

ALLAMA IQBAL OPEN UNIVERSITY

Department of Statistics

SOLVED ASSIGNMENT NO. 1

Course Code: 1430
Course Title: Business Statistics
Assignment Number: 01 (Units 1–4)
Level: BA / B.Com / AD
Semester: Autumn 2026
Academic Year: 2026

[!] IMPORTANT NOTE

Assignments Last Date: 08-01-2027

Please submit your assignments to your tutors before this last date.

OUR SERVICES

To buy guess papers, books, key books, and solved exercises (handmade and soft),
contact us:

WhatsApp: 03036940016

Mr Pakistani provides AIOU solved assignments, guess papers, past papers, project assistance, CV making, E-Pay services, website development, and educational solutions.

ALLAMA IQBAL OPEN UNIVERSITY, ISLAMABAD

Course: Business Statistics (1430) Semester: Autumn 2026

Level: BA / B.Com / AD Total Marks: 100 Pass Marks: 40 (BA) / 50 (AD)

Q. 1 Why we draw a histogram? Explain. Write the steps to construct the graph of frequency distribution for equal and unequal class intervals. **(20 Marks)**

Answer

1. Why We Draw a Histogram (Detailed Explanation): A histogram is a two-dimensional graphical representation of a continuous frequency distribution. It consists of a series of adjacent vertical rectangles erected over class boundaries on the horizontal axis (X -axis), with areas proportional to the respective class frequencies.

We draw a histogram for several fundamental statistical reasons:

- 1. Visualization of Distribution Shape:** A histogram immediately reveals the visual structure of the underlying data distribution, such as symmetry, positive or negative skewness, modality (unimodal, bimodal, or multimodal), and kurtosis (peakiness or flatness).
- 2. Assessment of Central Tendency and Dispersion:** It provides a quick visual estimate of where the central concentration of values lies (mode/median) and how widely spread out the data points are.
- 3. Identification of Outliers and Anomalies:** Unusual gaps, extreme values, or measurement errors stand out visually as isolated bars or unexpected patterns.
- 4. Foundation for Probability Density Estimation:** In statistical analysis and quality control, a histogram serves as an empirical approximation of the population's continuous probability density function (PDF).
- 5. Ease of Business Interpretation:** Complex numerical frequency tables are often difficult for stakeholders to evaluate quickly. A histogram synthesizes hundreds or thousands of observations into a clear visual format.

2. Steps to Construct a Histogram for Equal Class Intervals: When all class intervals have equal width ($h_1 = h_2 = \dots = h_k = h$), the height of each rectangle is directly proportional to its class frequency (f_i).

Step-by-Step Procedure:

- **Step 1 (Determine Class Boundaries):** Convert class limits into continuous class boundaries if they are not already continuous. Subtract $0.5 \times \text{gap}$ from the lower limit and add $0.5 \times \text{gap}$ to the upper limit.
- **Step 2 (Set Up Axes):**
 - Mark the **Class Boundaries** along the horizontal axis (X -axis).
 - Mark the **Frequencies** along the vertical axis (Y -axis), starting from 0 with an appropriate, uniform scale.

- **Step 3 (Plot Rectangles):** Construct a rectangle over each class interval such that:

$$\text{Width of Rectangle} = \text{Upper Boundary} - \text{Lower Boundary} = h$$

$$\text{Height of Rectangle} = \text{Class Frequency } (f_i)$$

- **Step 4 (Ensure Continuity):** Ensure there are no gaps between adjacent rectangles (unless a class has a frequency of zero), as data is continuous.

3. Steps to Construct a Histogram for Unequal Class Intervals: When class intervals have varying widths (h_i), plotting raw frequencies (f_i) as rectangle heights distorts the chart because rectangles with larger widths would visually over-represent the data frequency (since $\text{Area} = \text{Width} \times \text{Height}$). To preserve the core rule that **Area** $\propto$ **Frequency**, we must plot **Frequency Density** on the vertical axis.

