Statistical Modelling for Machine Learning (313307)

MSBTE Diploma · K-Scheme · Semester 3 (AI & ML — AN and allied) · Important Questions — unit-wise

Based on MSBTE papers W-2024, S-2025, W-2025 and S-2026 available on msbte.ac.in, and the K-scheme syllabus. Practice guidance only — not a prediction of the exam.
Probability: High = asked in 3 or 4 of the 4 papers · **Medium** = asked in 2 papers · **Low** = asked in only one paper, or in syllabus but not asked yet. These are trends from past papers, not certainties — study the full syllabus.

Unit	Title	Syllabus marks	Marks seen in papers (W-24 + S-25 + W-25 + S-26)
1	Statistical Techniques	20	26 + 30 + 28 + 28
2	Statistical Methods (curve fitting, correlation, regression)	14	22 + 18 + 20 + 20
3	Probability of Random Variable	8	14 + 14 + 14 + 14
4	Interpolation	14	20 + 14 + 20 + 20
5	Sampling Methods (testing of hypothesis)	14	20 + 26 + 20 + 20

Paper marks include optional questions (e.g. 7 questions printed in Q.1, any 5 to attempt), so they are higher than the syllabus marks. Every paper is almost fully numerical — practise the calculations.

Unit I — Statistical Techniques (syllabus: 20 marks)

Question / topic	Marks	Asked in	Count	Probability
Karl Pearson's coefficient of skewness from given mean / mode / median / SD (find the missing value) *W-24 Q.1c: SD 8.5, mean 20.6, mode 18.52 → Sk.* *S-25 Q.1d: Sk, mean, mode given → SD.* *S-25 Q.3b: Sk = 0.5, SD = 10, mean = 40 → mode and median.* *S-26 Q.2d: Define Sk; evaluate for median 70, SD 10.*	2, 4	S-26, S-25, W-24	4	**High**
Median of grouped (continuous) frequency distribution *W-24 Q.2a: Median: marks 0-70, f = 4,4,8,10,12,8,4.* *S-25 Q.2b: Median marks: 0-50, f = 2,12,22,8,6.* *S-26 Q.2a: Median marks: 0-50, f = 2,12,22,8,6 (repeat of S-25).*	4	S-26, S-25, W-24	3	**High**
Quartiles / deciles of raw data (Q3, D5) *W-24 Q.1a: Q3 of 20, 28, 31, 18, 19, 17, 32, 33, 22, 21.* *S-25 Q.1b: D5 of 15 values.* *S-26 Q.2c: Q3 of 33, 18, 24, 11, 59, 47, 61, 72.*	2, 4	S-26, S-25, W-24	3	**High**
Mode by graphical method (histogram), verify by formula *W-24 Q.2d: Mode of ages 10-90 by histogram.* *S-25 Q.2c: Mode graphically and verify analytically (0-40).* *W-25 Q.2d: Mode by histogram (marks 10-14 ... 45-49, inclusive classes).*	4	W-25, S-25, W-24	3	**High**

