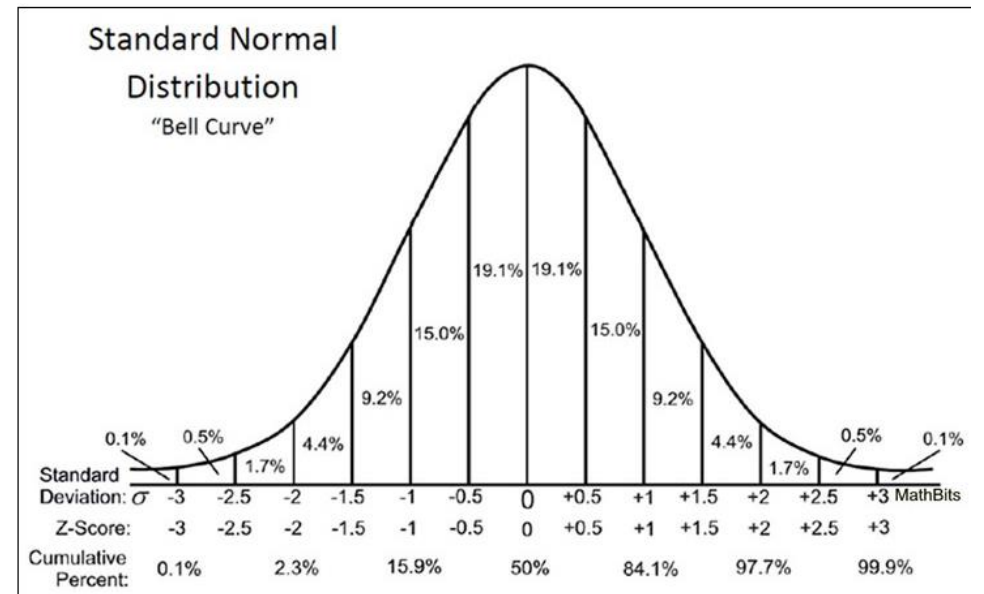


The distribution of the heights of college women long distance runners is approximately normal with a mean of 162.2 cm and a standard deviation of 3.4 cm. Which of the following is closest to the percentage of these runners who are between 155 and 163 cm tall?

- A 26.1
- B 42.4%
- C 57.6%
- D 73.9%
- E 84.8%



C

Students in a chemistry lab are measuring the temperature of a solution during a chemical reaction. The mean temperature of the class's temperature readings was  $125.7^{\circ}\text{F}$  and the standard deviation of their temperature readings was  $5.8^{\circ}\text{F}$ . Students were then asked to convert their temperatures to degrees Celsius by subtracting 32, then multiplying by  $5/9$ .

The mean and standard deviations of their temperatures in Celsius would be:

A  $\mu = 52.1^{\circ}\text{C}, \sigma = 3.2^{\circ}\text{C}$  ✓

B  $\mu = 52.1^{\circ}\text{C}, \sigma = 5.8^{\circ}\text{C}$

C  $\mu = 52.1^{\circ}\text{C}, \sigma = 37.8^{\circ}\text{C}$

D  $\mu = 69.8^{\circ}\text{C}, \sigma = 3.2^{\circ}\text{C}$

E  $\mu = 69.8^{\circ}\text{C}, \sigma = 37.8^{\circ}\text{C}$

A

A distribution of prices is approximately normal with a mean of \$5.60 and a standard deviation of \$0.80. Which of the following equations can be used to find the score  $x$  above which 45% of the prices fall?

(A)  $0.12 = \frac{x-5.60}{0.8^2}$

(B)  $0.45 = \frac{x-5.6}{0.8^2}$

(C)  $0.12 = \frac{x-5.6}{0.8}$

(D)  $0.45 = \frac{x-5.6}{0.8}$

(E)  $0.55 = \frac{x-5.6}{0.8}$

C

# Graphical representations of summary statistics

- Center — 中心
- Spread — 离散程度
- Shape — 分布形状
- Outlier — 离群值
- Distribution — 分布
- Mean — 均值
- Median — 中位数
- Interquartile Range — 四分位距
- Standard Deviation — 标准差
- Skewness — 偏态

Graphical representations of summary statistics help visualize key features of a data set.

**Boxplots** show the five-number summary (min, Q1, median, Q3, max) and highlight outliers.

**Histograms** display the shape, center, and spread, making it easier to see patterns and variability.

**Dotplots** represent individual data points and provide a clear picture of distribution and outliers.

All graphs should be clearly labeled, and conclusions drawn should reflect the context of the data.

## 1. Shape, Center, Spread (Multi-Graph Comparison)

Two distributions are shown:

- Group A: symmetric, unimodal, IQR = 12, mean = 50
- Group B: strongly right-skewed, IQR = 10, mean = 55

**Question:**

Without seeing raw data, determine which group likely has the higher median and justify your answer using properties of skewness and center.

---

## 2. Choosing the Best Graph (Justification)

A researcher collects:

- 200 exam scores (0–100)
- 5 summary statistics (min, Q1, median, Q3, max)

**Question:**

Explain why a box plot is more appropriate than a histogram for comparing this dataset across two schools.

