EnterMedSchool.org

Statistics and Probability

Study Guide — 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 — Study Guide with Answers

Free & Open-Source

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

▶ **Explanation:** By the complement rule, P(not A) = 1 - P(A) = 1 - 0.3 = 0.7. The sum of an event and its complement must equal 1.

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%

▶ **Explanation:** Using inclusion-exclusion: P(F or B) = P(F) + P(B) - P(F and B) = 60% + 40% - 25% = 75%. We subtract the intersection to avoid double-counting.

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 × 5/8

D 10/56

E 1/2

▶ **Explanation:** Without replacement: P(both red) = (5/8) × (4/7) = 20/56 = 5/14. After removing one red, there are 4 red left out of 7 total. Option A uses replacement (5/8 × 5/8); D is 10/56 which equals 5/28, not 5/14.

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)

▶ **Explanation:** For independent events, P(A and B) = P(A) × P(B) = 0.4 × 0.5 = 0.2. Independence means knowing B occurred doesn't change the probability of A.

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

▶ **Explanation:** Conditional probability: P(A|B) = P(A and B)/P(B) = 0.3/0.5 = 0.6. This tells us the probability of A given that B has occurred.

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

A P(A and B) = P(A) × 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

▶ **Explanation:** Mutually exclusive means P(A and B) = 0 (they cannot both occur). Therefore P(A or B) = P(A) + P(B) - 0 = P(A) + P(B). Option A describes independence, not mutual exclusivity.

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

▶ **Explanation:** Each flip is independent. Previous outcomes don't affect future ones. The probability remains 1/2. Thinking otherwise is the gambler's fallacy. Option D (1/64) would be the probability of 6 heads in a row before flipping.

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

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

▶ **Explanation:** Red and green are mutually exclusive (can't draw both at once), so P(red or green) = P(red) + P(green) = 4/9 + 2/9 = 6/9 = 2/3. Option D incorrectly multiplies the probabilities.

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

▶ **Explanation:** Using inclusion-exclusion: P(King or Heart) = P(King) + P(Heart) - P(King and Heart) = 4/52 + 13/52 - 1/52 = 16/52. We subtract the King of Hearts to avoid counting it twice.

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

▶ **Explanation:** There are 36 equally likely outcomes. Pairs summing to 7: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) = 6 outcomes. P = 6/36 = 1/6. Note that D is also 1/6, which is correct - B and D are equivalent.

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

▶ **Explanation:** With replacement: $(4/52)^3 = 64/140608$. Without replacement: $(4/52)(3/51)(2/50) = 24/132600$. The with-replacement probability is higher because removing aces reduces future chances.

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

A P(A and B) = 0

B P(A or B) = P(A) + P(B)

C **P(A|B) = P(A)** ✓

D A and B cannot both occur

E $P(A) = P(B)$

▶ **Explanation:** Independence means knowing B occurred doesn't change the probability of A: $P(A|B) = P(A)$. Options A and D describe mutually exclusive events. Option B is true only for mutually exclusive events.

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

▶ **Explanation:** Each question has $P = 1/2$ of being correct. For all 5: $(1/2)^5 = 1/32$. This assumes independent guessing.

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

▶ **Explanation:** P(all boys) = $(1/2)^3$ = 1/8. P(all girls) = 1/8. P(all same) = 1/8 + 1/8 = 2/8 = 1/4. These are mutually exclusive outcomes.

15 Events A and B are such that P(A) = 0.4, P(B) = 0.3, and P(A and B) = 0.12. Are A and B independent?

- A **Yes, because P(A and B) = P(A) × P(B)** ✓
- B No, because P(A)   P(B)
- C No, because P(A and B)   0
- D Yes, because P(A or B) = P(A) + P(B)
- E Cannot be determined

▶ **Explanation:** Events are independent if P(A and B) = P(A) × P(B). Here: 0.4 × 0.3 = 0.12 = P(A and B). So yes, they are independent. B confuses independence with equal probability; C confuses with mutual exclusivity.

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

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

▶ **Explanation:** This is a permutation of 5 distinct objects: 5! = 5 × 4 × 3 × 2 × 1 = 120. Order matters when arranging on a shelf.

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

▶ **Explanation:** Order doesn't matter for a committee, so we use combinations: C(7,3) = 7!/(3!×4!) = (7×6×5)/(3×2×1) = 35. Option C (210) is the permutation P(7,3).

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

▶ **Explanation:** Each of 4 positions can be any of 10 digits (0-9): $10^4 = 10{,}000$. Option B (5,040) would be without repetition: 10×9×8×7.

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

▶ **Explanation:** Order matters (different positions), so P(8,3) = 8×7×6 = 336. Option B (56) is C(8,3), which ignores that positions are distinct.

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

A 24

B **12** ✓

C 4

D 6

E 16

▶ **Explanation:** BOOK has 4 letters with O repeated twice. The formula is 4!/2! = 24/2 = 12. We divide by 2! because swapping the two O's doesn't create a new arrangement.

