| 20–29 | 18 |
| 30–39 | 26 |
| 40–49 | 9 |

The modal class is 30–39.
**Solution.** Here $L = 29.5$, $f_1 = 26$, $f_0 = 18$, $f_2 = 9$, $h = 10$. Hence

$$d_1 = 26 - 18 = 8, \qquad d_2 = 26 - 9 = 17.$$

So

$$\text{Mode} = 29.5 + \left(\frac{8}{8+17}\right) 10 = 29.5 + 3.2 = 32.7.$$

Thus the mode is approximately $\boxed{32.7}$.

## 3.4 Other Partition Values

### 3.4.1 Quartiles

Quartiles divide the data into four equal parts. $Q_1$ is the first quartile, $Q_2$ is the median and $Q_3$ is the third quartile.

### 3.4.2 Deciles and Percentiles

Deciles divide the data into ten equal parts while percentiles divide the data into one hundred equal parts.

**Worked Example 3.6: Quartiles of an ordered list.** For the ordered data

$$2, 4, 5, 7, 8, 10, 12, 13, 15, 18, 20,$$

find $Q_1, Q_2, Q_3$.
**Solution.** The median is the 6th item, so $Q_2 = 10$. The lower half is $2, 4, 5, 7, 8$ whose median is 5, so $Q_1 = 5$. The upper half is $12, 13, 15, 18, 20$ whose median is 15, so $Q_3 = 15$.

## 3.5 Past-question Style Problems

1. Find the mean, median and mode of the data: 6, 4, 7, 9, 2, 5, 4, 4, 8, 6, 3.
2. The frequency distribution of daily sales is given as follows: values 2, 3, 4, 5, 6, 7 with corresponding frequencies 1, 3, 5, 7, 4, 2. Find the mean and median.
3. Compute the grouped median and mode for the class intervals 10–19, 20–29, 30–39, 40–49, 50–59 with frequencies 4, 8, 15, 9, 4.
4. Explain when the mean is preferred to the median and when the median is preferred to the mean.

Chapter 4

# Measures of Dispersion

## 4.1 Range

The range is the difference between the largest and smallest observations:

$$\text{Range} = x_{\max} - x_{\min}.$$

**Worked Example 4.1.** For the data set

$$5, 3, 6, 4, 8, 2, 7, 8, 3, 1, 9, 6, 4, 2, 4,$$

the minimum is 1 and the maximum is 9. Therefore

$$\text{Range} = 9 - 1 = 8.$$

## 4.2 Variance and Standard Deviation

For a population,

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}, \qquad \sigma = \sqrt{\sigma^2}.$$

For a sample,

$$s^2 = \frac{\sum(x-\bar{x})^2}{n}, \qquad s = \sqrt{s^2},$$

which is the convention used in this note for descriptive computation.

**Worked Example 4.2: Variance and standard deviation.** For the data

$$5, 3, 6, 4, 8, 2, 7, 8, 3, 1, 9, 6, 4, 2, 4,$$

we already found that $\bar{x} = 4.8$.
Construct the following table:

| $x$ | $x - \bar{x}$ | $(x - \bar{x})^2$ | Remark |
|---|---|---|---|
| 5 | 0.2 | 0.04 | |
| 3 | -1.8 | 3.24 | |
| 6 | 1.2 | 1.44 | |
| 4 | -0.8 | 0.64 | |
| 8 | 3.2 | 10.24 | |
| 2 | -2.8 | 7.84 | |
| 7 | 2.2 | 4.84 | |
| 8 | 3.2 | 10.24 | |
| 3 | -1.8 | 3.24 | |
| 1 | -3.8 | 14.44 | |
| 9 | 4.2 | 17.64 | |
| 6 | 1.2 | 1.44 | |
| 4 | -0.8 | 0.64 | |
| 2 | -2.8 | 7.84 | |
| 4 | -0.8 | 0.64 | |
| $\sum(x - \bar{x})^2 =$ | | 84.40 | |

Therefore

$$s^2 = \frac{84.40}{15} = 5.6267, \qquad s = \sqrt{5.6267} \approx 2.372.$$

## 4.3 Mean Deviation

The mean deviation about the mean is

$$\mathrm{MD} = \frac{\sum |x - \bar{x}|}{n}.$$

Though less widely used than variance and standard deviation, it gives a direct average of absolute deviations.

## 4.4 Coefficient of Variation

The coefficient of variation compares dispersion relative to the mean:

$$\text{CV} = \frac{s}{\bar{x}} \times 100\%.$$

**Worked Example 4.3.** For the same data set, with $s \approx 2.372$ and $\bar{x} = 4.8$,

$$\text{CV} = \frac{2.372}{4.8} \times 100\% = 49.42\%.$$

This indicates that the spread is roughly half the magnitude of the mean.

## 4.5 Interquartile Range

The interquartile range is

$$\text{IQR} = Q_3 - Q_1.$$

It measures the spread of the middle 50% of the data and is less sensitive to extreme values than the range.

**Past-question Style Set.**

1. Compute the range, variance, standard deviation and coefficient of variation for the data 2, 4, 4, 5, 7, 8, 9.
2. Two classes have means and standard deviations $(58, 6)$ and $(72, 9)$ respectively. Which class is more consistent? Use the coefficient of variation.
3. For the grouped data with class marks 5, 15, 25, 35 and frequencies 4, 7, 12, 5, compute the mean and standard deviation.

## Chapter 5

# Skewness and Kurtosis

## 5.1 Skewness

Skewness measures asymmetry. A distribution may be:

- **positively skewed**: long tail to the right;
- **negatively skewed**: long tail to the left;
- **symmetric**: equal shape on both sides.

Pearson's first coefficient is

$$\text{Skewness} = \frac{\bar{x} - \text{Mode}}{s}.$$

Pearson's second coefficient is

$$\text{Skewness} = \frac{3(\bar{x} - \text{Median})}{s}.$$

**Worked Example 5.1.** Using the data set with mean 4.8, median 4, mode 4 and standard deviation 2.372:

$$\text{Pearson's first skewness} = \frac{4.8 - 4}{2.372} = 0.337.$$

Also,

$$\text{Pearson's second skewness} = \frac{3(4.8 - 4)}{2.372} = 1.012.$$

The positive values indicate a positively skewed distribution.

