MODULE 3

VARIABILITY

OUTLINE

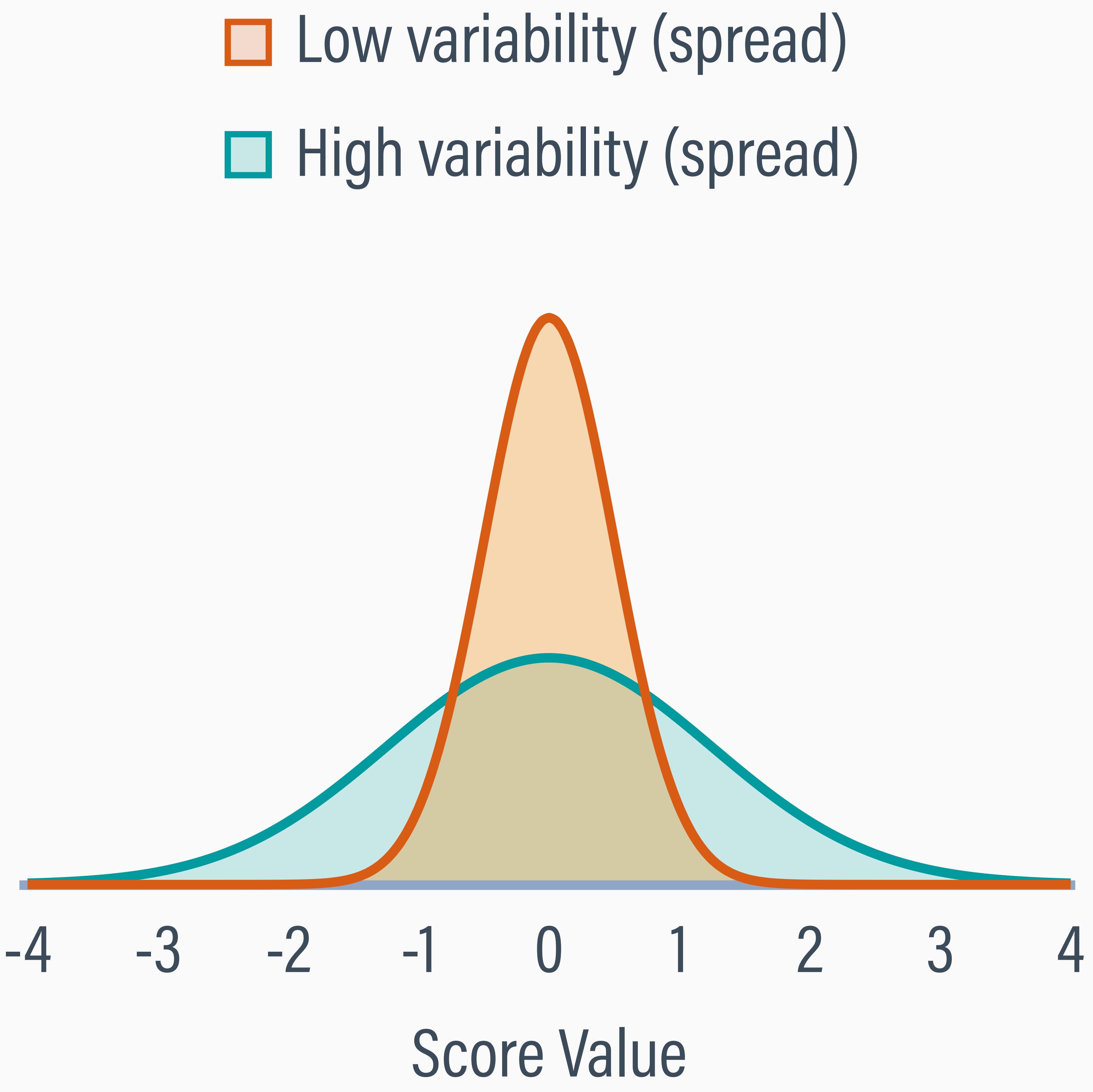
1. Measures of spread: Range, standard deviation, and variance
2. Statistical properties of the variance and standard deviation
3. Psychology application: Smoking cessation clinical trial
4. Standardizing: Expressing scores in standard deviation units
5. Study questions

OUTLINE

1. Measures of spread: Range, standard deviation, and variance
2. Statistical properties of the variance and standard deviation
3. Psychology application: Smoking cessation clinical trial
4. Standardizing: Expressing scores in standard deviation units
5. Study questions

VARIABILITY OR SPREAD

- Variability refers to the degree to which scores differ
- Low variability implies that scores are similar, whereas high variability means scores are different



MEASURES OF SPREAD

- Three definitions of variation: high vs. low score (the range), average distance from the scores to the mean (average absolute deviation and standard deviation), and average squared distance from the scores to the mean (the variance)
- The standard deviation is preferable for describing data, but the variance appears in formulas later in the class (t-tests, ANOVA)
- The variance is just the square of the standard deviation

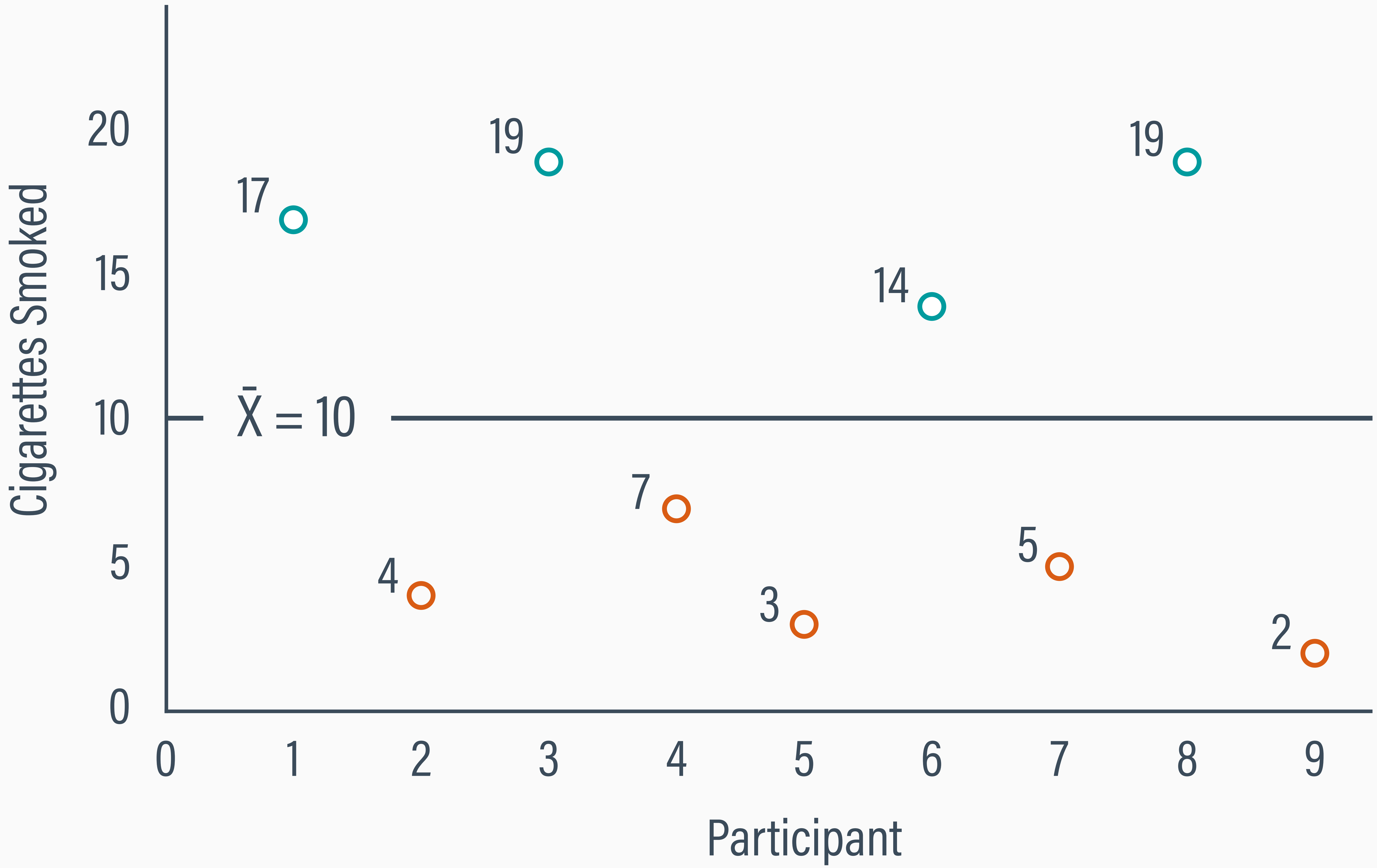
SMALL SAMPLE ILLUSTRATION

- Later in the slides, we will apply measures of variability to a smoking cessation clinical trial

- To illustrate how variation is defined and computed, I use a small sample where participants report the typical number of cigarettes that they smoked in the last week

X
17
4
19
7
3
14
5
19
2

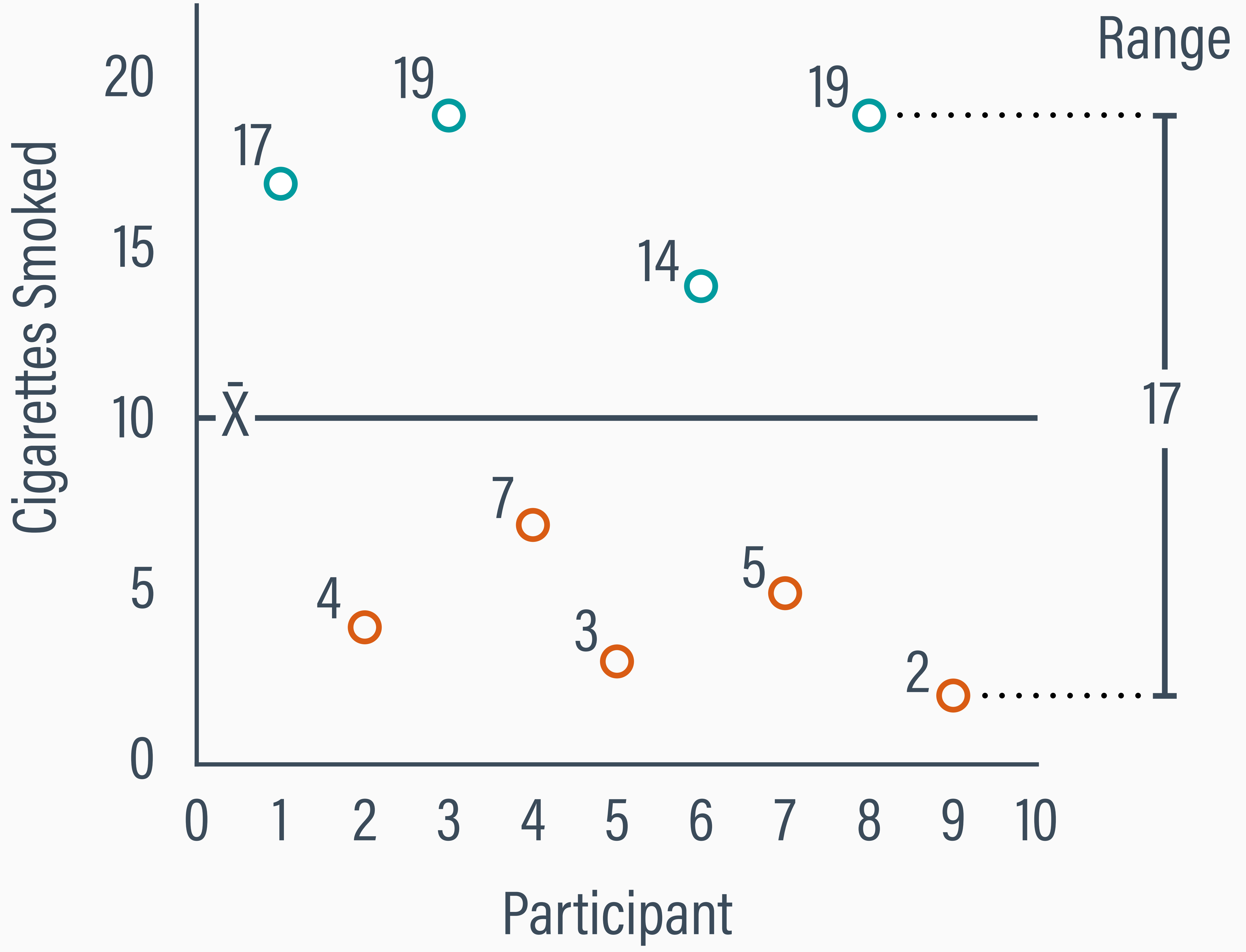
ILLUSTRATIVE DATA



X	$\bar{X}$
17	10
4	10
19	10
7	10
3	10
14	10
5	10
19	10
2	10

THE RANGE

- The range is the distance between the highest and lowest score
- Range = 19 – 2 = 17 cigarettes
- The range is not useful or reliable because it relies solely on two data points



STANDARD DEVIATION

- The standard deviation is average distance between the scores and the mean
- The average absolute deviation is a second measure that is uncommon in applications
- We will rely exclusively on the standard deviation because it is by far the predominant approach in practice

VARIATION RECIPE

1. Express scores as distances (deviations) from the mean
2. Square distances to eliminate negative values
3. Sum the squared distances
4. Average the squared distances to get the variance
5. Take the square root to get the standard deviation, which is the average distance from the mean

1. DISTANCE (DEVIATION) SCORES

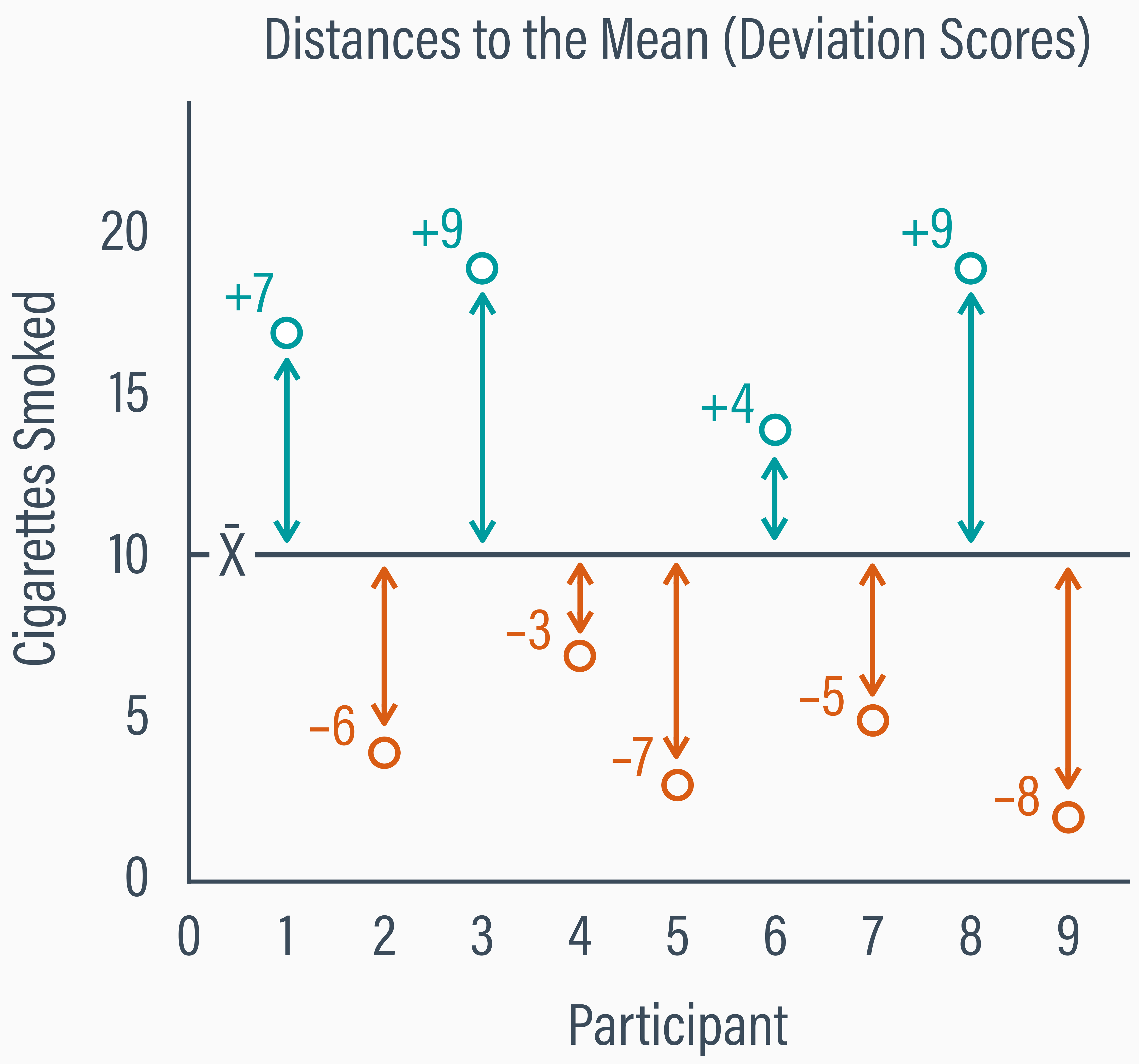
- The distance between a score X and the sample mean $\bar{X}$ is called a deviation score

$$d = X - \bar{X}$$

- Distance scores can be positive (score > mean), negative (score < mean), or zero (score = mean)