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

▶ **Explanation:** Choose 2 men from 4: C(4,2) = 6. Choose 2 women from 3: C(3,2) = 3. Total: 6 × 3 = 18. We multiply because these are independent choices.

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

A 4

B **5** ✓

C 10

D 20

E 3

▶ **Explanation:** C(n,2) = n(n-1)/2 = 10, so n(n-1) = 20. Testing: 5×4 = 20. Therefore n = 5. This is the formula for the number of handshakes among n people.

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%

▶ **Explanation:** Using Bayes' theorem: Of 10,000 people, 100 have disease (1%), 9,900 don't. True positives: 0.95×100 = 95. False positives: 0.10×9,900 = 990. P(disease|positive) = 95/(95+990) 8.7% 9%. The low base rate dramatically affects the result.

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

▶ **Explanation:** Sensitivity = P(positive test | disease) = true positive rate. Option B is the positive predictive value (PPV). Option C is specificity.

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

▶ **Explanation:** Specificity = P(negative test | no disease) = true negative rate. High specificity means few false positives. Option A describes sensitivity; C describes PPV.

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

▶ **Explanation:** This is the base rate fallacy. When disease prevalence is low, the large healthy population generates many false positives that outnumber true positives, reducing PPV.

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%

▶ **Explanation:** False positive rate = 1 - specificity = 1 - 0.95 = 0.05 = 5%. This is the probability of testing positive when you don't have the disease.

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

▶ **Explanation:** This is the base rate fallacy. PPV depends on prevalence: even with high accuracy, if the disease is rare, most positives may be false positives. P(disease|positive)   P(positive|disease).

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

▶ **Explanation:** Mean $= (20 + 22 + 25 + 28 + 35)/5 = 130/5 = 26$. The mean is the arithmetic average.

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

▶ **Explanation:** When data is ordered, the median is the middle value. With 5 values, the median is the 3rd value: 25.

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

▶ **Explanation:** The mode is the most frequent value. Here, 8 appears 3 times, more than any other value.

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

▶ **Explanation:** The mean is sensitive to extreme values: $(10+12+14+16+100)/5 = 30.4$. The median (14) is robust to outliers. Mode (no value repeats) is unaffected.

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

▶ **Explanation:** Adding a constant shifts all values equally, so the spread (distances from mean) doesn't change. The mean increases by 5, but the standard deviation remains unchanged.

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

▶ **Explanation:** Multiplying by a constant k multiplies the standard deviation by $|k|$. The variance would be multiplied by $k^2 = 9$, but SD $= \sqrt{}$variance is multiplied by 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

▶ **Explanation:** In a right-skewed distribution, the tail extends to the right (high values). These high values pull the mean up more than the median. So Mean > Median.

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

A 256

B 16

C 8

D **4** ✓

E 2

▶ **Explanation:** Standard deviation is the square root of variance: SD $= \sqrt{16} = 4$. The variance measures spread in squared units; SD converts back to original units.

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

▶ **Explanation:** z = (x - mean)/SD = (70 - 50)/10 = 20/10 = 2. This means 70 is 2 standard deviations above the mean.

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

▶ **Explanation:** The 75th percentile means 75% scored at or below this level. While D sounds similar, it says 'higher' which excludes ties. The precise definition is B.

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

▶ **Explanation:** By definition, the median splits the data so at least 50% is at or below. B is false (for even n, median may be average of two middle values). E fails if all values are equal.

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

▶ **Explanation:** IQR = Q3 - Q1 = 75th percentile - 25th percentile. It measures the spread of the middle 50% of data and is robust to outliers.

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 × IQR below Q1 or above Q3** ✓

C 2 × the range from the median

D Outside the box

E 3 standard deviations from the mean

▶ **Explanation:** The standard boxplot rule: outliers are points below Q1 - 1.5×IQR or above Q3 + 1.5×IQR. This is not based on standard deviations.

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

A 5

B 6

C **8** ✓

D 10

E 2

▶ **Explanation:** Range = maximum - minimum = 10 - 2 = 8. It's the simplest measure of spread but is sensitive to outliers.

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%

▶ **Explanation:** The 68-95-99.7 rule: about 68% of data falls within 1 SD, 95% within 2 SD, and 99.7% within 3 SD of the mean.

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%

▶ **Explanation:** 70 = 100 - 2(15) and 130 = 100 + 2(15), so this range is ±2 SD from the mean. By the 68-95-99.7 rule, about 95% of values fall here.

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

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

▶ **Explanation:** The standard normal distribution has mean 0 and standard deviation 1. Any normal distribution can be converted to this by standardizing: z = (x -)/ .

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

▶ **Explanation:** A negative z-score indicates the value is below the mean. z = -1.5 means 1.5 standard deviations below. The original value may or may not be 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%

▶ **Explanation:** 190 cm is 2 SDs above the mean ($z = 2$). By the 68-95-99.7 rule, 95% is within 2 SD, so 5% is outside. Half of that (2.5%) is above 190 cm.

