Allama Iqbal Open University AIOU BA/BCom/BBA Solved Assignment NO 1 Autumn 2025

Code 1430 Business Statistics

Q.1 (a) Discuss the importance of studying statistics. Provide at least three distinct reasons why statistical knowledge is valuable in various fields.

Introduction to Statistics

Statistics is a branch of mathematics concerned with the collection, organization, analysis, interpretation, and presentation of data. In the modern world, information is considered as valuable as material resources, and statistics provides the techniques to understand, summarize, and interpret this information. The study of statistics is not limited to scientists or researchers; it has become a necessity for students, professionals, policymakers, and even ordinary people in everyday life. Without statistics, the large amount of data produced in every sector would remain meaningless and unusable. It is through statistics that data is converted into knowledge, predictions, and strategies.

1. Informed Decision-Making

One of the most significant reasons for studying statistics is its role in decision-making. Every decision, whether personal or organizational, benefits from accurate data analysis. In business, for example, managers analyze sales statistics, customer feedback, and market trends to decide whether to expand operations, launch new products, or withdraw existing ones. If decisions are taken without data, there is a greater chance of failure. In the context of Pakistan, government departments rely on statistical surveys such as the census, labor force surveys, and household expenditure surveys to make policies related to health, education, poverty alleviation, and infrastructure development. For instance, when education enrollment statistics show a decline in rural areas, the government may allocate more resources to build schools and hire teachers in those areas.

2. Prediction and Forecasting

Another core reason to study statistics is its role in predicting the future by analyzing patterns in past and present data. Forecasting is especially important in economics, agriculture, and weather prediction. For example, Pakistan’s agriculture sector is highly dependent on weather conditions. Meteorological departments use statistical models to forecast rainfall, temperature changes, and potential floods. These forecasts allow farmers to plan irrigation schedules, sow crops at the right time, and protect their harvest from extreme weather conditions. Similarly, businesses use statistical forecasting to estimate future sales, profits, and production needs. In the stock market, investors analyze statistical trends to decide when to buy or sell shares. In healthcare, statistical forecasting helps predict the spread of diseases such as dengue or COVID-19, enabling timely precautionary measures.

3. Research and Scientific Development

Statistics plays a fundamental role in scientific research across disciplines. Researchers use statistics to collect data, test hypotheses, measure reliability, and validate their findings. In medical research, for instance, statistical methods are applied to test the effectiveness of new drugs or vaccines. Before a vaccine is approved for use, researchers conduct experiments on different groups of people, gather data, and apply statistical analysis to determine whether the results are significant or occurred by chance. In the field of education, researchers use statistics to compare the effectiveness of different teaching methods, evaluate curriculum outcomes, and understand learning patterns among students. Without statistics, research would lack accuracy and credibility, as it would be based only on assumptions.

4. Measuring Risk and Uncertainty

In many areas of life, people and organizations face uncertainty. Statistics provides methods to measure and reduce this uncertainty. For example, insurance companies in Pakistan use statistical models to calculate premiums by analyzing data such as age, medical history, and accident records. Similarly, banks use statistics to assess the risk of giving loans to individuals or businesses. Governments also use statistical risk analysis to plan for economic instability, inflation control, and natural disasters. For instance, during the COVID-19 pandemic, statistical modeling helped predict the spread of the virus, the number of hospital beds needed, and the required supply of vaccines. Thus, statistics is a powerful tool for minimizing risks and making safe, rational choices.

5. Role in Economic and Business Growth

In the economic field, statistics is essential for understanding trends in production, consumption, trade, and investment. National income, GDP, inflation rates, unemployment levels, and balance of trade are all statistical measures that show the health of an economy. In Pakistan, the Planning Commission and the State Bank rely heavily on statistical indicators to design fiscal and monetary policies. Businesses also use statistics to identify customer preferences, analyze competitor strategies, and improve their products. For example, a Pakistani textile company might analyze export statistics to decide which international market has the highest demand for its products.

6. Educational Applications of Statistics

In education, statistics helps measure student performance, teacher effectiveness, and overall system efficiency. Test scores, attendance records, and surveys are all analyzed using statistical methods to improve the quality of teaching and learning. For example, if the statistics show that a large number of students are underperforming in mathematics, the education department may design special training programs for teachers or introduce remedial classes for students. Statistics also helps in designing fair grading systems, conducting research on learning methods, and evaluating new curricula.