DEVIATION SCORES

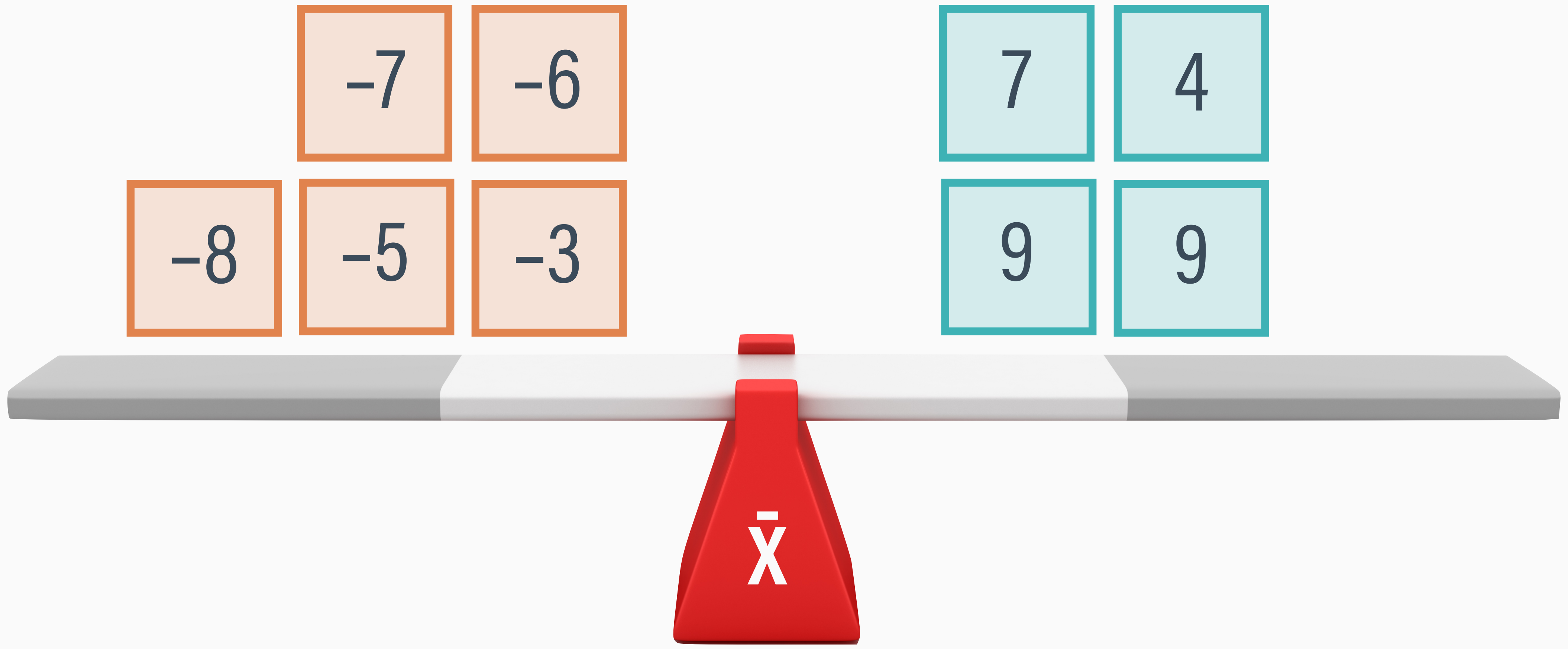
X	$\bar{X}$	d
19	10	9
19	10	9
17	10	7
14	10	4
7	10	-3
5	10	-5
4	10	-6
3	10	-7
2	10	-8



DEVIATION SCORES AVERAGE TO ZERO

X	$\bar{X}$	d
19	10	9
19	10	9
17	10	7
14	10	4
7	10	-3
5	10	-5
4	10	-6
3	10	-7
2	10	-8
	Sum =	**0**

- We cannot average deviation scores directly because the positive and negative distances from the mean always cancel out and must sum to zero



2. SQUARE DEVIATION SCORES

- Squaring distance scores eliminates negative values, so distances no longer cancel out

$$d^2 = (X - \bar{X})^2$$

- Achieves the same goal as the absolute value

X	$\bar{X}$	d	d^2
19	10	9	81
19	10	9	81
17	10	7	49
14	10	4	16
7	10	-3	9
5	10	-5	25
4	10	-6	36
3	10	-7	49
2	10	-8	64

3. SUM OF SQUARED DISTANCES

- Summing the squared distance scores gives a foundational measure called sum of squares (SS)

$$SS = \sum (X - \bar{X})^2 = \text{sum of squared distances}$$

- The sum of squares expresses the total amount of variability in the data as a lump sum

X	$\bar{X}$	d	d^2
19	10	9	81
19	10	9	81
17	10	7	49
14	10	4	16
7	10	-3	9
5	10	-5	25
4	10	-6	36
3	10	-7	49
2	10	-8	64
Sum of Squares =			**410**

4. AVERAGE SQUARED DISTANCE

- Averaging the squared distances between the scores and the mean gives a measure of variability called the variance

! This version of the formula is biased

$$s^2_{biased} = \frac{\sum(X - \bar{X})^2}{N} = \frac{SS}{N}$$

- The variance is just the arithmetic average (a sum divided by the number of observations) applied to squared scores

THE VARIANCE

- The average squared distance from the smoking scores to the mean is 45.6

$$s^2_{biased} = \frac{\sum(X - \bar{X})^2}{N} = \frac{SS}{N} = \frac{410}{9} = 45.6$$

- The variance will appear in equations later in the course, but we will not use it to describe data

X	$\bar{X}$	d	d^2
19	10	9	81
19	10	9	81
17	10	7	49
14	10	4	16
7	10	-3	9
5	10	-5	25
4	10	-6	36
3	10	-7	49
2	10	-8	64

Sum of squares = 410

Variance (average of squared distances) = 45.6

5. AVERAGE DISTANCE TO THE MEAN

- Taking the square root undoes the squared distances, giving the average distance from the scores to the mean

! This version of the formula is biased

$$s_{biased} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{s^2_{biased}}$$

- The standard deviation is average (typical) distance

THE STANDARD DEVIATION

- The standard deviation is the square root of the variance

$$s_{biased} = \sqrt{s^2_{biased}} = \sqrt{45.6} = 6.75$$

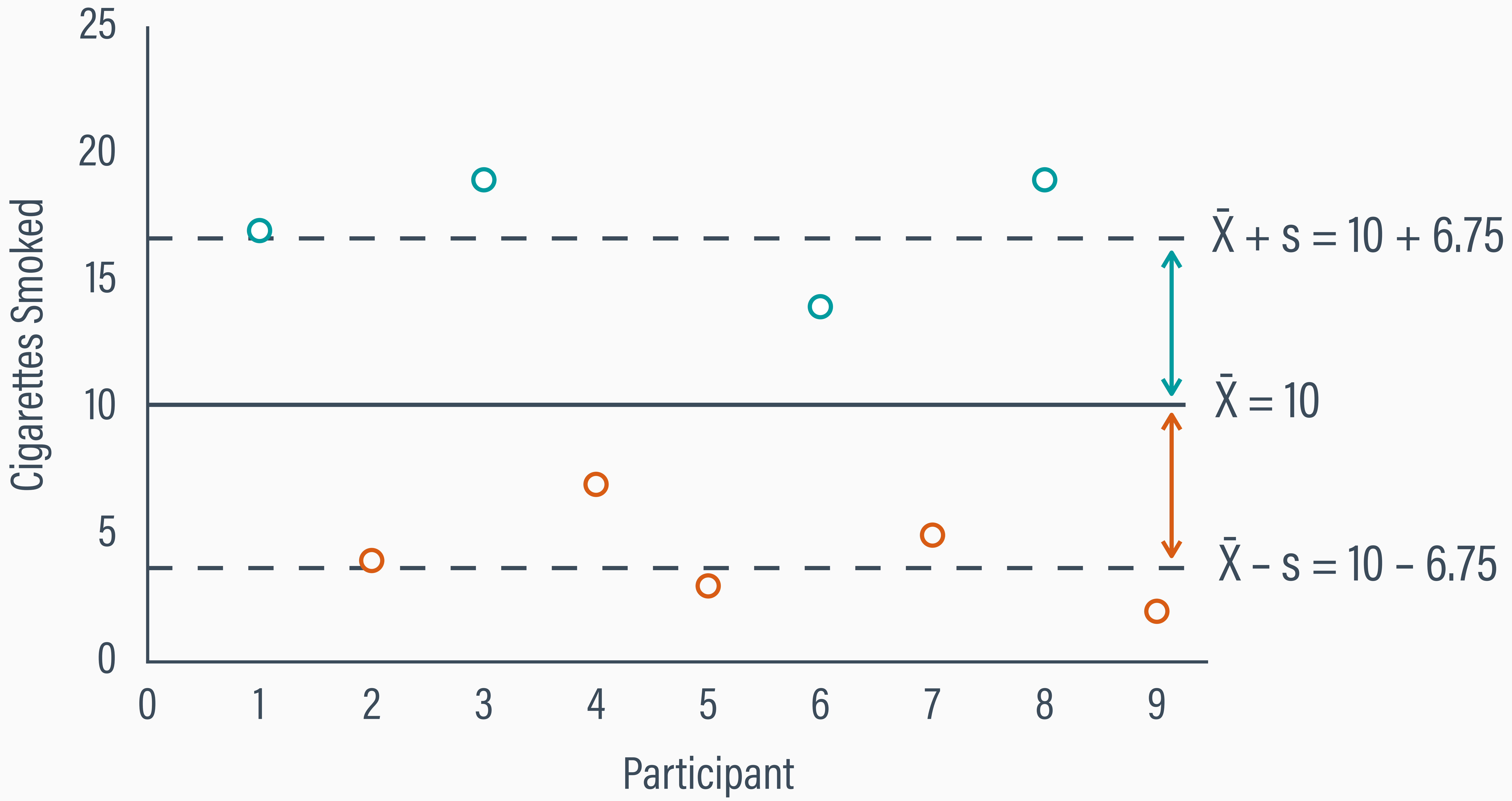
- The average (typical) distance from the smoking scores to the mean is 6.75 cigarettes

X	$\bar{X}$	d	d^2
19	10	9	81
19	10	9	81
17	10	7	49
14	10	4	16
7	10	-3	9
5	10	-5	25
4	10	-6	36
3	10	-7	49
2	10	-8	64

Sum of squares = 410

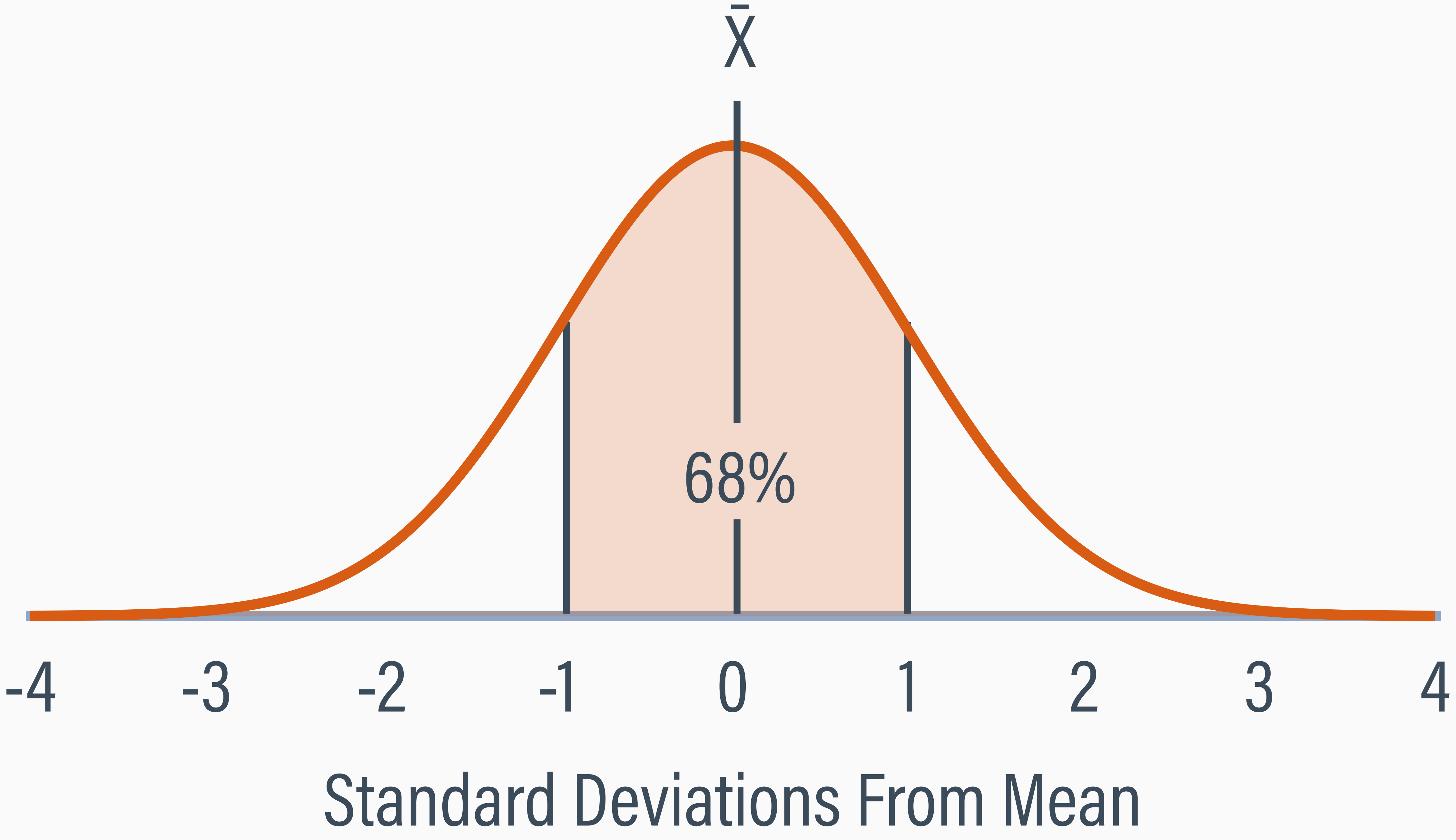
Variance (average of squared distances) = 45.6

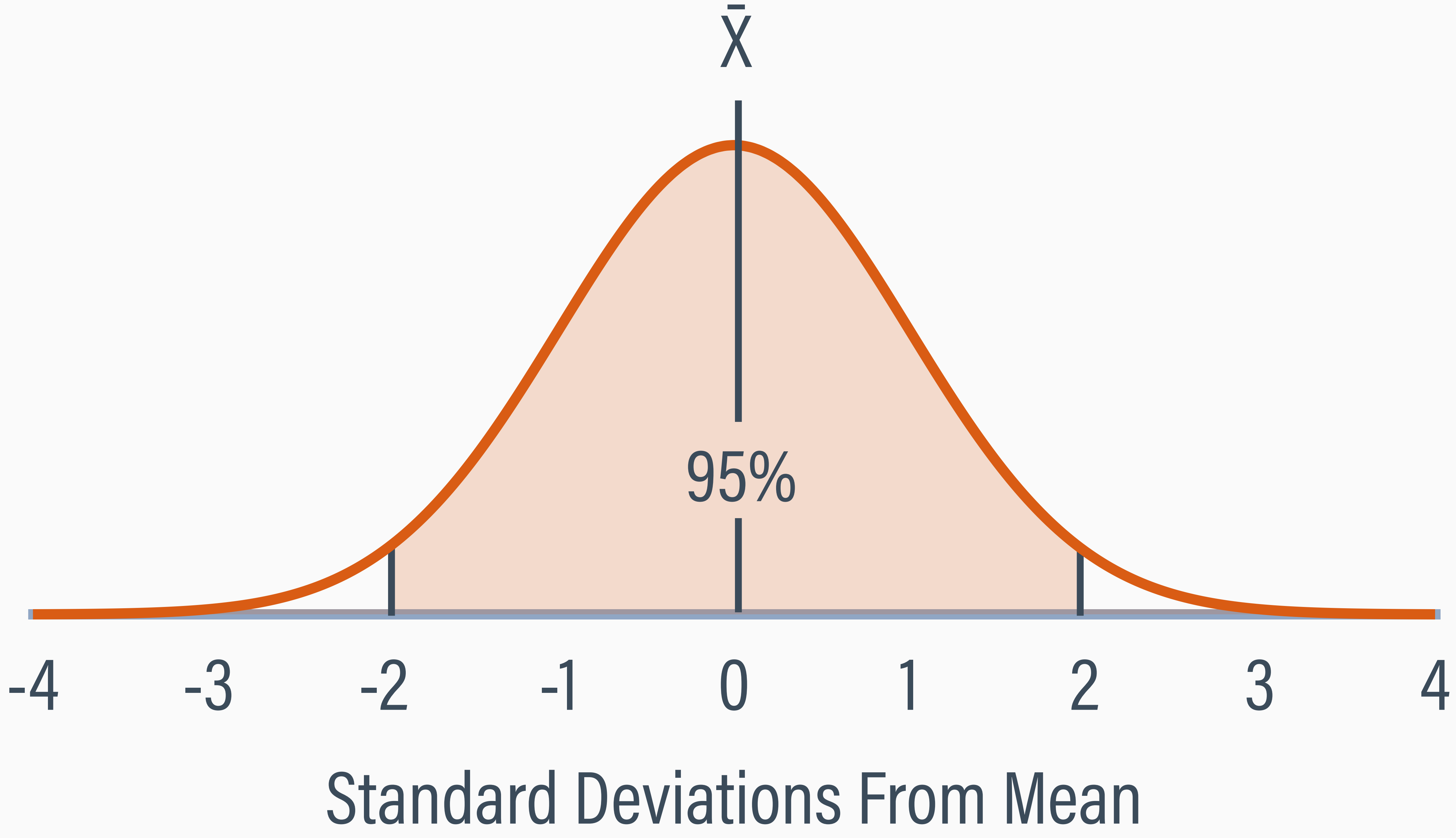
VISUALIZING THE STANDARD DEVIATION



RULE OF THUMB FOR NORMAL DATA

- In a normal curve, 68% of the scores are within ± 1 standard deviation of the mean, and 95% are within ≈ ± 2 standard deviations





OUTLINE

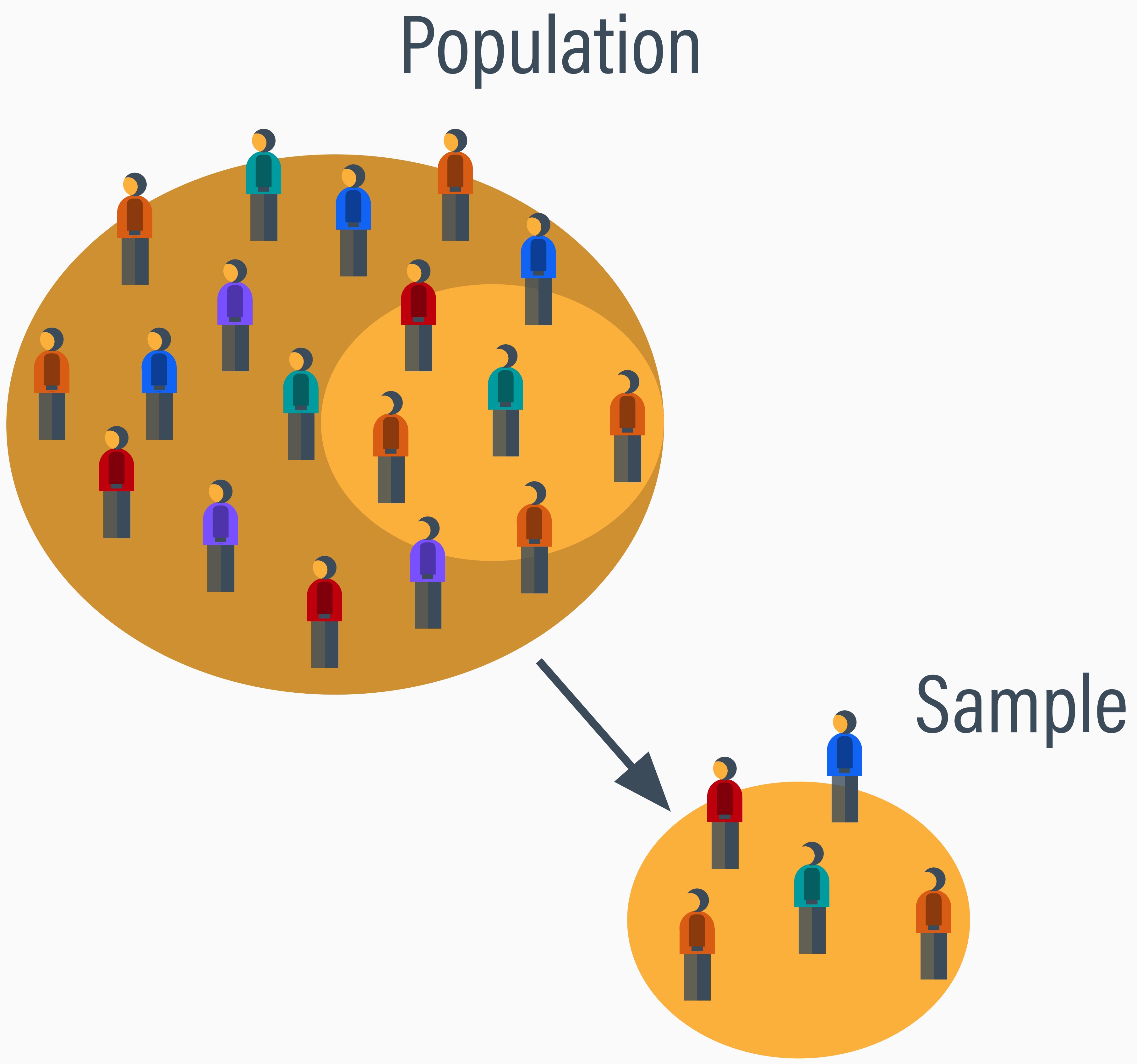
1. Measures of spread: Range, standard deviation, and variance
2. Statistical properties of the variance and standard deviation
3. Psychology application: Smoking cessation clinical trial
4. Standardizing: Expressing scores in standard deviation units
5. Study questions

