



# EnterMedSchool.org

---

## Statistics and Probability

### Exam — Statistics and Probability

Pre-med style conceptual questions on probability fundamentals (complement, inclusion-exclusion, independence, conditional probability), counting (permutations, combinations), Bayes' theorem and base rates, descriptive statistics (mean, median, mode, standard deviation), normal distributions, hypothesis testing, correlation vs causation, and experimental design.

75 items — Printable Exam

Free & Open-Source

Licensed under Creative Commons — Attribution required when sharing

Generated February 20, 2026

Scan to visit online





**1** The probability of an event occurring is 0.3. What is the probability that the event does NOT occur?



- A 0.3
- B 0.7
- C 0.03
- D 1.3
- E Cannot be determined without more information

**2** In a class, 60% of students play football, 40% play basketball, and 25% play both. What percentage of students play at least one of these sports?



- A 100%
- B 75%
- C 85%
- D 65%
- E 50%

**3** A bag contains 5 red and 3 blue marbles. If two marbles are drawn WITHOUT replacement, the probability of drawing two red marbles is:



- A  $25/64$
- B  $5/14$
- C  $5/8 \times 5/8$
- D  $10/56$
- E  $1/2$





4 Events A and B are independent. If  $P(A) = 0.4$  and  $P(B) = 0.5$ , what is  $P(A$  and  $B)$ ?



- A 0.9
- B 0.1
- C 0.2
- D 0.45
- E Cannot be calculated without knowing  $P(A|B)$

5 If  $P(A) = 0.6$ ,  $P(B) = 0.5$ , and  $P(A$  and  $B) = 0.3$ , what is  $P(A|B)$ ?



- A 0.3
- B 0.5
- C 0.6
- D 0.8
- E 1.1

6 Two events are mutually exclusive. Which statement must be true?



- A  $P(A$  and  $B) = P(A) \times P(B)$
- B  $P(A$  or  $B) = P(A) + P(B)$
- C  $P(A|B) = P(A)$
- D If one occurs, the other must also occur
- E They must have the same probability





**7** A fair coin is flipped 5 times and lands heads each time. What is the probability that the 6th flip will be heads?



- A Less than  $1/2$  because tails is 'due'
- B More than  $1/2$  because the coin is 'hot'
- C Exactly  $1/2$
- D  $1/64$
- E  $5/6$

**8** A jar contains 4 red, 3 blue, and 2 green marbles. If one marble is drawn at random, what is  $P(\text{red OR green})$ ?



- A  $4/9$
- B  $2/9$
- C  $6/9$
- D  $8/81$
- E  $1/9$

**9** In a standard deck of 52 cards, what is the probability of drawing a King OR a Heart?



- A  $17/52$
- B  $16/52$
- C  $4/52$
- D  $13/52$





E 1/52

10 A die is rolled twice. What is the probability of getting a sum of 7?



A 1/36

B 6/36

C 7/36

D 1/6

E 7/12

11 Three cards are drawn from a deck without replacement. Compared to drawing with replacement, the probability that all three are aces is:



A Higher with replacement

B Higher without replacement

C The same either way

D Cannot be compared

E Zero in both cases

12 If events A and B are independent, which must be true?



A  $P(A \text{ and } B) = 0$

B  $P(A \text{ or } B) = P(A) + P(B)$





- C  $P(A|B) = P(A)$
- D A and B cannot both occur
- E  $P(A) = P(B)$

**13** A test has 5 true/false questions. If a student guesses randomly on all questions, what is the probability of getting all 5 correct?



- A 1/5
- B 1/10
- C 1/25
- D 1/32
- E 1/2

**14** A family has 3 children. Assuming equal probability of boy or girl for each child, what is the probability that all three are the same gender?



- A 1/8
- B 1/4
- C 1/3
- D 1/2
- E 3/8





**15** Events A and B are such that  $P(A) = 0.4$ ,  $P(B) = 0.3$ , and  $P(A \text{ and } B) = 0.12$ . Are A and B independent?



- A Yes, because  $P(A \text{ and } B) = P(A) \times P(B)$
- B No, because  $P(A) \neq P(B)$
- C No, because  $P(A \text{ and } B) \neq 0$
- D Yes, because  $P(A \text{ or } B) = P(A) + P(B)$
- E Cannot be determined

**16** In how many ways can 5 different books be arranged on a shelf?



- A 5
- B 25
- C 120
- D 10
- E 32

**17** A committee of 3 people must be chosen from a group of 7. How many different committees are possible?



- A 21
- B 35
- C 210
- D 343
- E 7





**18** How many 4-digit PINs are possible if repetition of digits is allowed?



- A 40
- B 5,040
- C 10,000
- D 24
- E 1,000

**19** A president, vice president, and treasurer must be elected from 8 candidates. How many ways can these positions be filled?