Coefficient of skewness of a frequency distribution (Karl Pearson) *W-24 Q.2b: Skewness of wages of 230 persons (70-150).* *S-25 Q.3a: Skewness of ages of employees (20-55).* *W-25 Q.2b: Skewness of discrete x = 4, 7, 11, 16, 22.*	4	W-25, S-25, W-24	3	**High**
Bowley's (quartile) coefficient of skewness *W-24 Q.2c: Bowley's Sk of 0-60 distribution.* *S-25 Q.2d: Bowley Sk = 1.5 (given Q1+Q3 = 180, median 75): find Q1, Q3.* *W-25 Q.2c: Bowley's Sk of wages of 230 persons.*	4	W-25, S-25, W-24	3	**High**
Mean / median / mode of ungrouped data (incl. missing value k) *W-25 Q.1a: Mean of six values ×10⁴.* *S-26 Q.1a: Find k if mean is 2.5.* *S-26 Q.1b: Median of 8 values.* *S-26 Q.2b: Median and mode of daily earnings of 12 workers.*	2, 4	S-26, W-25	4	**Medium**
Measures of dispersion — SD, coefficient of variation (consistency), quartile deviation *W-25 Q.2a: Which football team is more consistent? (CV)* *W-25 Q.3a: Coefficient of Q.D. from less-than cumulative table.* *S-26 Q.3a: SD from "below" cumulative frequency table.*	4	S-26, W-25	3	**Medium**
Geometric mean / harmonic mean *W-24 Q.1b: H.M. of 90, 100, 150, 120, 200.* *W-25 Q.1b: G.M. of 10 values.*	2	W-25, W-24	2	**Medium**
Moments, β_1, β_2 and kurtosis *W-24 Q.3b: Kurtosis from moments, x = 2.0 … 5.0 with f.* *W-25 Q.3b: Moments about 2 are 1, 2.5, 5.5, 16 → central moments; comment on skewness and kurtosis.*	4	W-25, W-24	2	**Medium**
Combined mean *S-25 Q.1a: 40 boys mean 80, 20 girls mean 50.* *S-25 Q.2a: Combined mean 4496: find number of males and females.*	2, 4	S-25	2	Low
Histogram (unequal class widths) *S-26 Q.3b: Draw histogram, weight classes of unequal widths.*	4	S-26	1	Low
Ogive (less-than / more-than) to locate median graphically	—	Not yet	0	Low
Range, mean deviation and their coefficients	—	Not yet	0	Low
Percentiles of grouped data; types of skewness (with curves)	—	Not yet	0	Low

Unit II — Statistical Methods (curve fitting, correlation, regression) (syllabus: 14 marks)

Question / topic	Marks	Asked in	Count	Probability
Regression lines (y on x and x on y) from data *W-24 Q.4a: Regression lines, x = 2…10.* *W-25 Q.4a: Regression of Y on X (X = 6, 2, 10, 4, 8).* *S-26 Q.3c: Regression lines (same data as W-25).*	4	S-26, W-25, W-24	4	**High**

S-26 Q.4a: Regression of Y on X, 8 pairs.				
Spearman's rank correlation (ranks given / with tied ranks) *W-24 Q.3a: Rank correlation, 5 pairs.* *W-24 Q.4b: Two ladies rank lipsticks of 7 companies.* *S-25 Q.4a: Rank correlation with repeated values (68 thrice).* *W-25 Q.4b: Ranks of 10 students by two teachers.*	4	W-25, S-25, W-24	4	**High**
Karl Pearson's coefficient of correlation from data *W-24 Q.3d: x = 5, 9, 13, 17, 21; y = 12, 20, 25, 33, 35.* *S-25 Q.3d: Same data; explain type of correlation.* *W-25 Q.3d: From deviations from means.*	4	W-25, S-25, W-24	3	**High**
Fitting a straight line by least squares (trend line) *W-24 Q.3c: Fit y = ax + b, x = 0…7.* *S-25 Q.3c: Fit y = a + bx; estimate y at 2.5.* *W-25 Q.3c: Straight-line trend of passenger cars 2007–2012.*	4	W-25, S-25, W-24	3	**High**
Covariance *S-25 Q.1c: Covariance, x = 1…5.* *S-26 Q.1c: Covariance (same data).* *S-26 Q.1d: "Coefficient of covariance" of 4 pairs.*	2	S-26, S-25	3	**Medium**
Correlation coefficient from given sums (Σx, Σy, Σxy …) *W-24 Q.1d: n = 25 with sums.* *W-25 Q.1c: N = 7 with sums.*	2	W-25, W-24	2	**Medium**
Regression line from given sums; estimate *S-25 Q.4b: Price-supply, 10 observations; estimate y at x = 16.* *S-26 Q.3d: Same data.*	4	S-26, S-25	2	**Medium**
Properties of regression coefficients (r^2 = byx·bxy) *W-25 Q.1d: r = 0.5, byx = 0.5 → bxy.*	2	W-25	1	Low
Correlation and its types (theory) *S-26 Q.4b: Explain correlation and its types.*	4	S-26	1	Low
Fitting a second-degree parabola $y = a + bx + cx^2$ by least squares	—	Not yet	0	Low
Scatter diagram (types of correlation by graph)	—	Not yet	0	Low
Properties of correlation / regression; angle between regression lines	—	Not yet	0	Low