IMPORTANT TERMINOLOGY

Participants	Statistic
Population = all possible participants who share an attribute of interest	Parameter = a hypothetical statistic computed using the full population
Sample = the subset of people who participated in the study	Estimate = an observed statistic computed from the sample data

POPULATIONS VS. SAMPLES

- A population is the entire group of individuals that are of interest in a study
- Researchers almost exclusively work with a smaller subset called a sample
- The usual goal is use a sample statistic as a best guess about the population statistic

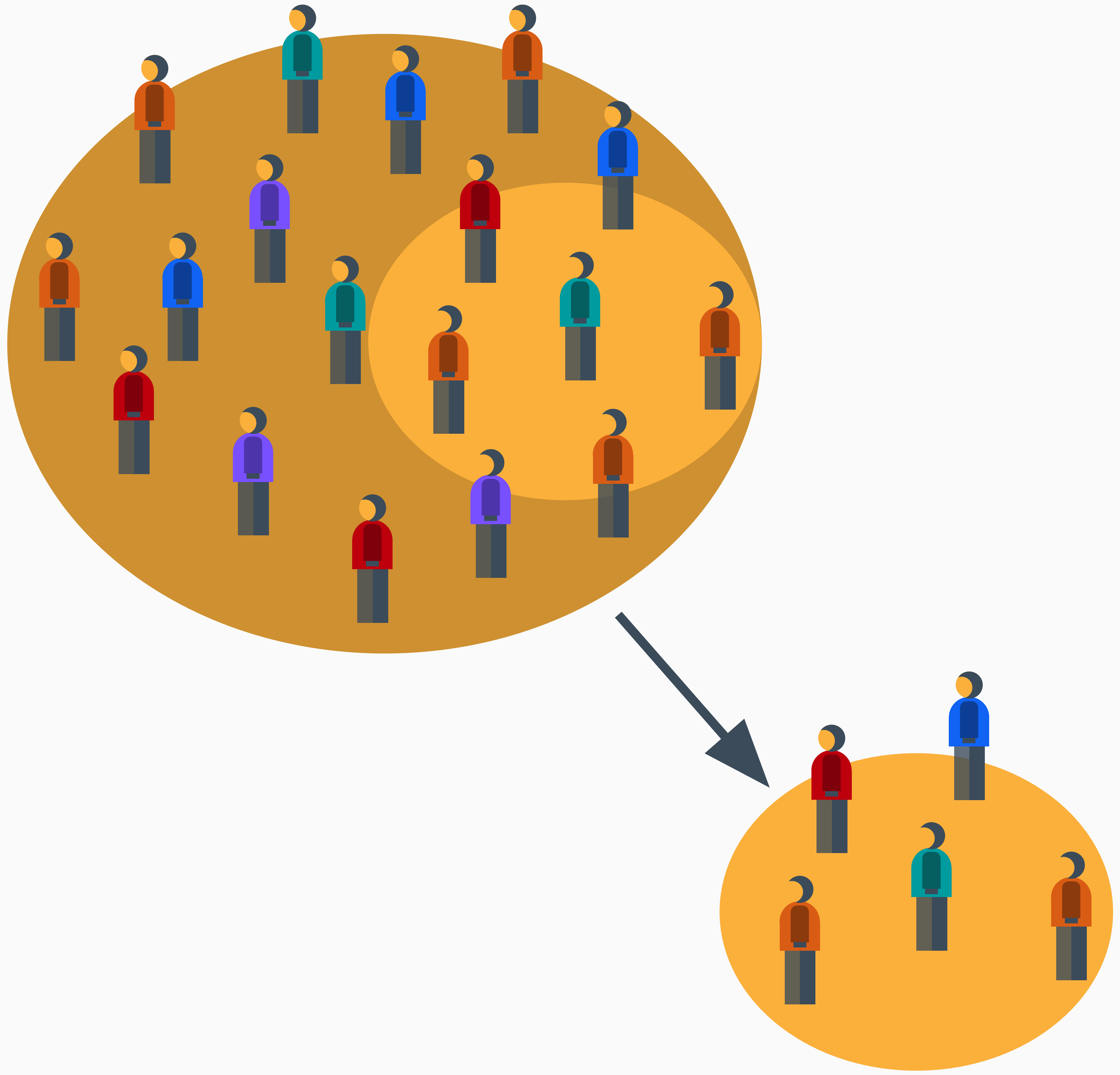


PARAMETERS VS. ESTIMATES

- Greek letters σ and σ^2 denote the true population-level standard deviation and variance (parameters)
- Roman letters s and s^2 reference a sample's standard deviation and variance (estimates)
- Ideally, the sample statistics will be close to the true values, but they will always differ

Population variance = σ^2

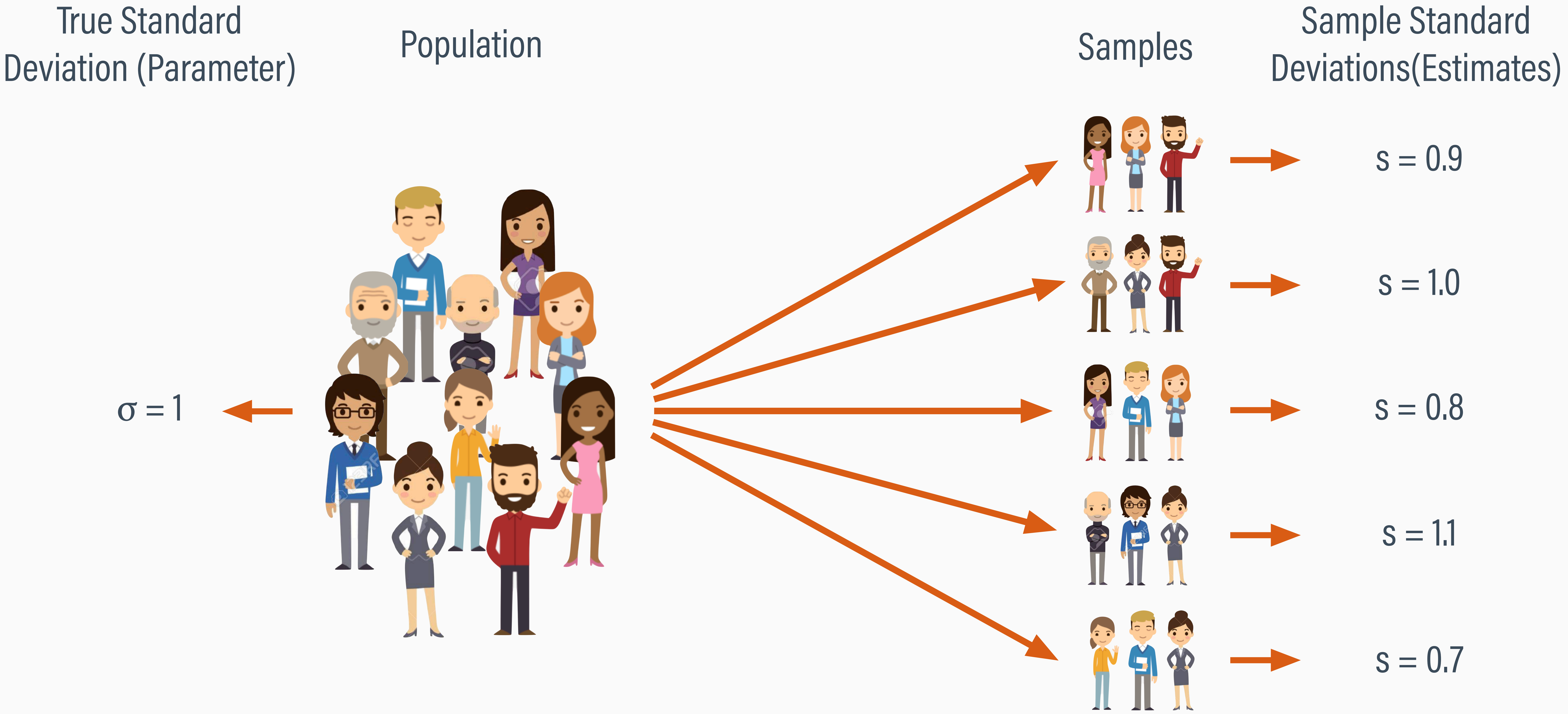
Population standard deviation = σ



Sample variance = s^2

Sample standard deviation = s

SAMPLE ESTIMATES DIFFER FROM TRUE VARIATION

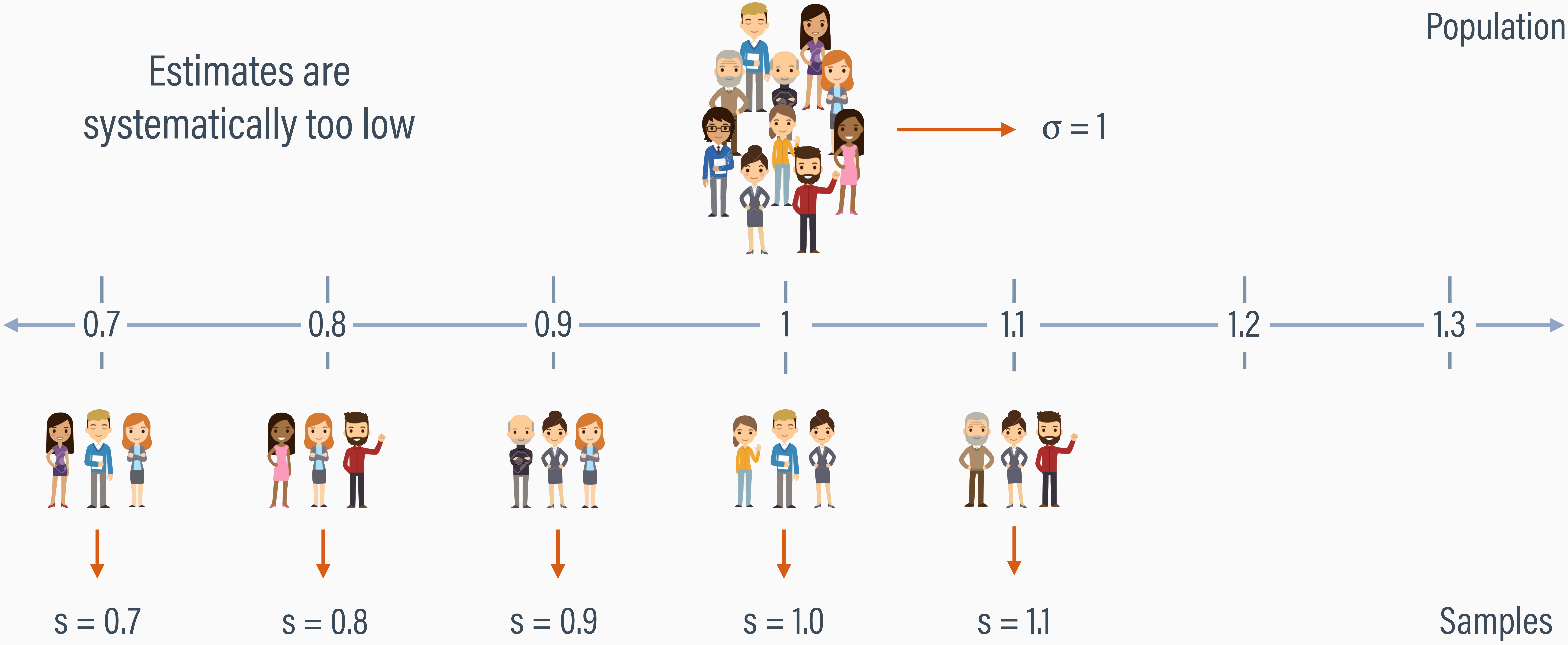


BIASEDNESS OF SAMPLE ESTIMATES

- Like the mean, the standard deviation and variance will vary from sample to sample
- The earlier equations give biased estimates of the true population variation: on average (across many samples), estimates systematically miss in one direction
- On average, sample estimates will be too low (they underestimate the true variation in the population)

BIASED ESTIMATES

Estimates are systematically too low

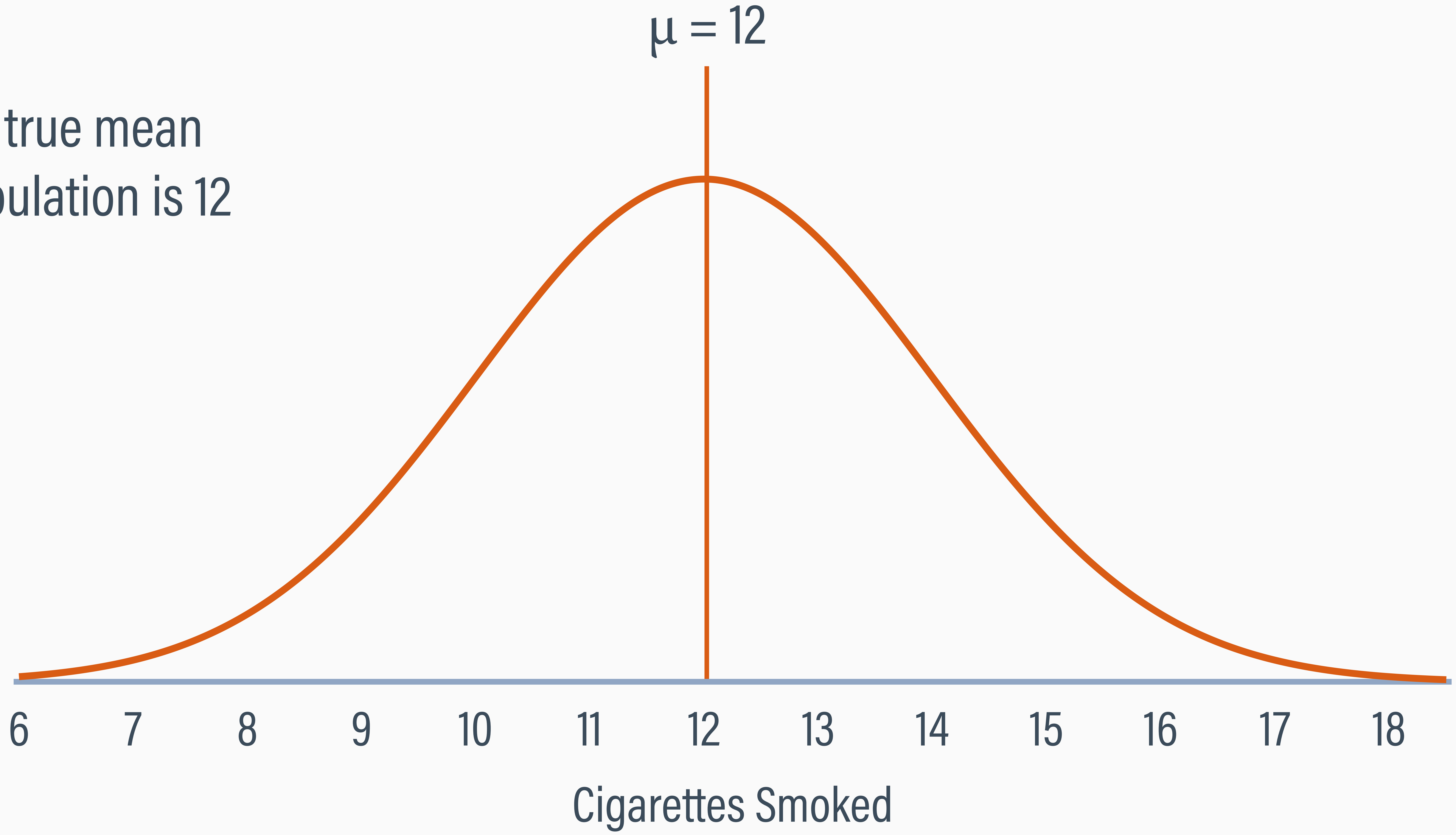


SOURCE OF BIAS

- We never know the true mean μ, so we must instead use the sample mean $\bar{X}$ to compute distance (deviation) scores
- Because the sample mean is the balancing point in the distribution, distances from $\bar{X}$ will be smaller than distances from any other value (including μ)
- This makes the numerator (sum of squares) too small, which reduces the value of the standard deviation and variance

