Researchers conducted a randomized controlled trial to assess the effectiveness of a new drug to reduce severe cutaneous reactions in patients with rheumatoid arthritis who are treated with adalimumab. Of 150 adults with rheumatoid arthritis who are treated with adalimumab, 75 received the new drug in addition to adalimumab, and 75 received a placebo in addition to adalimumab. Results show that 6 patients in the new drug group developed severe cutaneous reactions, compared to 9 in the placebo group. Which of the following represents the relative risk reduction for severe cutaneous reactions among patients in the new drug group?

- A. 0.08 (9%)
- B. 0.10 (2%)
- C. 0.12 (4%)
- ✓ D. 0.33 (61%)
- E. 0.67 (22%)

Omitted
Correct answer
D

61%
Answered correctly

08 secs
Time Spent

2023
Version

Explanation

Relative risk reduction (RRR) measures how much a given **treatment reduces the risk of an unfavorable outcome** relative to the control group. RRR may be defined as the difference in risk of the unfavorable outcome in the treatment group (eg, new drug) compared to the control group (eg, placebo):

$$RRR = (Risk_{control} - Risk_{treatment}) / Risk_{control}$$

$$RRR = (Risk_{control} - Risk_{treatment}) / Risk_{control}$$

In this example, $Risk_{control}$ (ie, risk of severe cutaneous reactions in the placebo group) is **9/75 = 0.12**, and $Risk_{treatment}$ (ie, risk of severe cutaneous reactions in the new drug group) is **6/75 = 0.08**. Therefore, RRR for severe cutaneous reactions among patients in the new drug group is:

$$\mathbf{RRR = (0.12 - 0.08) / 0.12 = 0.33}$$

Alternately, RRR may be calculated using relative risk (RR). When using RR, the formula for RRR is:

$$RRR = 1 - RR$$

For this example, RR is the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group, calculated as follows:

$$RR = (Risk_{treatment} / Risk_{control}) = (0.08/0.12) = 0.67$$

Therefore, RRR is:

$$RRR = 1 - 0.67 = 0.33$$

(Choice A) 0.08 represents the risk of severe cutaneous reactions in the new drug group (6/75 = 0.08, or 8%), not RRR.

(Choice B) 0.10 represents the risk of severe cutaneous reactions in the entire study (ie, both groups: [6 + 9] / [75 + 75] = 0.10, or 10%).

(Choice C) 0.12 represents the risk of severe cutaneous reactions in the placebo group (9/75 = 0.12, or 12%).

(Choice E) 0.67 represents the RR, the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group (0.08/0.12 = 0.67, or 67%).

Educational objective:

$$RR = (Risk_{treatment} / Risk_{control}) = (0.08/0.12) = 0.67$$

Therefore, RRR is:

$$RRR = 1 - 0.67 = 0.33$$

(Choice A) 0.08 represents the risk of severe cutaneous reactions in the new drug group (6/75 = 0.08, or 8%), not RRR.

(Choice B) 0.10 represents the risk of severe cutaneous reactions in the entire study (ie, both groups: [6 + 9] / [75 + 75] = 0.10, or 10%).

(Choice C) 0.12 represents the risk of severe cutaneous reactions in the placebo group (9/75 = 0.12, or 12%).

(Choice E) 0.67 represents the RR, the ratio of the risk of severe cutaneous reactions in the new drug group to the risk of severe cutaneous reactions in the placebo group (0.08/0.12 = 0.67, or 67%).

Educational objective:

Relative risk reduction (RRR) measures how much a treatment reduces the risk of an unfavorable outcome. Two methods to calculate it are:

- $RRR = (Risk_{control} - Risk_{treatment}) / Risk_{control}$
- RRR = 1 − relative risk

References

- Measures of effect in epidemiological research.
- Measures of effect: relative risks, odds ratios, risk difference, and 'number needed to treat.'

Biostatistics	Biostatistics & Epidemiology	Measures of association
Subject	System	Topic

A study assessed the association between a new vaccine and traveler's diarrhea (TD). Researchers selected a random sample of people who intended to travel to regions where they were at increased risk for TD and who had received the new vaccine and another independent random sample of people who intended to travel to the same regions and who had not received the new vaccine. These 2 samples of travelers were assessed for the occurrence of TD during the trip and for 7 days after returning home. Which of the following measures of association are the investigators most likely to report?

- A. Incidence (10%)
- B. Median survival time (0%)
- C. Odds ratio (20%)
- D. Prevalence (1%)
- ✓ E. Relative risk (66%)

Omitted
Correct answer
E

66%
Answered correctly

03 secs
Time Spent

2023
Version

Explanation

This case describes a cohort study in which participants are initially selected based on their **exposure status to a risk modifier**: exposed (eg, vaccinated) or nonexposed (eg, not vaccinated). Once participants are categorized based on exposure status, the occurrence (ie, incidence) of the outcome/disease of interest (ie, traveler's diarrhea [TD]) within a specific period is determined in each group. The **risk of disease** is finally compared between exposed and nonexposed groups to estimate the **association** between exposure to a risk

This case describes a cohort study in which participants are initially selected based on their **exposure status to a risk modifier**: exposed (eg, vaccinated) or nonexposed (eg, not vaccinated). Once participants are categorized based on exposure status, the occurrence (ie, incidence) of the outcome/disease of interest (ie, traveler's diarrhea [TD]) within a specific period is determined in each group. The **risk of disease** is finally compared between exposed and nonexposed groups to estimate the **association** between exposure to a risk modifier and disease occurrence. A common measure of association in cohort studies is the **relative risk (RR)**. The RR is calculated as follows:

RR = (risk of disease in exposed) / (risk of disease in nonexposed)

The exposed group could also have been defined as the unvaccinated group and the nonexposed group as the vaccinated group; in this case, the RR would be inversed (interpretation of RR depends on what is chosen to represent the exposure).

(Choices A and D) Incidence (ie, new occurrences of a disease in a population over a period) and prevalence (ie, existing cases of a disease in a population at a point in time) are measures of disease burden. They do not directly convey information regarding the relationship between exposure and disease.

(Choice B) Median survival time is calculated in cohort studies or clinical trials and is used to compare the median time to an event of interest (eg, TD) in ≥2 groups of individuals (eg, those not vaccinated and those vaccinated). The purpose of this study is to assess the association between exposure to a risk modifier (eg, vaccination) and disease occurrence (eg, TD), not to compare the median time to develop the disease (eg, TD) between exposed and nonexposed groups.

(Choice C) Odds ratio is calculated in case-control studies, where 2 groups of individuals are initially identified as "cases" or "controls" according to an outcome (eg, disease of interest), and then the frequency of exposure to the risk factors is compared between cases and controls to estimate the association between the risk factors and the outcomes.

RR = (risk of disease in exposed) / (risk of disease in nonexposed)

The exposed group could also have been defined as the unvaccinated group and the nonexposed group as the vaccinated group; in this case, the RR would be inversed (interpretation of RR depends on what is chosen to represent the exposure).

(Choices A and D) Incidence (ie, new occurrences of a disease in a population over a period) and prevalence (ie, existing cases of a disease in a population at a point in time) are measures of disease burden. They do not directly convey information regarding the relationship between exposure and disease.

(Choice B) Median survival time is calculated in cohort studies or clinical trials and is used to compare the median time to an event of interest (eg, TD) in ≥2 groups of individuals (eg, those not vaccinated and those vaccinated). The purpose of this study is to assess the association between exposure to a risk modifier (eg, vaccination) and disease occurrence (eg, TD), not to compare the median time to develop the disease (eg, TD) between exposed and nonexposed groups.

(Choice C) Odds ratio is calculated in case-control studies, where 2 groups of individuals are initially identified as "cases" or "controls" according to an outcome (eg, disease of interest), and then the frequency of exposure to the risk factors is compared between cases and controls to estimate the association between the risk factors and the outcomes.

Educational objective:
A cohort study is used to compare incidence of disease between exposed and nonexposed individuals. A common measure of association in cohort studies is the relative risk.

Biostatistics	Biostatistics & Epidemiology	Measures of association
Subject	System	Topic

A large study of serum folate levels in a sample of women age 16-45 reveals that this parameter is normally distributed with a mean of 5.0 ng/mL and a standard deviation of 0.5 ng/mL. According to the study results, 95% of serum folate observations in these patients will lie approximately between which of the following limits?

