Probability And Statistics Problems With Solutions

MASTERING PROBABILITY AND STATISTICS: ESSENTIAL PROBLEMS WITH STEP-BY-STEP SOLUTIONS

Probability and statistics form the backbone of data-driven decision-making in nearly every field today, from business analytics and healthcare to engineering and social sciences. Whether you are a student preparing for an exam, a professional brushing up on foundational concepts, or a curious learner, working through probability and statistics problems with solutions is one of the most effective ways to build genuine understanding. Theoretical knowledge alone is rarely enough; the real skill lies in applying concepts to practical scenarios, interpreting results, and drawing meaningful conclusions. This article presents a carefully curated selection of common problem types in probability and statistics, each accompanied by clear, step-by-step solutions. By working through these examples, you will strengthen your analytical thinking, become more comfortable with key formulas, and gain the confidence to tackle even complex data challenges on your own.

WHY PRACTICE PROBABILITY AND STATISTICS PROBLEMS WITH SOLUTIONS?

Before diving into specific problems, it is worth understanding why solving problems is so critical for mastering this subject. Probability and statistics are inherently applied disciplines. Reading definitions or memorizing formulas without context often leads to shallow learning that fades quickly. When you actively solve probability and statistics problems with solutions, you engage multiple cognitive processes: you must identify the correct approach, apply formulas accurately, interpret numerical results, and check your reasoning against the provided solution. This active engagement creates stronger neural connections and deeper long-term retention. Additionally, repeatedly encountering different problem structures helps you recognize patterns, so when you face a new problem in a real-world setting, you can quickly map it to a familiar framework. For students, practicing with solved problems is also an excellent exam preparation strategy, as it reveals common pitfalls, typical question formats, and the level of detail expected in answers.

FOUNDATIONAL PROBABILITY PROBLEMS

Probability is the language of uncertainty, and mastering its fundamentals is essential before moving to more advanced statistical topics. Below are several classic problems that illustrate core probability concepts.

Problem 1: Calculating Basic Probability

Question: A fair six-sided die is rolled once. What is the probability of rolling an even number or a number greater than 4?

Solution: First, list the sample space for a single die roll: {1, 2, 3, 4, 5, 6}. The event of rolling an even number is E = {2, 4, 6}. The event of rolling a number greater than 4 is G = {5, 6}. We need P(E ∪ G). Using the addition rule for non-mutually exclusive events: P(E ∪ G) = P(E) + P(G) − P(E ∩ G). P(E) = 3/6 = 1/2. P(G) = 2/6 = 1/3. The intersection E ∩ G = {6}, so P(E ∩ G) = 1/6. Therefore, P(E ∪ G) = 1/2 + 1/3 − 1/6 = 3/6 + 2/6 − 1/6 = 4/6 = 2/3. The probability is 2/3.

Problem 2: Conditional Probability

Question: In a class of 100 students, 60 study mathematics, 40 study physics, and 20 study both. If a student is selected at random and is known to study mathematics, what is the probability that they also study physics?

Solution: This is a classic conditional probability problem. Let M be the event that a student studies mathematics, and P be the event that a student studies physics. We are asked for P(P | M). Using the conditional probability formula: P(P | M) = P(P ∩ M) / P(M). We are given that P(P ∩ M) = 20/100 = 0.2, and P(M) = 60/100 = 0.6. Therefore, P(P | M) = 0.2 / 0.6 = 1/3 ≈ 0.3333. So, if a student studies mathematics, there is a 33.33% chance they also study physics.

Problem 3: Bayes' Theorem Application

Question: A medical test for a rare disease has a 99% sensitivity (true positive rate) and a 95% specificity (true negative rate). The disease prevalence in the population is 1%. If a randomly selected person tests positive, what is the probability they actually have the disease?

Solution: This problem demonstrates why even highly accurate tests can yield surprising results for rare conditions. Let D be the event of having the disease, and T+ be the event of testing positive. We want P(D | T+). Given: P(D) = 0.01, P(T+ | D) = 0.99, P(T+ | not D) = 1 − 0.95 = 0.05. Using Bayes' theorem: P(D | T+) = [P(T+ | D) × P(D)] / [P(T+ | D) × P(D) + P(T+ | not D) × P(not D)] = (0.99 × 0.01) / (0.99 × 0.01 + 0.05 × 0.99) = 0.0099 / (0.0099 + 0.0495) = 0.0099 / 0.0594 ≈ 0.1667. The probability is about 16.67%. This low value, despite the high test accuracy, is due to the low prevalence of the disease.

DESCRIPTIVE STATISTICS PROBLEMS

Descriptive statistics help us summarize and understand data before making inferences. These problems focus on measures of central tendency and dispersion.

Problem 4: Mean, Median, and Mode

Question: The following are the ages (in years) of 10 employees in a small company: 22, 25, 25, 27, 29, 31, 34, 37, 41, 53. Calculate the mean, median, and mode. Which measure best represents the center of this data?