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** ✓

▶ **Explanation:** A normal distribution can have any mean, including negative values. The tails extend infinitely in both directions. All other statements are true properties of normal distributions.

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

▶ **Explanation:** Expected winnings = 0.2(\$20) + 0.8(\$0) = \$4. Expected profit = Expected winnings - cost = \$4 - \$5 = -\$1. On average, you lose \$1 per game.

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

▶ **Explanation:** E(X) = (1+2+3+4+5+6)/6 = 21/6 = 3.5. Each outcome has probability 1/6. Note: you can never actually roll 3.5; it's the long-run average.

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

▶ **Explanation:** Odds of 3 to 1 against means 3 losses for every 1 win. Probability = wins/(wins + losses) = 1/(1+3) = 1/4 = 0.25.

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

▶ **Explanation:** A p-value is the probability of getting results as extreme as observed, assuming the null is true. It's NOT the probability that the null is true (a common misconception).

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

▶ **Explanation:** Type I error is a 'false positive': rejecting H0 when it's actually true. Type II error (option A) is 'false negative': failing to reject H0 when it's false.

54 **A researcher sets $= 0.05$. 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%

▶ **Explanation:** The significance level is the maximum acceptable probability of Type I error (rejecting a true null). Setting = 0.05 means accepting up to 5% chance of false positive.

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%

▶ **Explanation:** The frequentist interpretation: if we repeated this procedure many times, 95% of the resulting intervals would capture the true parameter. The true mean either is or isn't in this specific interval.

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

▶ **Explanation:** Larger samples provide more information, reducing the margin of error. The interval becomes narrower (more precise) while maintaining the same confidence level.

57 A study finds $p = 0.08$ with $= 0.05$. 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

▶ **Explanation:** Since $p = 0.08 > \;= 0.05$, we fail to reject H0. We never 'accept' H0 (B); we simply lack sufficient evidence against it. D is too strong - we can't prove the null.

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

▶ **Explanation:** This is a classic example of confounding. Hot weather (a lurking variable) increases both ice cream consumption and swimming activity (hence drownings). Correlation does not imply causation.

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

▶ **Explanation:** The sign indicates direction (negative = inverse relationship). The magnitude (0.9) indicates strength (close to -1 is strong). r = -0.9 means as one variable increases, the other strongly tends to decrease.

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

▶ **Explanation:** Observational correlations cannot establish causation due to potential confounders (e.g., wealth affects both). Only randomized experiments can control for confounders and demonstrate causality.

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

▶ **Explanation:** Simpson's paradox: a trend present in separate groups can reverse when groups are combined, due to a lurking variable. It highlights the importance of considering confounders.

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

▶ **Explanation:** This is Simpson's paradox. If Hospital A sees more severe cases (higher proportion), its overall rate can be higher even though it performs better within each category. Case mix is a confounding variable.

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

▶ **Explanation:** Double-blind means both participants and those administering/measuring don't know the assignment. This prevents placebo effect in participants and bias in researchers.

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

▶ **Explanation:** Random assignment distributes both known and unknown confounders approximately equally across treatment groups, allowing causal inference. Option A describes random sampling, not random assignment.

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

▶ **Explanation:** This is voluntary response bias (self-selection). People who feel strongly are more likely to respond, skewing results. Those with moderate opinions are underrepresented.

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

▶ **Explanation:** Stratified sampling divides the population into strata (subgroups like age ranges) and samples from each. This ensures representation of all subgroups. Option B describes simple random sampling.

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

▶ **Explanation:** Current customers may differ systematically from potential new customers. This sample doesn't capture views of people unfamiliar with the company who might still buy the new product.

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

▶ **Explanation:** The law of large numbers: as n→∞, the sample average converges to the expected value. This is why casinos profit long-term and why we trust large surveys more.

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

▶ **Explanation:** This is regression to the mean. Extreme performances are often partly due to chance factors that are unlikely to repeat. The next performance tends to be closer to the player's true 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

▶ **Explanation:** Those who score lowest may have had bad luck on the pretest. On any retest (with or without intervention), they would likely score closer to their true ability due to regression to the mean.

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

▶ **Explanation:** Parameters describe populations (e.g., population mean). Statistics describe samples (e.g., sample mean $\bar{x}$). We typically estimate unknown parameters using sample statistics.

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%

▶ **Explanation:** Relative risk reduction = (5% - 3%)/5% = 2%/5% = 40%. Relative measures can seem more impressive than absolute measures. A 40% reduction sounds bigger than a 2% reduction.

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

▶ **Explanation:** Strength of correlation is determined by the absolute value: $|-0.8| = 0.8$ is larger than $|0.6| = 0.6$. The sign only indicates direction, not strength.

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

▶ **Explanation:** The normal distribution is symmetric about the mean. Exactly half (50%) lies above the mean and half below, regardless of the mean or standard deviation values.

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%

▶ **Explanation:** The coefficient of determination $r^2 = (0.7)^2 = 0.49 = 49\%$. This represents the proportion of variance in the dependent variable explained by the independent variable.