- A. 3.5 and 6.0 ng/mL (0%)
- B. 3.5 and 6.5 ng/mL (5%)
- ✓ C. 4.0 and 6.0 ng/mL (79%)
- D. 4.0 and 5.5 ng/mL (1%)
- E. 4.5 and 5.5 ng/mL (12%)

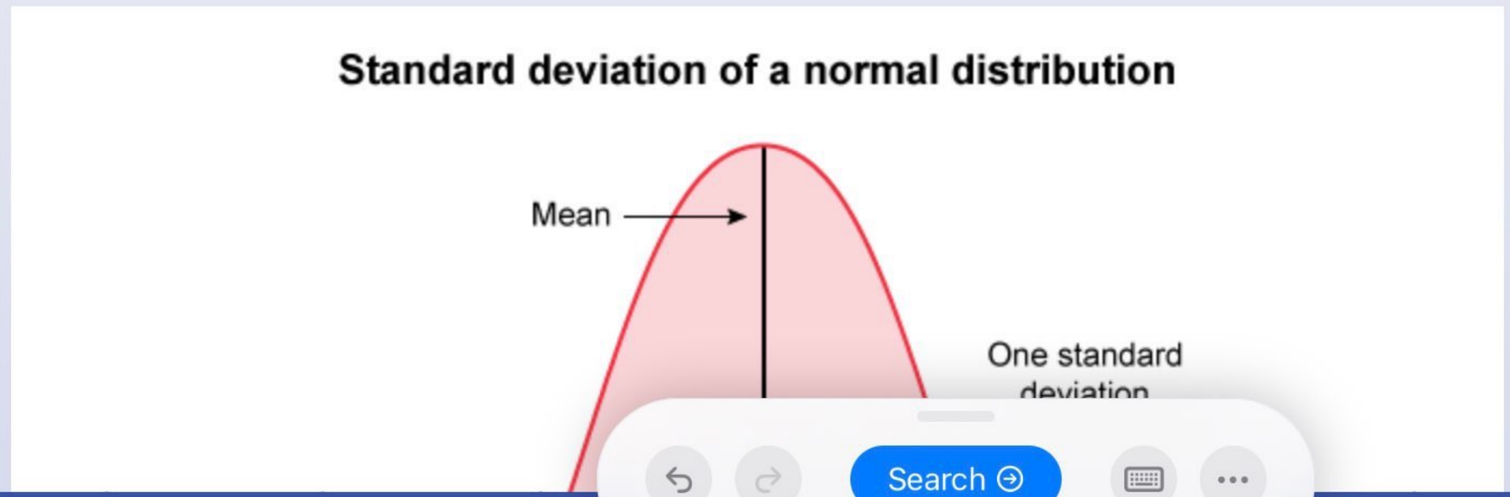
Omitted
Correct answer
C

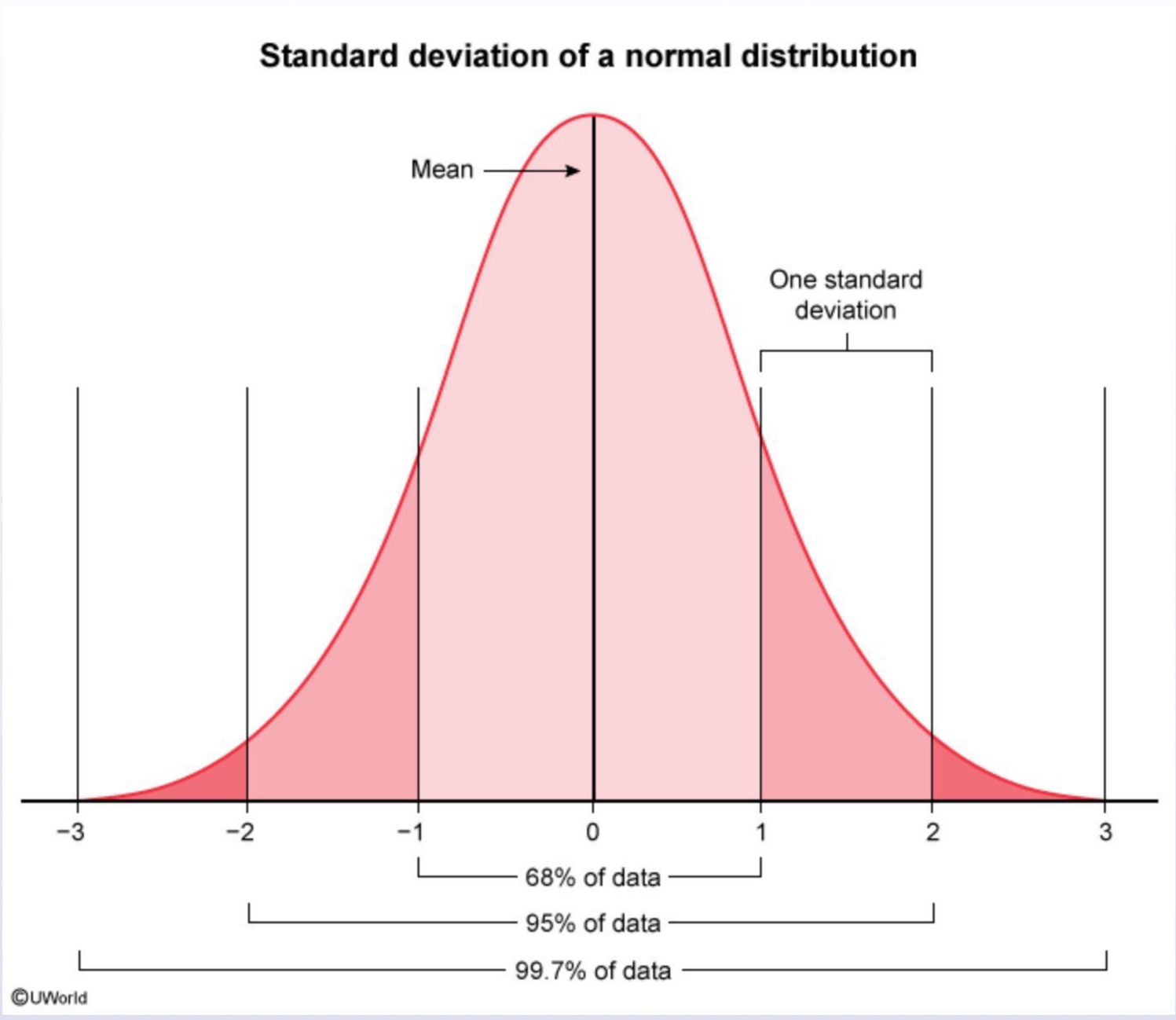
79%
Answered correctly

02 secs
Time Spent

2023
Version

Explanation





A **normal (Gaussian) distribution** is defined as a symmetrical, bell-shaped distribution curve. One of the most important attributes of the normal distribution is the **"68/95/99 rule,"** which states that 68% of all observations lie within 1 standard deviation (SD) of the mean, 95% within 2 SDs of the mean, and 99.7% within 3 SDs of the mean. The SD is a measure of dispersion or va rom one another). In this

A **normal (Gaussian) distribution** is defined as a symmetrical, bell-shaped distribution curve. One of the most important attributes of the normal distribution is the **"68/95/99 rule,"** which states that 68% of all observations lie within 1 standard deviation (SD) of the mean, 95% within 2 SDs of the mean, and 99.7% within 3 SDs of the mean. The SD is a measure of dispersion or variance (how far the measurements are from one another). In this example, the mean folate level is 5.0 ng/mL with a SD of 0.5 ng/mL. Therefore:

- 4.5-5.5 ng/mL is the range within +/- 1 SD of 5.0 ng/mL (given that 5.0 – 1*0.5 = 4.5 ng/mL and 5 + 1*0.5 = 5.5 ng/mL);
- 4.0-6.0 ng/mL is the range within +/- 2 SDs of 5.0 ng/mL (given that 5.0 – 2*0.5 = 4.0 ng/mL and 5.0 + 2*0.5 = 6.0 ng/mL)
- 3.5-6.5 ng/mL is the range within +/- 3 SDs of 5.0 ng/mL (given that 5.0 – 3*0.5 = 3.5 ng/mL and 5.0 + 3*0.5 = 6.5 ng/mL)

Based on the 68/95/99 rule, approximately 95% of observations lie within 2 SDs of the mean, so between 4.0 and 6.0 ng/mL in this case. These are helpful approximations; to be more precise, 95% of the observations lie within 1.96 SDs of the mean and 95.45% of the observations lie within 2 SDs of the mean.