Solution: The mean is the sum of all ages divided by 10: (22 + 25 + 25 + 27 + 29 + 31 + 34 + 37 + 41 + 53) / 10 = 324 / 10 = 32.4 years. To find the median, first order the data (already sorted). Since there are 10 observations (an even number), the median is the average of the 5th and 6th values: (29 + 31) / 2 = 30 years. The mode is the most frequent value: 25 years (appears twice). The data has a right skew due to the outlier age of 53. The median of 30 is less affected by this outlier than the mean of 32.4, so the median is likely the best measure of central tendency for this dataset.

Problem 5: Variance and Standard Deviation

Question: Calculate the sample variance and sample standard deviation for the following test scores: 70, 80, 80, 90, 100.

Solution: First, compute the sample mean: $\bar{x}$ = (70 + 80 + 80 + 90 + 100) / 5 = 420 / 5 = 84. Next, calculate the squared deviations from the mean: $(70 - 84)^2 = 196$, $(80 - 84)^2 = 16$, $(80 - 84)^2 = 16$, $(90 - 84)^2 = 36$, $(100 - 84)^2 = 256$. The sum of squared deviations is 196 + 16 + 16 + 36 + 256 = 520. The sample variance s^2 = sum of squared deviations / (n − 1) = 520 / 4 = 130. The sample standard deviation $s = \sqrt{130} \approx 11.40$. The variance is 130 points squared, and the standard deviation is about 11.40 points.

PROBABILITY DISTRIBUTIONS PROBLEMS

Understanding probability distributions is crucial for modeling random phenomena and making predictions. Here are two common distribution problems.

Problem 6: Binomial

Distribution

Question: A coin is flipped 8 times. What is the probability of getting exactly 5 heads? Assume the coin is fair.

Solution: This is a binomial experiment with $n = 8$ trials, probability of success $p = 0.5$, and we want $P(X = 5)$. The binomial probability formula is $P(X = k) = C(n, k) \times p^k \times (1 - p)^{(n - k)}$. Here, $C(8, 5) = 8! / (5! \times 3!) = 56$. So $P(X = 5) = 56 \times (0.5)^5 \times (0.5)^3 = 56 \times (0.5)^8 = 56 \times 0.00390625 = 0.21875$. The probability of getting exactly 5 heads in 8 flips is approximately 0.2188 or 21.88%.

Problem 7: Normal Distribution

Question: The heights of adult women in a certain population are normally distributed with a mean of 165 cm and a standard deviation of 6 cm. What percentage of women are between 159 cm and 177 cm tall?

Solution: We need to find $P(159 < X < 177)$. First, convert to z-scores. For $x = 159$, $z = (159 - 165) / 6 = -1.00$. For $x = 177$, $z = (177 - 165) / 6 = 2.00$. Using the standard normal distribution table or a calculator, $P(Z < -1.00) = 0.1587$, and $P(Z < 2.00) = 0.9772$. Therefore, $P(-1.00 < Z < 2.00) = 0.9772 - 0.1587 = 0.8185$. So about 81.85% of adult women are between 159 cm and 177 cm tall.

INFERENTIAL STATISTICS PROBLEMS

Inferential statistics allow us to draw conclusions about populations based on sample data. These problems cover confidence intervals and hypothesis testing.

Problem 8: Confidence Interval for a Mean

Question: A random sample of 25 students has an average test score of 78 points with a sample standard deviation of 10 points. Construct a 95% confidence interval for the population mean test score.

Solution: Since the population standard deviation is unknown and the sample size is relatively small ($n = 25$), we use the t-distribution. The degrees of freedom are $n - 1 = 24$. For a 95% confidence level, the critical t-value t^* for 24 df is approximately 2.064. The margin of error $E = t^* \times (s / \sqrt{n}) = 2.064 \times (10 / \sqrt{25}) = 2.064 \times 2 = 4.128$. The confidence interval is $\bar{x} \pm E = 78 \pm 4.128$, which gives (73.872, 82.128). We are 95% confident that the true population mean test score lies between approximately 73.87 and 82.13 points.

Problem 9: Hypothesis

Testing (Two-Sample t-Test)

Question: Two groups of plants are grown under different light conditions. Group A ($n = 10$) has a mean height of 24 cm with standard deviation 3 cm. Group B ($n = 12$) has a mean height of 28 cm with standard deviation 4 cm. Test at the 0.05 significance level whether the mean heights differ significantly. Assume equal variances.

Solution: This is a two-sample t-test for independent samples. Null hypothesis H_0: $\mu A = \mu B$. Alternative hypothesis H_1: $\mu A \neq \mu B$ (two-tailed). First, calculate the pooled standard deviation: $s_p = \sqrt{[((nA - 1) \times sA^2 + (nB - 1) \times sB^2) / (nA + nB - 2)]} = \sqrt{[((9 \times 9) + (11 \times 16)) / 20]} = \sqrt{[(81 + 176) / 20]} = \sqrt{(257 / 20)} = \sqrt{12.85} \approx 3.585$. The standard error of the difference is $SE = s_p \times \sqrt{(1/nA + 1/nB)} = 3.585 \times \sqrt{(1/10 + 1/12)} = 3.585 \times \sqrt{(0.1 + 0.08333)} = 3.585 \times \sqrt{0.183}$