

TOP 30 MOST IMPORTANT QUESTIONS

Statistics for Artificial Intelligence & Data Science

Compiled by: Nitin Gupta

Question Bank Analysis

Based on comprehensive analysis of 5 previous year question papers (2022-2025), the following 30 questions represent the most frequently asked and important topics for Statistics for AI & DS examination.

1. Hypothesis Testing [05 Marks]

Question: What is hypothesis testing? Explain type I and type II errors with examples. State null and alternative hypotheses for a given scenario.

Topic Coverage: Hypothesis testing fundamentals, Type I error (α), Type II error (β), null hypothesis (H_0), alternative hypothesis (H_1)

2. Linear Regression [10 Marks]

Question: Find the simple linear regression equation that fits the given data. Calculate the coefficient of determination (R^2). Explain applications of linear regression.

Topic Coverage: Simple linear regression, regression equation $y = mx + c$, R^2 , applications in AI/DS

3. Standard Deviation & Variance [05 Marks]

Question: Find the standard deviation and variance for a given dataset. Define standard deviation and its importance in statistics.

Topic Coverage: Measures of dispersion, standard deviation formula, variance, data spread

4. Correlation Coefficient [10 Marks]

Question: Find the correlation coefficient from given data (X vs Y). Explain what is correlation and how is it different from regression?

Topic Coverage: Pearson correlation coefficient, correlation vs regression, interpretation of r values

5. One-Way ANOVA [10 Marks]

Question: Explain why ANOVA is used. Solve a problem using one-way ANOVA method with given samples to identify differences between groups.

Topic Coverage: Analysis of Variance, F-statistic, between-group variance, within-group variance

6. Chi-Square Test [10 Marks]

Question: What is Chi-Square Test? Use Chi-Square Test for Independence to determine association between two categorical variables with contingency table at $\alpha=0.05$.

Topic Coverage: Chi-square test of independence, contingency tables, expected frequencies, test statistic

7. Normal Distribution Problems [10 Marks]

Question: X is normally distributed with mean μ and standard deviation σ . Find probabilities: $P(x < a)$, $P(x > b)$, $P(c < x < d)$. Solve real-world problems using normal distribution.

Topic Coverage: Normal distribution, z-score, probability calculations, standard normal table

8. Confidence Interval & Standard Error [10 Marks]

Question: Define Confidence Interval. Calculate standard error, mean, standard deviation, and 95% confidence interval for given sample data.

Topic Coverage: Confidence intervals, standard error, margin of error, population estimation

9. t-Test / t-Distribution [10 Marks]

Question: Calculate t-score for given sample mean, expected mean, standard deviation, and sample size. Conduct t-test to test a claim at given significance level.

Topic Coverage: Student's t-distribution, t-test, degrees of freedom, small sample testing

10. z-Score Calculation [05 Marks]

Question: A researcher is analyzing test scores. Calculate the z-score for the sample mean given population mean, sample mean, standard deviation, and sample size.

Topic Coverage: z-score formula, standardization, normal distribution applications

11. Sampling Methods [05 Marks]

Question: Explain the difference between Stratified and Cluster sampling. Write short notes on Random sampling methods.

Topic Coverage: Sampling techniques, probability sampling, stratified vs cluster sampling

12. Kruskal-Wallis Test [10 Marks]

Question: At 0.05 significance level, use Kruskal-Wallis test to determine if the scores/values differ significantly across different groups.

Topic Coverage: Non-parametric test, Kruskal-Wallis H test, rank-based methods

13. Box Plot & Scatter Plot [05 Marks]

Question: Explain the use of scatter plot and box plot with examples. Explain percentiles and boxplots with neat diagrams.

Topic Coverage: Data visualization, box plot components (Q1, Q2, Q3, IQR), scatter plots for correlation

14. Frequency Distribution Table [10 Marks]

Question: Create a frequency distribution table for given data. Answer questions about class mark, range, number of intervals, and frequency analysis.

Topic Coverage: Frequency tables, class intervals, class mark, cumulative frequency

15. Poisson Distribution [10 Marks]

Question: Solve Poisson distribution problems: Given average rate, find probability that none/some events occur in given time period. Calculate expected number.

Topic Coverage: Poisson distribution, λ (lambda), probability mass function, applications

16. Binomial Distribution [10 Marks]

Question: Explain Binomial distribution in detail. Solve problems calculating probabilities for all, at least, or exactly certain number of successes.

Topic Coverage: Binomial distribution, n trials, p probability, combination formula

17. Fisher's Exact Test [05 Marks]

Question: What is Fisher's exact test? When is it used instead of Chi-Square test?

Topic Coverage: Fisher's exact test, small sample categorical data, 2×2 contingency tables

18. p-value [05 Marks]

Question: Explain the concept of p-value in hypothesis testing. How is it used to make decisions about null hypothesis?

Topic Coverage: p-value interpretation, significance level α , decision rule

19. Central Limit Theorem [05 Marks]

Question: Illustrate central limit theorem with a neat diagram. Explain its importance in statistics and sampling.