Note that SD is used to describe individual observations in a dataset. A closely related measure called standard error (SE) is used to show how closely sample means are related to population means; SE = SD / $\sqrt{n}$, where n is the sample size. In this example, if SE had been given instead of SD, investigators would be 95% confident that the true mean folate level in the underlying population lies within the mean +/- 2 SE (which is also the range that would include 95% of sample means calculated from repeated samples of the same size taken from that population).

(Choices A and D) These limits are asymmetric and are therefore inconsistent with a normal distribution curve around the mean.

– 6.5 ng/mL)

Based on the 68/95/99 rule, approximately 95% of observations lie within 2 SDs of the mean, so between 4.0 and 6.0 ng/mL in this case. These are helpful approximations; to be more precise, 95% of the observations lie within 1.96 SDs of the mean and 95.45% of the observations lie within 2 SDs of the mean.

Note that SD is used to describe individual observations in a dataset. A closely related measure called standard error (SE) is used to show how closely sample means are related to population means; SE = SD / $\sqrt{n}$, where n is the sample size. In this example, if SE had been given instead of SD, investigators would be 95% confident that the true mean folate level in the underlying population lies within the mean +/- 2 SE (which is also the range that would include 95% of sample means calculated from repeated samples of the same size taken from that population).

(Choices A and D) These limits are asymmetric and are therefore inconsistent with a normal distribution curve around the mean.

(Choice B) Approximately 99.7% of all observations lie within 3 SDs of the mean (3.5-6.5 ng/mL).

(Choice E) Approximately 68% of all observations lie within 1 SD of the mean (4.5-5.5 ng/mL).

Educational objective:
In a normal (bell-shaped) distribution curve, 68% of observations lie within 1 standard deviation (SD) of the mean, 95% of observations lie within 2 SDs of the mean, and 99.7% of observations lie within 3 SDs of the mean.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

A new portable cholesterol-measuring device is being developed for use in medically underserved communities. During clinical trials of an early prototype, a patient's cholesterol level is found to be 200 mg/dL, 201 mg/dL, and 200 mg/dL on 3 separate measurements of the same blood sample. Using the gold standard measurement method, the same sample is found to have a cholesterol level of 260 mg/dL. Which of the following descriptions best characterizes the new cholesterol-measuring device?

- A. High accuracy; high precision (0%)
- B. High accuracy; low precision (7%)
- C. High sensitivity; low specificity (0%)
- ✓ D. Low accuracy; high precision (89%)
- E. Low accuracy; low precision (1%)
- F. Low sensitivity; high specificity (0%)

Omitted
Correct answer
D

89%
Answered correctly

01 sec
Time Spent

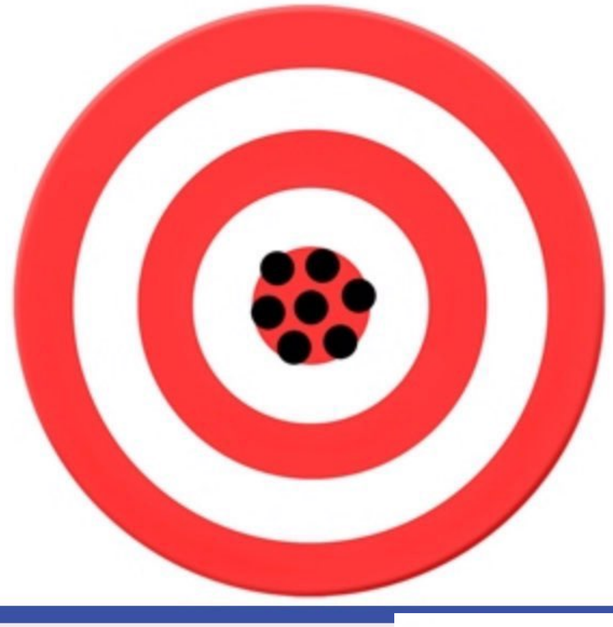
2023
Version

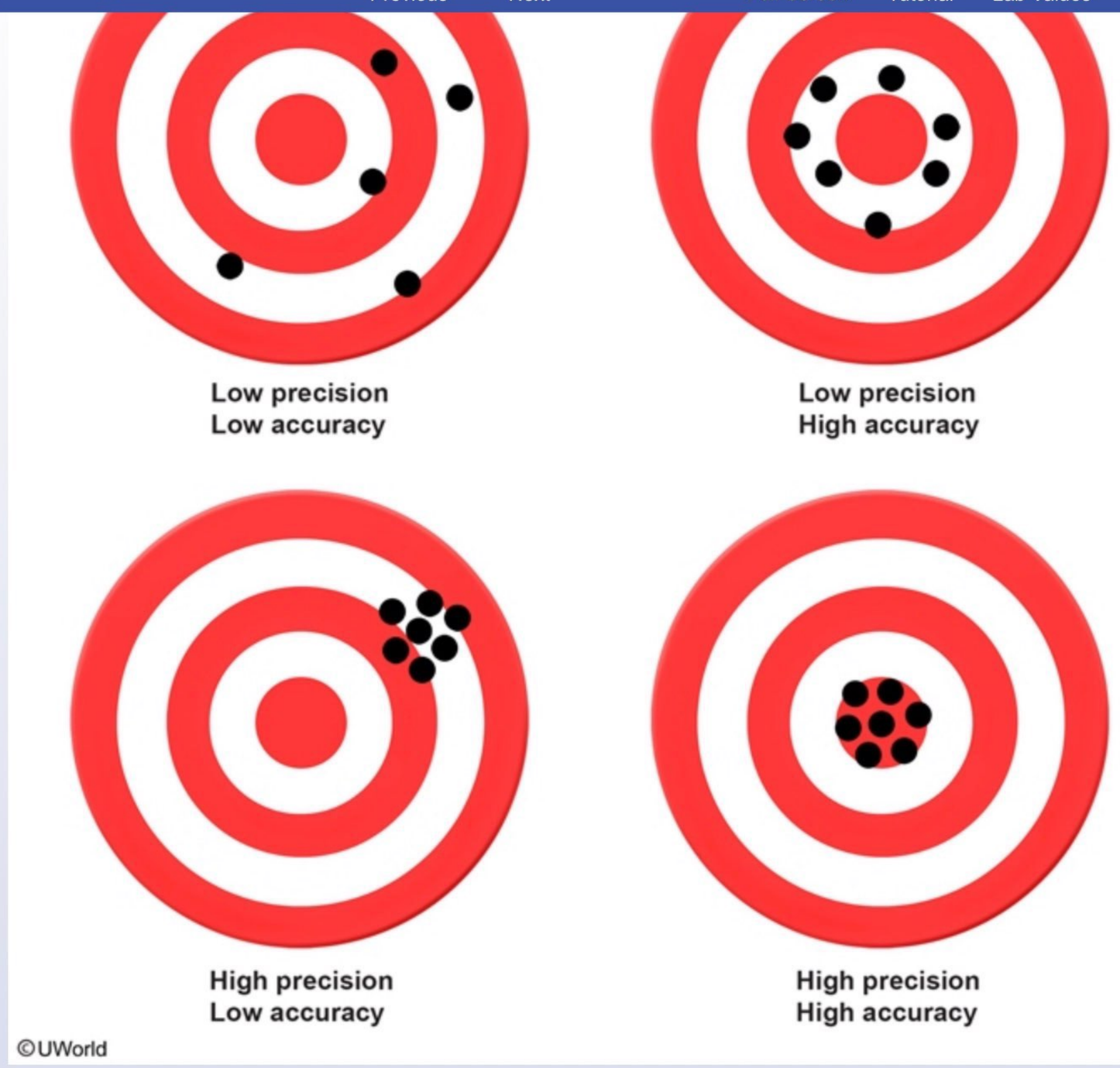
Explanation

Precision & accuracy

Low precision
Low accuracy

Low precision
High accuracy





- **Precision** (reliability) is the ability of a test to reproduce identical or similar results with repeated measurements (bottom half of figure)

- **Precision** (reliability) is the ability of a test to reproduce identical or similar results with repeated measurements (bottom half of figure).
- **Accuracy** (validity) is the ability of a test to measure what it is supposed to measure. For a new test to be accurate, its results should be equivalent to the results obtained with a gold standard (eg, best conventional test available) on the same individual (right half of figure).

