

MODULE 1

DESCRIBING DATA

OUTLINE

- 1 Measuring psychological constructs
- 2 Distribution shapes for numeric variables
- 3 Graphical displays for numeric variables
- 4 Graphical and tabular displays for discrete and categorical variables
- 5 Study questions

OUTLINE

- 1 Measuring psychological constructs
- 2 Distribution shapes for numeric variables
- 3 Graphical displays for numeric variables
- 4 Graphical and tabular displays for discrete and categorical variables
- 5 Study questions

UVEAL MELANOMA and DEPRESSION

Uveal melanoma, a rare eye cancer, presents potential vision loss and life threat. This prospective, longitudinal study interrogated the predictive utility of visual impairment, as moderated by optimism/pessimism, on depressive symptoms in 299 adults undergoing diagnostic evaluation.

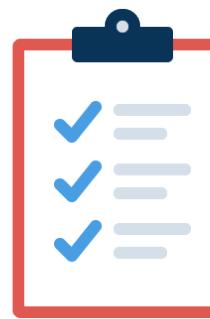


Annette
Stanton

James
MacDonald

MacDonald, J.J., Jorge-Miller, A., Enders, C.K., McCannel, T., Beran, T., & Stanton, A.L. (2021.. Perceived and objective visual impairment predicting depressive symptoms across one year in uveal melanoma diagnostic biopsy: Optimism and pessimism as moderators. *Health Psychology, 40*, 408-417.

KEY VARIABLES



Depressive Symptoms

The CES-D is a 20-item inventory that asks people to rate how often they experience depressive symptoms such as restless sleep, poor appetite, and feeling lonely.



Dispositional Optimism

The Life Orientation Test Revised (LOT-R) is a 3-item questionnaire that asks people to rate their tendency to expect positive outcomes in life.

MEASURING DEPRESSIVE SYMPTOMS

- The article's description of the measurement instrument:

The 20-item Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977; $\alpha = .89$. was completed at each assessment point. The CES-D has strong psychometric properties in cancer samples (Hann et al., 1999). Higher scores indicate greater depressive symptoms. An established cutoff score for clinically elevated depressive symptoms is ≥ 16 (Andresen et al., 1994).

CES-D INVENTORY

0 = Rarely or none of the time (less than 1 day)

1 = Some or a little of the time (1-2 days)

2 = Occasionally or a moderate amount of the time (3-4 days)

3 = Most or all of the time (5-7 days)

During the past week:

1. I was bothered by things that usually don't bother me 0 1 2 3

2. I did not feel like eating; my appetite was poor 0 1 2 3

3. I felt that I could not shake off the blues 0 1 2 3

4. I felt that I was just as good as other people 0 1 2 3

5. I had trouble keeping my mind on what I was doing 0 1 2 3

6. I felt depressed 0 1 2 3

7. I felt that everything I did was an effort 0 1 2 3

8. I felt hopeful about the future 0 1 2 3

9. I thought my life had been a failure 0 1 2 3

10. I felt fearful 0 1 2 3

11. My sleep was restless 0 1 2 3

12. I was happy 0 1 2 3

13. I talked less than usual 0 1 2 3

14. I felt lonely 0 1 2 3

15. People were unfriendly 0 1 2 3

16. I enjoyed life 0 1 2 3

17. I had crying spells 0 1 2 3

18. I felt sad 0 1 2 3

19. I felt that people disliked me 0 1 2 3

20. I could not get "going" 0 1 2 3

SUM SCORES

- Researchers routinely assign scores to a psychological construct by adding a participant's responses to questionnaire items measuring the same thing
- Referred to as a sum, scale, or composite score
- Sum scores have better reliability and validity, measure a broader range of attributes, and more closely approximate a numeric (interval or ratio) scale

RECENT METHODOLOGY WORK

- Liu, Y., & Pek, J. (2024). Summed versus estimated factor scores: Considering uncertainties when using observed scores. *Psychological Methods*, Advance online publication. <https://doi.org/dx.doi.org/10.1037/met0000644>
- McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52, 2287–2305. <https://doi.org/10.3758/s13428-020-01398-0>
- Rhemtulla, M., & Savalei, V. (2024). Estimated Factor Scores Are Not True Factor Scores. *Multivariate Behavioral Research*, Advanced online publication, 1–22. <https://doi.org/doi.org/10.1080/00273171.2024.2444943>

SUBCLINICAL SUM SCORE

Sum score = $0 + 1 + 1 + 0 + 0 + 2 + 0 + 1 + 0 + 0 + 1 + 1 + 0 + 0 + 1 + 0 + 0 + 0 + 1 = 9$

During the past week:

-
1. I was bothered by things that usually don't bother me 0 1 2 3
2. I did not feel like eating; my appetite was poor 0 1 2 3
3. I felt that I could not shake off the blues 0 1 2 3
4. I felt that I was just as good as other people 0 1 2 3
5. I had trouble keeping my mind on what I was doing 0 1 2 3
6. I felt depressed 0 1 2 3
7. I felt that everything I did was an effort 0 1 2 3
8. I felt hopeful about the future 0 1 2 3
9. I thought my life had been a failure 0 1 2 3
10. I felt fearful 0 1 2 3
-

11. My sleep was restless 0 1 2 3
12. I was happy 0 1 2 3
13. I talked less than usual 0 1 2 3
14. I felt lonely 0 1 2 3
15. People were unfriendly 0 1 2 3
16. I enjoyed life 0 1 2 3
17. I had crying spells 0 1 2 3
18. I felt sad 0 1 2 3
19. I felt that people disliked me 0 1 2 3
20. I could not get "going" 0 1 2 3
-

CLINICALLY ELEVATED SUM SCORE

Sum score = $1 + 2 + 2 + 1 + 1 + 3 + 1 + 2 + 1 + 1 + 2 + 2 + 1 + 1 + 2 + 1 + 1 + 1 + 1 + 2 = 29$

During the past week:

-
- | | |
|--|--|
| 1. I was bothered by things that usually don't bother me | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 2. I did not feel like eating; my appetite was poor | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |
| 3. I felt that I could not shake off the blues | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |
| 4. I felt that I was just as good as other people | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 5. I had trouble keeping my mind on what I was doing | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 6. I felt depressed | 0 <input type="radio"/> 1 <input type="radio"/> 2 <input checked="" type="radio"/> 3 |
| 7. I felt that everything I did was an effort | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 8. I felt hopeful about the future | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |
| 9. I thought my life had been a failure | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 10. I felt fearful | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |

-
- | | |
|------------------------------------|--|
| 11. My sleep was restless | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |
| 12. I was happy | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |
| 13. I talked less than usual | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 14. I felt lonely | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 15. People were unfriendly | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |
| 16. I enjoyed life | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 17. I had crying spells | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 18. I felt sad | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 19. I felt that people disliked me | 0 <input checked="" type="radio"/> 1 <input type="radio"/> 2 <input type="radio"/> 3 |
| 20. I could not get "going" | 0 <input type="radio"/> 1 <input checked="" type="radio"/> 2 <input type="radio"/> 3 |



The CES-D sum score ranges from 0 (“none of the time” for all items) to 60 (“all of the time” for all items). What level of measurement describes the CES-D items?