Unit III — Probability of Random Variable (syllabus: 8 marks)

Question / topic	Marks	Asked in	Count	Probability
Bayes' theorem (bags, machines, insurance, truth-teller) *W-24 Q.4e: Man speaks truth 2 out of 3 times; reports a four.* *S-25 Q.4e: Two bags, black ball drawn → bag 1.* *W-25 Q.4e: Machines A, B, C making bolts.* *S-26 Q.4d: Two bags (repeat of S-25).* *S-26 Q.4e: Scooter / car / truck drivers.*	4	S-26, W-25, S-25, W-24	5	**High**
Simple probability — coins, dice, numbered balls *W-24 Q.1e: 30 balls: multiple of 5 or 7.*	2, 4	S-26, W-25, S-25, W-24	5	**High**

S-25 Q.1e: Three coins: all heads. *W-25 Q.1e: Five coins: two tails.* *S-26 Q.1e: Coin tossed thrice: exactly two tails.* *S-26 Q.4c: Two dice: sum 6 or 10.*				
Drawing balls without replacement / conditional probability *W-24 Q.4d: 3 red, 7 black; second red → first red.* *S-25 Q.4c: 10 red, 5 white, 5 black: two not of same colour.* *W-25 Q.4d: 20 red, 10 blue: both red without replacement.*	4	W-25, S-25, W-24	3	**High**
Addition theorem (cards, "either–or", "at least one") *W-24 Q.4c: Diamond or a face card.* *S-25 Q.4d: HSC and IIT exam: P(IIT).* *W-25 Q.4c: Ace or red queen or king of hearts.*	4	W-25, S-25, W-24	3	**High**
Multiplication theorem for independent events	—	Not yet	0	Low
Definitions: sample space, mutually exclusive, exhaustive, independent events	—	Not yet	0	Low

Unit IV — Interpolation (syllabus: 14 marks)

Question / topic	Marks	Asked in	Count	Probability
Extrapolation (squares / cubes / beyond the table) *W-24 Q.5c(ii): y(19) from x = 8…18.* *S-25 Q.5b: Cubes: find 12^3.* *W-25 Q.6b(i): Value at x = 0.4 beyond table.* *S-26 Q.5b: Squares: find 13^2.*	3, 6	S-26, W-25, S-25, W-24	4	**High**
Forward / backward difference table (polynomial or given data) *W-24 Q.1f: Backward differences of $x^3 - 3x^2 + 5x + 7$.* *W-24 Q.5b(i): Forward table; write f(0), f(2) values.* *S-25 Q.1f: Forward table for 7 values.* *W-25 Q.1f: Forward table of $x^2 - 3x + 1$; second differences constant.* *W-25 Q.5b(i): Find a higher difference value from table.*	2, 3	W-25, S-25, W-24	5	**High**
Lagrange's interpolation formula *W-24 Q.5a: y at x = 10; f(2.5) from 3 points.* *S-25 Q.5a(i): f(13) from x = 0, 1, 2, 5.* *W-25 Q.5a: Find polynomial f(x); f(·) from 4 points.*	3, 6	W-25, S-25, W-24	3	**High**
Newton–Gregory forward interpolation *S-25 Q.5a(ii): Estimate f(8), x = 5…20.* *S-26 Q.1f: f(12) from x = 10, 15, 20.* *S-26 Q.5a: Find the function y = f(x) through 5 points.* *S-26 Q.6b: Population in 1955 from census.*	2, 3, 6	S-26, S-25	4	**Medium**
Newton–Gregory backward interpolation *W-24 Q.5c(i): f(1.28) from f(1.15) … f(1.30).* *W-25 Q.6b(ii): Same question repeated.*	3	W-25, W-24	2	**Medium**
Operators Δ, ∇, E — proofs / evaluations *W-24 Q.5b(ii): Prove $(\Delta^2/E)\, x^3 = 6x$.* *W-25 Q.5b(ii): Prove $\Delta^3[(1 - x)(1 - 2x)(1 - 3x)] = -36$ (h = 1).*	3	W-25, W-24	2	**Medium**