7. Everyday Importance of Statistics

Statistics is not only important for large institutions but also for daily life. Individuals use statistical concepts when comparing product prices, analyzing household budgets, or evaluating investment opportunities. For example, a person comparing different mobile packages relies on statistical data such as call rates, internet data limits, and monthly costs. Cricket fans in Pakistan use statistics to compare the performance of players such as batting averages, strike rates, and bowling economy. Similarly, families use price index data to understand how inflation affects their purchasing power. These examples show that statistical thinking is integrated into everyday decision-making.

8. Role in Social and Political Development

Statistics is vital for the development of social policies and political planning. Social surveys collect data about literacy rates, population growth, poverty levels, and employment opportunities. Political parties and governments use this data to understand public needs, evaluate the success of their programs, and design new initiatives. For example, population statistics guide the government in conducting constituency delimitation for elections, ensuring fair representation. Similarly, health statistics help policymakers allocate resources to areas with high disease prevalence.

9. Contribution to Technology and Innovation

In the era of technology, statistics is closely linked to innovation. Data science, artificial intelligence, and machine learning are built on statistical models. For example, in Pakistan’s e-commerce sector, companies like Daraz use statistical algorithms to recommend products to customers based on their past buying behavior. Similarly, YouTube and social media platforms use statistics to track user activity and suggest relevant videos. In this way, statistics contributes to both innovation and customer satisfaction.

Conclusion

The study of statistics is essential in the modern world because it provides tools for informed decision-making, forecasting, research, risk assessment, business growth, education, social development, and technological innovation. Its value is visible in every sector of life, from government planning to daily household management. By learning statistics, individuals and organizations are better equipped to understand data, reduce uncertainty, and achieve progress in an increasingly complex and data-driven world.

Q.1 (b) Define and briefly explain the following fundamental statistical terms: Population, Sample, Variable, Observation

Population

In statistics, the term *population* refers to the complete set of individuals, objects, or data that share a common characteristic of interest to the researcher. It includes every possible observation related to the study. For example, if a researcher wants to study the literacy rate in Pakistan, the population would include all people living in Pakistan. Populations can be finite (such as the number of students in a specific university) or infinite (such as the number of possible outcomes when tossing a coin).

Understanding the population is crucial because statistical studies aim to describe and make inferences about it.

Sample

A *sample* is a smaller group selected from the population to represent it. Since studying an entire population is often difficult, time-consuming, and costly, researchers use samples to make estimates or predictions about the whole population. A sample must be chosen carefully to ensure it reflects the characteristics of the population. For example, if a researcher wants to know the average income of families in Lahore, it is impossible to ask every family. Instead, a sample of families is selected, and their average income is calculated to represent the whole. A well-selected sample gives accurate and reliable results.

Variable

A *variable* is any characteristic, number, or quantity that can change or take on different values among individuals in a population or sample. Variables are classified mainly into two types:

- **Quantitative variables** (numerical), such as age, height, weight, or income, which can be measured in numbers.

- **Qualitative variables** (categorical), such as gender, nationality, or religion, which describe categories or groups.

  For example, in a survey of students, age is a quantitative variable, while gender is a qualitative variable. Variables are important because they are the basis of statistical analysis.

Observation

An *observation* is the value or measurement obtained for a particular variable from an individual unit of the population or sample. In simple terms, it is the data collected from one unit. For example, if the variable is age and we collect data from 5 students (20, 21, 22, 19, 23), then each number is an observation. Observations form the raw data that is analyzed in statistics. Without observations, statistical summaries such as averages or percentages cannot be calculated.

Q.2 Transmission ABC stores recorded the number of service tickets submitted by 50 stores last month as follows:

823 648 321 634 752 669 427 555 904 586 722 360 468 847 641 217 588 349 308 766 114 163 150 718 687 763 607 441 305 662 227 624 791 960 334 163 550 842 860 413 439 981 416 115 810 957 919 846

(a) Arrange the data using the Data Array and the Frequency Distribution

Step 1: Data Array (Ascending Order)

Arranging the given 50 values from the smallest to the largest:

114, 115, 150, 163, 163, 217, 227, 305, 308, 321,

334, 349, 360, 413, 416, 427, 439, 441, 468, 550,

555, 586, 588, 607, 624, 634, 641, 648, 662, 669, 687, 718, 722, 752, 763, 766, 791, 810, 823, 842,

846, 847, 860, 904, 919, 957, 960, 981

Step 2: Frequency Distribution Table

We will round this to **100** for simplicity.

