**Questions Set Statistics**

1. Define data. How is data different from information?
2. What is a line chart, and when would you use it?
3. Explain the difference between a scatter chart and a line chart.
4. What is the purpose of a bar graph? How does it differ from a histogram?
5. Describe when you would use a pie chart and its potential limitations.
6. Define and differentiate between the mean, median, and mode.
7. What are measures of dispersion, and why are they important?
8. Explain the concept of skewness in a dataset.
9. What is a uniform distribution?
10. Describe the binomial distribution and provide an example of its application.
11. What is a standard normal distribution?
12. What is random sampling and how does it differ from stratified sampling?
13. Explain the Central Limit Theorem and its significance.
14. What is hypothesis testing, and how do you interpret p-values?
15. Define Type 1 and Type 2 errors.
16. Describe different data collection techniques and their pros and cons.
17. Explain the process of data cleaning and why it is crucial.
18. Explain the central limit theorem
19. How would you describe a 'p-value
20. What is sampling?
21. What is linear regression?
22. What is the normal distribution?
23. What is bias in data science?
24. Where is inferential statistics used?
25. Difference between population and sample
26. What does it mean by inlier
27. What does standard deviation mean?
28. What is a statistical interaction?
29. What is an outlier?
30. What is selection bias?
31. What is the confidence interval?
32. What is the Pareto Principle?
33. Describe hypothesis testing
34. What are descriptive statistics?
35. What are left-skewed and right-skewed distributions
36. What is Bessel's correction?
37. Exploratory data analysis
38. What is root cause analysis?
39. What is the assumption of normality?
40. What is the binomial distribution formula?
41. How do data scientists use statistics
42. Explain the central limit theorem and give examples of when you can use it in a real-world problem?
43. Briefly explain the A/B testing and its application? What are some common pitfalls encountered in A/B testing?
44. Describe briefly the hypothesis testing and p-value in layman’s term? And give a practical application for them?
45. Given a left-skewed distribution that has a median of 60, what conclusions can we draw about the mean and the mode of the data?
46. What is the meaning of selection bias and how to avoid it?
47. Explain the long-tailed distribution and provide three examples of relevant phenomena that have long tails. Why are they important in classification and regression problems?
48. What is the meaning of KPI in statistics
49. Say you flip a coin 10 times and observe only one head. What would be the null hypothesis and p-value for testing whether the coin is fair or not?
50. You are testing hundreds of hypotheses, each with a t-test. What considerations would you take into account when doing this?
51. What general conditions must be satisfied for the central limit theorem to hold?
52. What is skewness discuss two methods to measure it?
53. You sample from a uniform distribution [0, d] n times. What is your best estimate of d?
54. Discuss the Chi-square, ANOVA, and t-test
55. Say you have two subsets of a dataset for which you know their means and standard deviations. How do you calculate the blended mean and standard deviation of the total dataset? Can you extend it to K subsets?
56. What is the relationship between the significance level and the confidence level in Statistics?
57. What is the Law of Large Numbers in statistics and how it can be used in data science ?
58. What is the difference between a confidence interval and a prediction interval, and how do you calculate them?
59. What are the differences between the z-test and t-test?
60. When to use a z-test Vs a t-test?
61. Given a specific dataset, how do you calculate t-statistic or z-statistics?