In this case, repeated measurements of the same blood sample yielded **nearly identical values**; therefore, the new device has **high precision**. However, none of the measurements with the new device were equivalent to the results obtained with the gold standard (ie, the measurements are **incorrect**), meaning the new device has **low accuracy**.

(Choices C and F) Sensitivity is defined as the percentage of diseased individuals who are correctly identified (ie, proportion of subjects who have a positive test among all those with disease), whereas specificity is defined as the percentage of healthy individuals who are correctly identified (ie, proportion of subjects who have a negative test among all those without disease). Poor accuracy or precision can lead to low test sensitivity and specificity. However, no information is provided to determine the sensitivity or specificity of the new device compared to the gold standard test.

Educational objective:
A precise tool is one that consistently provides very similar or the same value when measuring a fixed quantity. An accurate tool is one that provides a measurement identical or similar to the actual value (as reflected in a gold standard measurement).

Biostatistics	Biostatistics & Epidemiology	Accuracy and precision
Subject	System	Topic

A 65-year-old man with a history of congestive heart failure is hospitalized with chest pain and hypotension requiring admission to the cardiac care unit. An intra-arterial line is placed for direct blood pressure monitoring. Consecutive readings of his intra-arterial blood pressure are 75, 110, 80, 90, 75, and 110 mm Hg. Which of the following represents the median of these blood pressure readings?

- A. 80 mm Hg (2%)
- ✓ B. 85 mm Hg (88%)
- C. 90 mm Hg (7%)
- D. 100 mm Hg (0%)
- E. 110 mm Hg (0%)

Omitted
Correct answer
B

88%
Answered correctly

02 secs
Time Spent

2023
Version

Explanation

It is important to understand the difference between the median, mean, and mode, which are 3 measures of central tendency. The **median** of an ordered dataset is the number that separates the right half of the data from the left half. The dataset MUST be ordered before the median is determined. If the number of observations in the dataset is even, finding the median requires adding the middle 2 values together and dividing that sum by 2.

The ordered dataset in this case is {75, 75, **80, 90**, 110, 110}. There are 6 observations, which is an even

It is important to understand the difference between the median, mean, and mode, which are 3 measures of central tendency. The **median** of an ordered dataset is the number that separates the right half of the data from the left half. The dataset MUST be ordered before the median is determined. If the number of observations in the dataset is even, finding the median requires adding the middle 2 values together and dividing that sum by 2.

The ordered dataset in this case is {75, 75, **80, 90**, 110, 110}. There are 6 observations, which is an even number of observations. The median value splits the dataset in half; it lies between 80 and 90 (3 values on the left and 3 values on the right). The median is (80 + 90) / 2 = 85 mm Hg.

Now, assume one of the values is missing and the ordered dataset includes the following 5 observations (odd number of observations): {75, 80, **90**, 110, 110}. In this case, the median value would be 90 mm Hg, which splits the dataset in half (2 values on the left and 2 values on the right).

(Choices A and D) The values 80 and 100 mm Hg are not measures of the center of the dataset.

(Choice C) To find the **mean** of a dataset, add all of the observations and divide that sum by the number of observations. In this case, the mean is (75 + 75 + 80 + 90 + 110 + 110) / 6 = 90 mm Hg.

(Choice E) Another measure of the center of a dataset is the mode. Finding the **mode** is easy because it is the most frequent value of a dataset. In our scenario, the dataset is "bimodal" because 75 and 110 mm Hg are each listed twice.

Educational objective:
The median is the value that is located in the precise center of an ordered dataset. It separates the right half of the data from the left half.

Biostatistics	Biostatistics & Epidemiology	Central tendency (mean, median, mode, outliers)
Subject	System	Topic

An intern in the endocrinology department is asked to complete a chart review of cholesterol levels of all patients hospitalized with diabetes mellitus-related cardiovascular complications in the past 6 months. The intern's preliminary data analysis reveals that there were 400 such patients, and that serum cholesterol levels were normally distributed among the patients with a mean of 220 mg/dL and a standard deviation of 10 mg/dL. Based on these results, how many patients in this study would be expected to have serum cholesterol ≥240 mg/dL?

- A. 2 (9%)
- ✓ B. 10 (65%)
- C. 20 (18%)
- D. 64 (4%)
- E. 128 (1%)

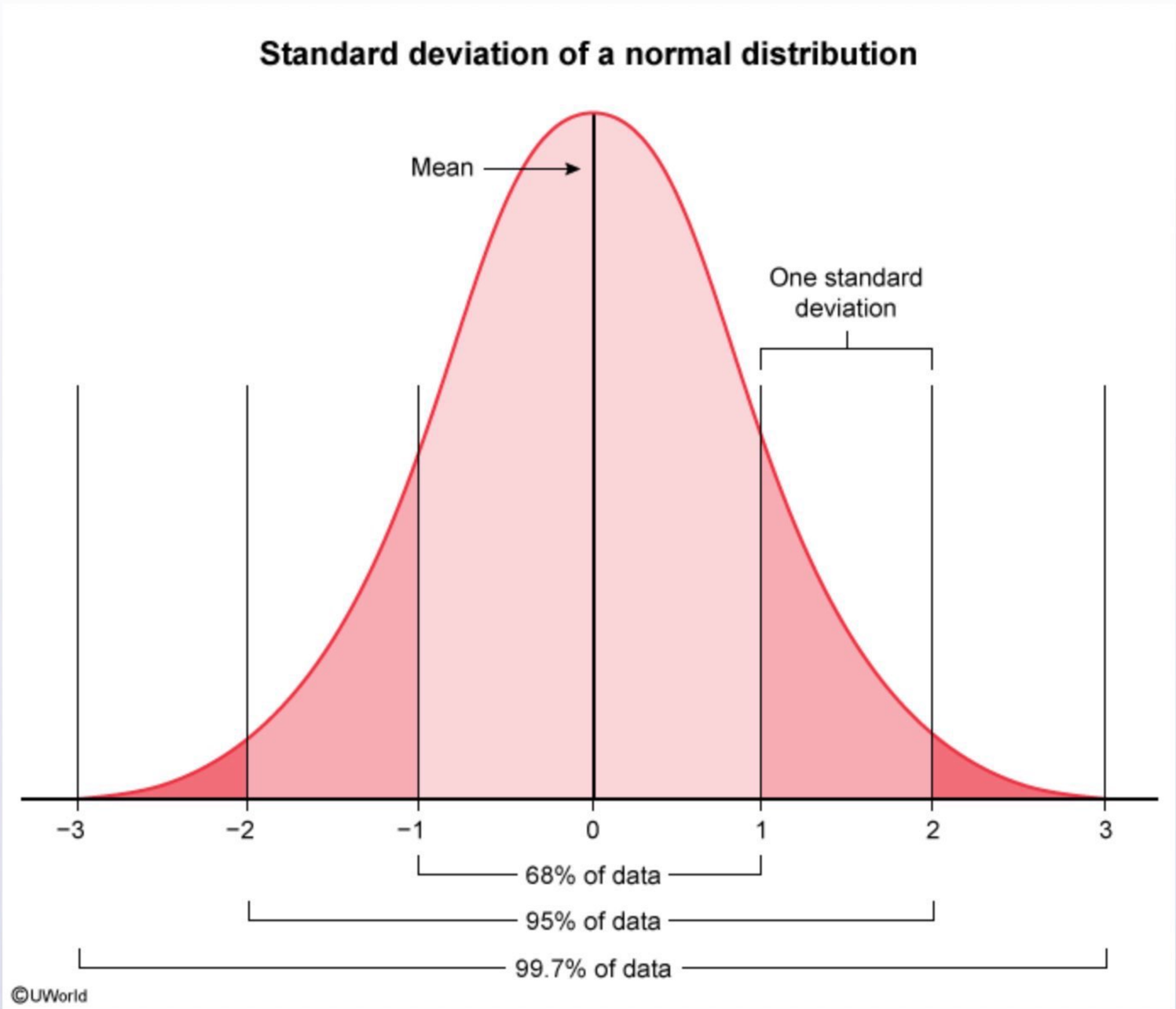
Omitted
Correct answer
B

65%
Answered correctly

02 secs
Time Spent

2023
Version

Explanation



A **normal distribution** refers to a symmetrical, bell-shaped distribution pattern with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)**

A **normal distribution** refers to a symmetrical, bell-shaped distribution pattern with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and is a measure of the **degree of dispersion** from the mean. When the SD is small, the observations tend to have minimal variation and are tightly clustered around the mean. In contrast, a large SD implies that the observations are spread over a large range. According to the properties of normal distribution, 68% of all observations lie within 1 SD of the mean, with half of those above and half of those below the mean. The remaining 32% of observations lie outside 1 SD from the mean, with half of those remaining observations above and the other half below 1 SD from the mean. In addition, **95% of all observations lie within 2 SDs** of the mean, and 99.7% of all observations lie within 3 SDs from the mean. This is commonly known as the "**68-95-99.7 rule**."