- A 24
- B 56
- C 336
- D 512
- E 8

**20** How many ways can the letters in 'BOOK' be arranged?



- A 24
- B 12
- C 4
- D 6
- E 16





21 From a group of 4 men and 3 women, a committee of 2 men and 2 women is to be formed. How many such committees are possible?



- A 12
- B 18
- C 24
- D 35
- E 6

22 If  $C(n,2) = 10$ , what is  $n$ ?



- A 4
- B 5
- C 10
- D 20
- E 3

23 A disease affects 1% of a population. A test for the disease has 95% sensitivity (true positive rate) and 90% specificity (true negative rate). If a randomly selected person tests positive, what is the approximate probability they have the disease?



- A About 95%
- B About 90%
- C About 50%
- D About 9%





E About 1%

24 Sensitivity of a medical test refers to:



- A The probability of testing positive given you have the disease
- B The probability of having the disease given a positive test
- C The probability of testing negative given you don't have the disease
- D The overall accuracy of the test
- E The probability of having the disease

25 A test has 99% specificity. This means:



- A 99% of people with the disease test positive
- B 99% of people without the disease test negative
- C 99% of positive tests are true positives
- D The test is 99% accurate overall
- E 1% of people have the disease

26 Why does a positive test for a rare disease often have a low positive predictive value (PPV)?



- A Because rare diseases are harder to detect
- B Because even a small false positive rate creates many false positives when most people don't have the disease





- C Because doctors order too many tests
- D Because sensitivity is always low for rare diseases
- E Because the test equipment is unreliable

**27** In a population where 5% have a disease, a test with 90% sensitivity and 95% specificity is used. The false positive rate is:



- A 5%
- B 10%
- C 90%
- D 95%
- E 50%

**28** A doctor says: 'This test is 99% accurate, so if you test positive, there's a 99% chance you have the disease.' This reasoning is flawed because:



- A 99% accuracy is too low to be useful
- B It ignores the base rate (prevalence) of the disease
- C Doctors cannot interpret probabilities
- D The test must be wrong
- E Accuracy and positive predictive value are always equal

**29** The ages of 5 people are: 20, 22, 25, 28, 35. What is the mean age?





- A 25
- B 26
- C 27
- D 28
- E 22

**30** The ages of 5 people are: 20, 22, 25, 28, 35. What is the median age?



- A 22
- B 25
- C 26
- D 28
- E 27.5

**31** A dataset is: 3, 5, 5, 7, 8, 8, 8, 10. What is the mode?



- A 5
- B 6.75
- C 7
- D 8
- E There is no mode





**32** A dataset has values: 10, 12, 14, 16, 100. Which measure of center is most affected by the outlier?



- A Mode
- B Median
- C Mean
- D All are affected equally
- E None are affected

**33** If every value in a dataset is increased by 5, what happens to the standard deviation?



- A It increases by 5
- B It increases by 25
- C It stays the same
- D It decreases by 5
- E It becomes 5

**34** If every value in a dataset is multiplied by 3, what happens to the standard deviation?



- A It is multiplied by 3
- B It is multiplied by 9
- C It stays the same
- D It is divided by 3
- E It becomes 3





**35** A distribution is right-skewed (positively skewed). Which relationship between mean and median is typically true?



- A Mean < Median
- B Mean = Median
- C Mean > Median
- D Cannot determine without the mode
- E Mean and median are unrelated to skewness

**36** The variance of a dataset is 16. What is the standard deviation?



- A 256
- B 16
- C 8
- D 4
- E 2

**37** A dataset has mean 50 and standard deviation 10. A value of 70 has a z-score of:



- A -2
- B 0
- C 2
- D 7
- E 20





**38** A student scores at the 75th percentile on a test. This means:



- A The student got 75% of questions correct
- B 75% of test-takers scored at or below this student's score
- C The student scored 75 points
- D 25% of students scored higher
- E Both B and D

**39** Which is always true about the median of any dataset?



- A It equals the mean
- B It is a value in the dataset
- C At least half the values are at or below it
- D It is less affected by outliers than the mode
- E It is greater than the first quartile