## 5.2 Kurtosis

Kurtosis describes the peakedness or flatness of a distribution relative to the normal curve.

- **Leptokurtic**: more peaked than normal.

- **Mesokurtic**: same peakedness as normal.
- **Platykurtic**: flatter than normal.

In elementary courses, kurtosis is often treated descriptively rather than through moment calculations.

## 5.3 Interpretation of Shape from Mean, Median and Mode

A useful classroom rule is:

- for a positively skewed distribution: Mean $>$ Median $>$ Mode;
- for a negatively skewed distribution: Mean $<$ Median $<$ Mode;
- for a symmetric distribution: Mean $=$ Median $=$ Mode.

1. Explain the difference between positive and negative skewness.
2. State whether each of the following likely describes a positively skewed or negatively skewed distribution: personal income, age at death in a retirement home, examination scores from an extremely difficult test.
3. For a data set with mean 36, median 32 and standard deviation 10, compute Pearson's second coefficient of skewness.

## Chapter 6

# Elementary Probability

## 6.1 Basic Concepts

Probability measures the likelihood of an event. If all outcomes in a sample space $S$ are equally likely and $A$ is an event, then

$$P(A) = \frac{n(A)}{n(S)}.$$

A probability always lies between 0 and 1.

## 6.2 Experiments, Outcomes and Sample Space

An experiment is a process whose outcome cannot be predicted with certainty. The set of all possible outcomes is called the sample space.

Examples:

- Tossing one coin: $S = \{H, T\}$.
- Tossing two coins: $S = \{HH, HT, TH, TT\}$.
- Throwing one die: $S = \{1, 2, 3, 4, 5, 6\}$.

### 6.2.1 Events

An event is any subset of the sample space. For example, when a die is thrown,

$$A = \{2, 4, 6\}$$

is the event of obtaining an even number.

**Worked Example 6.1: One die.** If a fair die is thrown once, find the probabilities of obtaining:

(a) a 4,

(b) an even number,

(c) a number greater than 4.

**Solution.** Since $n(S) = 6$:

$$P(4) = \frac{1}{6}.$$

Even numbers are $\{2, 4, 6\}$, so

$$P(\text{even}) = \frac{3}{6} = \frac{1}{2}.$$

Numbers greater than 4 are $\{5, 6\}$, so

$$P(> 4) = \frac{2}{6} = \frac{1}{3}.$$

## 6.3 Addition Law

If $A$ and $B$ are two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If $A$ and $B$ are mutually exclusive, then $P(A \cap B) = 0$, so

$$P(A \cup B) = P(A) + P(B).$$

## 6.4 Multiplication Law

For two independent events $A$ and $B$,

$$P(A \cap B) = P(A)P(B).$$

For dependent events,

$$P(A \cap B) = P(A)P(B \mid A).$$

## 6.5 Tree Diagrams

Tree diagrams help represent sequential experiments such as repeated coin tosses or drawing balls from a bag.

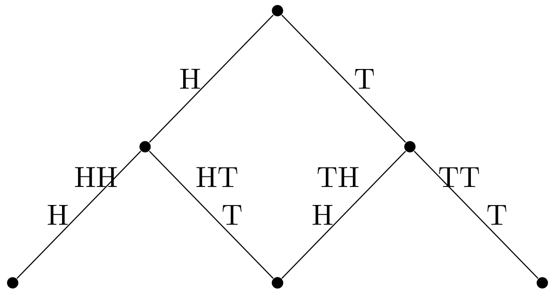


Figure 6.1: Tree diagram for tossing a coin twice.

**Worked Example 6.2: Two coins tossed simultaneously.** Find the probability that:

(a) both show heads,

(b) exactly one head occurs,

(c) both show the same face.

**Solution.** The sample space is

$$S = \{HH, HT, TH, TT\}.$$

Hence

$$P(\text{both heads}) = \frac{1}{4}.$$

Exactly one head occurs in $\{HT, TH\}$, so

$$P(\text{exactly one head}) = \frac{2}{4} = \frac{1}{2}.$$

Both show the same face in $\{HH, TT\}$, so

$$P(\text{same face}) = \frac{2}{4} = \frac{1}{2}.$$

## 6.6 Sampling With and Without Replacement

**Worked Example 6.3: Drawing balls from a bag.** A bag contains 3 black, 2 blue and 4 red balls. Two balls are drawn one after the other.

**(a) With replacement.** The total number of balls is 9. Because the first ball is replaced, the second draw has the same probabilities as the first.

$$P(\text{both black}) = \frac{3}{9} \times \frac{3}{9} = \frac{1}{9}.$$

$$P(\text{both blue}) = \frac{2}{9} \times \frac{2}{9} = \frac{4}{81}.$$

$$P(\text{first blue, second red}) = \frac{2}{9} \times \frac{4}{9} = \frac{8}{81}.$$

**(b) Without replacement.** Now the composition changes after the first draw.

$$P(\text{both black}) = \frac{3}{9} \times \frac{2}{8} = \frac{1}{12}.$$

$$P(\text{both blue}) = \frac{2}{9} \times \frac{1}{8} = \frac{1}{36}.$$

$$P(\text{first blue, second red}) = \frac{2}{9} \times \frac{4}{8} = \frac{1}{9}.$$

**Worked Example 6.4: Coloured balls.** A bag contains 3 red, 9 white, 6 blue and 3 orange balls. One ball is picked at random. The total is $3 + 9 + 6 + 3 = 21$.

$$P(\text{red}) = \frac{3}{21} = \frac{1}{7}, \qquad P(\text{white}) = \frac{9}{21} = \frac{3}{7},$$

$$P(\text{blue}) = \frac{6}{21} = \frac{2}{7}, \qquad P(\text{not orange}) = 1 - \frac{3}{21} = \frac{18}{21} = \frac{6}{7}.$$

## 6.7 Probability Review Questions

1. A card is drawn from a standard deck of 52 cards. Find the probability that it is (i) a king, (ii) a red card, (iii) a queen or a heart.
2. Two dice are thrown. Find the probability that the sum is 7.
3. In a class of 20 students, 12 are male and 8 are female. Two students are selected without replacement. Find the probability that both are female.
4. State the difference between independent events and mutually exclusive events.