Topic Coverage: Central Limit Theorem (CLT), sampling distribution, normal approximation

20. F-Test [10 Marks]

Question: What is F-Test? Given F statistic and degrees of freedom, determine whether to reject null hypothesis at given significance level using F-table.

Topic Coverage: F-distribution, variance ratio test, F-table, ANOVA applications

21. Two-Way ANOVA [10 Marks]

Question: Explain the concept of two-way ANOVA. How does it differ from one-way ANOVA? Describe assumptions and when to use Friedman's test as non-parametric alternative.

Topic Coverage: Two-way ANOVA, interaction effects, main effects, assumptions, Friedman test

22. Weibull Distribution [05 Marks]

Question: Write short notes on Weibull distribution. Explain its applications and when it is used.

Topic Coverage: Weibull distribution, reliability analysis, failure rate, shape parameter

23. QQ Plot [05 Marks]

Question: Explain QQ plots in detail. Show how scatterplots explore relationships between variables.

Topic Coverage: Quantile-Quantile plots, normality testing, distribution comparison

24. Histogram [05 Marks]

Question: What is Histogram? Give its applications. Explain with examples and neat diagrams.

Topic Coverage: Histogram construction, frequency visualization, data distribution

25. Bootstrapping vs Re-sampling [05 Marks]

Question: Discuss Bootstrapping vs. re-sampling methods. Explain their differences and applications.

Topic Coverage: Bootstrap methods, resampling techniques, standard error estimation

26. Least Square Regression [10 Marks]

Question: Explain linear least square regression (LLSR) along with its advantages and disadvantages. When is it preferred?

Topic Coverage: Least squares method, minimizing residuals, linear regression fitting

27. Friedman Test [10 Marks]

Question: Explain Friedman Test as a non-parametric alternative to repeated measures ANOVA. Solve given problem using Friedman test.

Topic Coverage: Friedman test, rank-based analysis, repeated measures, blocking

28. Outlier Detection [05 Marks]

Question: How to detect outliers? Explain different methods for outlier detection with examples.

Topic Coverage: Outlier identification, IQR method, z-score method, box plot outliers

29. Numerical vs Categorical Data [05 Marks]

Question: Explain Numerical and Categorical data types with appropriate examples. Give differences and when each is used.

Topic Coverage: Data types, continuous vs discrete, nominal vs ordinal, ratio vs interval

30. Stem & Leaf Plot [05 Marks]

Question: Write short notes on Stem & Leaf Plot. Explain with examples and its advantages over histograms.

Topic Coverage: Stem-and-leaf display, data representation, maintaining original values

Preparation Strategy

High Priority Topics (Appeared 10+ times):

1. **Hypothesis Testing** - Type I & II errors, null/alternative hypotheses
2. **Standard Deviation** - Calculation, interpretation, applications

High Priority Topics (Appeared 5-7 times):

3. **Linear Regression** - Equation, R^2 , applications
4. **Correlation** - Coefficient calculation, interpretation
5. **ANOVA** - One-way, two-way, when to use
6. **Chi-Square Test** - Independence test, contingency tables
7. **Probability Calculations** - Normal, Poisson, Binomial distributions

Medium Priority Topics (Appeared 3-4 times):

8. **Standard Error & Confidence Intervals**
9. **Sampling Methods**
10. **Box Plot & Scatter Plot**
11. **Normal Distribution**

Important Topics (Appeared 2 times):

12. t-Test/t-Distribution
13. Kruskal-Wallis Test
14. Binomial & Poisson Distributions
15. F-Test
16. Fisher's Exact Test
17. Frequency Distribution
18. Central Limit Theorem
19. Friedman Test

Exam Pattern Summary

- **Total Marks:** 80
- **Duration:** 3 Hours
- **Question Pattern:**
 - Q1: Compulsory - Attempt any 4 out of 5/6 sub-questions (4×5 marks = 20 marks)
 - Q2-Q6: Attempt any 3 out of 5 (3 questions \times 20 marks = 60 marks)

- Each question Q2-Q6 has 2 sub-parts of 10 marks each

Important Formulas to Remember:

Measures of Central Tendency & Dispersion:

- Mean: $\bar{x} = \frac{\sum x}{n}$
- Standard Deviation: $s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$
- Variance: s^2

Probability Distributions:

- Normal: $z = \frac{x-\mu}{\sigma}$
- Binomial: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Poisson: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Regression & Correlation:

- Correlation: $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$
- Regression: $y = mx + c$ where $m = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$

Hypothesis Testing:

- z-score: $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$
- t-score: $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
- Chi-square: $\chi^2 = \sum \frac{(O-E)^2}{E}$
- F-statistic: $F = \frac{MS_{between}}{MS_{within}}$

Confidence Interval:

- CI: $\bar{x} \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$

Note: Practice numerical problems extensively. Most questions involve calculations with real data. Always show step-by-step working and state formulas used.