**40** The interquartile range (IQR) measures:



- A The range of the entire dataset
- B The difference between the 75th and 25th percentiles
- C The mean of the quartiles
- D The standard deviation divided by 4
- E The number of quartiles





41 In a boxplot, a point is typically considered an outlier if it lies more than:



- A 1 standard deviation from the mean
- B  $1.5 \times \text{IQR}$  below Q1 or above Q3
- C  $2 \times$  the range from the median
- D Outside the box
- E 3 standard deviations from the mean

42 A sample has values: 2, 4, 6, 8, 10. The range is:



- A 5
- B 6
- C 8
- D 10
- E 2

43 In a normal distribution, approximately what percentage of data falls within 1 standard deviation of the mean?



- A 50%
- B 68%
- C 95%
- D 99.7%
- E 100%





**44** A normal distribution has mean 100 and standard deviation 15. What percentage of values fall between 70 and 130?



- A 68%
- B 95%
- C 99.7%
- D 50%
- E 34%

**45** In a standard normal distribution (z-distribution), the mean is:



- A 1
- B 100
- C 0
- D Variable depending on data
- E -1

**46** A value has a z-score of -1.5. This means the value is:



- A 1.5 standard deviations above the mean
- B 1.5 standard deviations below the mean
- C 1.5 times the mean
- D In the top 1.5% of the distribution
- E Negative





47 If heights of adults are normally distributed with mean 170 cm and SD 10 cm, approximately what percentage of adults are taller than 190 cm?



- A About 50%
- B About 16%
- C About 5%
- D About 2.5%
- E About 0.15%

48 Which statement about the normal distribution is FALSE?



- A It is symmetric about the mean
- B Mean = Median = Mode
- C It is bell-shaped
- D Exactly 50% of values are above the mean
- E All values must be positive

49 A game costs \$5 to play. You win \$20 with probability 0.2 and win nothing otherwise. What is the expected value (profit) per game?



- A \$4
- B \$1
- C -\$1
- D \$0





E -\$5

50 A fair six-sided die is rolled. What is the expected value of the outcome?



- A 3
- B 3.5
- C 4
- D 6
- E 21

51 The odds of winning a bet are 3 to 1 against you. What is the probability of winning?



- A  $1/3$
- B  $1/4$
- C  $3/4$
- D  $1/2$
- E  $3/1$

52 A p-value of 0.03 means:



- A There is a 3% probability the null hypothesis is true
- B There is a 3% probability the alternative hypothesis is true





- C If the null hypothesis is true, there is a 3% probability of observing results this extreme or more extreme
- D The effect size is 3%
- E 97% of the data supports the hypothesis

**53** A Type I error occurs when:



- A We fail to reject a false null hypothesis
- B We reject a true null hypothesis
- C Our sample size is too small
- D The p-value is greater than 0.05
- E We accept the null hypothesis

**54** A researcher sets  $\alpha = 0.05$ . This means:



- A They accept a 5% chance of Type II error
- B They accept a 5% chance of Type I error
- C The test has 5% power
- D 95% of samples will give significant results
- E The effect size must be at least 5%

**55** A 95% confidence interval for a population mean is (42, 58). Which interpretation is correct?





- A 95% of the population falls between 42 and 58
- B There is a 95% probability the true mean is between 42 and 58
- C If we repeated sampling many times, about 95% of such intervals would contain the true mean
- D The sample mean is 95% accurate
- E The margin of error is 95%

**56** Increasing sample size (other things equal) will make a confidence interval:



- A Wider
- B Narrower
- C More likely to be wrong
- D Have a lower confidence level
- E Unchanged

**57** A study finds  $p = 0.08$  with  $\alpha = 0.05$ . The correct conclusion is:



- A Reject the null hypothesis
- B Accept the null hypothesis
- C Fail to reject the null hypothesis
- D The study proves the null is true
- E The study has no conclusion





**58** Ice cream sales and drowning deaths are positively correlated. This most likely means:



- A Ice cream causes drowning
- B Drowning causes increased ice cream sales
- C A third variable (like hot weather) affects both
- D The data must be wrong
- E There is no real relationship

**59** A correlation coefficient of  $r = -0.9$  indicates:



- A A weak positive relationship
- B A strong positive relationship
- C A weak negative relationship
- D A strong negative relationship
- E No relationship

**60** A study finds that countries with more Nobel laureates also have higher chocolate consumption. To conclude chocolate causes intelligence:



- A Is valid because the correlation is positive
- B Requires a randomized controlled experiment
- C Is impossible to determine statistically
- D Just requires a larger sample
- E Is valid if the correlation is above 0.5