ILLUSTRATION

Pretend the true mean
in the full population is 12



VARIABILITY RELATIVE TO TRUE MEAN

This is hypothetical, as we never know the value of μ

X	μ	d	d^2
11	12	-1	1
12	12	0	0
14	12	2	4
15	12	3	9

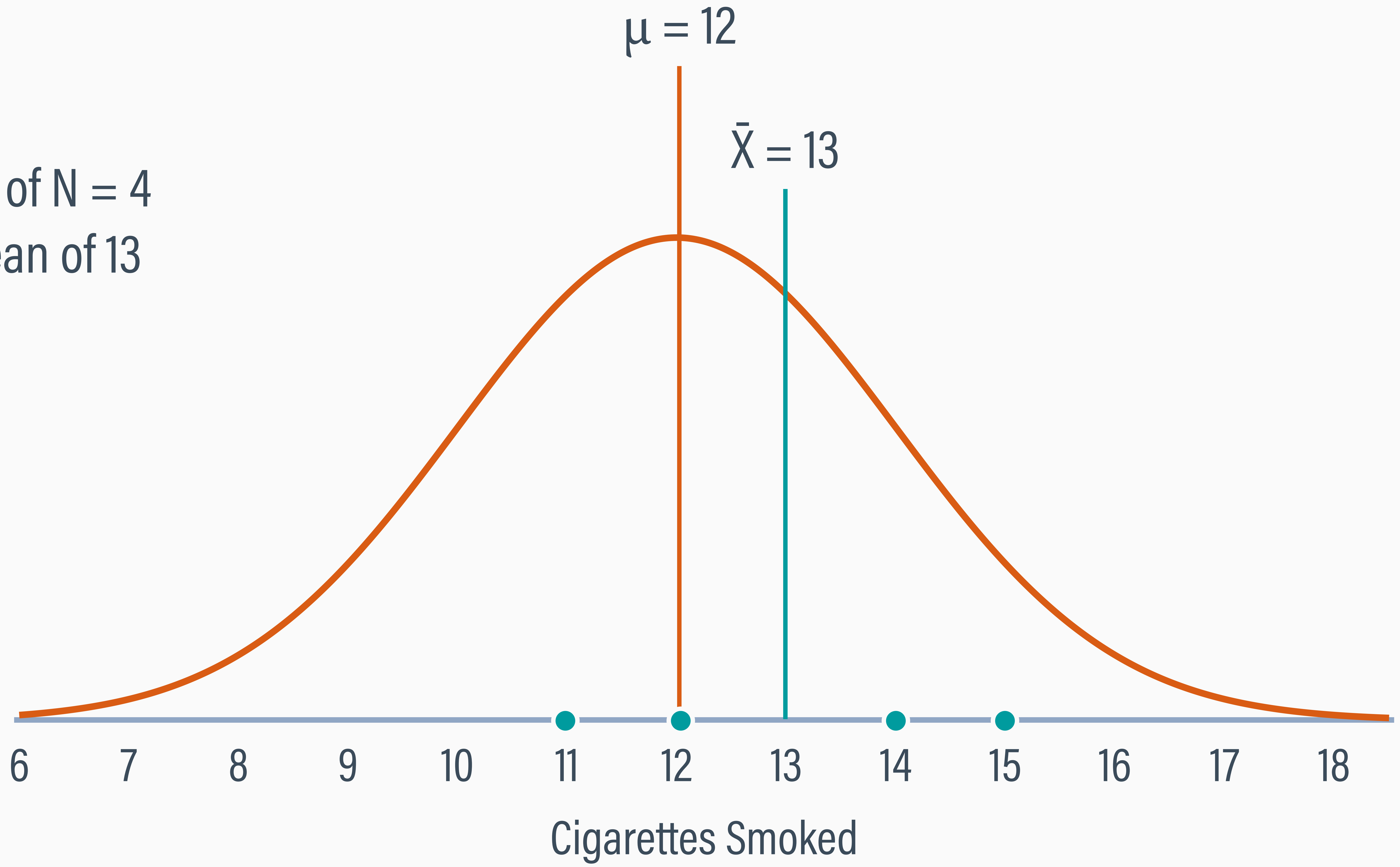
Sum of distances = 4

Sum of squared distances = **14**

$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}} = \sqrt{\frac{14}{4}} = 1.87$$

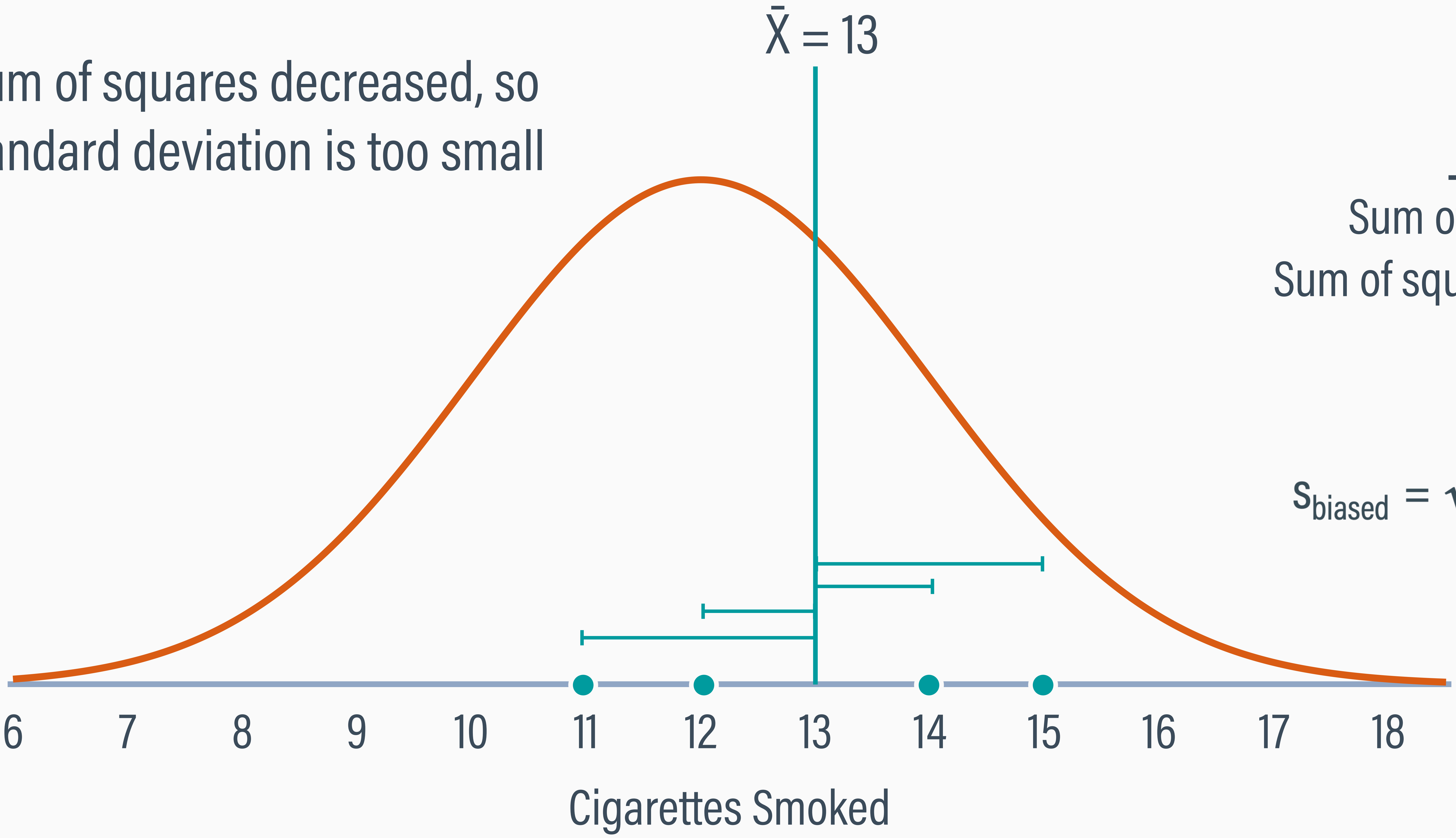
SAMPLE FROM THE POPULATION

A sample of N = 4
has a mean of 13



VARIABILITY RELATIVE TO SAMPLE MEAN

The sum of squares decreased, so the standard deviation is too small



X	$\bar{X}$	d	d^2
11	13	-2	4
12	13	-1	1
14	13	1	1
15	13	2	4

Sum of distances = 0

Sum of squared distances = **10**

$$s_{biased} = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{\frac{10}{4}} = 1.58$$

UNBIASED ESTIMATES

- Averaging the squared distances by N – 1 (called the "degrees of freedom") eliminates bias in the estimates

$$s^2 = \frac{\sum(X - \bar{X})^2}{N - 1} = \frac{SS}{N - 1} \qquad s = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}} = \sqrt{s^2}$$

- On average (across many samples), s^2 and s do not systematically miss in one direction

UNBIASED ESTIMATES

Estimates do not systematically miss in one direction

Population

$\sigma = 1$

0.7 — 0.8 — 0.9 — 1 — 1.1 — 1.2 — 1.3

s = 0.8 | s = 0.9 | s = 1.0 | s = 1.1 | s = 1.2

Samples

THE VARIANCE

- The average squared distance from the smoking scores to the mean is 51.25

$$s^2 = \frac{\sum(X - \bar{X})^2}{N - 1} = \frac{SS}{N - 1} = \frac{410}{8} = 51.25$$

- The biased formula gave a smaller estimate (45.6) because it is systematically to low

X	$\bar{X}$	d	d^2
19	10	9	81
19	10	9	81
17	10	7	49
14	10	4	16
7	10	-3	9
5	10	-5	25
4	10	-6	36
3	10	-7	49
2	10	-8	64

Sum of squares = 410

Variance (average of squared distances) = 51.25

THE STANDARD DEVIATION

- The average (typical) distance from the smoking scores to the mean is 7.16 cigarettes

$$s = \sqrt{s^2} = \sqrt{51.25} = 7.16$$

- The biased formula gave a smaller estimate (6.75) because it is systematically to low

X	$\bar{X}$	d	d^2
19	10	9	81
19	10	9	81
17	10	7	49
14	10	4	16
7	10	-3	9
5	10	-5	25
4	10	-6	36
3	10	-7	49
2	10	-8	64

Sum of squares = 410

Variance (average of squared distances) = 51.25

DEGREES OF FREEDOM

- The $N - 1$ term is called the degrees of freedom
- The degrees of freedom is the adjusted sample size needed to get an unbiased estimate of variation
- Degrees of freedom adjustment terms are common in statistical formulas (e.g., t-tests, ANOVA, regression)

DF AS NUMBER OF UNIQUE DATA POINTS

- The degrees of freedom can also be understood as the number of unique data points that contribute to an estimate
- After estimating the sample mean (and plugging it into the sum of squares formula), we have N – 1 unique data points with which to estimate variation
- Using the sample mean in an equation reduces the amount of available information in the data

DEGREES OF FREEDOM EXAMPLE

- Suppose that a sample of $N = 3$ scores has a mean of $\bar{X} = 10$ (there are $N - 1 = 2$ degrees of freedom)
- Two of the scores $(N - 1)$ can take on any value, but the third is fully determined after we estimate $\bar{X}$ (the amount of unique information available in the data has been reduced)
- If two of the scores are 8 and 11, then the third must be 11; if two of the scores are 12 and 17, then the third must be 1; if two of the scores are 8 and 9, then the third must be 13

STATISTICAL MONEY ANALOGY

- The sample data are statistical “money” in a bank account
- Whenever we estimate a parameter like $\bar{X}$, we must pay for that estimate with a single degree of freedom
- The value $N - 1$ is the amount of statistical money left in our bank account that we can spend on estimating additional parameters, such as the variance or standard deviation

OUTLINE

1. Measures of spread: Range, standard deviation, and variance
2. Statistical properties of the variance and standard deviation
3. Psychology application: Smoking cessation clinical trial
4. Standardizing: Expressing scores in standard deviation units
5. Study questions

SMOKING AND DRINKING CESSATION TRIAL

Pharmacological treatments that can concomitantly address cigarette smoking and heavy drinking stand to improve health care delivery for these highly prevalent co-occurring conditions. This superiority trial compared the combination of varenicline and naltrexone against varenicline alone for smoking cessation and drinking reduction among heavy-drinking smokers.

Lara Ray

ReJoyce Green

Ray, L.A., Green, R., Enders, C., et al. (2021). Efficacy of combining varenicline and naltrexone for smoking cessation and drinking reduction: A randomized clinical trial. *American Journal of Psychiatry*, *178*, 818–828.

KEY VARIABLES

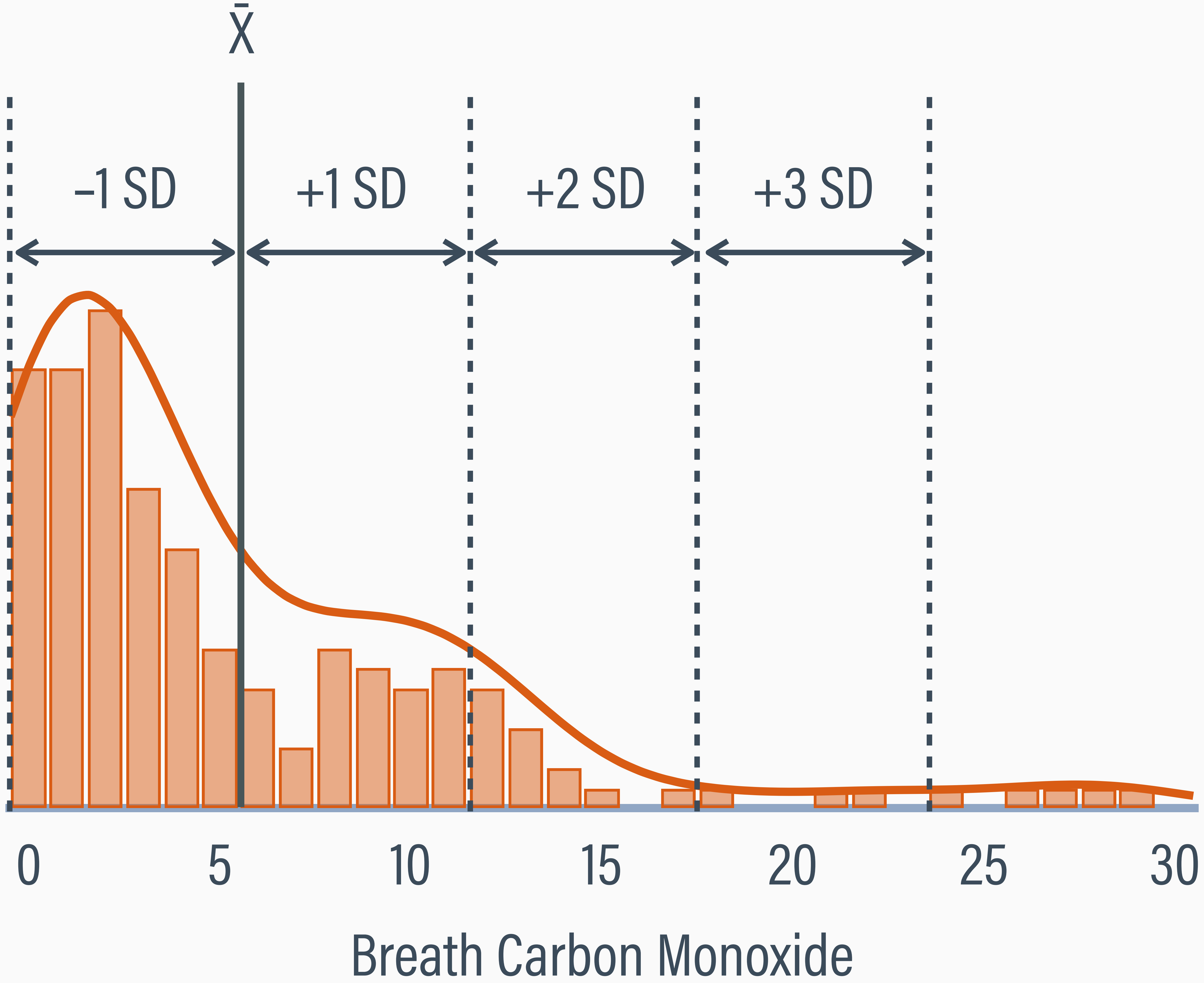
Breath (alveolar) carbon monoxide

A measure of carbon monoxide in the lungs. Breath carbon monoxide is a biomarker of smoking behavior common in clinical trials. Higher scores reflect more frequent smoking.

Medication arm

Participants were randomly assigned to receive one of two meds: varenicline plus naltrexone or varenicline plus placebo pills

DESCRIPTIVE STATISTICS



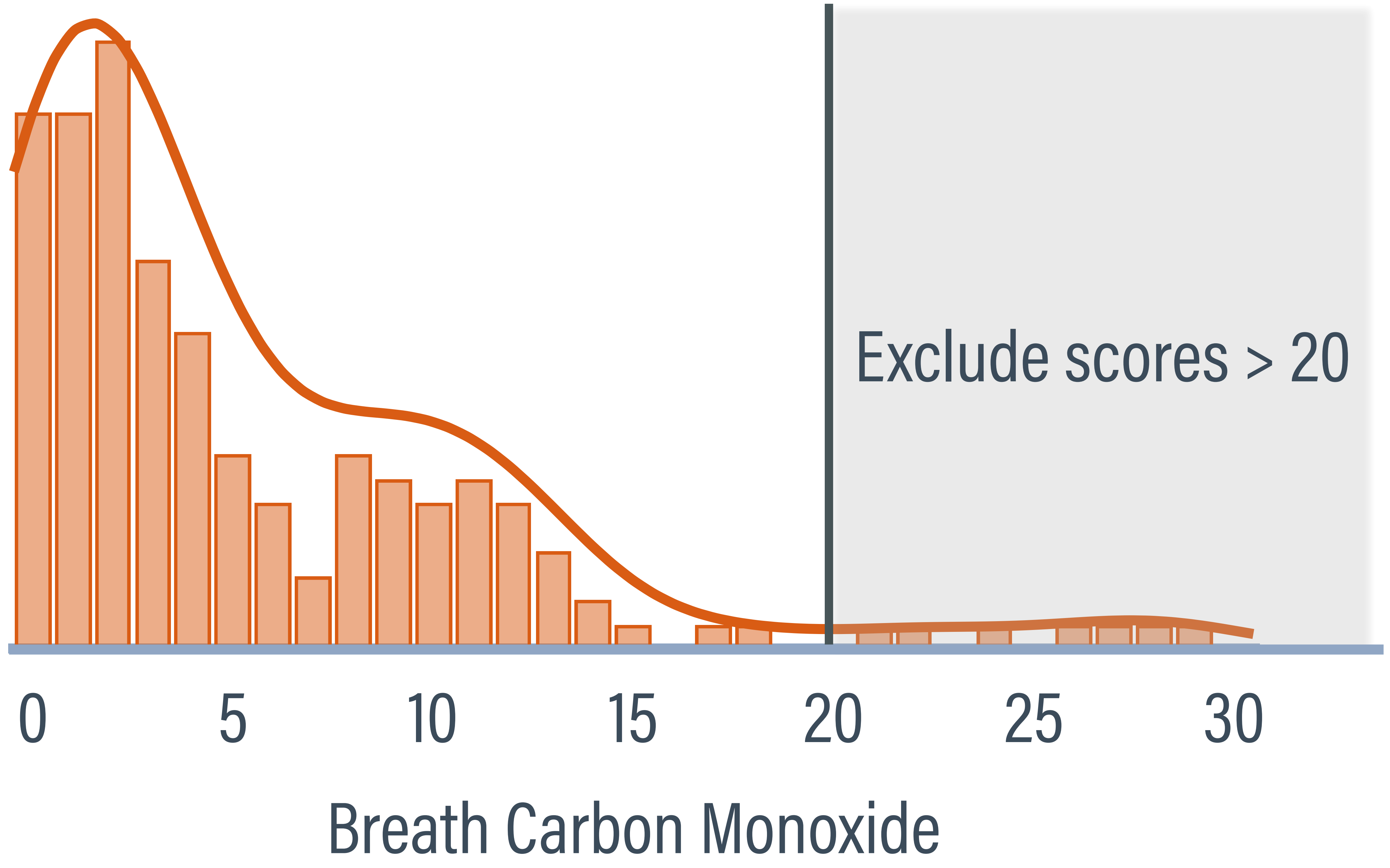
Mean ($\bar{X}$)	5.53
Std. Dev. (s)	5.96
N	165

