Statistics & Probability Practice Test for Data Analysts

This is a concept-driven, practical practice test designed specifically for Data Analyst roles.
No unnecessary theory — every question reflects real-world analyst thinking.

Instructions:
• Attempt without using formulas initially; rely on intuition first.
• Use Python / Excel only where explicitly mentioned.
• If you can confidently solve 80–85%, you have a strong grip on statistics & probability.

Section 1: Descriptive Statistics

1. A dataset has extreme outliers on the higher end. Which is more appropriate to report — mean or median? Why?

2. Can the mean be smaller than the median? Describe a real business example.

3. A company tracks customer spend. Most customers spend ■200–■300, but a few spend ■50,000+. What happens to mean, median, and standard deviation?

4. Two datasets have the same mean but very different standard deviations. What does this imply practically?

5. When would mode be more useful than mean or median in analytics?

6. If variance is zero, what does it tell you about the dataset?

Section 2: Data Distribution & Shape

7. Explain right-skewed vs left-skewed distribution using income data.

8. In a right-skewed distribution, how are mean, median, and mode ordered?

9. Why is normal distribution important for analytics, even though real data is often not normal?

10. Can a dataset be symmetrical but not normal? Explain.

11. What does kurtosis tell you in simple terms?

Section 3: Probability Fundamentals

12. Difference between independent and mutually exclusive events (with example).

13. A website has a 2% conversion rate. What does this probability really mean?

14. Why is probability always between 0 and 1?

15. If $P(A) = 0.6$, what is $P(\text{not } A)$? Why?

16. Explain conditional probability using an email spam filter example.

Section 4: Bayes' Theorem

17. Why is Bayes' theorem important in decision-making?

18. A disease affects 1% of the population. Test accuracy is 95%. Why can positive test results still be misleading?

19. In analytics terms, explain prior, likelihood, and posterior.

20. Give one business use case where Bayes' theorem is applied.

Section 5: Sampling & Bias

21. Difference between population and sample.

22. Why do analysts prefer random sampling?

23. Explain sampling bias with a real-world example.

24. What is selection bias, and how does it affect insights?

25. Why does increasing sample size reduce variability?

Section 6: Hypothesis Testing

26. Define null hypothesis and alternative hypothesis in business terms.

27. What is a p-value, in simple language?

28. Why do analysts commonly use 0.05 as significance level?

29. Difference between Type I and Type II errors with a business example.

30. What happens if we set alpha = 0.10 instead of 0.05?

31. Can we prove the null hypothesis? Why or why not?

Section 7: Confidence Intervals

32. What does a 95% confidence interval actually mean?

33. Why is a confidence interval often more useful than a point estimate?

34. What happens to confidence interval width if sample size or confidence level increases?

35. Can two groups have overlapping confidence intervals but still be statistically significant?

Section 8: Statistical Tests

36. When would you use a t-test instead of a z-test?

37. Difference between one-tailed and two-tailed tests.

38. When is ANOVA preferred over multiple t-tests?

39. What does chi-square test check in analytics?

40. Give an example of categorical vs categorical analysis.

Section 9: Correlation

41. Difference between correlation and causation.

42. Can correlation be negative? Give a business example.

43. Why does correlation not capture non-linear relationships?

44. What does correlation = 0 mean?

Section 10: Real-World Scenarios

45. An A/B test shows a 3% uplift but is not statistically significant. What should you do?

46. Your dataset is highly skewed. What transformations might you consider and why?

47. Sales dropped last month. Why is jumping to conclusions statistically dangerous?

48. Why should outliers be investigated before removal?

49. How does seasonality affect statistical analysis?

50. How do you decide which statistical method to use for a problem?