MEASURING DISPOSITIONAL OPTIMISM

- The article's description of the optimism instrument:

Dispositional optimism and pessimism were assessed prior to diagnosis with the Life Orientation Test—Revised (LOT-R; Scheier et al., 1994; $\alpha = .82$). The LOT-R consists of three optimism items, three pessimism items, and four fillers. Five-point Likert scale items range from I agree a lot to I disagree a lot. Summed scores for optimism ($\alpha = .74$) and pessimism ($\alpha = .80$) were calculated. Each scale's potential range is 0–12; higher scores represent higher levels of the attribute.

LOT-R OPTIMISM QUESTIONNAIRE

4 = I agree a lot

3 = I agree a little

2 = I neither agree nor disagree

1 = I disagree a little

0 = I disagree a lot

In uncertain times, I usually expect the best.

0 1 2 3 4

I'm always optimistic about my future.

0 1 2 3 4

Overall, I expect more good things to happen to me than bad

0 1 2 3 4

OUTLINE

- 1 Measuring psychological constructs
- 2 Distribution shapes for numeric variables
- 3 Graphical displays for numeric variables
- 4 Graphical and tabular displays for discrete and categorical variables
- 5 Study questions

DISTRIBUTIONS

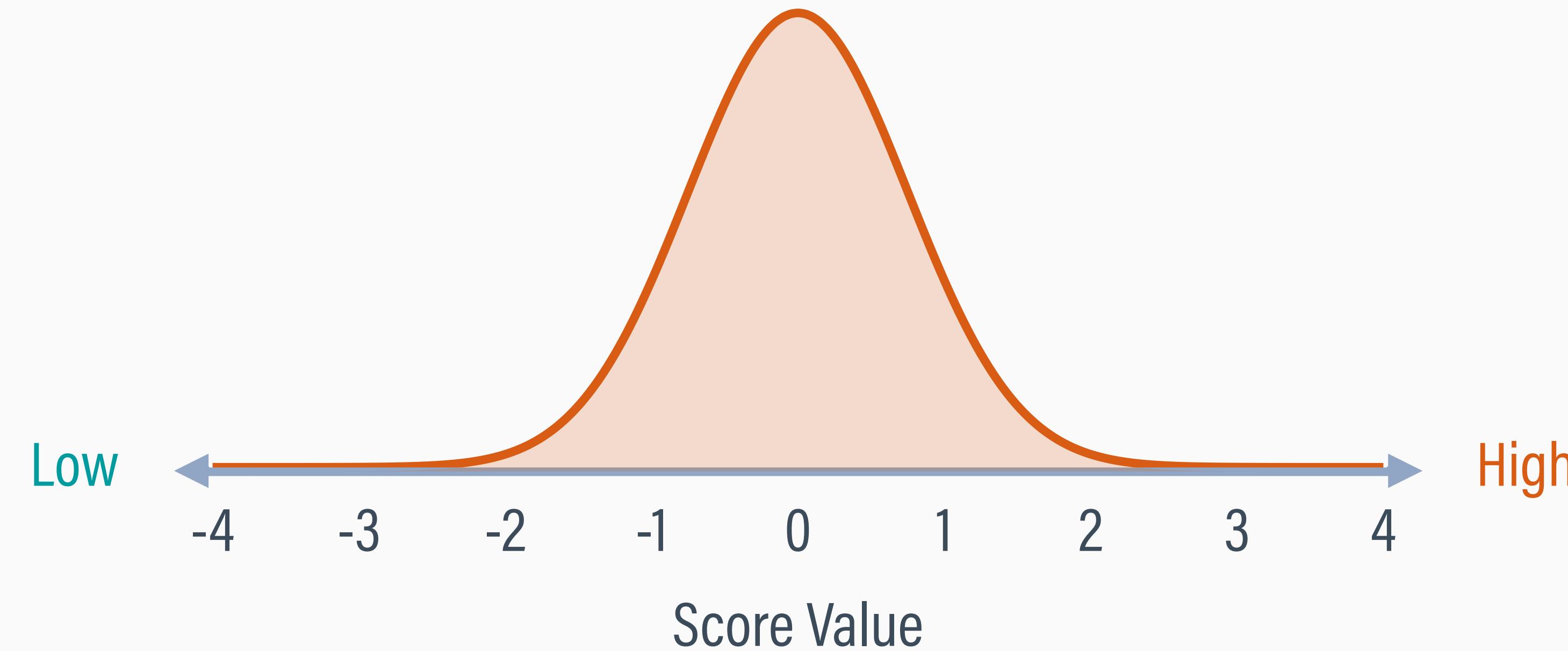
- A distribution describes the concentration of score values
- Distributions can be summarized with graphs, tables, or mathematical functions
- The type of summary depends on whether a variable is numeric (interval or ratio) or categorical (nominal or ordinal)

NUMERIC (CONTINUOUS) VARIABLES

- Interval and ratio scales refer to numeric (continuous) variables that approximate a number line (e.g., response time, age, GRE)
- Any two adjacent score values reflect the same amount of the variable (e.g., the difference between CESD = 5 and 6 is the same amount of depression as CESD = 20 and 21)
- Many behavioral, emotional, and physiological variables in psychology approximate interval-level data (but not perfectly)