Chapter 7

# Random Variables and Probability Distributions

## 7.1 Random Variables

A random variable is a numerical quantity whose value depends on chance. If the number of heads in three coin tosses is denoted by $X$, then $X$ can take the values 0, 1, 2, 3.

### 7.1.1 Discrete and Continuous Random Variables

A **discrete** random variable takes countable values. A **continuous** random variable takes values on an interval.

## 7.2 Probability Distribution of a Discrete Random Variable

A probability distribution lists all possible values of the random variable together with their probabilities.

**Worked Example 7.1: Number of heads in three coin tosses.** Let $X$ be the number of heads obtained when a fair coin is tossed three times. The sample space has 8 equally likely outcomes.

| $X$ | Outcomes | $P(X = x)$ |
|---|---|---|
| 0 | TTT | 1/8 |
| 1 | HTT,THT,TTH | 3/8 |
| 2 | HHT,HTH,THH | 3/8 |
| 3 | HHH | 1/8 |

Hence

$$P(X = 0) = \frac{1}{8},\ P(X = 1) = \frac{3}{8},\ P(X = 2) = \frac{3}{8},\ P(X = 3) = \frac{1}{8}.$$

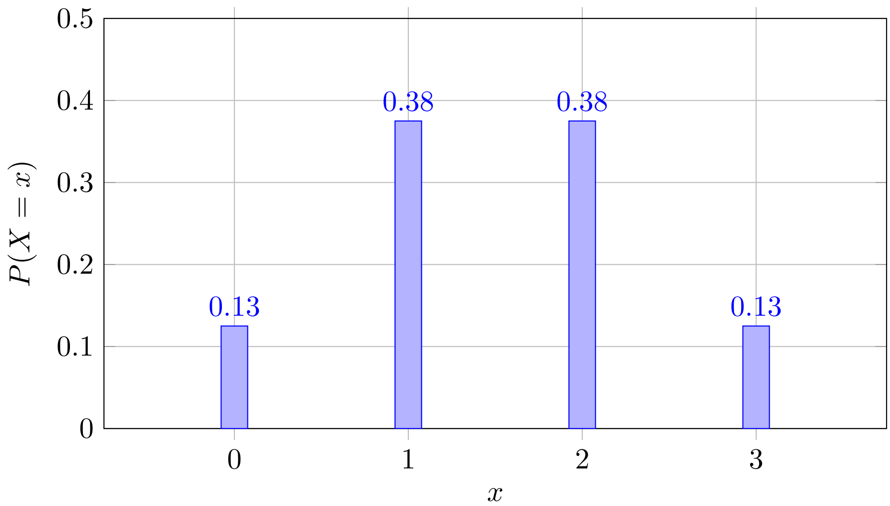


Figure 7.1: Probability distribution for number of heads in three tosses.

## 7.3 Expectation and Variance of a Discrete Random Variable

The expected value is

$$E(X) = \sum xP(X = x),$$

and the variance is

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2.$$

**Worked Example 7.2.** For the distribution above,

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5.$$

Also,

$$E(X^2) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = \frac{24}{8} = 3.$$

Hence

$$\mathrm{Var}(X) = 3 - (1.5)^2 = 0.75.$$

Chapter 8

# Binomial Distribution

## 8.1 Definition

A binomial experiment has the following properties:

1. a fixed number $n$ of trials;
2. each trial has only two possible outcomes: success or failure;
3. the probability of success $p$ is constant from trial to trial;
4. the trials are independent.

If $X$ is the number of successes in $n$ trials, then

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, \ldots, n,$$

where $q = 1 - p$.

## 8.2 Mean and Variance

For $X \sim B(n, p)$,

$$E(X) = np, \qquad \mathrm{Var}(X) = npq.$$

**Worked Example 8.1: Basketball free throws.** A basketball player makes three-fourths of his free throws. Assuming independence, find the probability that he makes all of the next four free throws.

**Solution.** Here $n = 4$, $p = 3/4$, and we want $P(X = 4)$.

$$P(X = 4) = \binom{4}{4}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^0 = \left(\frac{3}{4}\right)^4 = \frac{81}{256} = 0.3164.$$

**Worked Example 8.2: Exactly two successes.** A student guesses on five true-false questions. If the probability of a correct guess is 1/2, find the probability of exactly two correct answers.
**Solution.** Here $n = 5$, $p = 1/2$, $q = 1/2$ and $x = 2$.

$$P(X = 2) = \binom{5}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 = \binom{5}{2}\left(\frac{1}{2}\right)^5 = 10 \cdot \frac{1}{32} = \frac{5}{16}.$$

**Worked Example 8.3: At least one defect.** If 10% of bulbs are defective and 4 bulbs are selected independently, find the probability that at least one is defective.
**Solution.** Let success be defective. Then $n = 4$, $p = 0.1$, $q = 0.9$. Using the complement,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0.9)^4 = 1 - 0.6561 = 0.3439.$$

## 8.3 Binomial Probability Table of a Typical Example

For $n = 4$ and $p = 0.75$, the distribution is:

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(X = x)$ | 0.0039 | 0.0469 | 0.2109 | 0.4219 | 0.3164 |