Based on the above example:

- 68% of observations lie within 1 SD: 220 mg/dL ± 10 mg/dL = 210-230 mg/dL
- 95% of observations lie within 2 SDs: 220 mg/dL ± 20 mg/dL = 200-240 mg/dL
- 99.7% of observations lie within 3 SDs: 220 mg/dL ± 30 mg/dL = 190-250 mg/dL

Because the cutoff point of ≥240 mg/dL is 2 SDs from the mean, it follows that 2.5% of the observations must lie below 200 mg/dL and **2.5% above 240 mg/dL**. Therefore, of 400 patients included in the analysis, **10 patients** (ie, 2.5%) would have a cholesterol level ≥240 mg/dL.

(Choice A) No more than 1-2 patients (0.25%-0.5% of the study population) would be expected to have cholesterol levels that lie 3 SDs above or below the mean (ie, ≤190 mg/dL or ≥250 mg/dL).

(Choice C) 20 patients (5% of the study population) would be expected to have cholesterol levels that lie outside 2 SDs from the mean. Only half of these observations would lie above 2 SDs (ie, ≥240 mg/dL) from the mean.

- 95% of observations lie within 2 SDs: 220 mg/dL ± 20 mg/dL = 200-240 mg/dL
- 99.7% of observations lie within 3 SDs: 220 mg/dL ± 30 mg/dL = 190-250 mg/dL

Because the cutoff point of ≥240 mg/dL is 2 SDs from the mean, it follows that 2.5% of the observations must lie below 200 mg/dL and **2.5% above 240 mg/dL**. Therefore, of 400 patients included in the analysis, **10 patients** (ie, 2.5%) would have a cholesterol level ≥240 mg/dL.

(Choice A) No more than 1-2 patients (0.25%-0.5% of the study population) would be expected to have cholesterol levels that lie 3 SDs above or below the mean (ie, ≤190 mg/dL or ≥250 mg/dL).

(Choice C) 20 patients (5% of the study population) would be expected to have cholesterol levels that lie outside 2 SDs from the mean. Only half of these observations would lie above 2 SDs (ie, ≥240 mg/dL) from the mean.

(Choice D) 64 patients (16% of the study population) would be expected to have cholesterol levels that lie 1 SD above the mean (ie, ≥230 mg/dL).

(Choice E) 128 patients (32% of the study population) would be expected to have cholesterol levels that lie 1 SD above or below the mean (ie, ≥230 mg/dL or ≤210 mg/dL).

Educational objective:
In a normal (bell-shaped) distribution: 68% of all values are within 1 standard deviation from the mean; 95% of all values are within 2 standard deviations from the mean; 99.7% of all values are within 3 standard deviations from the mean.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

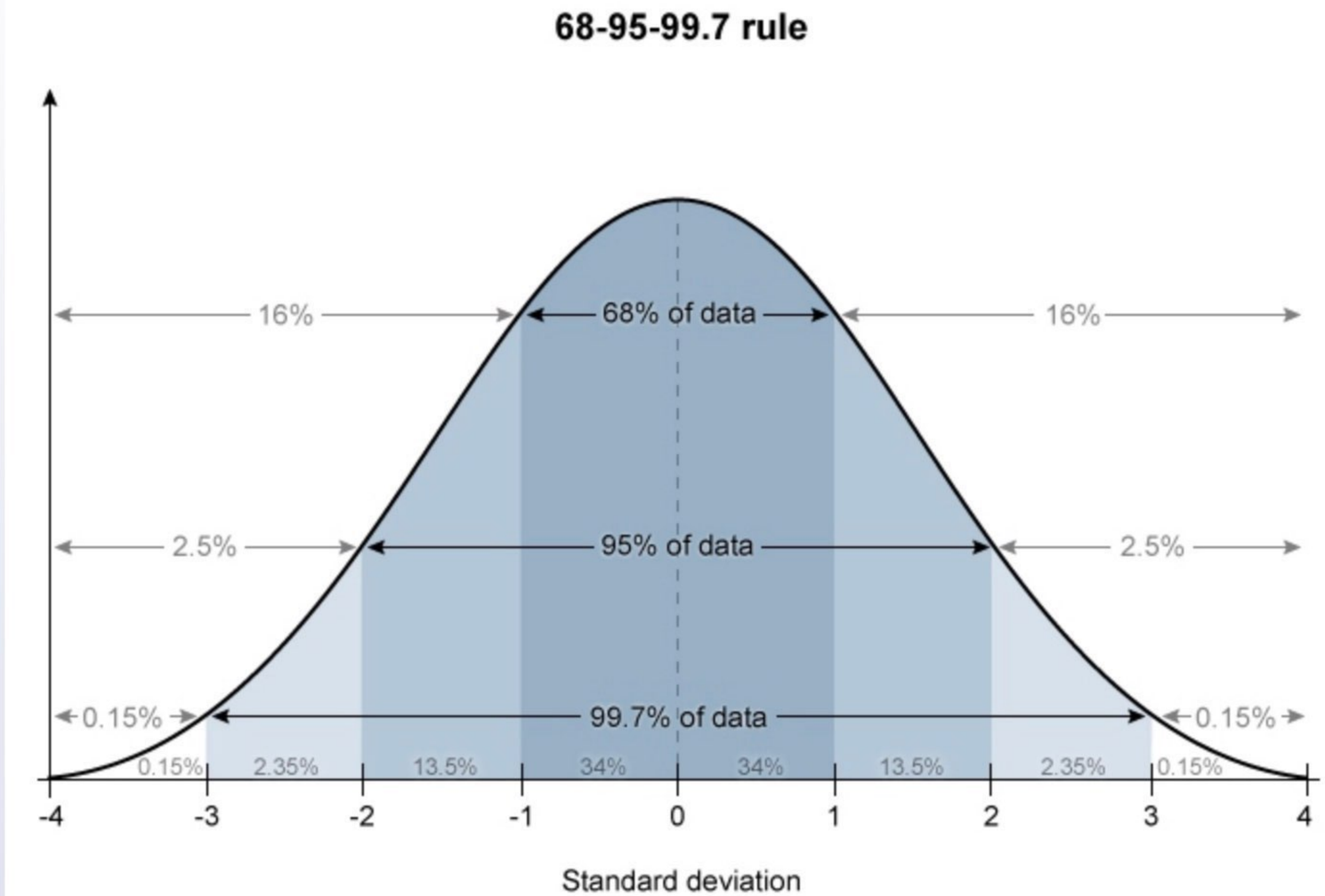
Blood pressure measurements are obtained from a sample of individuals with no known medical conditions. For systolic blood pressure (SBP), the mean measurements and associated standard deviations (SDs) are shown by age group for men and women:

	Men	Women
Age group	Mean SBP (mm Hg) ± SD	Mean SBP (mm Hg) ± SD
35-44	120 ± 20	124 ± 18
45-54	131 ± 21	137 ± 24
55-64	141 ± 19	140 ± 20

If hypertension is defined as SBP >140 mm Hg, approximately what percentage of men age 35-44 in this sample will be classified as having hypertension, assuming a normal (Gaussian) distribution?