Frequency Distribution Table (Class Interval = 100):

Class Interval	**Frequency**
100 – 199	3

200 – 299	2
300 – 399	5
400 – 499	4
500 – 599	4
600 – 699	6

700 – 799	5
800 – 899	5
900 – 999	4
Total	**38**

(b) Which arrangement do you prefer? Explain

- The **Data Array** is useful for seeing exact values in ascending order. It helps identify the minimum, maximum, median, and can be directly used for further descriptive statistics like quartiles or percentiles. However, it is difficult to quickly observe patterns or trends when the dataset is large.

- The **Frequency Distribution** groups data into intervals and shows how many values fall into each range. This makes it much easier to see the distribution of service tickets, identify clusters, detect where most stores fall (e.g., between 600–700), and compare different ranges at a glance.

Therefore, Frequency Distribution is usually preferred because it provides a clearer picture of the overall pattern and makes statistical analysis (like drawing histograms or calculating mode classes) easier, especially when the dataset is large.

Q.3(a) Describe different measures of central tendency

Introduction

Measures of central tendency are methods used in statistics to find the central or average value of a dataset. They help summarize large amounts of information into a single value. The most common measures are Mean, Median, and Mode, while Geometric Mean and Harmonic Mean are also used in specific cases.

---

1. Mean (Arithmetic Mean)

- Definition: The mean is the average of all values.

- Method: Add all the values and divide by the number of values.
- Example: The marks of five students are 10, 15, 20, 25, and 30. The mean is (10 + 15 + 20 + 25 + 30) ÷ 5 = 20.
- Use: Useful for continuous data and when all values are equally important.
- Limitation: Affected by extreme values such as very high or very low numbers.

---

2. Median

- Definition: The median is the middle value when all numbers are arranged in order.
- Method: If the number of values is odd, the middle value is the median. If it is even, take the average of the two middle values.
- Example: For 5, 8, 12, 20, 25 → Median = 12. For 5, 8, 12, 20, 25, 30 → Median = (12 + 20) ÷ 2 = 16.
- Use: Best when data is skewed or has outliers.
- Limitation: Ignores the influence of other data points except the middle ones.

3. Mode

- Definition: The mode is the value that occurs most often.
- Example: In 4, 6, 8, 8, 10, 12 → Mode = 8.
- Use: Best for categorical data (e.g., favorite color, most purchased product).
- Limitation: Data may have no mode or more than one mode.

4. Geometric Mean

- Definition: The geometric mean is used for values related to percentages, ratios, or growth rates.

- Example: If growth rates are 10%, 20%, and 30%, the geometric mean gives the average rate of growth over time.

- Use: Common in business, economics, and finance.

- Limitation: Cannot be used with negative or zero values.

---

5. Harmonic Mean

- Definition: The harmonic mean is used for values given in terms of rates like speed or efficiency.

- Example: A car travels the same distance at 30 km/h and 60 km/h. The harmonic mean speed = 40 km/h.

- Use: Helpful in time, distance, and rate problems.

- Limitation: Sensitive to very small values.

Conclusion

- Mean is the most common but affected by extreme values.

- Median is best for skewed data.
- Mode is useful for categorical data.
- Geometric Mean works well for percentages and growth rates.
- Harmonic Mean is effective for averages involving speed, time, or rates.

Each measure has its own importance depending on the type of data and situation.

Q.3(b) Calculate the Mean and Median number of persons per house from the following data.

No. of persons per house (x) and No. of houses (f):

1 = 26

2 = 113

3 = 120

4 = 95

5 = 60

6 = 42

7 = 21

8 = 14

9 = 5

10 = 4

Step 1: Mean

Formula: Mean = Σ(f × x) ÷ Σf

Now multiply each x with f and add:

(1×26) + (2×113) + (3×120) + (4×95) + (5×60) + (6×42) + (7×21) + (8×14) + (9×5) + (10×4)

= 26 + 226 + 360 + 380 + 300 + 252 + 147 + 112 + 45 + 40

= 1888

Σf = total houses = 26 + 113 + 120 + 95 + 60 + 42 + 21 + 14 + 5 + 4 = 600

So, Mean = 1888 ÷ 600 = 3.15 persons (approx).

Step 2: Median

Formula: Median = value of (N/2)th item, where N = Σf.

Here N = 600.

N/2 = 600 ÷ 2 = 300.

So, the median is the 300th item in the ordered data.

Now find the cumulative totals:

For 1 person = 26

For 2 persons = 26 + 113 = 139

For 3 persons = 139 + 120 = 259

For 4 persons = 259 + 95 = 354

The 300th item lies between 259 and 354.