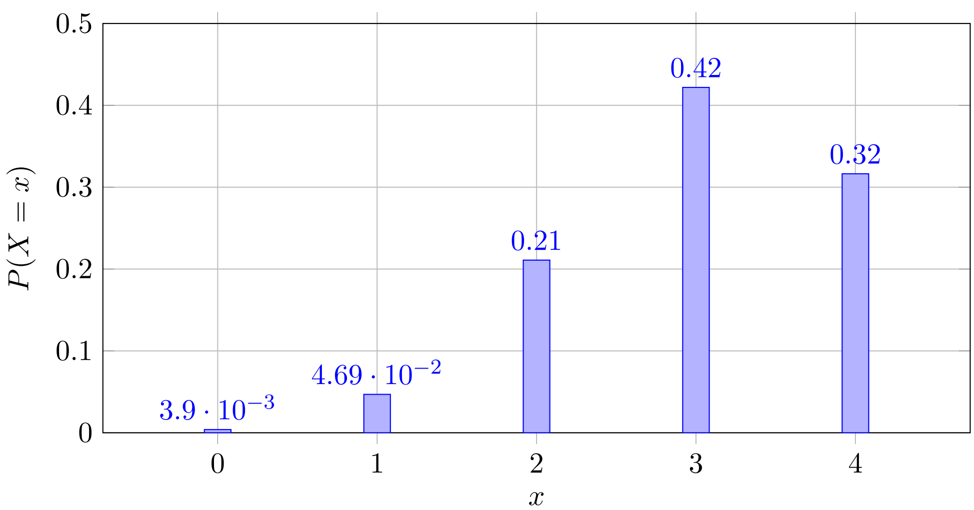


Figure 8.1: Binomial distribution for $n = 4$, $p = 0.75$.

1. A fair coin is tossed 6 times. Find the probability of getting exactly 4 heads.
2. A candidate has probability 0.8 of passing an oral interview. If 5 candidates are assessed independently, find the probability that exactly 3 pass.
3. A student answers 8 multiple-choice questions by guessing. Each question has four options. Find the probability that the student gets at least one answer correct.

Chapter 9

# Poisson Distribution

## 9.1 Definition

A random variable $X$ has the Poisson distribution with parameter $\lambda > 0$ if

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

The Poisson distribution models the number of occurrences of rare events in a fixed interval of time, length, area or volume.

For a Poisson random variable,

$$E(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda.$$

**Worked Example 9.1: Telephone calls.** If a help desk receives on average 3 calls every 10 minutes, find the probability of receiving exactly 2 calls in a 10-minute period.
**Solution.** Here $\lambda = 3$ and $x = 2$.

$$P(X = 2) = \frac{e^{-3}3^2}{2!} = \frac{9e^{-3}}{2} \approx \frac{9(0.04979)}{2} = 0.2240.$$

**Worked Example 9.2: At most one arrival.** Using the same mean $\lambda = 3$, find the probability of at most one call.
**Solution.**

$$P(X \leq 1) = P(0) + P(1) = e^{-3} + 3e^{-3} = 4e^{-3} \approx 4(0.04979) = 0.1992.$$

## 9.2 Poisson as Approximation to Binomial

When $n$ is large and $p$ is small, with $np = \lambda$, the Poisson distribution approximates the binomial distribution.

**Worked Example 9.3.** If a factory produces items with defect probability 0.01 independently and a batch of 100 items is inspected, then $X \sim B(100, 0.01)$ may be approximated by Poisson($\lambda = 1$). Thus

$$P(X = 0) \approx e^{-1} = 0.3679,$$

which is the approximate probability of no defect.

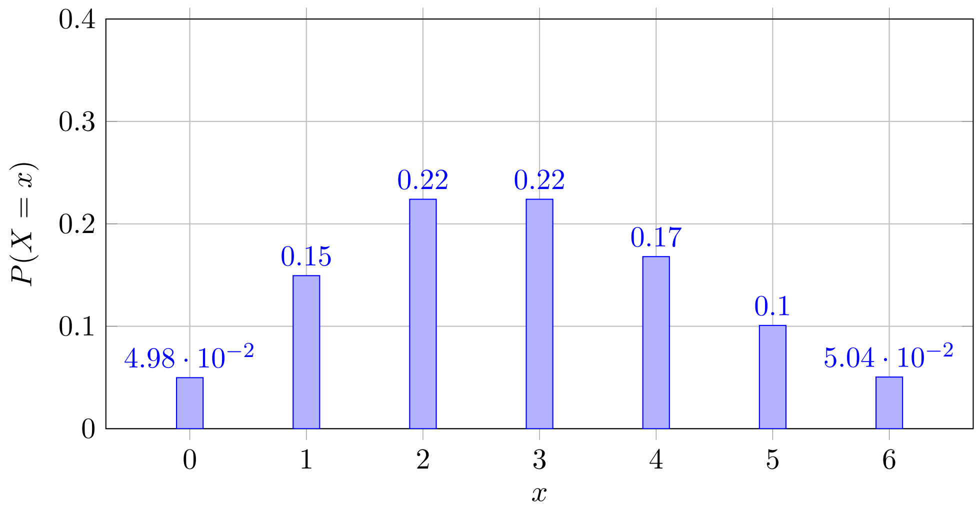


Figure 9.1: Poisson distribution with parameter $\lambda = 3$ (rounded values).

1. The average number of road accidents at a junction is 2 per week. Assuming a Poisson model, find the probability of exactly 3 accidents in a given week.
2. Find the probability of at least one accident in the situation above.
3. A page of a book has on average 0.4 typographical errors. Find the probability that a page contains no error.

Chapter 10

# Normal Distribution

## 10.1 The Normal Curve

The normal distribution is a continuous probability distribution with bell shape, symmetric about its mean. If $X \sim N(\mu, \sigma^2)$, then its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Important properties:

- the mean, median and mode are equal;
- the curve is symmetric about $x = \mu$;
- total area under the curve is 1;
- the tails approach but never touch the horizontal axis.