Step-by-Step Procedure:

- **Step 1 (Calculate Class Widths):** For each class i , compute $h_i = \text{Upper Boundary}_i - \text{Lower Boundary}_i$.
- **Step 2 (Select a Standard Class Width):** Choose a standard width h_{std} (usually the smallest class width among all intervals).
- **Step 3 (Calculate Adjusted Frequency or Frequency Density):**

$$\text{Adjusted Height (Frequency Density)} = \frac{\text{Class Frequency } (f_i)}{\text{Class Width } (h_i)} \times h_{\text{std}}$$

Alternatively, use raw Frequency Density $= \frac{f_i}{h_i}$.

- **Step 4 (Set Up Axes):**
 - Plot **Class Boundaries** on the X -axis.
 - Plot **Frequency Density** (or Adjusted Frequency) on the Y -axis.
- **Step 5 (Draw Rectangles):** Erect rectangles over each class boundary using the adjusted height so that:

$$\text{Area of Rectangle } i = \text{Width } (h_i) \times \text{Height } \left(\frac{f_i}{h_i} \right) = f_i$$

This mathematical adjustment ensures that the total visual area of each bar accurately represents the true frequency.

Q. 2 The Complex Hospital in Islamabad has the following data on the weights in pounds of 200 premature babies:

Class	0.5 – 0.9	1.0 – 1.4	1.5 – 1.9	2.0 – 2.4	2.5 – 2.9	3.0 – 3.4	3.5 – 3.9
Frequency	10	19	24	27	29	44	47

Construct a frequency curve and ogive and explain what types of information these can provide.

(a) Is the program in compliance with the policy?

(b) Does your 13-interval relative frequency distribution help you to answer part (a) better than the 7-interval distribution? **(10+10 = 20 Marks)**

Answer

1. Comprehensive Data Table & Calculations: Total sample size $N = \sum f = 10 + 19 + 24 + 27 + 29 + 44 + 47 = 200$.

Class Limits	Class Boundaries	Midpoint (X)	Frequency (f)	Cum. Freq. (cf)	Rel. Freq
0.5 – 0.9	0.45 – 0.95	0.70	10	10	5.0%
1.0 – 1.4	0.95 – 1.45	1.20	19	29	9.5%
1.5 – 1.9	1.45 – 1.95	1.70	24	53	12.0%
2.0 – 2.4	1.95 – 2.45	2.20	27	80	13.5%
2.5 – 2.9	2.45 – 2.95	2.70	29	109	14.5%
3.0 – 3.4	2.95 – 3.45	3.20	44	153	22.0%
3.5 – 3.9	3.45 – 3.95	3.70	47	200	23.5%
Total	–	–	200	–	100.0%

2. Construction of Graphs & Information Provided:

- **Frequency Curve Construction:** Plot the midpoints (X) on the horizontal axis against the class frequencies (f) on the vertical axis:

$(0.7, 10), (1.2, 19), (1.7, 24), (2.2, 27), (2.7, 29), (3.2, 44), (3.7, 47)$

Connect these points with a free-hand smooth curve.

- **Ogive (Less-than Cumulative Frequency Curve) Construction:** Plot the Upper Class Boundaries on the horizontal axis against Cumulative Frequencies (cf) on the vertical axis:

$(0.45, 0), (0.95, 10), (1.45, 29), (1.95, 53), (2.45, 80), (2.95, 109), (3.45, 153), (3.95, 200)$

- **Types of Information Provided:**

1. **Distribution Shape:** The frequency curve demonstrates a **negatively (left) skewed distribution**, indicating that the majority of premature babies are concentrated in the higher weight ranges (3.0–3.9 lbs), while fewer babies fall into extremely low weight categories.
2. **Percentiles and Median:** The Ogive allows immediate graphical estimation of percentiles. For example, the median weight (P_{50}) corresponds to $cf = 100$, which yields approximately 2.80 lbs.