R OUTPUT

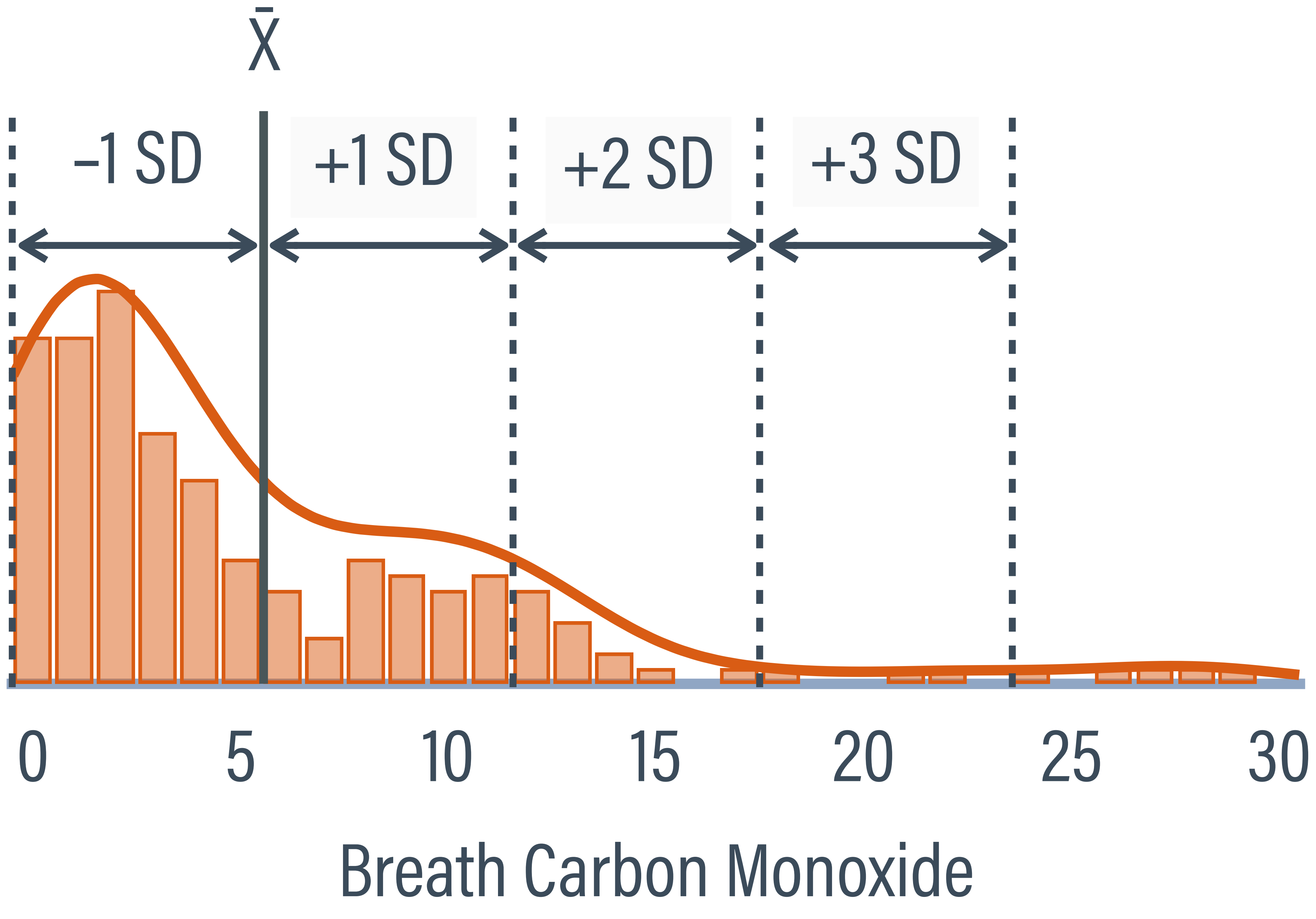
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Participant	1	165	83.00	47.78	83	83.00	60.79	1	165	164	0.00	-1.22	3.72
Condition*	2	165	1.50	0.50	2	1.50	0.00	1	2	1	-0.01	-2.01	0.04
Gender*	3	165	1.61	0.49	2	1.63	0.00	1	2	1	-0.43	-1.83	0.04
COWeek0	4	165	10.55	6.86	9	9.82	5.93	0	46	46	1.41	3.60	0.53
COWeek4	5	165	5.46	5.10	4	4.69	4.45	0	24	24	1.24	0.96	0.40
COWeek8	6	165	5.53	5.96	3	4.55	2.97	0	29	29	1.75	3.35	0.46
QuitCigsWeek4*	7	165	1.62	0.49	2	1.65	0.00	1	2	1	-0.48	-1.78	0.04
QuitCigsWeek8*	8	165	1.64	0.48	2	1.68	0.00	1	2	1	-0.59	-1.66	0.04
DrinksWeek0	9	165	6.40	4.42	6	5.77	2.97	1	35	34	2.44	10.46	0.34
DrinksWeek4	10	165	3.59	2.98	3	3.30	2.97	0	13	13	0.80	0.30	0.23
DrinksWeek8	11	165	3.23	2.68	3	2.97	2.97	0	13	13	0.89	0.67	0.21
CigsWeek0	12	165	14.22	8.22	12	13.05	5.93	3	51	48	1.59	3.20	0.64
CigsWeek4	13	165	4.18	5.70	2	3.08	2.97	0	41	41	3.01	13.13	0.44
CigsWeek8	14	165	3.16	4.75	2	2.25	2.97	0	35	35	3.44	16.81	0.37

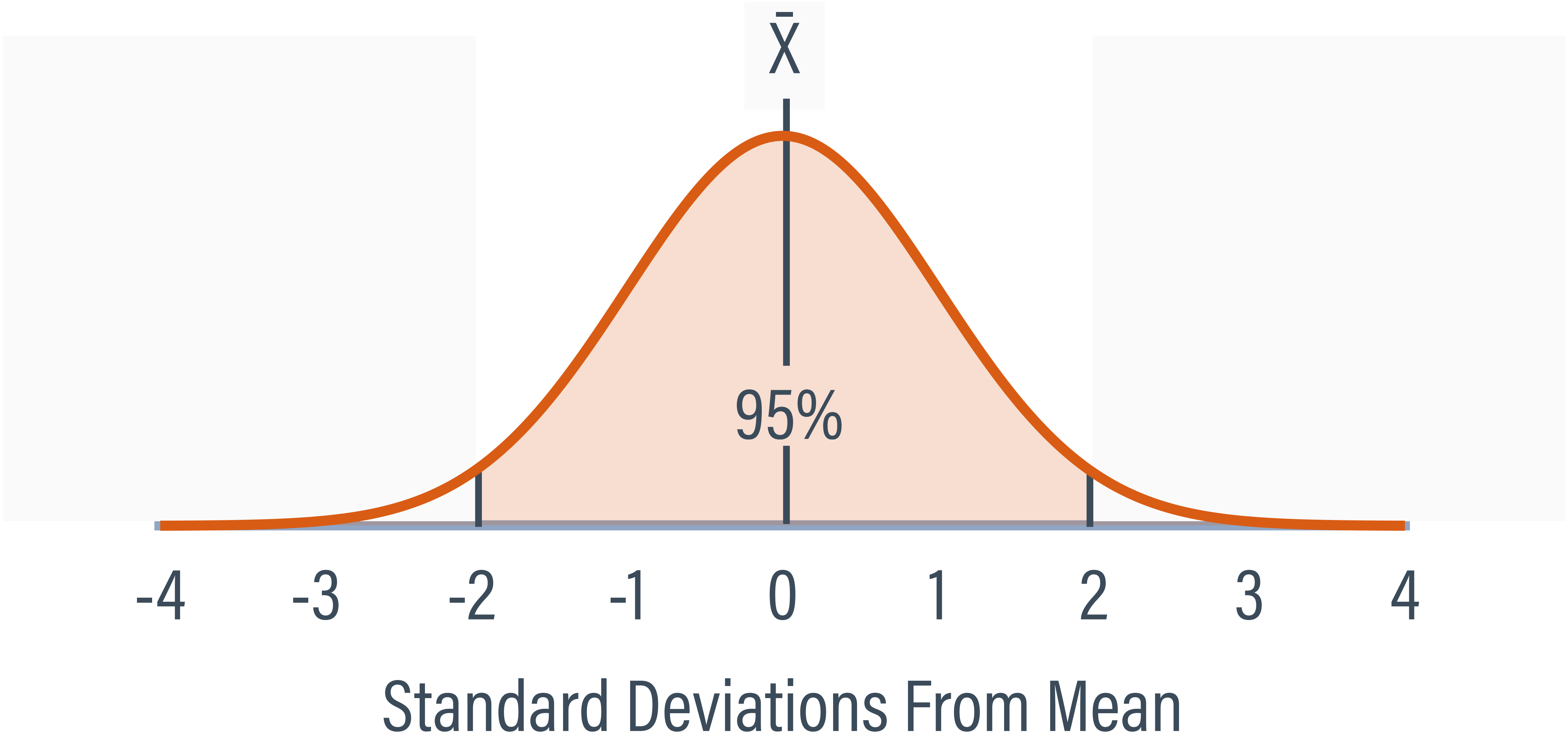
In small groups of two or three, interpret the standard deviation in practical terms. Your explanation should use the numeric value from the example, and it should be understandable to a non-technical audience without losing accuracy.

In small groups of two or three, discuss what would happen to the standard deviation if you excluded outlier participants with breath CO scores > 20?



In small groups of two or three, discuss the utility of the rule of thumb that 95% of scores fall within ± 2 standard deviations of the mean

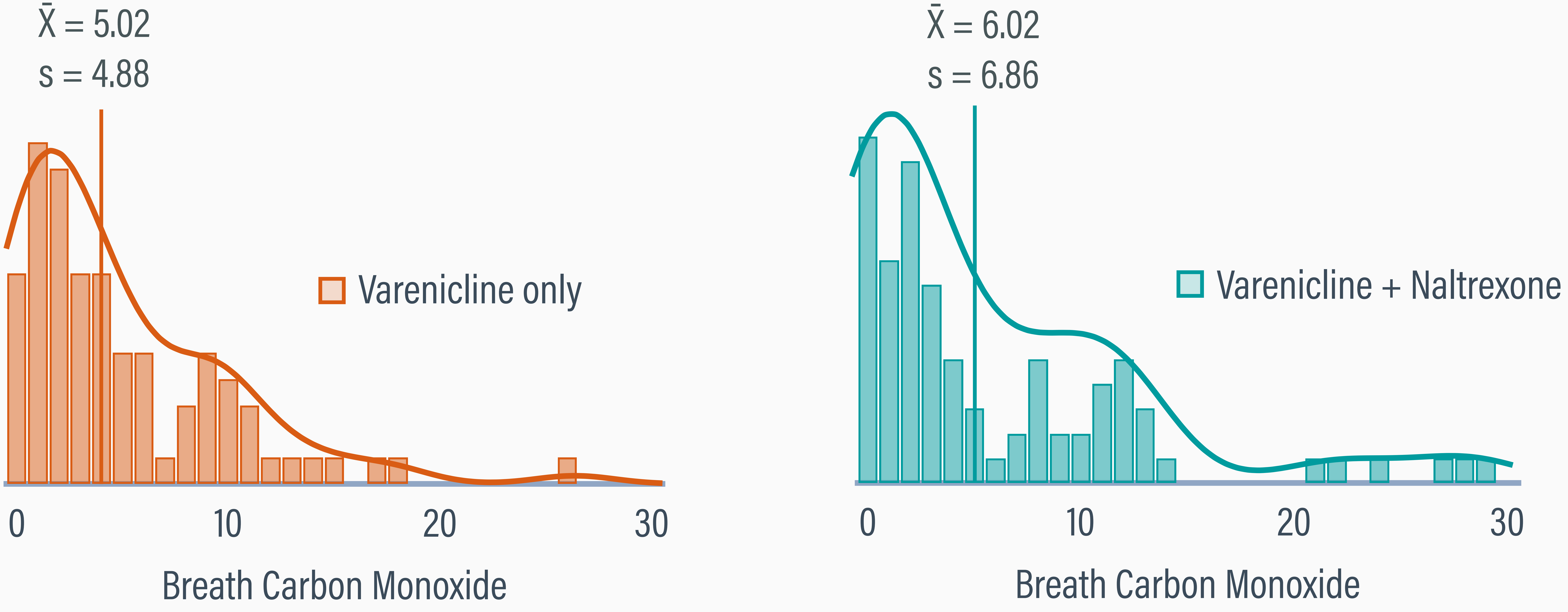




COMPARATIVE RESEARCH QUESTIONS

- Comparative research questions ask whether two or more groups (or occasions) differ from one another
- Question: Do participants in the two treatment groups differ in their smoking levels?
- We can answer this question by comparing descriptive statistics