- ✓ A. 16% (83%)
- B. 34% (8%)
- C. 50% (1%)
- D. 68% (2%)
- E. 84% (1%)
- F. 95% (2%)

Omitted



A **normal (Gaussian) distribution** refers to a symmetrical, bell-shaped distribution with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and reflects the **degree of dispersion** from the mean. According to the properties of this distribution, 68% of observations lie within 1 SD on either side of the mean, with half (ie, 68/2 = 34%) above and half (34%) below the mean. The remaining 32% (= 100% − 68%) lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above

A **normal (Gaussian) distribution** refers to a symmetrical, bell-shaped distribution with a fixed proportion of observations lying within a certain distance of the mean. This distance is called the **standard deviation (SD)** and reflects the **degree of dispersion** from the mean. According to the properties of this distribution, 68% of observations lie within 1 SD on either side of the mean, with half (ie, 68/2 = 34%) above and half (34%) below the mean. The remaining 32% (= 100% − 68%) lie outside 1 SD from the mean, with half (ie, 32/2 = 16%) above and the other half (16%) below 1 SD from the mean. In addition, 95% of all observations lie within 2 SDs of the mean, and 99.7% of all observations lie within 3 SDs of the mean. This is the **68-95-99.7 rule**.

For men age 35-44 in this sample, the mean systolic blood pressure (SBP) is 120 mm Hg with an SD of 20 mm Hg. Therefore, subjects with SBP >140 mm Hg (to fit the proposed definition of hypertension) represent observations >1 SD above the mean. Based on the 68-95-99.7 rule, 32% of observations (= 100% − 68%) lie outside 1 SD from the mean, with **half (16%) above** and half (16%) below 1 SD from the mean. Therefore, 16% of men age 35-44 have SBP >140 mm Hg.

(Choices B, C, D, and E) Based on the 68-95-99.7 rule, 34% of subjects lie either 1 SD above the mean or 1 SD below the mean. In men age 35-44:

- 34% have SBP 100-120 mm Hg, 34% have SBP 120-140 mm Hg, and, 68% have SBP 100-140 mm Hg.
- Half (50%) have SBP above the mean (ie, >120 mm Hg) and half have SBP below the mean (<120 mm Hg).
- Given that 84% = 50% + 34%, this could represent the sum of those with SBP 100-120 mm Hg (34%) and those with SBP >120 mm Hg (50%): in other words, those with SBP >100 mm Hg.

(Choice F) Based on the 68-95-99.7% rule, 95% of observations lie within 2 SDs (eg, 2 × 20 = 40 mm Hg) of the mean (eg, 120 mm Hg) (ie, SBP between 80 mm Hg [= 120 − 40] and 160 mm Hg [= 120 + 40]).

Educational objective:

Hg. Therefore, subjects with SBP >140 mm Hg (to fit the proposed definition of hypertension) represent observations >1 SD above the mean. Based on the 68-95-99.7 rule, 32% of observations (= 100% − 68%) lie outside 1 SD from the mean, with **half (16%) above** and half (16%) below 1 SD from the mean. Therefore, 16% of men age 35-44 have SBP >140 mm Hg.

(Choices B, C, D, and E) Based on the 68-95-99.7 rule, 34% of subjects lie either 1 SD above the mean or 1 SD below the mean. In men age 35-44:

- 34% have SBP 100-120 mm Hg, 34% have SBP 120-140 mm Hg, and, 68% have SBP 100-140 mm Hg.
- Half (50%) have SBP above the mean (ie, >120 mm Hg) and half have SBP below the mean (<120 mm Hg).
- Given that 84% = 50% + 34%, this could represent the sum of those with SBP 100-120 mm Hg (34%) and those with SBP >120 mm Hg (50%): in other words, those with SBP >100 mm Hg.

(Choice F) Based on the 68-95-99.7% rule, 95% of observations lie within 2 SDs (eg, 2 × 20 = 40 mm Hg) of the mean (eg, 120 mm Hg) (ie, SBP between 80 mm Hg [= 120 − 40] and 160 mm Hg [= 120 + 40]).

Educational objective:
In a normal (bell-shaped) distribution, 68% of all values are within 1 standard deviation (SD) of the mean; 95% are within 2 SD of the mean; and 99.7% are within 3 SD of the mean.

References

- Our obsession with normal values.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

A researcher selects a random sample of 100 men age 18-24 and determines their mean serum total cholesterol is 180 mg/dL with a standard deviation of 40 mg/dL. According to national statistics, cholesterol levels for the population of men age 18-24 follow a normal (gaussian) distribution. Based on this information, approximately 50% of the men in the sample will have which of the following serum total cholesterol levels?

- A. 100-260 mg/dL (3%)
- B. <140 mg/dL or >220 mg/dL (9%)
- C. 140-220 mg/dL (45%)
- ✓ D. <180 mg/dL (41%)
- E. >260 mg/dL (0%)

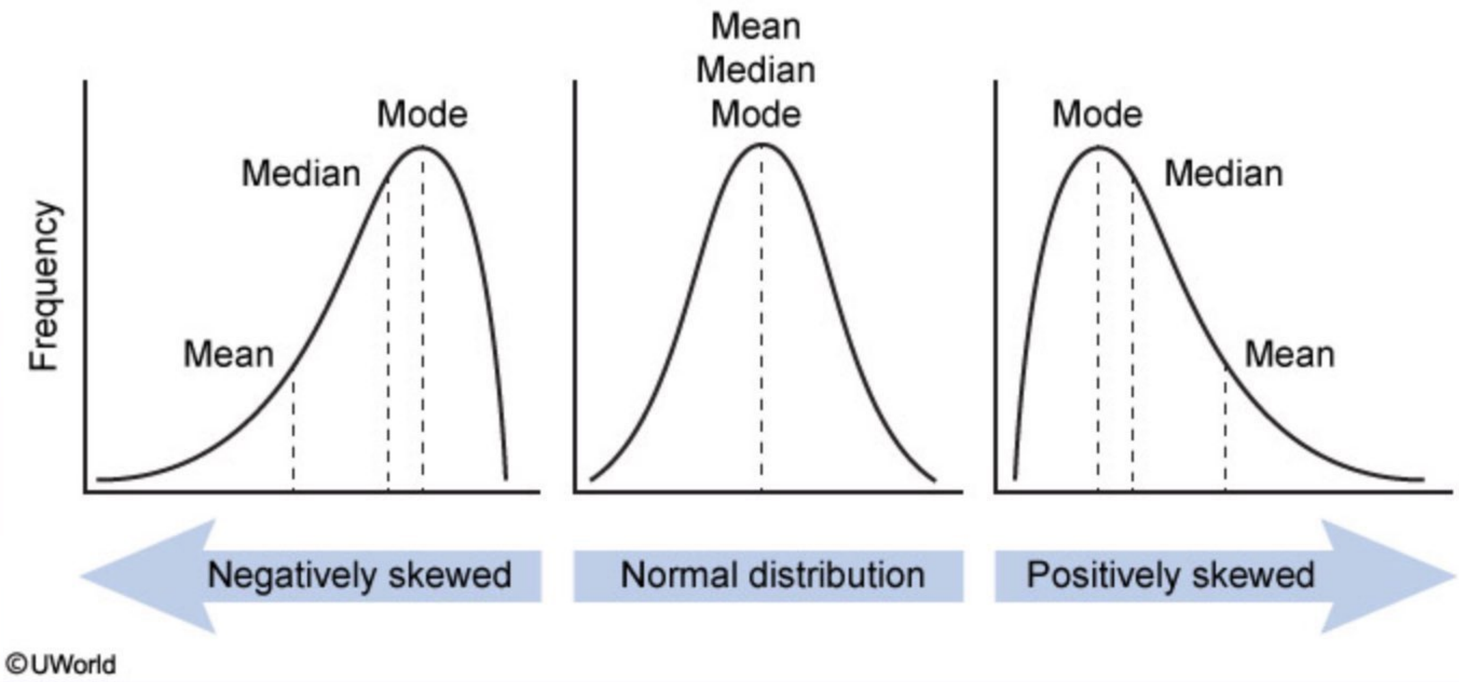
Omitted
Correct answer
D

41%
Answered correctly

04 secs
Time Spent

2023
Version

Explanation



A **normal distribution** refers to a symmetric, bell-shaped distribution pattern with a fixed proportion of observations lying within specific distances from the mean. For a normal distribution, mean equals median. The median divides a distribution of data in half, so for a normal distribution about **50% of observations** lie **below the mean** and 50% of observations lie above the mean.