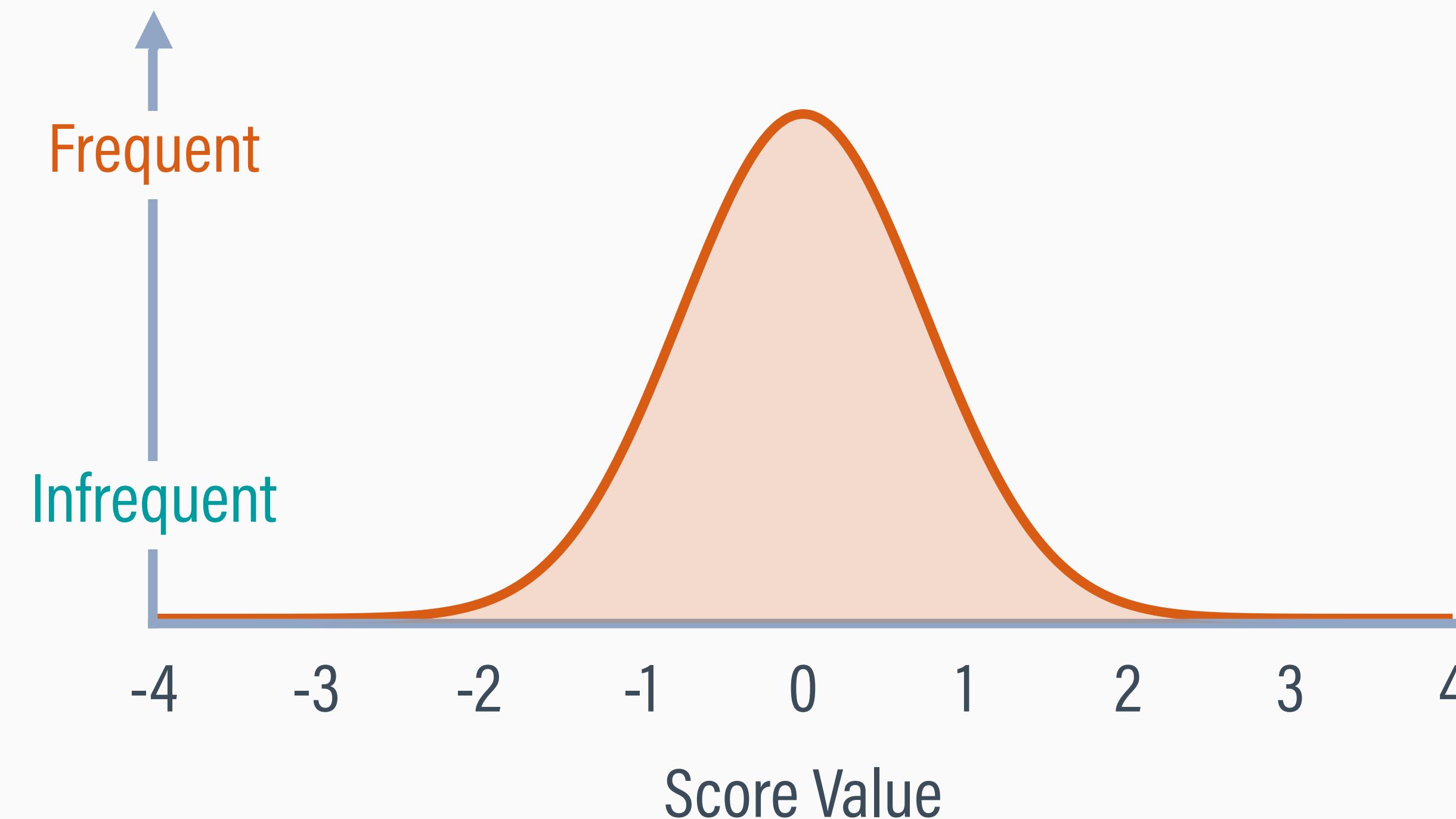
HORIZONTAL AXIS (SCORE RANGE)

- Scores follow a number line and range from low to high along the horizontal axis



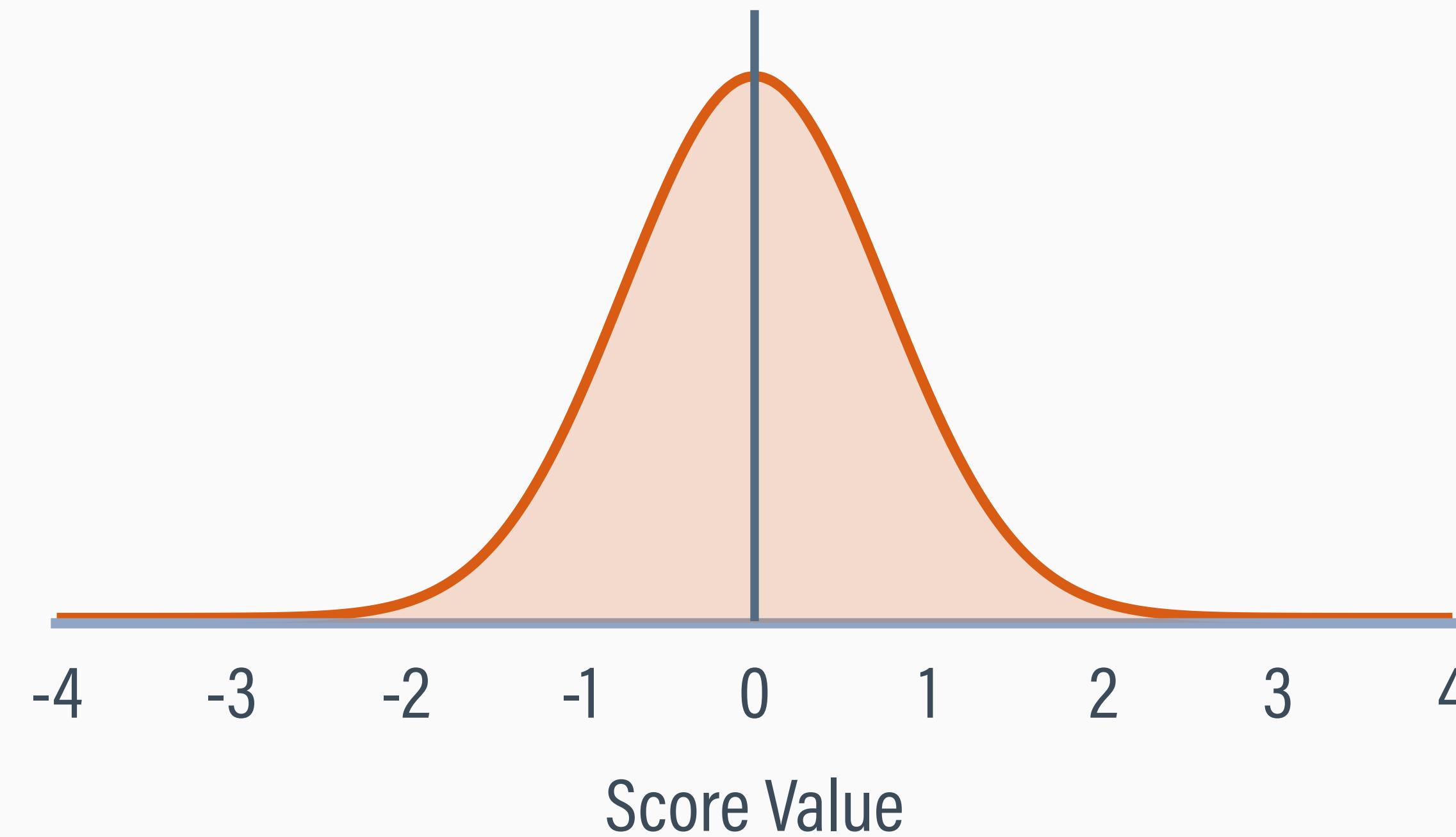
VERTICAL AXIS (FREQUENCY)