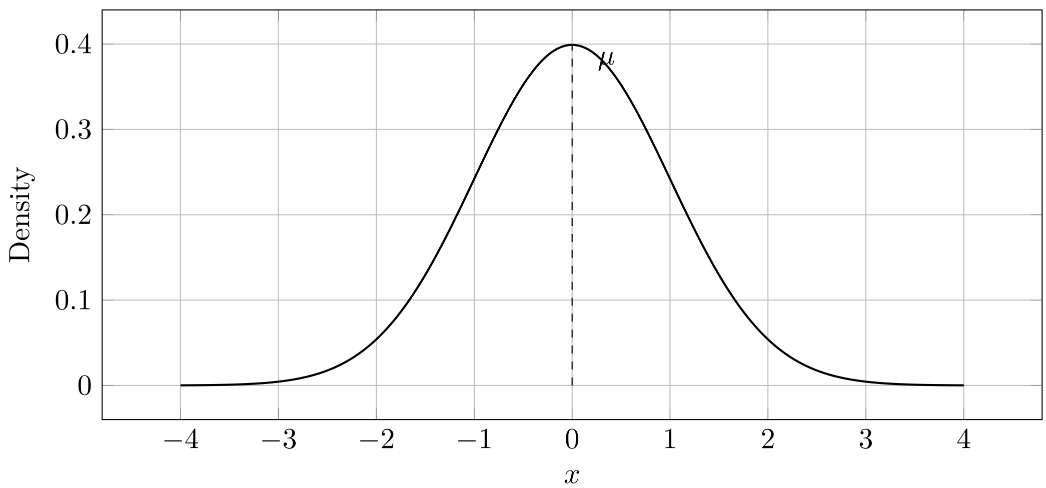


Figure 10.1: Standard normal curve.

## 10.2 Standardization

If $X \sim N(\mu, \sigma^2)$, then the standardized variable

$$Z = \frac{X - \mu}{\sigma}$$

has the standard normal distribution $N(0, 1)$.

**Worked Example 10.1.** Suppose test scores are normally distributed with mean 60 and standard deviation 8. Find the $z$-score corresponding to a score of 76.
**Solution.**

$$z = \frac{76 - 60}{8} = 2.$$

So a score of 76 is two standard deviations above the mean.

**Worked Example 10.2.** In the same distribution, find the score corresponding to $z = -1.5$.
**Solution.**

$$X = \mu + z\sigma = 60 + (-1.5)(8) = 60 - 12 = 48.$$

## 10.3 Empirical Rule

For a normal distribution:

- about 68% of observations lie within $\mu \pm \sigma$;
- about 95% lie within $\mu \pm 2\sigma$;
- about 99.7% lie within $\mu \pm 3\sigma$.

**Worked Example 10.3.** If body weights have mean 70 kg and standard deviation 5 kg, approximately 95% of the weights lie between

$$70 - 2(5) = 60 \text{ kg and } 70 + 2(5) = 80 \text{ kg}.$$

1. For $X \sim N(50, 4^2)$, find the $z$-score of $x = 58$.
2. If a distribution has mean 100 and standard deviation 15, what raw score corresponds to $z = 1.2$?
3. State three properties of the normal distribution.
4. Explain why the mean, median and mode coincide for a normal distribution.

Chapter 11

# Regression and Correlation

## 11.1 Scatter Diagram

A scatter diagram is a graph of paired observations $(x, y)$. It helps us see whether a relationship exists between two variables.

**Worked Example 11.1: Scatter plot of JAMB score and test score.** Consider the pairs:

$$(40, 7), (70, 11), (50, 8), (60, 10), (80, 14), (60, 12), (10, 8), (20, 9), (10, 6).$$

The scatter plot appears in Figure 11.1. The points show a general upward trend, indicating positive association.

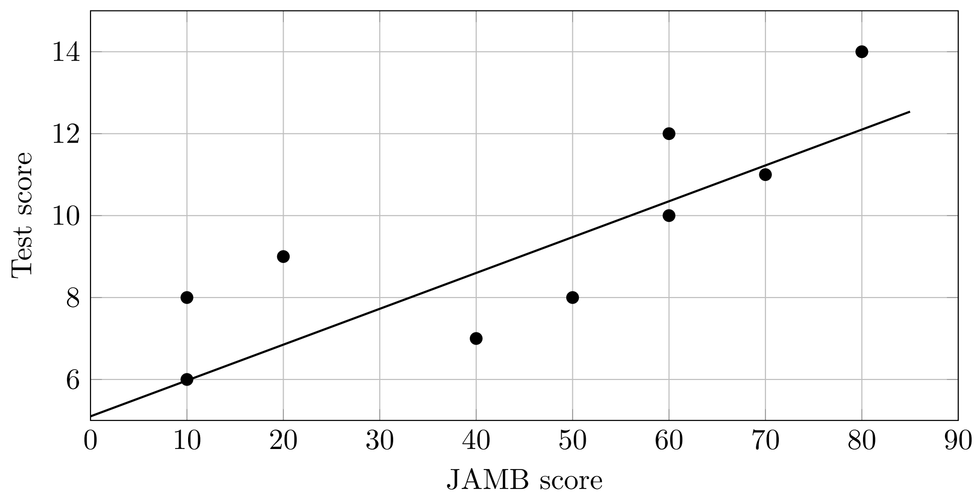


Figure 11.1: Scatter plot with fitted line for JAMB score and test score.

## 11.2 Simple Linear Regression

The regression line of $Y$ on $X$ has the form

$$Y = a + bX,$$

where $a$ is the intercept and $b$ is the slope. The least squares formulas are

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}, \qquad a = \bar{Y} - b\bar{X}.$$

**Worked Example 11.2: Fitting a regression line.** Fit the regression line of $Y$ on $X$ for the data:

| $X$ | 1 | 3 | 4 | 6 | 8 | 9 | 11 | 14 |
|---|---|---|---|---|---|---|---|---|
| $Y$ | 1 | 2 | 4 | 4 | 5 | 7 | 8 | 9 |

Construct the table:

| $X$ | $Y$ | $XY$ | $X^2$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 3 | 2 | 6 | 9 |
| 4 | 4 | 16 | 16 |
| 6 | 4 | 24 | 36 |
| 8 | 5 | 40 | 64 |
| 9 | 7 | 63 | 81 |
| 11 | 8 | 88 | 121 |
| 14 | 9 | 126 | 196 |
| $\sum X = 56$ | $\sum Y = 40$ | $\sum XY = 364$ | $\sum X^2 = 524$ |

Now $n = 8$, so

$$b = \frac{8(364) - 56(40)}{8(524) - 56^2} = \frac{2912 - 2240}{4192 - 3136} = \frac{672}{1056} = 0.6364.$$

Also,

$$\bar{X} = \frac{56}{8} = 7, \qquad \bar{Y} = \frac{40}{8} = 5.$$

Hence

$$a = \bar{Y} - b\bar{X} = 5 - 0.6364(7) = 5 - 4.4548 = 0.5452.$$

Therefore the regression equation is

$$\boxed{Y = 0.545 + 0.636X.}$$

## 11.3 Interpretation of Regression Coefficient

The slope $b = 0.636$ means that, on average, a one-unit increase in $X$ is associated with an increase of about 0.636 units in $Y$.