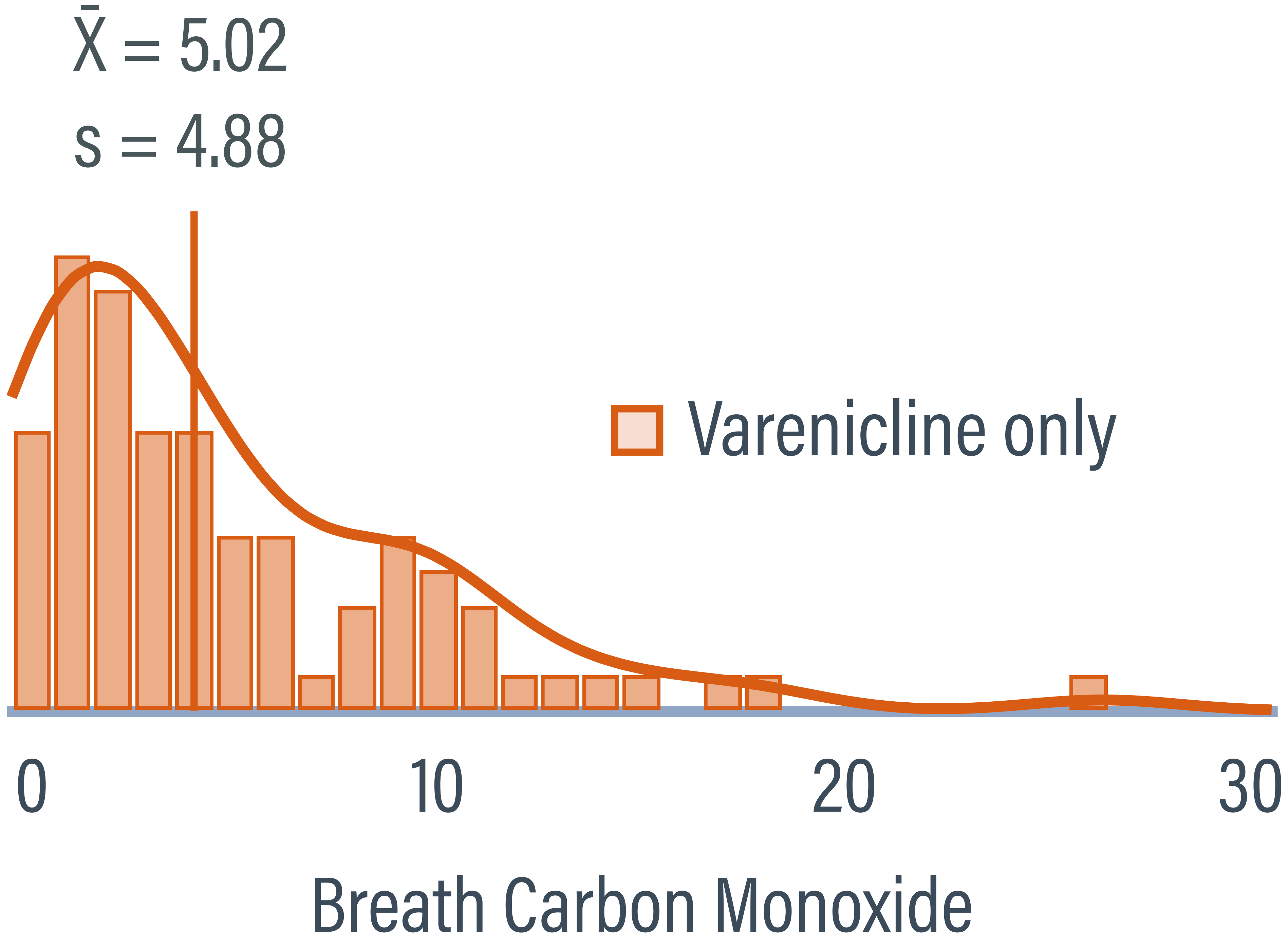
GROUP STATISTICS

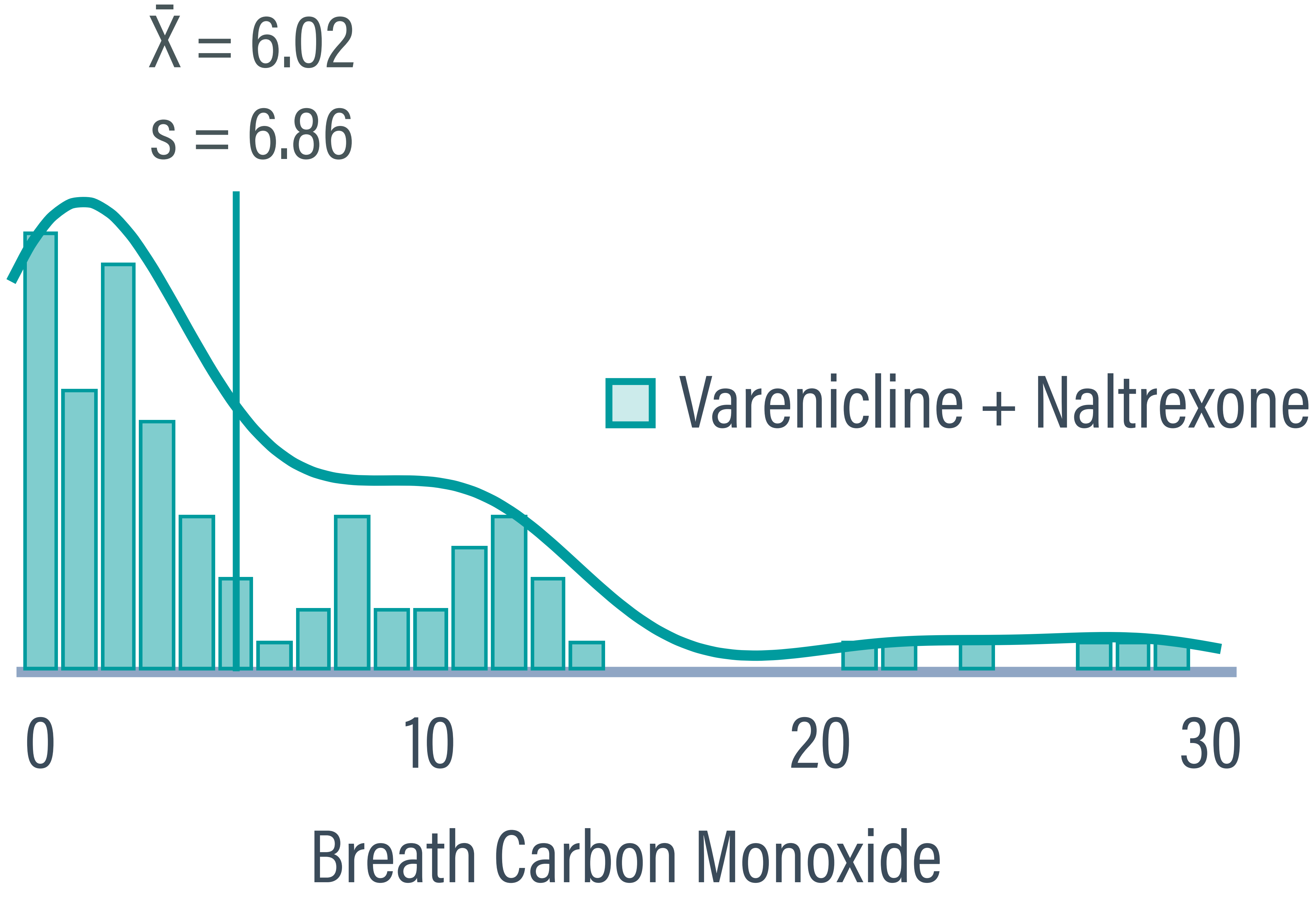


R OUTPUT

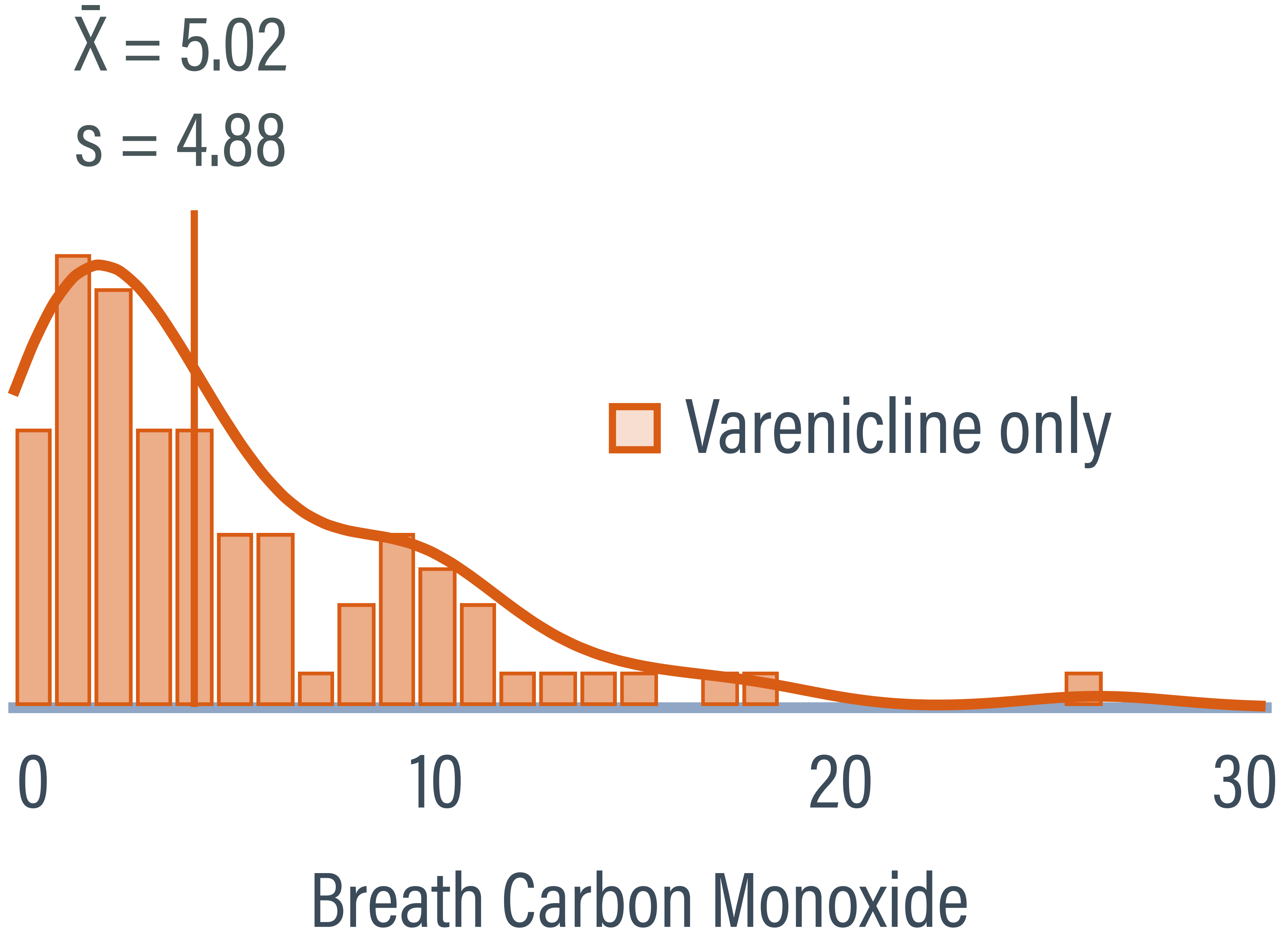
```
 Descriptive statistics by group
Condition: Varenicline
        vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
COWeek8    1 82 5.02 4.88    3.5    4.33 3.71   0  26    26 1.57     3.09 0.54
--------------------------------------------------------------------------------
Condition: Varenicline + Naltrexone
        vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
COWeek8    1 83 6.02 6.86      3    4.81 4.45   0  29    29 1.62     2.31 0.75
```

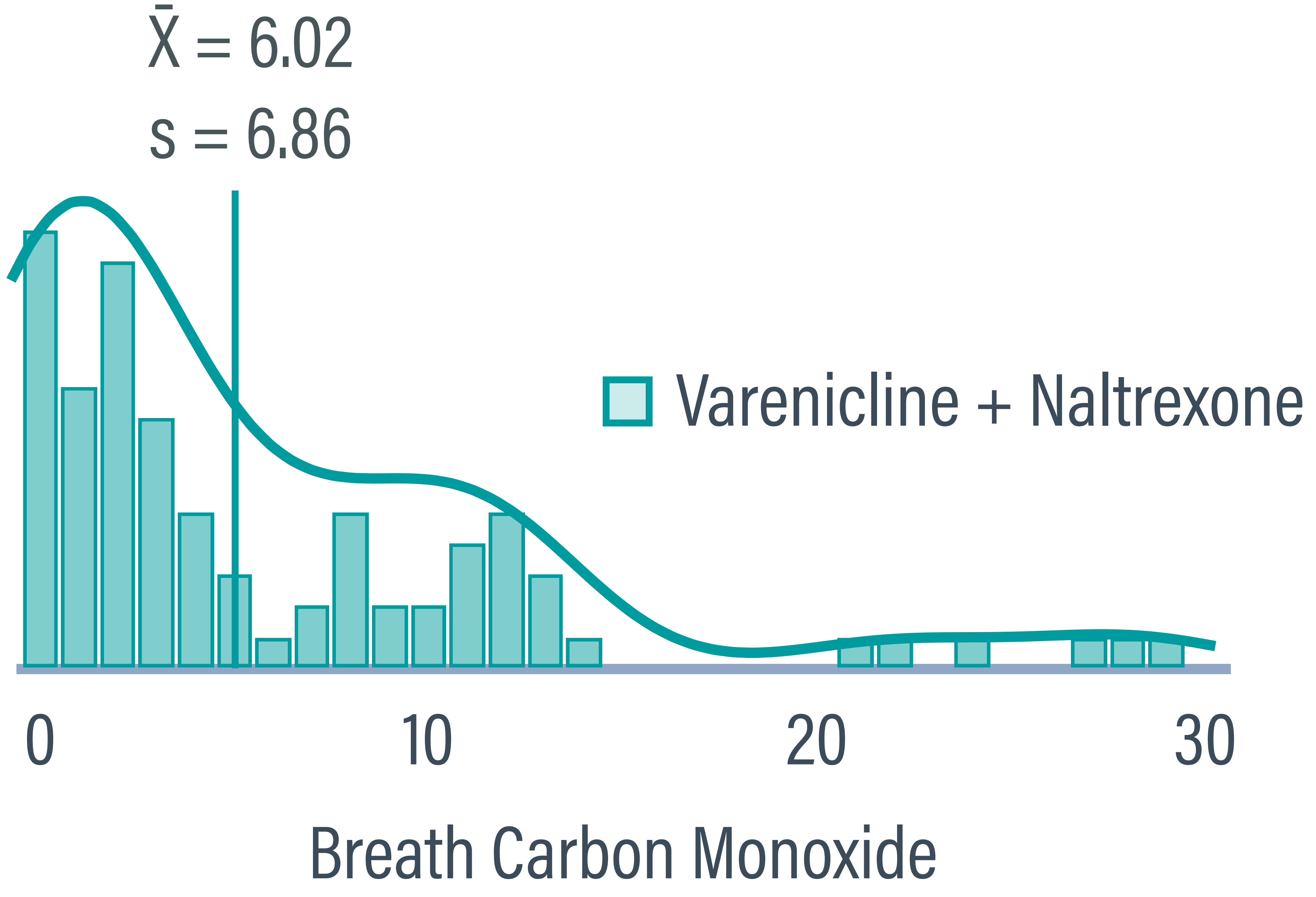
In small groups of two or three, discuss the standard deviation differences. What feature of the distribution causes one group's standard deviation to be smaller than the other's?





Later analyses we will learn (t-tests, ANOVA) assume that all groups have the same variation. In small groups of two or three, discuss whether the difference in standard deviations is small enough to treat them as approximately equal. This is a subjective evaluation, provide a rationale.





OUTLINE

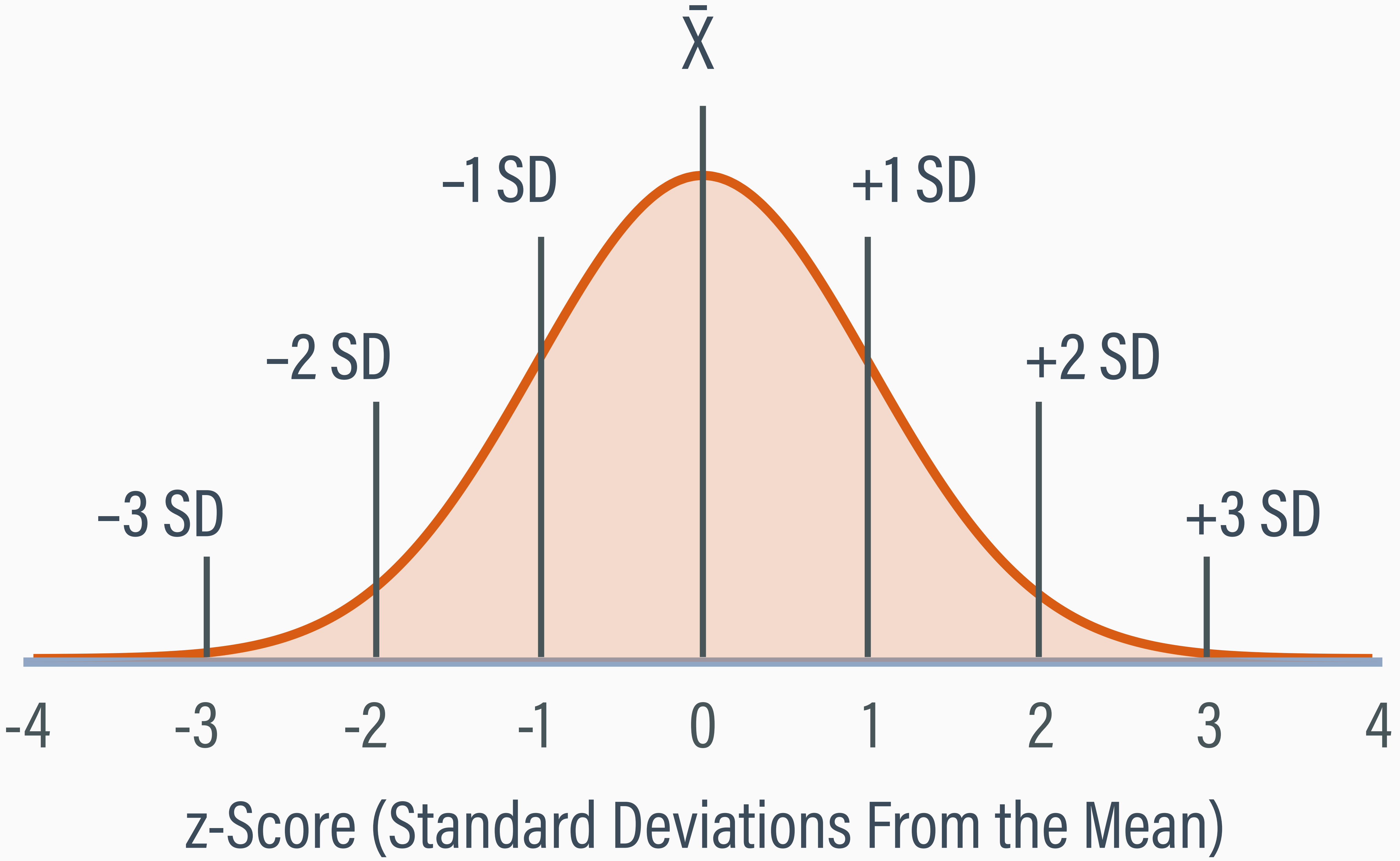
1. Measures of spread: Range, standard deviation, and variance
2. Statistical properties of the variance and standard deviation
3. Psychology application: Smoking cessation clinical trial
4. Standardizing: Expressing scores in standard deviation units
5. Study questions

STANDARDIZED SCORES

- Many physical variables (e.g., age, height, weight) have natural metrics that are easy to understand
- Psychological variables (e.g., depression, narcissism) are often use sum scores that do not have inherent metrics
- To facilitate interpretation, researchers often use standard scores with “universal” metrics (e.g., T-scores, IQ, GRE)

z-SCORES

- The z-score scale is a common standardized metric that expresses scores as standard deviation units from the mean

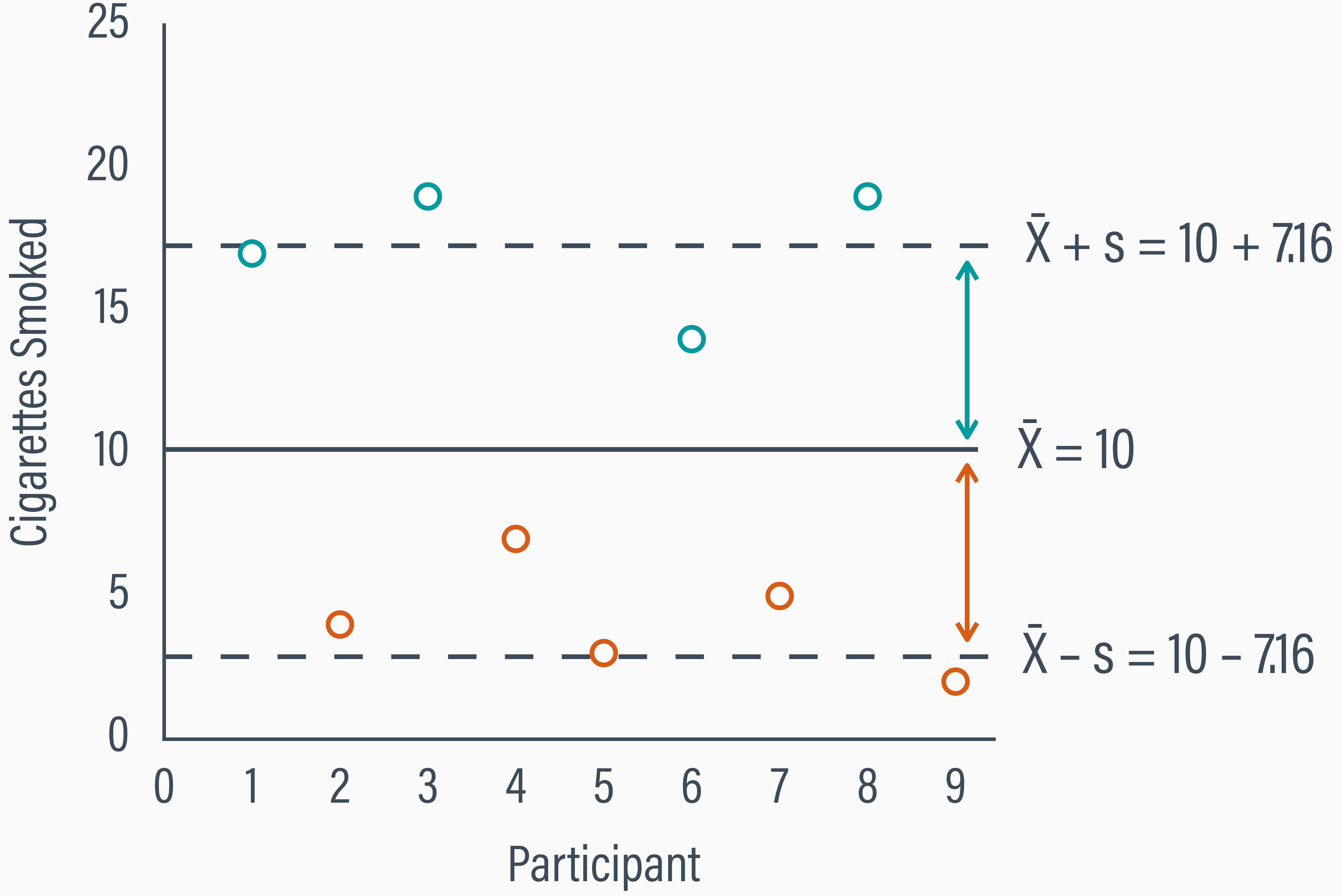


SMALL SAMPLE ILLUSTRATION

- To illustrate standardization, return to the earlier example where N = 9 participants reported the typical number of cigarettes that they smoked in the last week
- The mean was $\bar{X} = 10$, and the standard deviation was $s = 7.16$