In this case, a researcher took a random sample of men age 18-24 and estimated their **mean** serum total cholesterol at **180 mg/dL** with a standard deviation of 40 mg/dL. Therefore, approximately 50% of men age 18-24 in the sample will have serum total cholesterol levels **<180 mg/dL**.

(Choices A, B, C, and E) According to the 68-95-99.7 rule for normal distributions, the distribution of serum total cholesterol levels should have the following pattern:

- 68% of serum total cholesterol values will be within 1 SD from the mean, calculated as being between [180 − (1 × 40)] = 140 mg/dL and [180 + (1 × 40)] = 220 mg/dL, and (100% − 68%) = 32% will be <140 mg/dL or

the mean and 50% of observations lie above the mean.

In this case, a researcher took a random sample of men age 18-24 and estimated their **mean** serum total cholesterol at **180 mg/dL** with a standard deviation of 40 mg/dL. Therefore, approximately 50% of men age 18-24 in the sample will have serum total cholesterol levels **<180 mg/dL**.

(Choices A, B, C, and E) According to the 68-95-99.7 rule for normal distributions, the distribution of serum total cholesterol levels should have the following pattern:

- 68% of serum total cholesterol values will be within 1 SD from the mean, calculated as being between [180 − (1 × 40)] = 140 mg/dL and [180 + (1 × 40)] = 220 mg/dL, and (100% − 68%) = 32% will be <140 mg/dL or >220 mg/dL.
- 95% of serum total cholesterol values will be within 2 SDs from the mean, calculated as being between [180 − (2 × 40)] = 100 mg/dL and [180 + (2 × 40)] = 260 mg/dL, and (100% − 95%) = 5% will be <100 mg/dL or >260 mg/dL.
- 99.7% of serum total cholesterol will be within 3 SDs from the mean, calculated as being between [180 − (3 × 40)] = 60 mg/dL and [180 + (3 × 40)] = 300 mg/dL, and (100% − 99.7%) = 0.3% will be <60 mg/dL or >300 mg/dL.

Educational objective:
In a normal distribution, the mean is equal to the median; 50% of observations lie below the mean and 50% of observations lie above the mean.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Pulmonary capillary wedge pressure (PCWP) measurements can be used to estimate left atrial pressure; the normal range is between 6-12 mm Hg, and recorded values are whole numbers. A patient in the intensive care unit has 20 serial PCWP measurements taken over the course of 2 hours. Among these 20 observations, the maximal recorded value is 12 mm Hg and the minimal recorded value is 10 mm Hg. If the next measurement is 26 mm Hg, which of the following is most likely to remain unchanged?

- A. Mean (3%)
- ✓ B. Mode (89%)
- C. Range (1%)
- D. Standard deviation (3%)
- E. Variance (1%)

Omitted
Correct answer
B

89%
Answered correctly

13 secs
Time Spent

2023
Version

Explanation

An **outlier** is defined as an extreme and unusual value observed in a dataset. It may be the result of a recording error, a measurement error, or a natural phenomenon. In this case, the value of 26 mm Hg is an outlier as all the other values lie between 10 and 12 mm Hg. An outlier can affect measures of central tendency (mean, median, mode) as well as measures of dispersion (standard deviation, variance).

This patient had 20 initial measurements that were all between 10 and 12 mm Hg. Although it is not required to answer this question correctly, a numerical example can help to explain the effect of an outlier such as 26 mm Hg. Assume that the patient's 20 initial measurements were: 9 measurements of 10 mm Hg, 6 measurements of 11 mm Hg, and 5 measurements of 12 mm Hg. The **mode** is the most frequently observed data point, so the mode for this initial dataset is 10 mm Hg (because this value is observed 9 times whereas the other 2 values are observed 6 times and 5 times). If the next measurement is 26 mm Hg, the mode would remain 10 mm Hg because this would still be the most commonly observed value. Therefore, the mode would remain unchanged. The mode tends to be resistant to outliers because outliers are not likely to affect the most frequently observed value in a dataset.

(Choice A) The **mean** (or average) can be quite sensitive to outliers, especially in the case of small datasets and extreme outliers. For instance, the mean of the 20 initial measurements is 10.8 mm Hg (obtained by calculating [9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 20). With an additional measurement of 26 mm Hg, the mean would become 11.5 mm Hg (obtained by calculating [1 × 26 mm Hg + 9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 21).

(Choice C) The **range** is equal to the maximal value minus the minimal value and would clearly be affected by outliers. With the 20 initial measurements, the range is 12 mm Hg – 10 mm Hg = 2 mm Hg. If the next measurement is 26 mm Hg, then the range will be 26 mm Hg – 10 mm Hg = 16 mm Hg.

(Choices D and E) The standard deviation and variance are measures of dispersion, which reflect how spread out the values in a dataset are from one another. The standard deviation and variance tend to be sensitive to outliers because outliers increase the dispersion of datasets.

Educational objective:
An outlier is defined as an extreme and unusual observed value in a dataset. It can affect measures of central

mode for this initial dataset is 10 mm Hg (because this value is observed 9 times whereas the other 2 values are observed 6 times and 5 times). If the next measurement is 26 mm Hg, the mode would remain 10 mm Hg because this would still be the most commonly observed value. Therefore, the mode would remain unchanged. The mode tends to be resistant to outliers because outliers are not likely to affect the most frequently observed value in a dataset.

(Choice A) The **mean** (or average) can be quite sensitive to outliers, especially in the case of small datasets and extreme outliers. For instance, the mean of the 20 initial measurements is 10.8 mm Hg (obtained by calculating [9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 20). With an additional measurement of 26 mm Hg, the mean would become 11.5 mm Hg (obtained by calculating [1 × 26 mm Hg + 9 × 10 mm Hg + 6 × 11 mm Hg + 5 × 12 mm Hg] / 21).

(Choice C) The **range** is equal to the maximal value minus the minimal value and would clearly be affected by outliers. With the 20 initial measurements, the range is 12 mm Hg – 10 mm Hg = 2 mm Hg. If the next measurement is 26 mm Hg, then the range will be 26 mm Hg – 10 mm Hg = 16 mm Hg.

(Choices D and E) The standard deviation and variance are measures of dispersion, which reflect how spread out the values in a dataset are from one another. The standard deviation and variance tend to be sensitive to outliers because outliers increase the dispersion of datasets.

Educational objective:
An outlier is defined as an extreme and unusual observed value in a dataset. It can affect measures of central tendency (mean, median, mode) as well as measures of dispersion (standard deviation, variance). Modes tend to be resistant to outliers.

Biostatistics Biostatistics & Epidemiology Central tendency (mean, median, mode, outliers)

A medical student is conducting a chart review of patients admitted to the local emergency department with acute pancreatitis. He is interested in comparing blood glucose levels between mild and severe cases. Before attempting to compare the mean blood glucose levels between the 2 groups of patients, he decides to conduct a descriptive analysis of the variables in each group. Blood glucose levels in patients with severe acute pancreatitis were found to have a strong positively skewed distribution. Which of the following is most likely to be true regarding the data for the severe acute pancreatitis group?

- A. The mean is equal to the median (0%)
- B. The mean is equal to the mode (0%)
- ✓ C. The mean is greater than the median (76%)
- D. The median is greater than the mean (14%)
- E. The mode is greater than the mean (7%)