## 11.4 Correlation Coefficient

Pearson's product moment correlation coefficient is

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2]\,[n\sum Y^2 - (\sum Y)^2]}}.$$

Its value lies between $-1$ and 1.

- $r = 1$: perfect positive correlation;
- $r = -1$: perfect negative correlation;
- $r = 0$: no linear correlation.

**Worked Example 11.3: Pearson correlation.** Using the same data, first compute

$$\sum Y^2 = 1^2 + 2^2 + 4^2 + 4^2 + 5^2 + 7^2 + 8^2 + 9^2 = 256.$$

Then

$$r = \frac{8(364) - 56(40)}{\sqrt{[8(524) - 56^2][8(256) - 40^2]}} = \frac{672}{\sqrt{(1056)(448)}}.$$

Since

$$(1056)(448) = 473088, \qquad \sqrt{473088} \approx 687.81,$$

we obtain

$$r \approx \frac{672}{687.81} = 0.977.$$

This indicates a very strong positive linear relationship.

## 11.5 Spearman's Rank Correlation

If raw measurements are not available or the data are naturally ordinal, Spearman's rank correlation may be used:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)},$$

where $d$ is the difference between paired ranks.

1. Fit the regression line of $Y$ on $X$ for the data: $(3, 2), (5, 3), (6, 4), (8, 6), (9, 5), (11, 8)$.
2. Explain the meaning of a negative correlation coefficient.
3. If $r = 0.92$, comment on the strength and direction of the linear relationship.
4. Distinguish between correlation and regression.

## Chapter 12

# Comprehensive Solved Past-Question Style Problems

This chapter presents longer university-style problems similar to continuous assessment and final examination questions.

## 12.1 Presentation of Data Problems

**Problem 12.1.** The numbers of students admitted into six departments are Mathematics 9, Chemistry 12, English 4, Economics 11, Home Economics 16 and Music 6.

(a) Construct a table.

(b) Draw a bar chart.

(c) State which department has the highest enrolment and by how much it exceeds Mathematics.

**Solution.** The table is immediate from the given data. The bar chart would have six separate bars with heights 9, 12, 4, 11, 16 and 6 respectively. Home Economics has the highest enrolment, namely 16, which exceeds Mathematics by $16 - 9 = 7$ students.

**Problem 12.2.** The grouped data below show the marks of 50 students:

| Class | 0–9 | 10–19 | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 |
|---|---|---|---|---|---|---|---|
| Frequency | 3 | 5 | 8 | 12 | 10 | 7 | 5 |

Construct the cumulative frequency table and state the median class.

**Solution.** The cumulative frequencies are 3, 8, 16, 28, 38, 45, 50. Since $N = 50$, we compute $N/2 = 25$. The first cumulative frequency greater than 25 is 28, corresponding to class 30–39. Therefore the median class is 30–39.

## 12.2 Measures of Location and Dispersion Problems

**Problem 12.3.** For the data set

$$12, 15, 14, 10, 9, 11, 13, 12, 16, 18, 17, 12,$$

find the mean, median, mode, range and standard deviation.
**Solution.**

- Sum = 159, so mean = 159/12 = 13.25.
- Ordered list: 9, 10, 11, 12, 12, 12, 13, 14, 15, 16, 17, 18. Median is the average of the 6th and 7th terms:
$$\frac{12+13}{2} = 12.5.$$
- Mode = 12.
- Range = 18 - 9 = 9.
- The squared deviations sum to 86.25, so variance = 86.25/12 = 7.1875 and standard deviation ≈ 2.681.

## 12.3 Probability Problems

**Problem 12.4.** A class has 7 male and 5 female students. Two students are selected at random without replacement.

(a) Find the probability that both are female.

(b) Find the probability that one is male and one is female.

**Solution.**

$$P(FF) = \frac{5}{12} \cdot \frac{4}{11} = \frac{20}{132} = \frac{5}{33}.$$

For one male and one female,

$$P(MF) + P(FM) = \frac{7}{12} \cdot \frac{5}{11} + \frac{5}{12} \cdot \frac{7}{11} = \frac{70}{132} = \frac{35}{66}.$$

## 12.4 Distribution Problems

**Problem 12.5.** A machine produces defective items with probability 0.2 independently. If 6 items are selected, find the probability that exactly 2 are defective.
**Solution.** The model is binomial with $n = 6$, $p = 0.2$, $q = 0.8$.

$$P(X = 2) = \binom{6}{2}(0.2)^2(0.8)^4 = 15(0.04)(0.4096) = 0.24576.$$

**Problem 12.6.** Cars arrive at a fuel station according to a Poisson process with mean 4 per hour. Find the probability that exactly 5 cars arrive in an hour.
**Solution.**

$$P(X = 5) = \frac{e^{-4}4^5}{5!} = \frac{1024e^{-4}}{120} \approx 0.1563.$$

## 12.5 Regression and Correlation Problems

**Problem 12.7.** Explain why a high correlation does not necessarily imply causation.
**Solution.** Correlation measures association, not cause and effect. Two variables may be strongly correlated because each responds to a third factor, because of coincidence in a small sample, or because of structural constraints in the data. Statistical evidence of causation requires more than a large $r$-value.

Chapter 13

# Extended Revision Problems and Model Solutions

This chapter adds more university-style revision questions so that students can practise repeatedly before tests and final examinations. Each question is followed by a compact model solution.