---

## 3. Outliers and Effect on Summary Statistics

A dataset has:

- Median = 40
- Mean = 52
- One extreme value = 200

**Question:**

Explain how the outlier affects:

1. Mean
2. Median
3. Standard deviation

and justify which measure of center is more reliable.

---

## 4. Interpreting Box Plots (Higher Order)

Two box plots show test scores:

- School A: tight IQR, one extreme low outlier
- School B: wide IQR, no outliers

**Question:**

Which school has more consistent performance? Justify using variability and outliers.

---

## 5. Transformation Reasoning

A dataset is multiplied by 3 and then increased by 10.

**Question:**

Describe the effect on:

- Mean
- Median
- IQR
- Standard deviation

and explain which measures change and which do not in terms of transformation rules.

---

## 6. Comparing Distributions in Context

Two countries' incomes:

- Country X: median higher, but very skewed right
- Country Y: lower median, but symmetric distribution

**Question:**

Which country better represents "typical income" and why might mean be misleading in one case?

## 1. Median comparison

Group A median  $\approx$  50 (symmetric).

Group B median  $<$  mean (right-skewed), so median  $<$  55.

👉 Group A likely has higher median.

---

## 2. Why box plot

Summarises distribution (median, IQR, outliers), good for **comparing groups**; histogram is less efficient for side-by-side summary comparison.

---

## 3. Outlier effects

- Mean: increases strongly (pulled up)
  - Median: little/no change
  - SD: increases
- 👉 Best centre: **median**

## 4. Consistency

School A more consistent (smaller IQR), despite one outlier.

👉 Spread (IQR) matters more than single extreme value.

---

## 5. Transformation ( $\times 3$ , $+10$ )

- Mean:  $\times 3 + 10$
- Median:  $\times 3 + 10$
- IQR:  $\times 3$
- SD:  $\times 3$

👉 Adding constant doesn't change spread; multiplication scales all.

---

## 6. Typical income

Country X mean misleading due to skew; median better measure.

👉 Country Y more representative of "typical" due to symmetry.

# Comparing Distributions

- Center — 中心
- Spread — 离散程度
- Shape — 分布形状
- Outlier — 离群值
- Distribution — 分布
- Symmetry — 对称性
- Skewness — 偏态
- Variability — 变异性
- Median — 中位数
- Interquartile Range — 四分位距
- Standard Deviation — 标准差
- Overlap — 重叠
- Shift — 平移 (整体偏移)
- Clustering — 聚集
- Gap — 间隙
- Unimodal — 单峰
- Bimodal — 双峰
- Uniform — 均匀分布

1. **Shape** – Compare the form of each distribution (skewed, symmetric, unimodal, bimodal).
2. **Center** – Compare typical values using mean or median, depending on the shape.
3. **Spread** – Compare variability using IQR, range, or standard deviation.
4. **Outliers** – Identify any outliers and note how they affect the comparison.
5. **Context + Comparison Statement** – Always describe differences clearly in context using comparative language (e.g., higher, more variable, less spread out).

A researcher compares the effectiveness of two online revision platforms, **Platform X** and **Platform Y**, by recording the test scores of students after using each platform.

The summary statistics are shown below:

- **Platform X:**  
median = 82, mean = 80, IQR = 6, standard deviation = 5, skewed left, no outliers
  - **Platform Y:**  
median = 78, mean = 81, IQR = 14, standard deviation = 12, slightly right-skewed, one low outlier at 50
- 

**(a) Compare the distributions of test scores for the two platforms.**

**(b) Which platform appears to be more consistent in improving student performance?  
Justify your answer.**

**(c) What effect, if any, does the outlier in Platform Y have on your comparison?**

### **(a) Comparison**

The distributions differ in center, spread, shape, and outliers.

In terms of **center**, Platform X has a higher median (82 vs 78), suggesting higher typical performance.

However, Platform Y has a slightly higher mean (81 vs 80), likely influenced by extreme values.

For **spread**, Platform Y has much greater variability (IQR 14 and SD 12) compared to Platform X (IQR 6 and SD 5), meaning Platform X scores are more consistent.

Regarding **shape**, Platform X is skewed left while Platform Y is slightly skewed right.

In terms of **outliers**, Platform Y has a low outlier at 50, while Platform X has none.

Overall, Platform X shows higher and more consistent performance, while Platform Y shows more variability.

---

### **(b) Consistency**

Platform X appears more consistent because it has a smaller IQR and standard deviation, indicating less spread in student performance.

---

### **(c) Effect of Outlier**

The low outlier in Platform Y pulls the mean down and increases the standard deviation, making Platform Y appear more variable and slightly lowering its average performance compared to the median.

# TRUE or FALSE

- A skewed distribution usually has mean  $\neq$  median.

TRUE

# TRUE or FALSE

- The interquartile range (IQR) is resistant to outliers.

TRUE

# TRUE or FALSE

- A skewed distribution usually has mean  $\neq$  median.

TRUE