X
19
19
17
14
7
5
4
3
2

VISUALIZING z-SCORES



Standard deviation units from the mean

X	z
19	1.26
19	1.26
17	0.98
14	0.56
7	−0.42
5	−0.7
4	−0.84
3	−0.98
2	−1.12

z-SCORE FORMULA

- The numerator captures the raw score's distance (deviation) from the sample mean $\bar{X}$

$$z = \frac{X - \bar{X}}{s} = \frac{\text{distance from mean}}{\text{standard deviation}}$$

- Dividing by the standard deviation converts the distance to standard deviation units

FORMULA ILLUSTRATION

- A score of 17 is about 0.98 standard deviations above the sample mean ($\bar{X} = 10$), and a score of 4 is roughly 0.84 standard deviation units below the mean

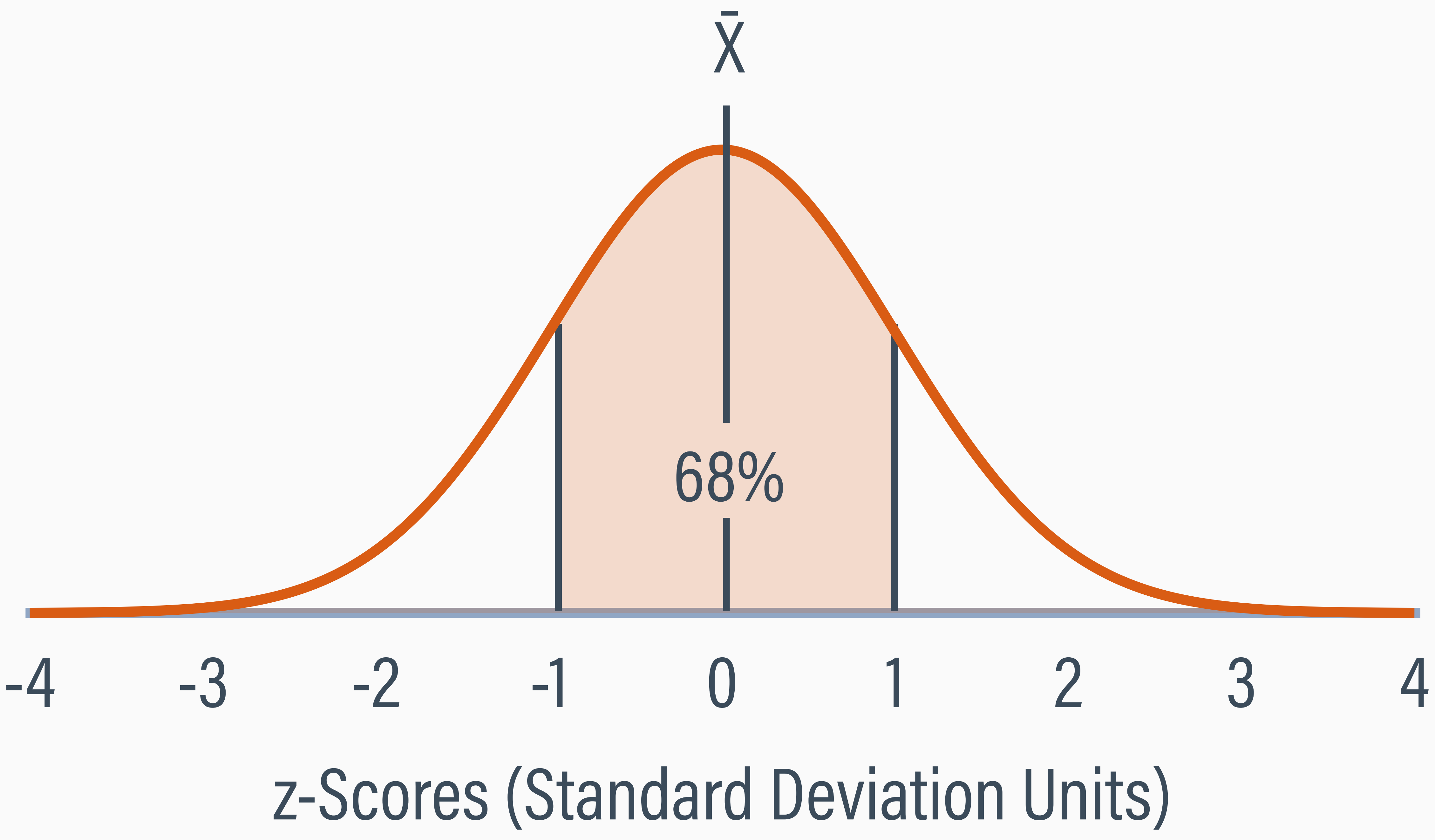
$$z = \frac{X - \bar{X}}{s} = \frac{17 - 10}{7.16} = \frac{+7}{7.16} = +0.98$$

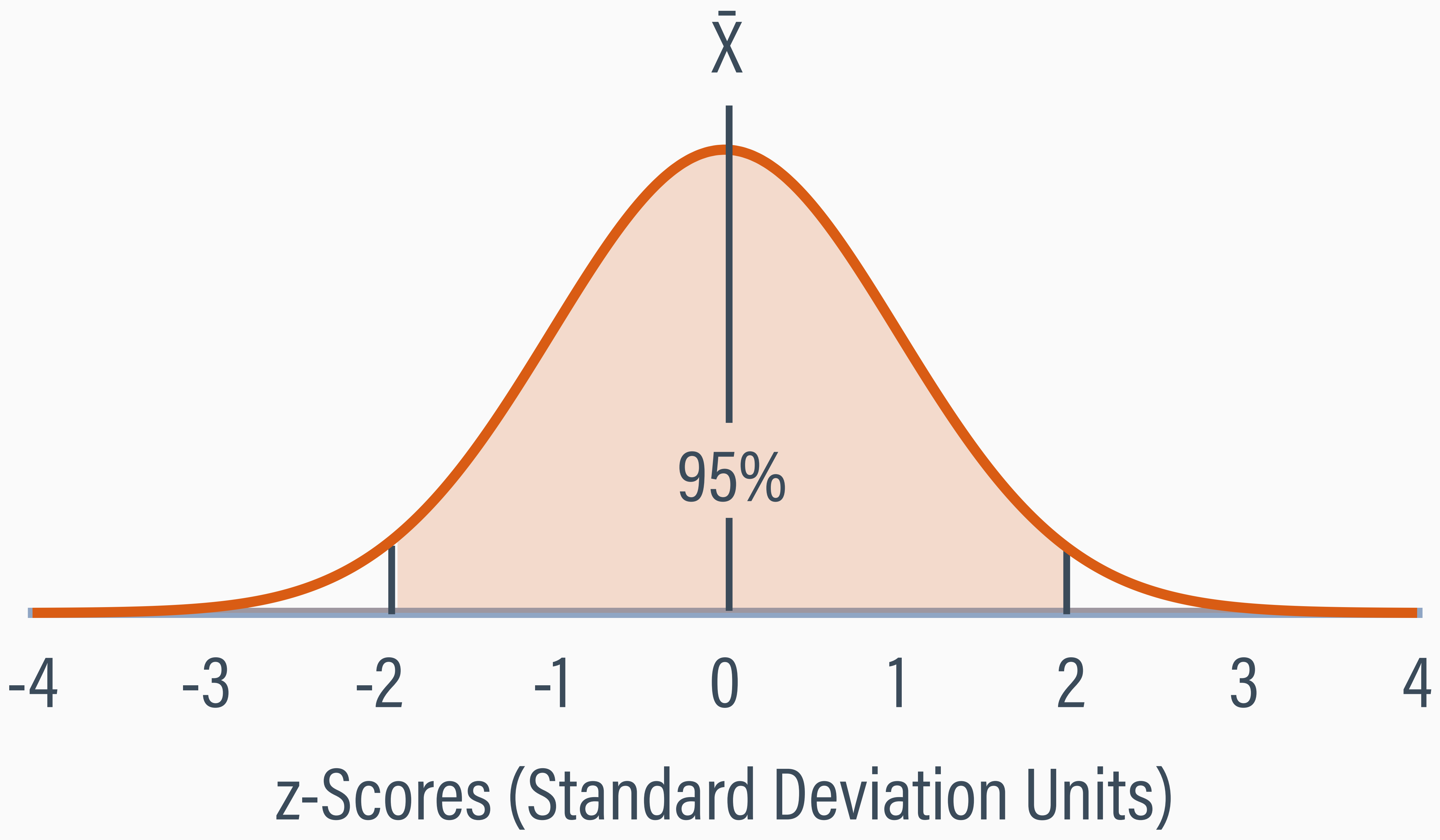
$$z = \frac{X - \bar{X}}{s} = \frac{4 - 10}{7.16} = \frac{-6}{7.16} = -0.84$$

X	z
19	1.26
19	1.26
17	0.98
14	0.56
7	−0.42
5	−0.7
4	−0.84
3	−0.98
2	−1.12

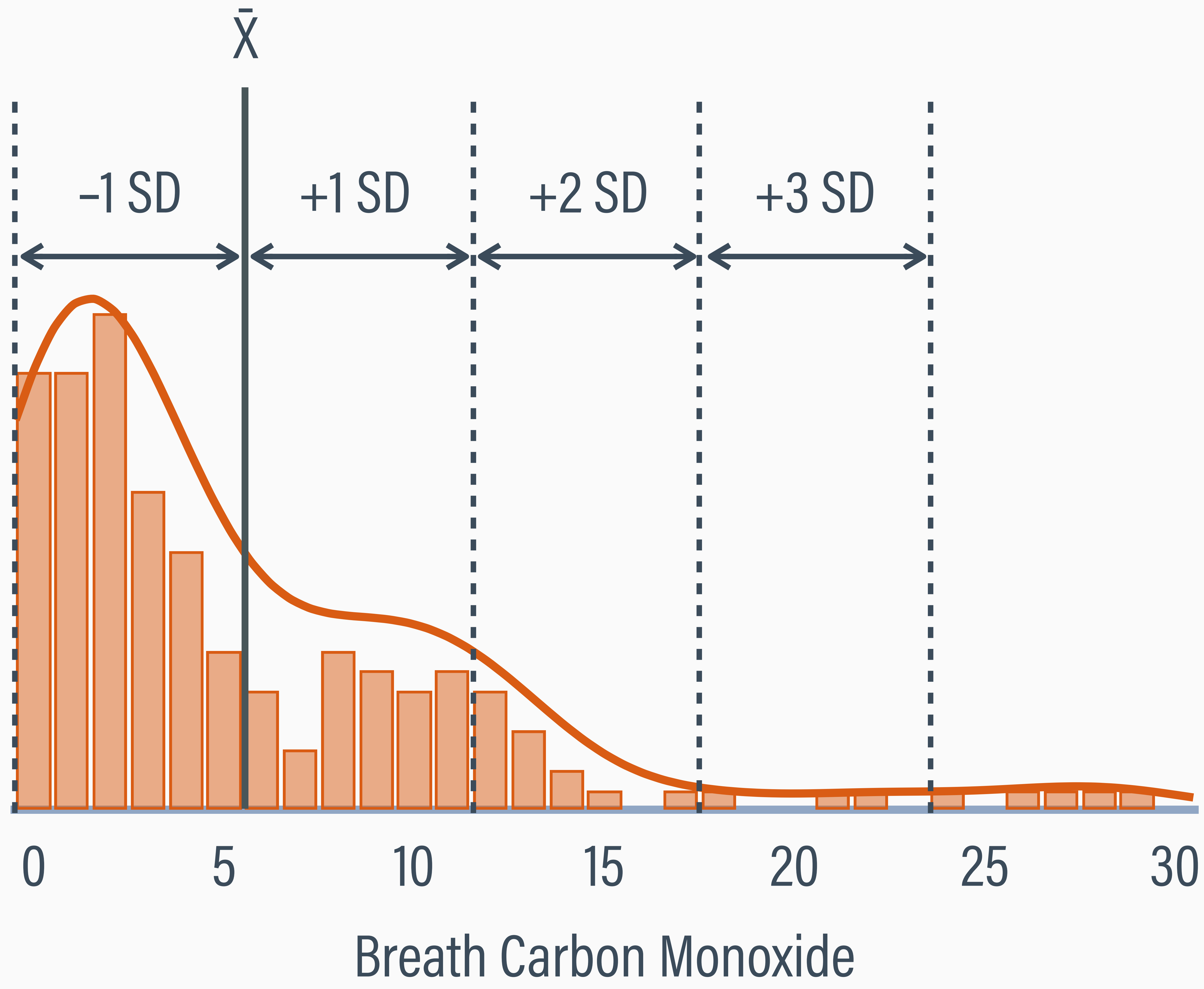
RULE OF THUMB FOR NORMAL DATA

- In a normal curve, 68% of the scores are within ± 1 z-scores (standard deviation units) of the mean, and 95% are within ≈ ± 2





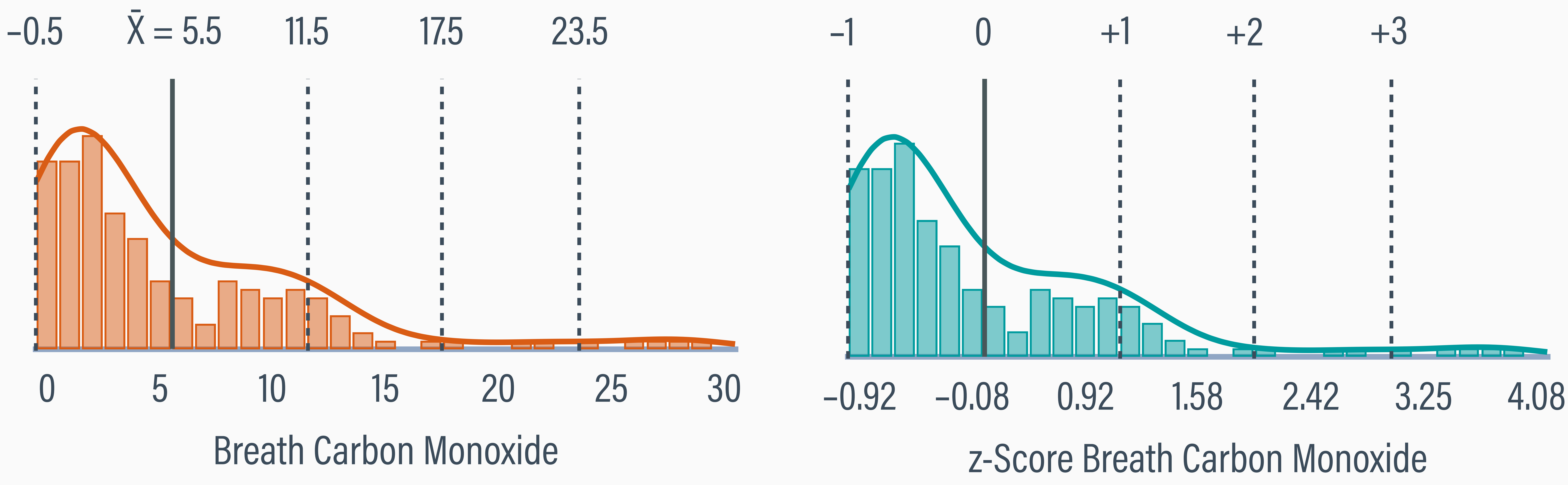
CLINICAL TRIAL DESCRIPTIVES REVISITED



Mean ($\bar{X}$)	5.5
Std. Dev. (s)	6
N	165

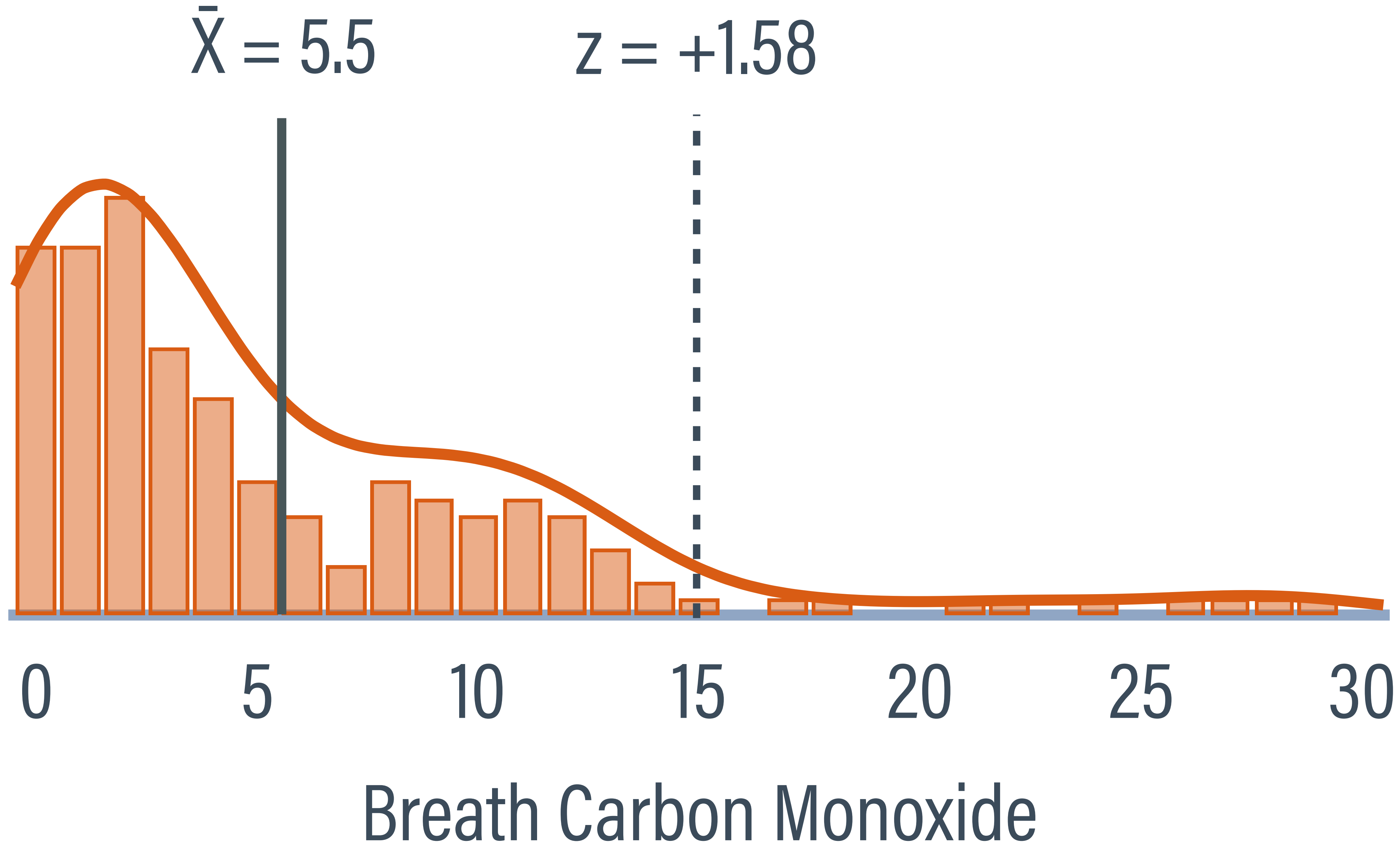
z-SCORES AND DISTRIBUTION SHAPE

- The z-score transformation changes the mean to 0 and standard deviation to 1, but it does not change the distribution's shape



R OUTPUT

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
COWeek8	1	165	5.53	5.96	3.00	4.55	2.97	0.00	29.00	29.00	1.75	3.35	0.46
zCOWeek8	2	165	0.00	1.00	-0.42	-0.16	0.50	-0.93	3.94	4.87	1.75	3.35	0.08



The mean and standard deviation are $\bar{X} = 5.5$ and $s = 6$. The z-score for someone with a breath CO value of 15 is about $z = 1.58$. In small groups of two or three, interpret the z-score in practical terms. Your explanation should reference both the sign and the magnitude of the value. How extreme is this score?