So, the Median = 4 persons.

Final Answer:

Mean number of persons per house = 3.15

Median number of persons per house = 4

Q.4 Calculate the standard deviation and coefficient of variation for the following distribution of lengths of 200 metal bars.

X	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
f	4	8	2	3	6	4	1	4	1	1
			3	5	2	4	8			

Step 1: Find Mean

N = Σf = 200

Σ(f·x) = 6780

Mean ($\bar{x}$) = Σ(f·x) / N = 6780 / 200 = 33.9

Step 2: Calculate $(x - \bar{x})$, $(x - \bar{x})^2$ and $f \cdot (x - \bar{x})^2$

$x = 30$, $f = 4 \rightarrow (x - \bar{x}) = -3.9 \rightarrow (x - \bar{x})^2 = 15.21 \rightarrow f \cdot (x - \bar{x})^2 = 60.84$

$x = 31$, $f = 8 \rightarrow (x - \bar{x}) = -2.9 \rightarrow (x - \bar{x})^2 = 8.41 \rightarrow f \cdot (x - \bar{x})^2 = 67.28$

$x = 32$, $f = 23 \rightarrow (x - \bar{x}) = -1.9 \rightarrow (x - \bar{x})^2 = 3.61 \rightarrow f \cdot (x - \bar{x})^2 = 83.03$

$x = 33$, $f = 35 \rightarrow (x - \bar{x}) = -0.9 \rightarrow (x - \bar{x})^2 = 0.81 \rightarrow f \cdot (x - \bar{x})^2 = 28.35$

$x = 34$, $f = 62 \rightarrow (x - \bar{x}) = 0.1 \rightarrow (x - \bar{x})^2 = 0.01 \rightarrow f \cdot (x - \bar{x})^2 = 0.62$

$x = 35$, $f = 44 \rightarrow (x - \bar{x}) = 1.1 \rightarrow (x - \bar{x})^2 = 1.21 \rightarrow f \cdot (x - \bar{x})^2 = 53.24$

$x = 36$, $f = 18 \rightarrow (x - \bar{x}) = 2.1 \rightarrow (x - \bar{x})^2 = 4.41 \rightarrow f \cdot (x -$

$\bar{x})^2 = 79.38$

$x = 37, f = 4 \rightarrow (x - \bar{x}) = 3.1 \rightarrow (x - \bar{x})^2 = 9.61 \rightarrow f \cdot (x - \bar{x})^2 = 38.44$

$x = 38, f = 1 \rightarrow (x - \bar{x}) = 4.1 \rightarrow (x - \bar{x})^2 = 16.81 \rightarrow f \cdot (x - \bar{x})^2 = 16.81$

$x = 39, f = 1 \rightarrow (x - \bar{x}) = 5.1 \rightarrow (x - \bar{x})^2 = 26.01 \rightarrow f \cdot (x - \bar{x})^2 = 26.01$

$\Sigma[f \cdot (x - \bar{x})^2] = 453.00$

Step 3: Variance and Standard Deviation

Variance = $\Sigma[f \cdot (x - \bar{x})^2] / N = 453.00 / 200 = 2.27$

Standard Deviation (σ) = $\sqrt{2.27} = 1.51$

Step 4: Coefficient of Variation

CV = (σ / Mean) × 100 = (1.51 / 33.9) × 100 = 4.44%

Final Answer:

Mean = 33.9

Variance = 2.27

Standard Deviation = 1.51

Coefficient of Variation = 4.44%

Q.5 (a) Discuss the basic concepts in the hypothesis-testing procedure.

Hypothesis testing is one of the most important concepts in statistics because it provides a systematic way to make decisions about data. It helps researchers, analysts, and decision-makers determine whether the evidence in a sample of data is strong enough to draw conclusions about a population. To understand hypothesis testing, let us break down its key concepts step by step.

---

1. Hypothesis

A hypothesis is a statement or assumption that we want to test using data. It is about the population parameter (like mean, proportion, or variance). There are always two hypotheses:

- **Null Hypothesis (H_0):** This is the assumption that there is no effect, no difference, or no relationship. It is considered the default statement. For example, H_0: The average exam score of students = 50.

- **Alternative Hypothesis (H_1 or Ha):** This is the statement that contradicts the null hypothesis. It suggests that there is an effect, a difference, or a relationship. For example, H_1: The average exam score ≠ 50.