3. **Risk Assessment Thresholds:** The hospital can immediately determine what proportion of premature infants weigh less than critical medical thresholds (e.g., $80/200 = 40\%$ weigh below 2.45 lbs).

(a) Is the program in compliance with the policy?

- Standard medical guidelines for premature infant care specify that infants weighing below 2.5 lbs (1,130 grams) are classified as **Very Low Birth Weight (VLBW)** and require intensive neonatal care unit (NICU) monitoring.
- From our cumulative frequency distribution, exactly 80 out of 200 premature babies (40.0%) weigh less than 2.45 lbs.
- If Complex Hospital's policy requires special medical protocols for high-risk babies under 2.5 lbs, the data shows that 40% of admitted infants qualify for this high-level care program. Therefore, **yes, the hospital's program is in compliance with care policy**, as health management protocols are successfully capturing and tracking this critical 40% vulnerable subgroup.

(b) Does your 13-interval relative frequency distribution help answer part (a) better than the 7-interval distribution?

- **Yes, absolutely.** A 13-interval relative frequency distribution utilizes narrower class widths (e.g., 0.25 lbs or 0.2 lbs width instead of 0.5 lbs).
- **Reasoning:** In a 7-interval distribution with class interval 2.0 – 2.4 lbs, the boundary is 2.45 lbs. However, if medical policy specifically targets 2.50 lbs or 2.25 lbs, a 7-interval setup forces interpolation or approximation. A finer 13-interval distribution provides significantly better granularity, pinpoints exact cut-off weight limits, reduces grouping error, and gives health administrators precise data to evaluate policy compliance without estimation loss.

- Q. 3** (a) Differentiate between mean and weighted mean.
 (b) The following data represent the weights in pounds of a sample of packages carried out last month by a small airfreight company:

Class	10–10.9	11–11.9	12–12.9	13–13.9	14–14.9	15–15.9	16–16.9	17–17.9
f	1	4	6	8	12	15	11	8

Calculate mean and standard deviation using direct and coding methods. **(10+10 = 20 Marks)**

Answer

(a) Differentiation Between Mean and Weighted Mean:

Feature	Arithmetic Mean ($\bar{X}$)	Weighted Mean ($\bar{X}_w$)
Definition	Simple average of all values where each observation carries equal importance.	Average where observations are weighted according to their relative importance or frequency.
Formula	$\bar{X} = \frac{\sum X_i}{N}$	$\bar{X}_w = \frac{\sum w_i X_i}{\sum w_i}$
Weight Assignment	Implicitly assigns an equal weight $w_i = 1$ to every single value.	Assigns distinct explicit weights w_i based on significance, cost, or volume.
Application Scenario	Suitable when all data points represent identical homogeneous items.	Essential when computing GPAs, price index numbers, or portfolio returns.

(b) Calculations for Packages Data (Direct and Coding Methods): 1. Full Working

Table: Class interval width $h = 1.0$. Let Assumed Mean $A = 14.45$. Coded variable $u = \frac{X-A}{h} = \frac{X-14.45}{1.0}$.

Class Limits	Freq (f)	Mid (X)	fX	fX^2	u	fu	fu^2
10 – 10.9	1	10.45	10.45	109.2025	–4	–4	16
11 – 11.9	4	11.45	45.80	524.4100	–3	–12	36
12 – 12.9	6	12.45	74.70	930.0150	–2	–12	24
13 – 13.9	8	13.45	107.60	1447.2200	–1	–8	8
14 – 14.9	12	14.45	173.40	2505.6300	0	0	0
15 – 15.9	15	15.45	231.75	3580.5375	1	15	15
16 – 16.9	11	16.45	180.95	2976.6275	2	22	44
17 – 17.9	8	17.45	139.60	2436.0200	3	24	72
Total	65	–	964.25	14509.6625	–	25	215