- The height of the vertical axis – called **frequency** or **count** or **density** – describes how often each score occurs



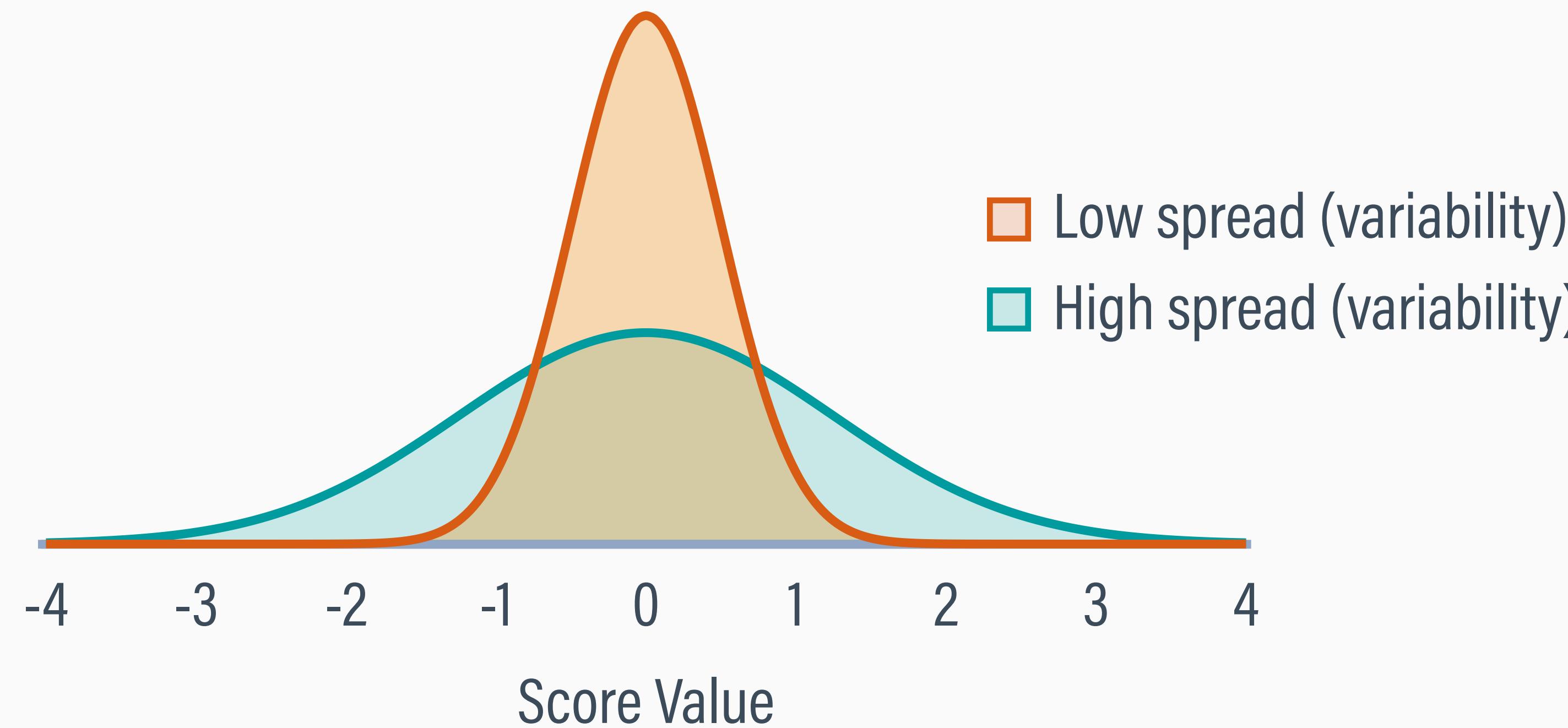
DISTRIBUTION CENTER

- A distribution's center (e.g., mean) is located at a value where scores are highly concentrated