Relations between operators ($E = 1 + \Delta$, $\nabla = 1 - E^{-1}$, $\Delta = E\nabla$ …)	—	Not yet	0	Low

Unit V — Sampling Methods (testing of hypothesis) (syllabus: 14 marks)

Question / topic	Marks	Asked in	Count	Probability
Chi-square test of goodness of fit (uniform / binomial) *W-24 Q.6b: 320 families with 5 children: boys and girls equally probable?* *S-25 Q.5c: Aircraft accidents on days of week.* *W-25 Q.6c: 320 families (repeat of W-24).* *S-26 Q.5c: Accidents uniformly distributed over the week?*	6	S-26, W-25, S-25, W-24	4	**High**
Large-sample test of proportion (one / two proportions) *W-24 Q.1g: Rice vs wheat eaters, 280 of 500.* *W-24 Q.6a: Machine defectives before/after; die thrown 1500 times.* *S-25 Q.6a(ii): Smokers in two cities.* *W-25 Q.6a(ii): Coin 400 tosses, 216 heads.*	2, 3, 6	W-25, S-25, W-24	4	**High**
Student's t-test (single mean / paired) *W-24 Q.6c: 9 items vs population mean 47.5.* *S-25 Q.6a(i): 9 values vs population mean 65.* *W-25 Q.5c: Increase in blood pressure of 12 patients.*	3, 6	W-25, S-25, W-24	3	**High**
Large-sample test of mean (one mean / two means) *S-25 Q.1g: Z for sample of 50.* *S-25 Q.6c: Electric bulbs 1570 h vs claimed 1600 h.* *W-25 Q.1g: Mean scores of girls and boys.*	2, 6	W-25, S-25	3	**Medium**
Chi-square test of independence (contingency table) *S-25 Q.6b: IQ vs economic condition, 2 × 2.* *S-26 Q.6a: Quinine and fever, 2 × 2.* *S-26 Q.6c: Soft drinks vs employee category, 3 × 3.*	6	S-26, S-25	3	**Medium**
Chi-square test for population variance *W-25 Q.6a(i): n = 12, s = 11; is σ = 10?*	3	W-25	1	Low
Random and non-random sampling (definitions) *S-26 Q.1g: Define random and non-random sampling.*	2	S-26	1	Low
Null / alternative hypothesis, Type I and Type II errors, level of significance, critical region	—	Not yet	0	Low
Standard error, parameter vs statistic, degrees of freedom	—	Not yet	0	Low
Large-sample test for difference of two standard deviations	—	Not yet	0	Low

Paper pattern (same in all four papers)

Q.	Attempt	Marks each	Total	Units usually asked
1	Any FIVE of 7	2	10	All units (short numericals)
2	Any THREE of 4	4	12	Unit I
3	Any THREE of 4	4	12	Unit I and II
4	Any THREE of 5	4	12	Unit II and III
5	Any TWO of 3	6	12	Unit IV (and V)
6	Any TWO of 3	6	12	Unit V (and IV)
	3 Hours		**70**	Non-programmable calculator allowed

Note: This is an **important-questions list prepared by Ananda Technologies** only for practice and exam preparation, based on the pattern of previous MSBTE question papers and the K-scheme syllabus. It is **not an official MSBTE paper**, and it does **not** mean that these questions will come in the exam. Answers follow standard textbook points; marks shown are indicative. Free for students — anandatechnologies.in