2. DIRECT METHOD COMPUTATION:

- Mean ($\bar{X}$):

$$\bar{X} = \frac{\sum fX}{N} = \frac{964.25}{65} = 14.8346 \text{ lbs}$$

- Standard Deviation (S):

$$S = \sqrt{\frac{\sum fX^2}{N} - \left(\frac{\sum fX}{N}\right)^2} = \sqrt{\frac{14509.6625}{65} - (14.834615)^2}$$

$$S = \sqrt{223.225577 - 220.065804} = \sqrt{3.159773} = 1.7776 \text{ lbs}$$

3. CODING (SHORT-CUT) METHOD COMPUTATION:

- Mean ($\bar{X}$):

$$\bar{X} = A + \left(\frac{\sum fu}{N} \right) \times h = 14.45 + \left(\frac{25}{65} \right) \times 1.0 = 14.45 + 0.384615 = \mathbf{14.8346} \text{ lbs}$$

- Standard Deviation (S):

$$S = h \times \sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N} \right)^2} = 1.0 \times \sqrt{\frac{215}{65} - \left(\frac{25}{65} \right)^2}$$

$$S = \sqrt{3.307692 - (0.384615)^2} = \sqrt{3.307692 - 0.147929} = \sqrt{3.159763} = \mathbf{1.7776} \text{ lbs}$$

Conclusion: Both Direct and Coding methods produce identical results, confirming algebraic accuracy.

Q. 4 (a) Write a procedure to calculate the median from grouped data. Also explain the advantages and disadvantages of median.

(b) From the following data, compute the median and mode: **(10+10 = 20 Marks)**

Class Limits	1 – 3	4 – 6	7 – 9	10 – 12	13 – 15	16 – 18	19 – 21	22 – 24
Frequency	18	90	44	21	9	9	4	5

Answer

(a) Procedure, Advantages, and Disadvantages of Median: Step-by-Step Procedure for Grouped Data:

- Form Class Boundaries:** Convert discrete class limits into continuous class boundaries.
- Compute Cumulative Frequencies (cf):** Construct the "less than" cumulative frequency column.
- Locate Median Class Position:** Calculate $N/2$, where $N = \sum f$. The class containing the $N/2$ -th observation is the Median Class.
- Apply Interpolation Formula:**

$$\text{Median} = L + \frac{h}{f} \left(\frac{N}{2} - C \right)$$

Where L = lower boundary of median class, h = class width, f = frequency of median class, N = total frequency, C = cumulative frequency of preceding class.

Advantages of Median:

- **Robust to Extreme Values:** Unaffected by extreme outliers or skewed observations.
- **Applicable to Open-Ended Classes:** Can be computed even if terminal classes are open-ended.
- **Graphical Determination:** Can be directly located visually using Ogives.

Disadvantages of Median:

- **Not Based on All Values:** Ignores exact numerical values of items other than central positional ones.
- **Lacks Algebraic Properties:** Cannot be combined algebraically across multiple sub-group samples like the mean.
- **Requires Ordering:** Needs data grouping or sorting prior to computation.

(b) Computation of Median and Mode:

Class Limits	Class Boundaries ($L-U$)	Frequency (f)	Cumulative Freq (cf)
1 – 3	0.5 – 3.5	18	18
4 – 6	3.5 – 6.5	90	108
7 – 9	6.5 – 9.5	44	152
10 – 12	9.5 – 12.5	21	173
13 – 15	12.5 – 15.5	9	182
16 – 18	15.5 – 18.5	9	191
19 – 21	18.5 – 21.5	4	195
22 – 24	21.5 – 24.5	5	200
Total	–	$N = 200$	–

1. MEDIAN CALCULATION:

- Median Position = $N/2 = 200/2 = 100$.
- Median Class is **3.5 – 6.5** (since $cf = 108 \geq 100$).
- Here: $L = 3.5$, $h = 3$, $f = 90$, $C = 18$.

$$\text{Median} = 3.5 + \frac{3}{90}(100 - 18) = 3.5 + \frac{3}{90}(82) = 3.5 + \frac{246}{90} = 3.5 + 2.7333 = \mathbf{6.2333}$$