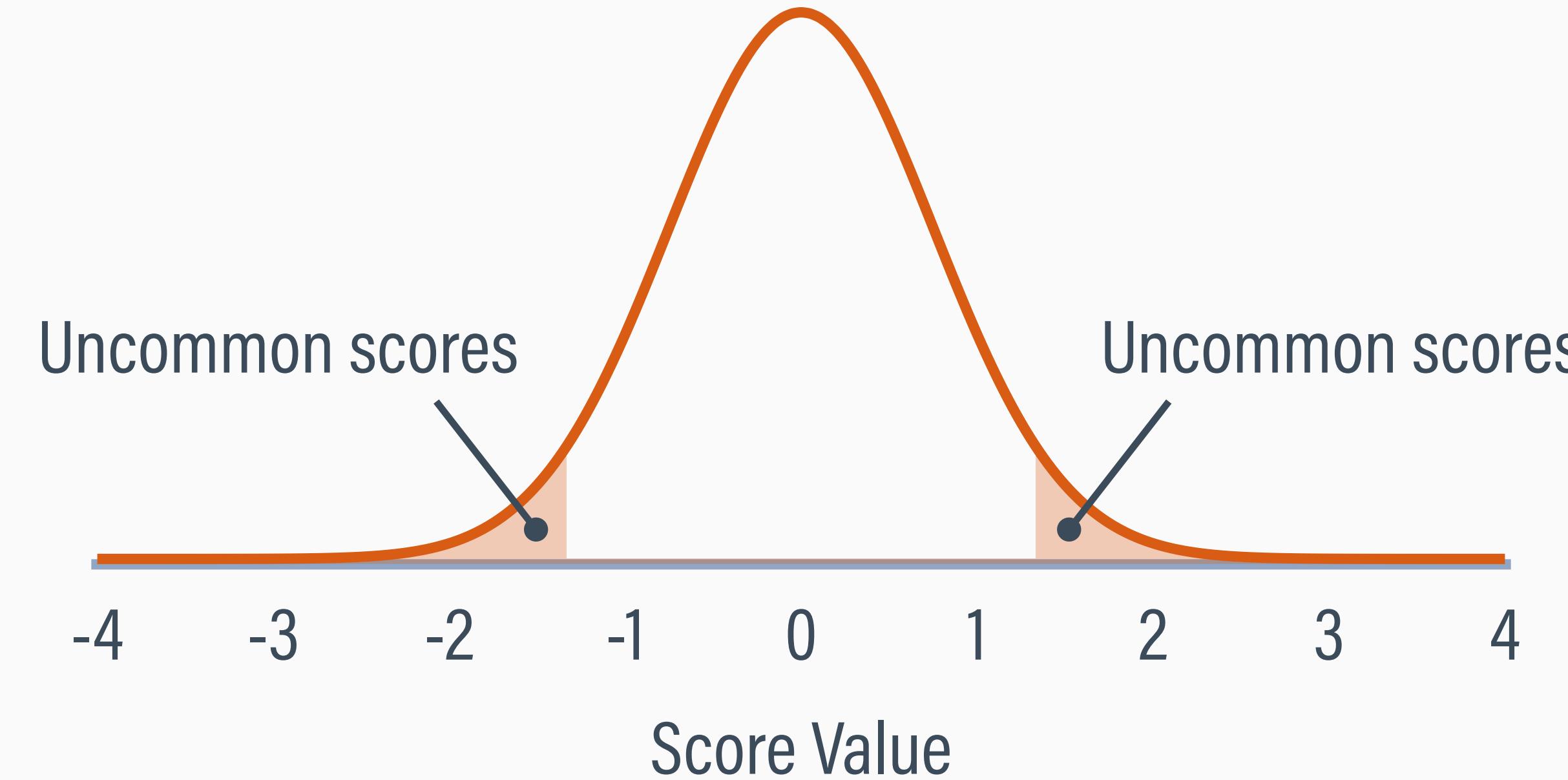
VARIABILITY OR SPREAD

- High variability (spread) indicates greater score differences, low spread indicates similar scores



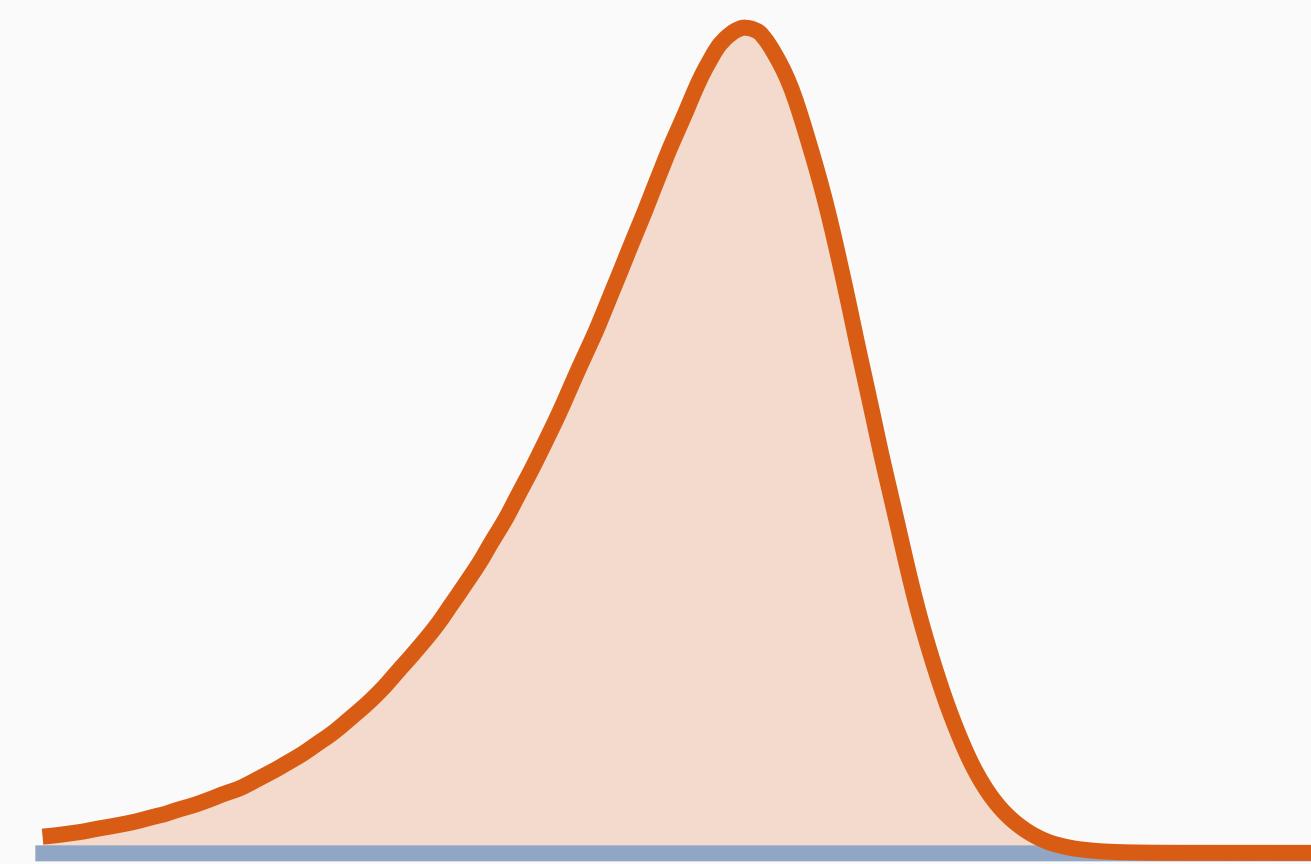
DISTRIBUTION TAILS

- A distribution's tails contain unusual or infrequent score values (outliers)