The whole testing procedure is about deciding whether to reject H_0 in favor of H_1, or to not reject H_0.

---

2. Level of Significance (α)

The level of significance (denoted by α) is the probability of rejecting the null hypothesis when it is actually true. It shows how much risk the researcher is willing to take. Commonly used values are 0.05 (5%) or 0.01 (1%).

- If α = 0.05, it means there is a 5% chance of making a wrong decision by rejecting H_0 when it is true.

---

3. Test Statistic

A test statistic is a numerical value calculated from sample data. It measures how far the observed data are from what we would expect under H_0. The choice of test statistic depends on the type of data and test being used. For example:

- For testing a mean (when variance is known): Z-test
- For testing a mean (when variance is unknown): t-test
- For categorical data: Chi-square test
- For comparing variances: F-test

This statistic is then compared with a critical value or used to calculate a p-value.

---

4. Critical Region and Decision Rule

The critical region is the area of values of the test statistic where we would reject H_0. It is based on the significance level (α).

- If the test statistic falls in the critical region, reject H_0.
- If the test statistic does not fall in the critical region, do not reject H_0.

For example, in a two-tailed Z-test with $\alpha = 0.05$, the critical region is when $|Z| > 1.96$.

---

5. P-value

The p-value is the probability of obtaining results as extreme as (or more extreme than) the observed data, assuming H_0 is true.

- If p-value $\leq \alpha \rightarrow$ Reject H_0

- If p-value > α → Do not reject H_0

The p-value approach is widely used because it directly shows how strong the evidence is against H_0.

---

6. Errors in Hypothesis Testing

In hypothesis testing, there is always a chance of making errors. These are:

- **Type I Error:** Rejecting H_0 when it is true (false positive). Probability of this error = α.

- **Type II Error:** Failing to reject H_0 when it is false (false negative). Probability of this error = β.

Good hypothesis testing tries to minimize both errors.

---

7. Power of a Test

The power of a test is the probability of correctly rejecting H_0 when it is false. Power = 1 – β. A higher power means the test is more effective in detecting true differences. Researchers usually prefer power of at least 80%.

---

8. One-tailed vs. Two-tailed Tests

- **One-tailed test:** Tests for a difference in one direction only (e.g., H_1: mean > 50).

- **Two-tailed test:** Tests for a difference in both directions (e.g., H_1: mean $\neq$ 50).

The choice depends on the research question.

---

9. Steps in Hypothesis Testing

The general procedure can be summarized as:

1. State the null hypothesis (H_0) and alternative hypothesis (H_1).

2. Select the level of significance (α).

3. Choose the appropriate test and compute the test statistic.

4. Determine the critical region or calculate the p-value.

5. Make a decision: Reject or fail to reject H_0.

6. Interpret the result in the context of the problem.

In short:

Hypothesis testing is a structured decision-making process that uses sample data to make conclusions about population parameters. Its core ideas include forming hypotheses, setting a significance level, calculating a test statistic, comparing with critical values or p-values, and then deciding whether the evidence is strong enough to reject the null hypothesis.

Question: 5

The average commission charged by full-service brokerage firms on a sale of common stock is $144, and the standard deviation is $52. Joel Freelander has taken a random sample of 121 trades by his clients and determined that they paid an average commission of $151. At a 0.10 significance level, can Joel conclude that his clients' commissions are higher than the industry average?

Answer:

Step 1 – State the Hypotheses

H_0: $\mu = 144$

H_1: $\mu > 144$ (one-tailed, right test)

Step 2 – Calculate the Test Statistic

Population mean (μ_0) = 144

Sample mean ($\bar{x}$) = 151

Standard deviation (σ) = 52

Sample size (n) = 121

Standard error (SE) = $\sigma / \sqrt{n} = 52 / \sqrt{121} = 52 / 11 = 4.727$

$z = (\bar{x} - \mu_0) / SE = (151 - 144) / 4.727 = 7 / 4.727 = 1.481$

Step 3 – Find Critical Value and p-value

At $\alpha = 0.10$ (one-tailed), critical $z = 1.2816$

Observed $z = 1.481 > 1.2816$ → falls in rejection region

p-value = $P(Z \geq 1.481) = 1 - \Phi(1.481) \approx 1 - 0.9307 = 0.0693$

Step 4 – Decision

Since p-value (0.0693) < α (0.10), reject H0.

Conclusion

At the 0.10 significance level, there is enough evidence to conclude that Joel’s clients pay higher commissions than the industry average of $144.