**61 Simpson's paradox refers to:**



- A** A trend that appears in subgroups but disappears or reverses when groups are combined
- B** The paradox of infinite sample sizes
- C** When mean equals median
- D** When two variables have zero correlation
- E** When p-values are exactly 0.05

**62 Hospital A has a higher overall death rate than Hospital B. However, for each disease category, Hospital A has a LOWER death rate. This is possible because:**



- A** The data must contain an error
- B** Hospital A treats a higher proportion of severe cases
- C** Sample sizes are too small
- D** Death rates cannot be compared between hospitals
- E** The diseases are not comparable

**63 In a double-blind experiment:**



- A** Neither participants nor researchers know who receives treatment
- B** Participants are randomly assigned to groups
- C** Two groups receive the same treatment
- D** The experiment is repeated twice
- E** Results are analyzed by two statisticians





**64** Random assignment in an experiment helps to:



- A** Ensure the sample represents the population
- B** Balance potential confounding variables across groups
- C** Increase the sample size
- D** Eliminate the need for a control group
- E** Make the experiment double-blind

**65** A website asks visitors to complete a survey about product satisfaction. The results may be biased because:



- A** Only those with strong opinions (positive or negative) may respond
- B** The sample is too large
- C** Websites cannot collect valid data
- D** Product satisfaction cannot be measured
- E** The survey is double-blind

**66** A study uses stratified random sampling. This means:



- A** The population is divided into subgroups, and random samples are taken from each
- B** Every member of the population has an equal chance of selection
- C** Participants are assigned to treatment groups randomly
- D** The sample is taken in layers over time
- E** Only the top stratum of the population is sampled





**67** A company surveys only its current customers about a proposed new product. This sample may not represent:



- A Potential new customers who might be interested
- B The company's current customers
- C The company's employees
- D The researchers conducting the survey
- E The product itself

**68** The law of large numbers states that:



- A Large samples always have no variance
- B As sample size increases, the sample mean approaches the population mean
- C Large numbers are harder to work with
- D Every sample will equal the population mean
- E Probability increases with more trials

**69** A basketball player scores 40 points (well above average) in one game. In the next game, they are most likely to:



- A Score even higher due to momentum
- B Score closer to their season average
- C Score exactly 40 again
- D Score zero points





- E Score exactly their season average

**70** Students who score lowest on a pretest show the most improvement on a posttest. A teacher concludes their intervention was most effective for weak students. This conclusion:



- A Is definitely correct
- B May be confounded by regression to the mean
- C Is wrong because weak students cannot improve
- D Requires a larger sample to confirm
- E Is impossible without a control group, but unrelated to regression

**71** A population parameter is:



- A A characteristic of a sample
- B A characteristic of an entire population
- C Always known exactly
- D Always estimated from a sample
- E The same as a sample statistic

**72** The absolute risk reduction is 2% (from 5% to 3%). The relative risk reduction is:



- A 2%





- B 40%
- C 60%
- D 3%
- E 8%

**73** Which of the following correlations represents the strongest relationship?



- A  $r = 0.6$
- B  $r = -0.8$
- C  $r = 0.5$
- D  $r = -0.3$
- E  $r = 0$

**74** In a normal distribution, what percentage of data lies above the mean?



- A 68%
- B 95%
- C 50%
- D 34%
- E It depends on the standard deviation





**75** A researcher finds a correlation of  $r = 0.7$  between study hours and exam scores. What percentage of variation in exam scores is explained by study hours?

- A 7%
- B 30%
- C 49%
- D 70%
- E 100%







#	Ans	Answer Text
	B	
2	B	75%
	B	
4	C	0.2
	C	
6	B	$P(A \text{ or } B) = P(A) + P(B)$
	C	
8	C	6/9
	C	
10	B	6/36
	B	
12	A	
	C	$P(A B) = P(A)$
	D	
14	B	1/4
	B	
16	A	
	C	120
	B	
18	C	10,000
	C	
20	B	12
	B	
22	B	5
	B	
24	D	
	A	The probability of testing positive given you have the disease
	B	
26	B	Because even a small false positive rate creates many false positives wh...
	B	
	A	
28	B	It ignores the base rate (prevalence) of the disease
	B	
30	B	25
	B	
	D	
32	C	Mean
	C	
34	A	It is multiplied by 3
	A	
	C	
36	D	4
	D	
	C	
38	B	75% of test-takers scored at or below this student's score
	B	