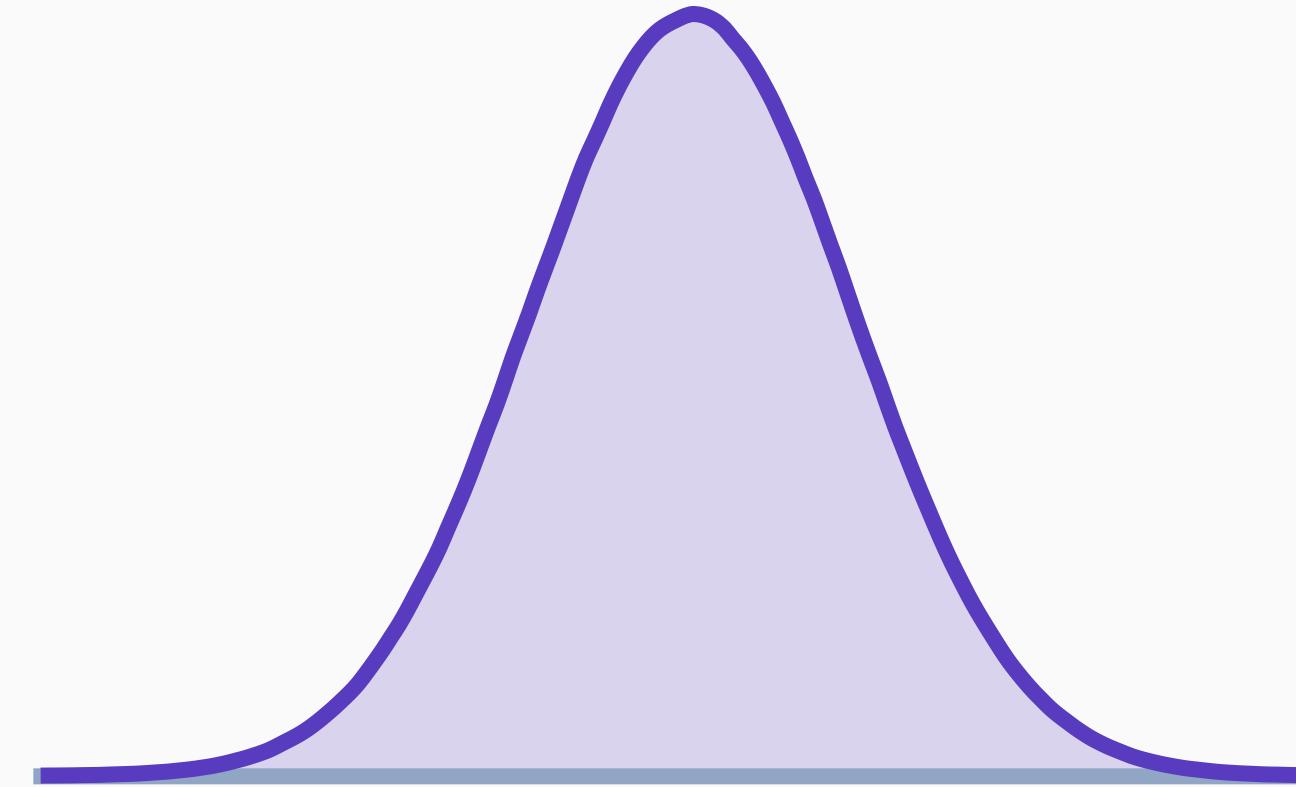


DISTRIBUTION SHAPE

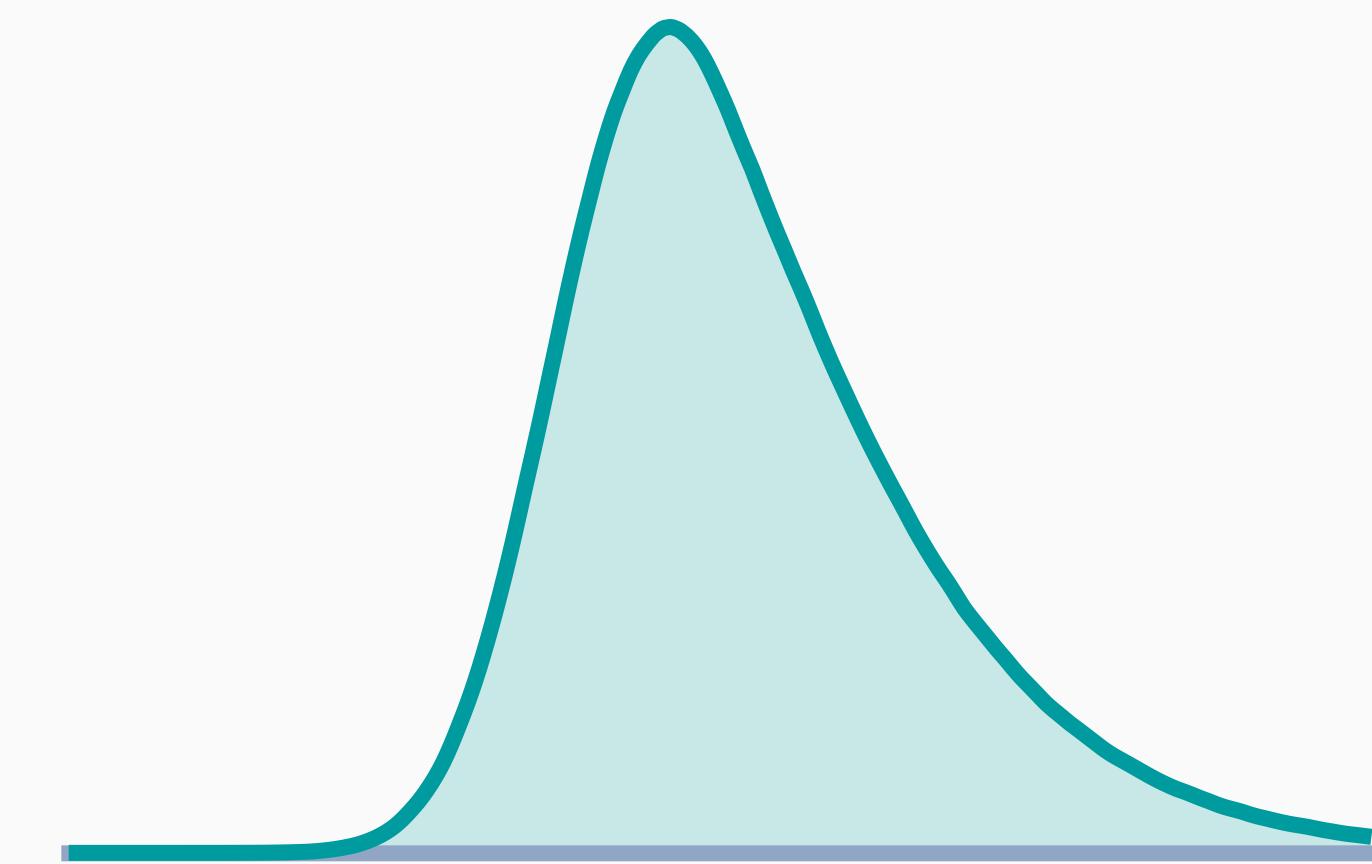
- Asymmetric distributions can be characterized as positively or negatively skewed