## 13.1 Presentation of Data Revision Set

**Revision Problem 1.** The table below gives the weekly patronage of a departmental cybercafe.

| Day | Mon | Tue | Wed | Thu | Fri | Sat |
|---|---|---|---|---|---|---|
| Users | 42 | 48 | 55 | 51 | 63 | 39 |

Draw a bar chart and comment on the busiest day.

**Model Solution.** The tallest bar corresponds to Friday with 63 users. Saturday recorded the least value, 39. The general pattern shows rising patronage from Monday to Friday and a drop on Saturday.

**Revision Problem 2.** The age distribution of 40 workers is given below.

| Age class | 20–24 | 25–29 | 30–34 | 35–39 | 40–44 |
|---|---|---|---|---|---|
| Frequency | 5 | 9 | 13 | 8 | 5 |

State the modal class, the median class and sketch the expected histogram.

**Model Solution.** Since the highest frequency is 13, the modal class is 30–34. The cumulative frequencies are 5, 14, 27, 35, 40. Since $N/2 = 20$, the median class is also 30–34.

## 13.2 Measures of Location and Dispersion Revision Set

**Revision Problem 3.** Find the mean, median and range of 11, 15, 14, 18, 11, 13, 12, 17, 16.
**Model Solution.** The sum is 127, so the mean is $127/9 = 14.11$. The ordered values are 11, 11, 12, 13, 14, 15, 16, 17, 18, so the median is 14. The range is $18 - 11 = 7$.

**Revision Problem 4.** For the grouped distribution with class marks 5, 15, 25, 35 and frequencies 2, 6, 10, 4, compute the grouped mean.
**Model Solution.** We compute

$$\sum fx = 2(5) + 6(15) + 10(25) + 4(35) = 10 + 90 + 250 + 140 = 490.$$

Since $\sum f = 22$, the mean is

$$\bar{x} = \frac{490}{22} = 22.27.$$

## 13.3 Probability and Distribution Revision Set

**Revision Problem 5.** A fair coin is tossed four times. Find the probability of at most one head.
**Model Solution.** Let $X$ be the number of heads. Then $X \sim B(4, 0.5)$. Hence

$$P(X \leq 1) = P(0) + P(1) = \binom{4}{0}\left(\frac{1}{2}\right)^4 + \binom{4}{1}\left(\frac{1}{2}\right)^4 = \frac{1+4}{16} = \frac{5}{16}.$$

**Revision Problem 6.** Calls arrive at a help desk at mean rate 2 per minute. Under a Poisson model, find the probability of no call in a minute.
**Model Solution.** If $X \sim \text{Poisson}(2)$, then

$$P(X = 0) = e^{-2} = 0.1353.$$

**Revision Problem 7.** In a normal distribution with mean 45 and standard deviation 6, find the $z$-score of 57.
**Model Solution.**

$$z = \frac{57 - 45}{6} = 2.$$

So the observation lies two standard deviations above the mean.

## 13.4 Regression and Correlation Revision Set

**Revision Problem 8.** State the meaning of each of the following values of Pearson's correlation coefficient: (i) $r = 0.98$, (ii) $r = -0.52$, (iii) $r = 0$.
**Model Solution.** (i) very strong positive linear relationship; (ii) moderate negative linear relationship; (iii) no linear correlation.

**Revision Problem 9.** Explain in one paragraph why the least squares regression line is called a line of best fit.
**Model Solution.** The least squares line is called the line of best fit because among all possible straight lines, it minimizes the sum of the squares of the vertical deviations of the observed points from the line. Squaring ensures that positive and negative deviations do not cancel, and larger errors are penalized more heavily.

Chapter 14

# Assignments for Students

This chapter is designed as a take-home practice set.

## 14.1 Assignment A: Presentation of Data

1. The following marks were obtained by 30 students: 21, 24, 29, 23, 25, 28, 26, 22, 24, 20, 27, 29, 21, 25, 24, 23, 26, 28, 20, 22, 23, 25, 27, 29, 20, 21, 24, 26, 28, 27.
   (a) Construct an ungrouped frequency table.
   (b) Compute the cumulative frequency.
   (c) Draw a bar chart.
2. Group the following 24 observations into classes of width 10 and draw a histogram and ogive: 14, 18, 25, 31, 22, 46, 52, 61, 73, 44, 38, 27, 35, 49, 58, 63, 70, 41, 29, 33, 55, 47, 21, 66.

## 14.2 Assignment B: Measures of Central Tendency and Dispersion

1. Find the mean, median, mode, variance and standard deviation of 3, 5, 7, 7, 8, 9, 10, 12.
2. Compute the grouped mean and median for the class intervals 10–19, 20–29, 30–39, 40–49 with frequencies 5, 9, 7, 4.
3. Two data sets have means 24 and 60 with standard deviations 4 and 8 respectively. Compare their relative dispersions using the coefficient of variation.

## 14.3 Assignment C: Probability and Distributions

1. A coin is tossed five times. Find the probability of getting at least three heads.
2. A bag contains 4 white, 3 black and 5 red balls. Two balls are selected without replacement. Find the probability that both are red.
3. If $X \sim B(7, 0.3)$, find $P(X = 0)$, $P(X = 1)$ and $P(X \geq 2)$.
4. If $X \sim \text{Poisson}(2.5)$, find $P(X = 0)$ and $P(X \leq 2)$.
5. For $X \sim N(40, 5^2)$, find the $z$-score of 47.

## 14.4 Assignment D: Regression and Correlation

1. Fit the regression line of $Y$ on $X$ for the pairs $(2, 1), (4, 3), (5, 4), (7, 5), (8, 7)$.
2. Explain the practical meaning of an $r$-value close to zero.
3. For paired ranks with differences $d = 1, -1, 2, -2, 0$, compute Spearman's rank correlation.