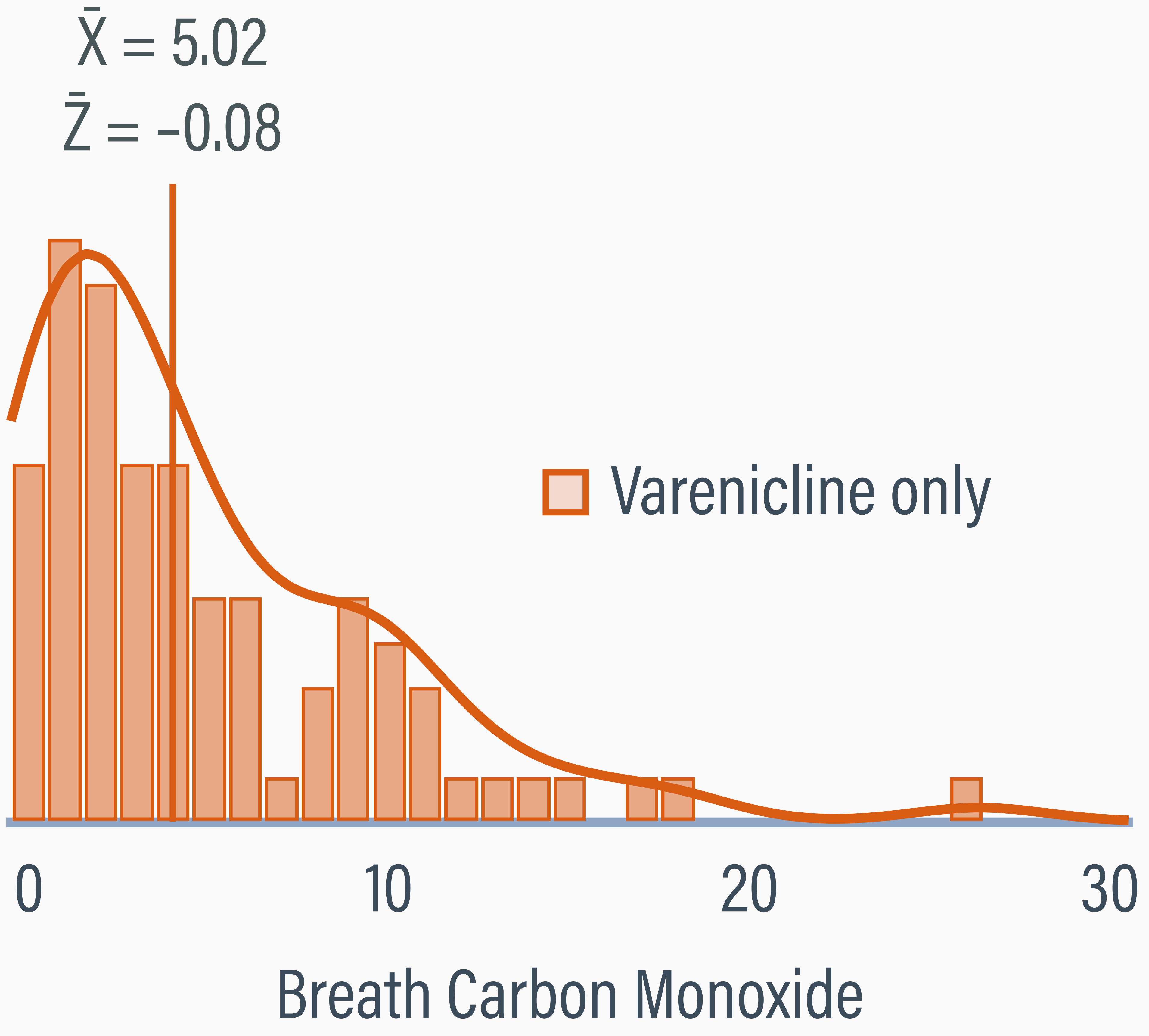
COMPARATIVE RESEARCH QUESTIONS

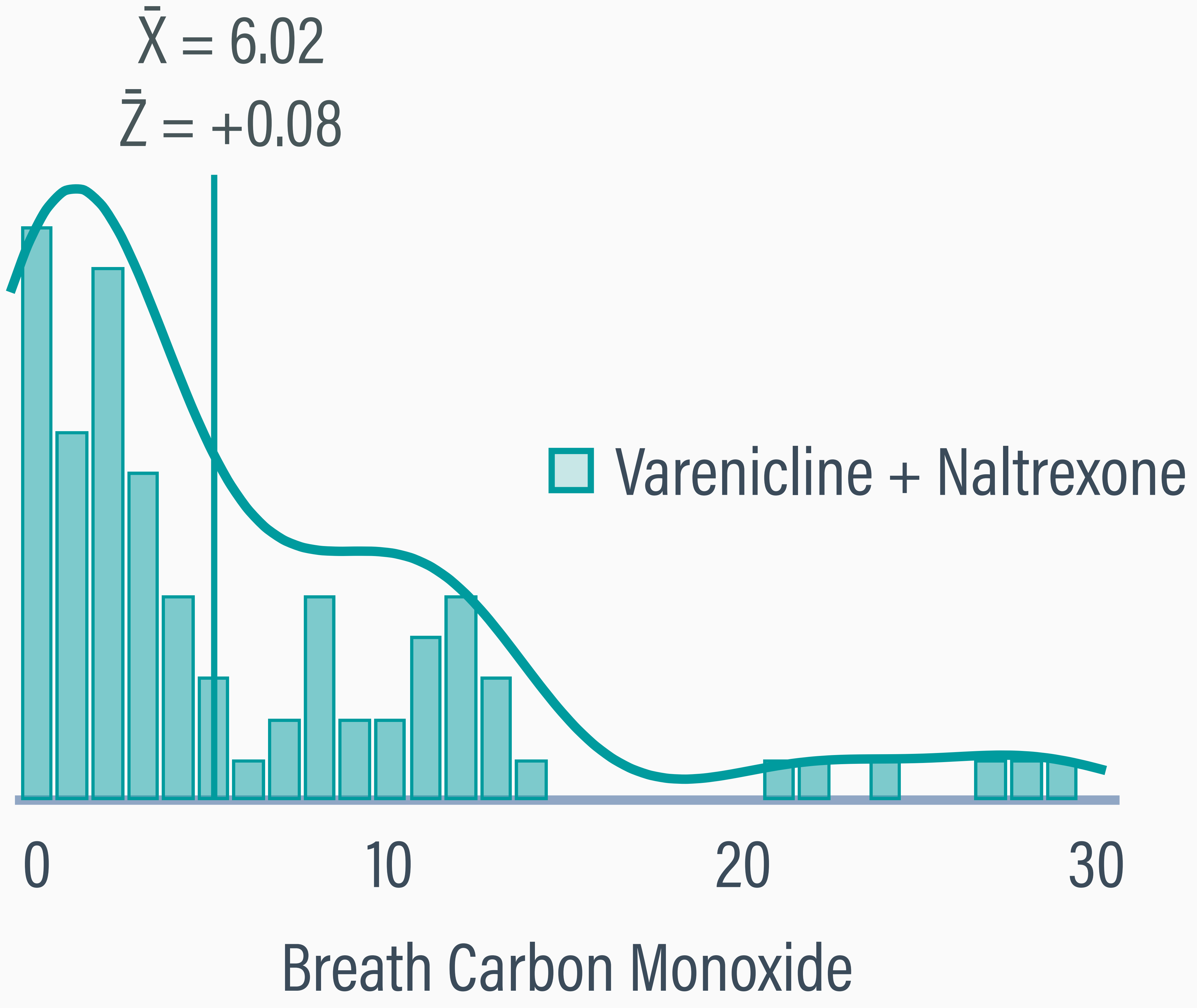
- Comparative research questions ask whether two or more groups (or occasions) differ from one another
- Question: Do participants in the two treatment groups differ in their smoking levels?
- We can answer this question by comparing descriptive statistics

STANDARDIZED MEAN DIFFERENCE PREVIEW

- It is often difficult to assess the magnitude of a group mean difference on the raw metric of the data
- A common approach is to express the mean difference in standard deviation (z-score) units
- The mean difference in z-score units is called the standardized mean difference effect size

GROUP MEANS AS z-SCORES

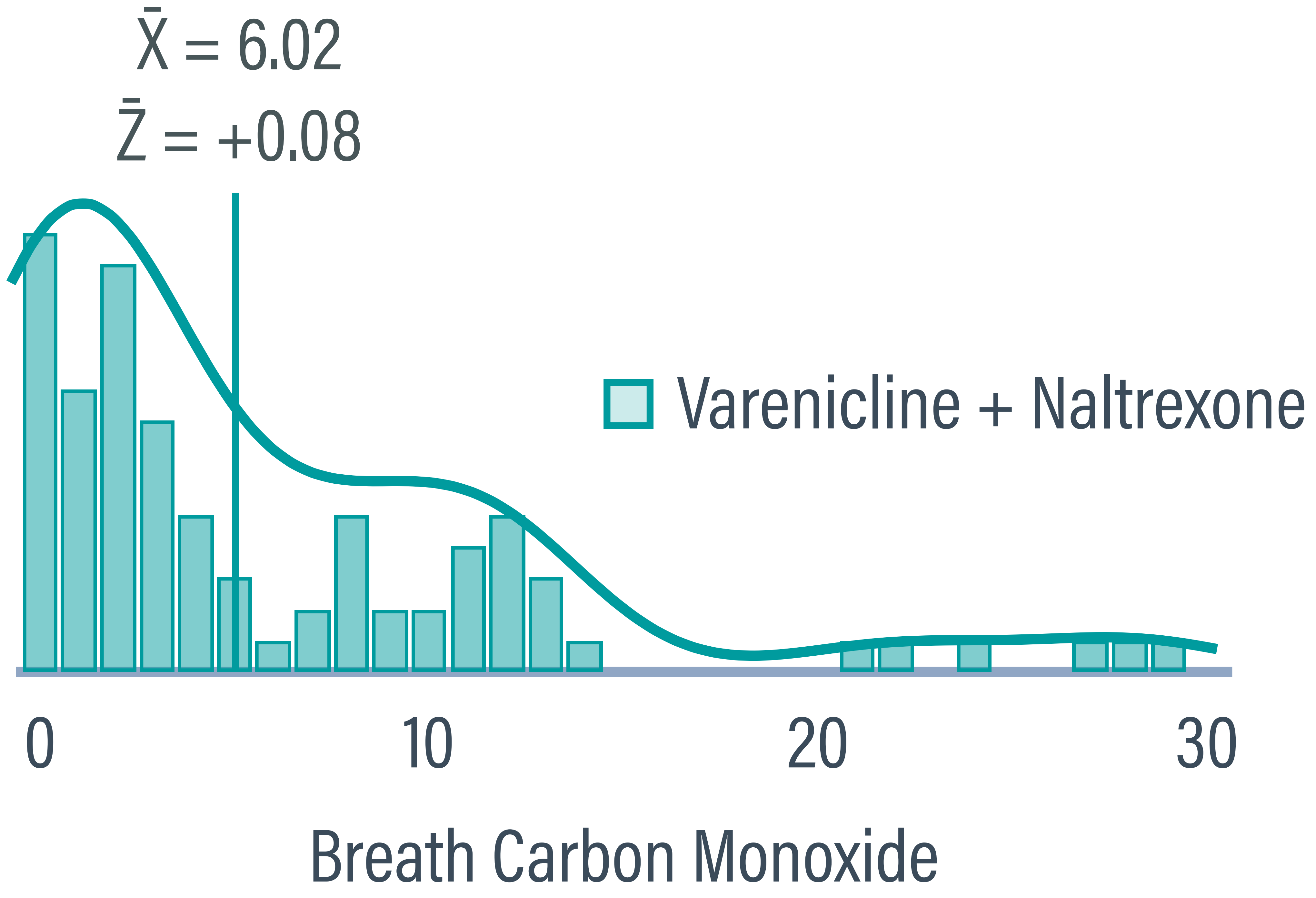


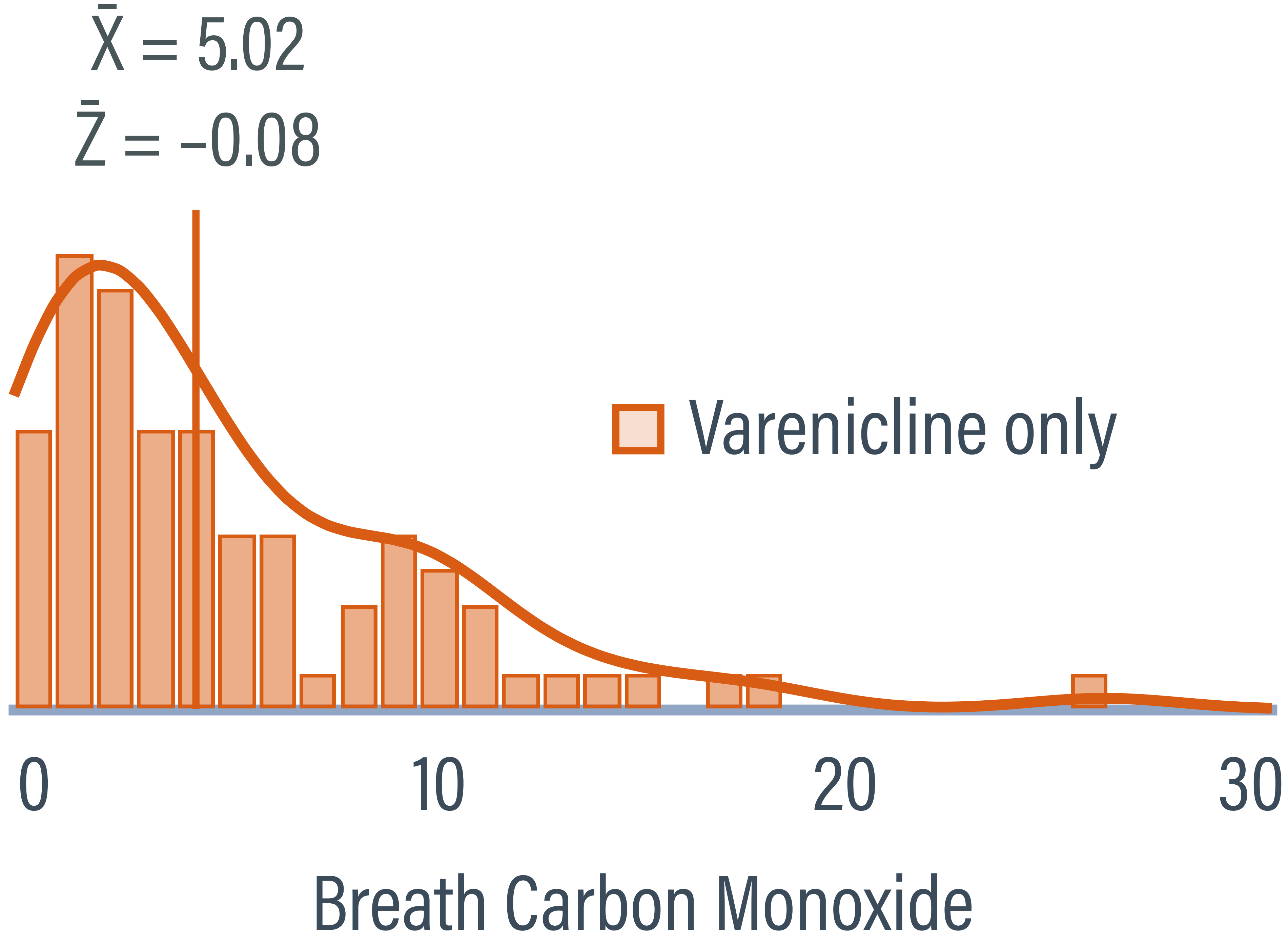


R OUTPUT

```
 Descriptive statistics by group
Condition: Varenicline
         vars  n  mean   sd median trimmed  mad   min  max range skew kurtosis   se
zCOWeek8    1 82 -0.08 0.82  -0.34    -0.2 0.62 -0.93 3.44  4.36 1.57     3.09 0.09
---------------------------------------------------------------------------------
Condition: Varenicline + Naltrexone
         vars  n mean   sd median trimmed  mad   min  max range skew kurtosis   se
zCOWeek8    1 83 0.08 1.15  -0.42   -0.12 0.75 -0.93 3.94  4.87 1.62     2.31 0.13
```

One of the main goals of the research study is to determine whether the treatment groups differ. In small groups of two or three, discuss the size of the mean difference on the z-score metric (about 0.16 standard deviation units). How would you gauge the magnitude of the standardized difference? Is it meaningful? This is a subjective evaluation, so provide a rationale.





OUTLINE

1. Measures of spread: Range, standard deviation, and variance
2. Statistical properties of the variance and standard deviation
3. Psychology application: Smoking cessation clinical trial
4. Standardizing: Expressing scores in standard deviation units
5. Study questions

STUDY QUESTIONS (1)

1. Make up a sample of five age values that exhibit high variability.

2. Make up a sample of five age values that exhibit very little variability.

3. Make up a sample of five age values with a standard deviation equal to zero.

STUDY QUESTIONS (2)

4. What is a distance or deviation score? Suppose that the mean of a sample of depression scores is 19. Illustrate a deviation score for an individual with a high depression score. Do the same for a low depression score.

5. Why do we need to square deviation scores when computing variability?

STUDY QUESTIONS (3)

6. The Beck Depression Inventory scoring manual reports that the mean and standard deviation of a mildly depressed normative sample are 19 and 6, respectively. Provide an interpretation of the standard deviation. The variance of depression scores is $6^2 = 36$. Provide an interpretation of the variance.

STUDY QUESTIONS (4)

7. The standard deviation estimate computed by dividing by N in the denominator is biased. Describe the concept of bias in this context.

8. Describe the degrees of freedom adjustment and its purpose in the standard deviation formula.

STUDY QUESTIONS (5)

9. The Beck Depression Inventory scoring manual reports that the mean and standard deviation of a mildly depressed normative sample are 19 and 6, respectively. How would you characterize the depression for someone with a z-score of -0.10? Interpret the z-score, and describe what its sign and magnitude tell you about the location of the depression score in the distribution.