Negatively Skewed



Symmetric



Positively Skewed

SKEWNESS AND KURTOSIS

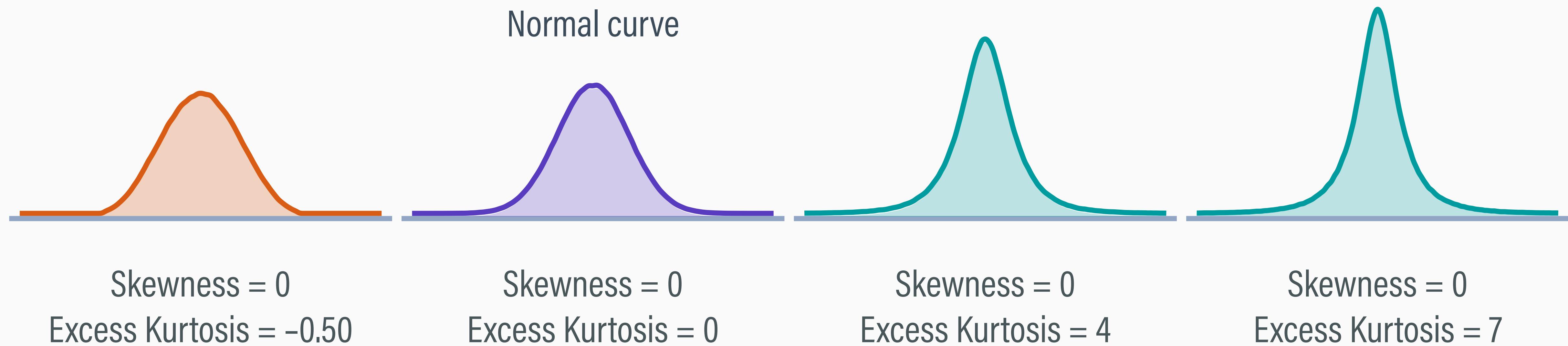
- **Skewness** is a numerical measure of asymmetry in a distribution (a normal curve has skewness = 0)
- **Kurtosis** is often described as peakedness but, more accurately, it reflects the proportion of extreme outliers in the tails relative to a normal curve
- Both measures are standardized, such that their numeric values don't depend on a variable's scale

TWO MEASURES OF KURTOSIS

- The kurtosis value for a normal curve is 3
- Researchers often use “excess” kurtosis where the value for a normal curve is 0 (i.e., $kurtosis - 3$)
- Software packages are not uniform, always check which value is getting reported (we will use the psych package, which reports excess kurtosis)

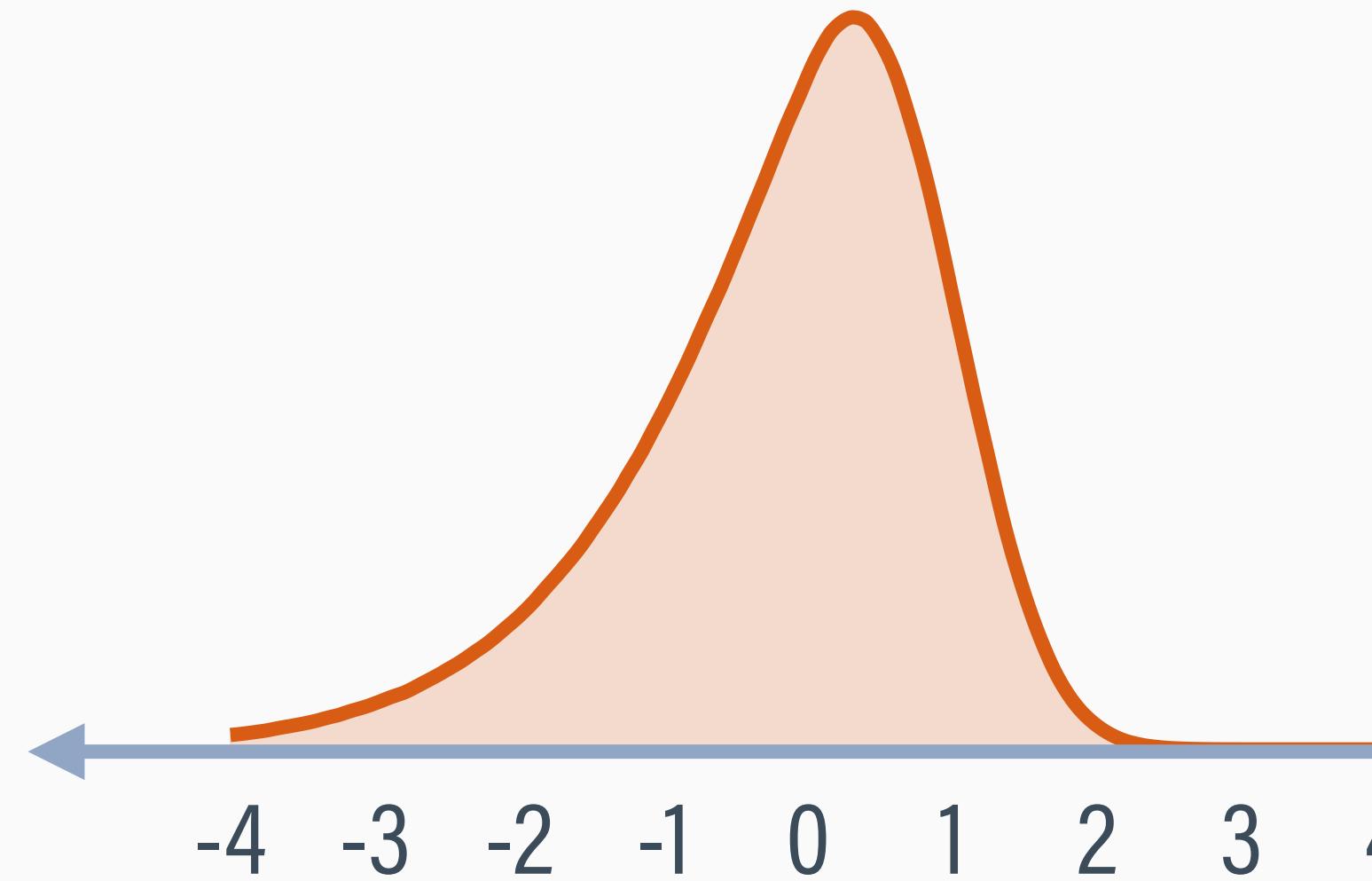
EXCESS KURTOSIS

- As kurtosis increases, more scores migrate from the center of the distribution to the tails (thus increasing peakedness)