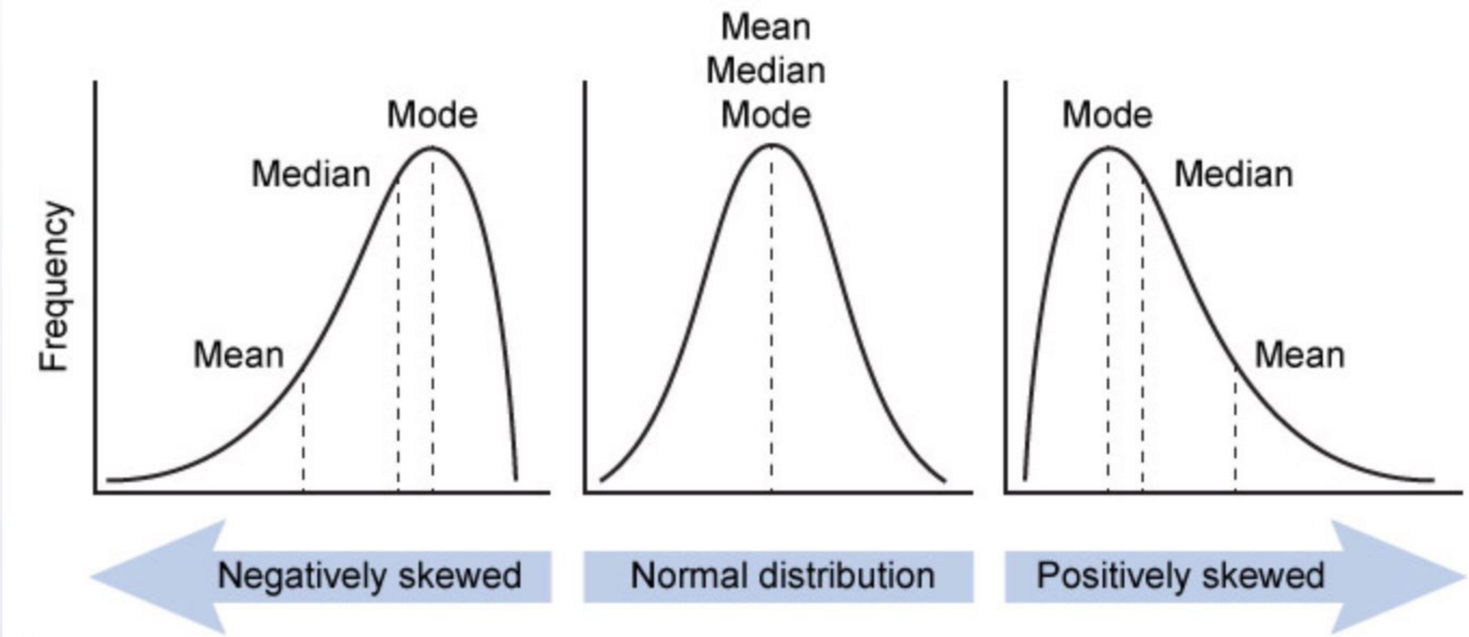
Omitted
Correct answer
C

76%
Answered correctly

06 secs
Time Spent

2023
Version

Explanation



The symmetrical bell-shaped curve is usually cited as the prototypical curve describing the distribution of a continuous variable. However, in actual research scenarios, continuous variables often have **asymmetrical distributions**. When a distribution curve is asymmetric, it is either positively or negatively skewed:

- In a **positively skewed** distribution, there is an increased number of observations with **larger-than-expected magnitudes** (ie, extreme values) that shift the mean toward the **right**, producing a longer slope of the curve ("tail") on the positive side of the distribution.
- In a **negatively skewed** distribution, there is an increased number of observations with **smaller-than-expected magnitudes** that shift the mean toward the **left**, producing a longer slope of the curve ("tail") on the negative side of the distribution.

In general, in a **positively** skewed distribution, the mean is the most shifted in the positive direction, followed by

- In a **positively skewed** distribution, there is an increased number of observations with **larger-than-expected magnitudes** (ie, extreme values) that shift the mean toward the **right**, producing a longer slope of the curve ("tail") on the positive side of the distribution.
- In a **negatively skewed** distribution, there is an increased number of observations with **smaller-than-expected magnitudes** that shift the mean toward the **left**, producing a longer slope of the curve ("tail") on the negative side of the distribution.

In general, in a **positively** skewed distribution, the mean is the most shifted in the positive direction, followed by the median and then the mode (**mode < median < mean**). In such a situation, the median often reflects a central tendency better than the mean does.

(Choices A and B) In perfectly normal distribution curves where there is no skew, all 3 measures of central tendency are precisely equal: mean = median = mode. If a minor skew is present, the 3 measures are approximately equal.

(Choices D and E) In general, in a negatively skewed distribution (with the "tail" on the left), the mean is the most shifted in the negative direction, followed by the median and then the mode.

Educational objective:
In general, in a positively skewed distribution, the mean is the most shifted in the positive direction (to the right), followed by the median and then the mode. In such a situation, the median often reflects a central tendency better than the mean does.

Biostatistics	Biostatistics & Epidemiology	Normal distribution
Subject	System	Topic

Epidemiology and population health

A randomized controlled trial is conducted to compare transcatheter aortic valve replacement (TAVR), using a self-expanding transcatheter aortic valve bioprosthesis, with surgical aortic valve replacement in patients with severe aortic stenosis. The objective is to determine the difference in risk of death from any cause. Seven hundred forty-seven patients with severe aortic stenosis who are at increased surgical risk are recruited and randomly assigned to groups for treatment with TAVR or surgical valve replacement. Follow-up assessments are performed at discharge and at 1 month, 6 months, and 1 year. The 1-year follow-up results are as follows:

	Death from any cause		
	Yes	**No**	**Total**
TAVR	57	333	390
Surgical valve replacement	68	289	357
Total	125	622	747

Which of the following best represents the absolute risk reduction for death from any cause in patients treated with TAVR compared to those treated with surgical replacement?

- A. (68/357) − (57/390) = 0.044
- B. [(68/357) − (57/390)] / (68/357) = 0.233
- C. (57/390) / (68/357) = 0.767
- D. (68/357) / (57/390) = 1.303
- E. 1 / [(68/357) − (57/390)] = 22.562

A group of researchers conducted a case-control study to estimate the association between exposure to a certain pesticide and head and neck squamous cell carcinomas. Controls were matched to cases by age and gender. Exposure status was determined by interviewing the subjects. On matched-pair analysis, the odds ratio is 7.5 (95% confidence interval: 2.3-14.8). Based on this information, which of the following is most likely to affect the validity of this study?

- A. Confounding bias by age
- B. Healthy worker bias
- C. Misclassification bias
- D. Nonresponse bias
- E. Placebo effect

Proceed To Next Item

A randomized controlled study investigated the effects of single-dose dexmedetomidine on emergence delirium (ED) in children who were under anesthesia induced with sevoflurane. A total of 2,500 patients age 2-7 who were scheduled for tonsillectomy were enrolled in the study; 80% were randomized to receive dexmedetomidine 0.5 µg/kg or volume-matched normal saline over 10 minutes after induction of anesthesia. The primary outcome was the incidence of ED within 30 minutes after extubating. The results show an incidence of ED of 300 cases per 1,000 patients for dexmedetomidine and 500 cases per 1,000 patients for saline. Which of the following is the best estimate of the absolute risk reduction for ED following dexmedetomidine compared to saline?

- A. $0.80 \times 2{,}500$
- B. $(500 / 1{,}000) - (300 / 1{,}000)$
- C. $(300 / 1{,}000) / (500 / 1{,}000)$
- D. $1 / [(500 / 1{,}000) - (300 / 1{,}000)]$
- E. $[(500 / 1{,}000) - (300 / 1{,}000)] / (500 / 1{,}000)$

Proceed To Next Item

A 73-year-old man comes to the office for follow-up. He was diagnosed with chronic lymphocytic leukemia (CLL) 3 years ago when routine laboratory testing revealed a markedly elevated leukocyte count. He feels well. On examination, he has stable lymphadenopathy. He has been reading about management options should his CLL progress and inquires about an experimental drug that selectively binds malignant lymphocytes. The drug has been shown to significantly prolong survival in patients with stage 3 and 4 CLL, without curing the malignancy. If this new drug were widely used, what changes would be expected in the number of incident and prevalent cases of CLL?

- A. The number of incident cases will decrease, the number of prevalent cases will decrease
- B. The number of incident cases will increase, the number of prevalent cases will not change
- C. The number of incident cases will decrease, the number of prevalent cases will increase
- D. The number of incident cases will not change, the number of prevalent cases will increase
- E. The number of incident cases will not change, the number of prevalent cases will not change

Proceed To Next Item

The specificity of the standard test to detect disease X in the general population is 73%. A team of researchers set forth a goal to increase specificity for detecting disease X. They develop a new test and conduct a study to evaluate its performance on a random sample from the general population. The results of their study are shown below.

	Patients with disease X	Patients without disease X
Positive	270	45
Negative	30	255

According to these results, have the researchers achieved their goal?

- A. Cannot be determined because the prevalence of disease X is not provided
- B. No, the researchers' new test has about 9% lower specificity than the standard test
- C. No, the researchers' results lead to nearly the same specificity as the standard test
- D. Yes, the researchers achieved an increase in specificity of about 12%
- E. Yes, the researchers achieved an increase in specificity of about 15%

Proceed To Next Item