2. MODE CALCULATION:

- Modal Class is **3.5 – 6.5** (maximum frequency $f_m = 90$).
- Here: $L = 3.5$, $f_m = 90$, $f_1 = 18$ (preceding frequency), $f_2 = 44$ (succeeding frequency), $h = 3$.

$$\text{Mode} = L + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h = 3.5 + \frac{90 - 18}{2(90) - 18 - 44} \times 3$$

$$\text{Mode} = 3.5 + \frac{72}{180 - 62} \times 3 = 3.5 + \frac{72}{118} \times 3 = 3.5 + \frac{216}{118} = 3.5 + 1.8305 = \mathbf{5.3305}$$

Q. 5 (a) Discuss the five-step process for hypothesis testing using a standardized scale.
 (b) From a total of 10,200 loans made by a state employees' credit union in the most recent 5-year period, 350 were sampled to determine what proportion was made to women. This sample showed that 39% of the loans were made to women employees. A complete census of loans 5 years ago showed that 41% of the borrowers then were women. At a significance level of 0.02, can you conclude that the proportion of loans made to women has changed significantly in the past five years? **(10+10 = 20 Marks)**

Answer

(a) The Five-Step Process for Hypothesis Testing:

- Step 1: Formulate Hypotheses (H_0 and H_1):** Set up the Null Hypothesis (H_0 , default baseline of no change) and Alternative Hypothesis (H_1 , research claim of change or effect).
- Step 2: Select Level of Significance (α):** Choose the probability of Type I error (e.g., $\alpha = 0.05, 0.02, 0.01$).
- Step 3: Compute Standardized Test Statistic:** Convert sample data into a standardized test statistic (such as Z or t):

$$Z = \frac{\text{Sample Statistic} - \text{Hypothesized Parameter}}{\text{Standard Error}}$$

- Step 4: Formulate Decision Rule (Find Critical Values):** Determine critical values Z_{crit} from probability tables based on α and test directionality (one-tailed or two-tailed). Define rejection region.
- Step 5: Make Statistical Decision & Business Conclusion:** If calculated test statistic falls into rejection region, reject H_0 ; otherwise fail to reject H_0 . State practical interpretation.

(b) Hypothesis Test for Loan Proportion to Women: Given Parameters:

- Historical Census Proportion (p_0): 0.41
- Sample Size (n): 350
- Sample Proportion ($\hat{p}$): 0.39
- Significance Level (α): 0.02

Step-by-Step Test Execution:

1. Hypotheses:

$$H_0 : p = 0.41 \quad (\text{Proportion of loans to women has not changed})$$

$$H_1 : p \neq 0.41 \quad (\text{Proportion of loans to women has changed significantly})$$

(This is a two-tailed test).

2. Significance Level:

$$\alpha = 0.02 \implies \frac{\alpha}{2} = 0.01 \text{ in each tail.}$$

3. **Standard Error & Test Statistic (Z_{calc}):**

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.41 \times (1-0.41)}{350}} = \sqrt{\frac{0.41 \times 0.59}{350}} = \sqrt{\frac{0.2419}{350}} = \sqrt{0.00069114} = 0.0262896$$
$$Z_{\text{calc}} = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.39 - 0.41}{0.0262896} = \frac{-0.02}{0.0262896} = -\mathbf{0.7608}$$

4. **Critical Value (Z_{crit}):** For a two-tailed test at $\alpha = 0.02$, standard normal table gives:

$$Z_{\text{crit}} = \pm 2.326$$

Rejection Region: Reject H_0 if $Z_{\text{calc}} < -2.326$ or $Z_{\text{calc}} > +2.326$.

5. **Decision & Conclusion:** Since $|Z_{\text{calc}}| = 0.7608 < 2.326$, the calculated test statistic falls inside the non-rejection region.

Decision: Fail to Reject H_0 .

Conclusion: At a 0.02 level of significance, there is no statistically significant evidence to conclude that the proportion of loans made to women employees has changed over the past five years.