Chapter 15

# Test and Examination Practice Section

## 15.1 Section A: Objective-style Questions

1. The average of 8, 10, 12 and 14 is:

   A. 9

   B. 10

   C. 11

   D. 12

2. The graph suitable for grouped continuous data is:

   A. Pie chart

   B. Histogram

   C. Simple bar chart

   D. Pictogram

3. If $P(A) = 0.3$ and $P(B) = 0.4$ for mutually exclusive events, then $P(A \cup B) =$

   A. 0.12

   B. 0.70

   C. 0.10

   D. 0.50

4. For a binomial distribution, the number of trials is:

   A. variable

   B. fixed

   C. zero

   D. continuous

5. A correlation coefficient near $-1$ indicates:

   A. no relationship

   B. weak positive relationship

   C. strong negative relationship

   D. exact equality

## 15.2 Section B: Theory and Computation

1. Explain the difference between a histogram and a bar chart.
2. Define primary and secondary data and give two examples of each.
3. Compute the mean and standard deviation of 4, 6, 8, 8, 9, 10.
4. A fair die is thrown twice. Find the probability that the sum is at least 10.
5. Fit the regression line of $Y$ on $X$ for the data given in Chapter 11.

Chapter A

# Solutions to Selected Exercises and Assignments

## A.1 Solutions from Chapter 1

**Review Exercise 1.** Descriptive statistics summarizes observed data, while inferential statistics uses sample information to draw conclusions about a population.

**Review Exercise 2.** In a faculty with 1,500 students, the population is all 1,500 students. If 100 are selected for study, those 100 form the sample.

## A.2 Solutions from Chapter 2

**Transport problem.** Total frequency $= 24 + 10 + 12 + 18 + 16 = 80$. Pie chart angles are:

$$\text{Bus} = \frac{24}{80} \times 360 = 108^\circ, \quad \text{Taxi} = 45^\circ, \quad \text{Bike} = 54^\circ,$$

$$\text{Tricycle} = 81^\circ, \quad \text{Trekking} = 72^\circ.$$

## A.3 Solutions from Chapter 3

For the data 6, 4, 7, 9, 2, 5, 4, 4, 8, 6, 3, the ordered list is 2, 3, 4, 4, 4, 5, 6, 6, 7, 8, 9. Thus

$$\bar{x} = \frac{58}{11} = 5.273, \quad \text{Median} = 5, \quad \text{Mode} = 4.$$

## A.4 Solutions from Chapter 4

For 2, 4, 4, 5, 7, 8, 9, the mean is

$$\bar{x} = \frac{39}{7} = 5.571.$$

The range is $9 - 2 = 7$. Students should compute the variance by first forming deviations from the mean.

## A.5 Solutions from Chapter 6

For the deck of cards problem:

$$P(\text{king}) = \frac{4}{52} = \frac{1}{13}, \quad P(\text{red}) = \frac{26}{52} = \frac{1}{2}.$$

The event queen or heart has size $4 + 13 - 1 = 16$, so

$$P(\text{queen or heart}) = \frac{16}{52} = \frac{4}{13}.$$

## A.6 Solutions to Assignment B

**Question 1.** For 3, 5, 7, 7, 8, 9, 10, 12,

$$\bar{x} = \frac{61}{8} = 7.625.$$

Median is the average of the 4th and 5th terms: $(7 + 8)/2 = 7.5$. Mode is 7. Variance and standard deviation follow from squared deviations from 7.625.

**Question 3.** The coefficients of variation are:

$$\text{CV}_1 = \frac{4}{24} \times 100 = 16.67\%, \qquad \text{CV}_2 = \frac{8}{60} \times 100 = 13.33\%.$$

Therefore the second data set is more consistent.

## A.7 Solutions to Assignment C

**Question 1.** For five fair coin tosses,

$$P(X \geq 3) = P(3) + P(4) + P(5) = \frac{\binom{5}{3} + \binom{5}{4} + \binom{5}{5}}{2^5} = \frac{10 + 5 + 1}{32} = \frac{1}{2}.$$

**Question 2.** Total balls $= 12$, so

$$P(\text{both red}) = \frac{5}{12} \cdot \frac{4}{11} = \frac{5}{33}.$$

**Question 3.** If $X \sim B(7, 0.3)$,

$$P(X = 0) = 0.7^7 = 0.08235,$$

$$P(X = 1) = \binom{7}{1}(0.3)(0.7)^6 = 7(0.3)(0.117649) = 0.24706,$$

so

$$P(X \geq 2) = 1 - 0.08235 - 0.24706 = 0.67059.$$

**Question 4.** If $X \sim \text{Poisson}(2.5)$,

$$P(X = 0) = e^{-2.5} = 0.08208,$$

$$P(X \leq 2) = P(0) + P(1) + P(2) = e^{-2.5}\left(1 + 2.5 + \frac{2.5^2}{2}\right) = 0.5438.$$

**Question 5.** For $X \sim N(40, 5^2)$ and $x = 47$,

$$z = \frac{47 - 40}{5} = 1.4.$$

Chapter B

# Additional Drill Problems with Short Answers

1. Find the class mark of 30–39. **Answer:** 34.5
2. If a die is thrown once, what is the probability of an odd number? **Answer:** 1/2
3. For $X \sim B(4, 0.2)$, find $E(X)$. **Answer:** 0.8
4. For a Poisson variable with mean 6, what is the variance? **Answer:** 6
5. If mean = 24 and standard deviation = 3, what is the coefficient of variation? **Answer:** 12.5%
6. If $r = -0.88$, describe the correlation. **Answer:** Strong negative linear correlation
7. What graph is used to display cumulative frequencies? **Answer:** Ogive
8. The median class of a grouped distribution with $N = 40$ contains which cumulative position? **Answer:** the 20th value
9. Compute $P(X = 0)$ for $X \sim \text{Poisson}(1)$. **Answer:** $e^{-1} \approx 0.3679$
10. For scores with mean 60 and standard deviation 8, what is the $z$-score of 68? **Answer:** 1

# Concluding Note

The best way to learn statistics is to compute repeatedly, draw the graphs by hand at least once, and always interpret the answer in words. A student who merely memorizes formulas without understanding the meaning of the results will struggle in applications. These notes are therefore designed to be read slowly, worked through line by line and used together with active problem solving.