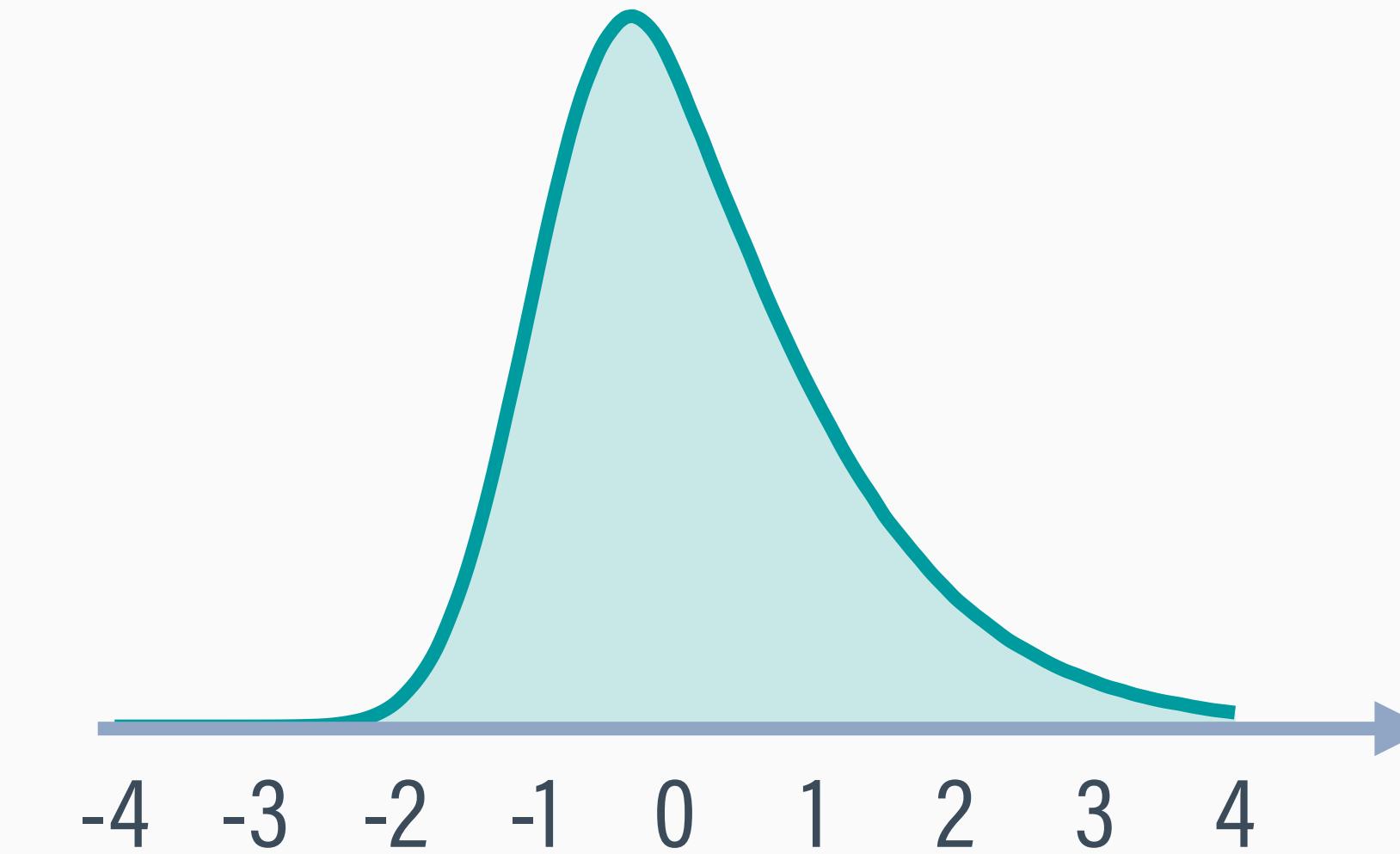


POSITIVE VERSUS NEGATIVE SKEW

- A positively skew has a long tail pointing toward the positive side of the number line, and negative skew has the opposite



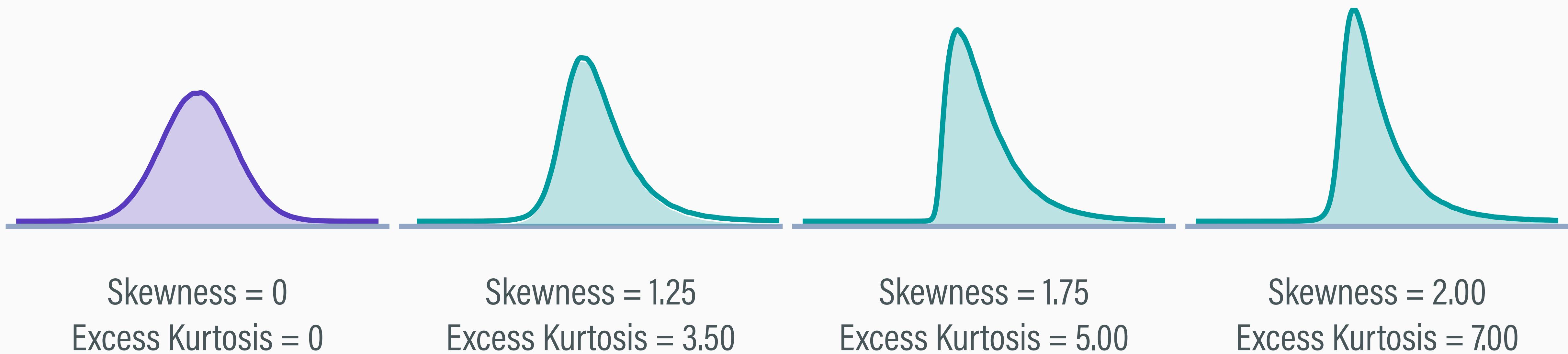
Negatively Skewed



Positively Skewed

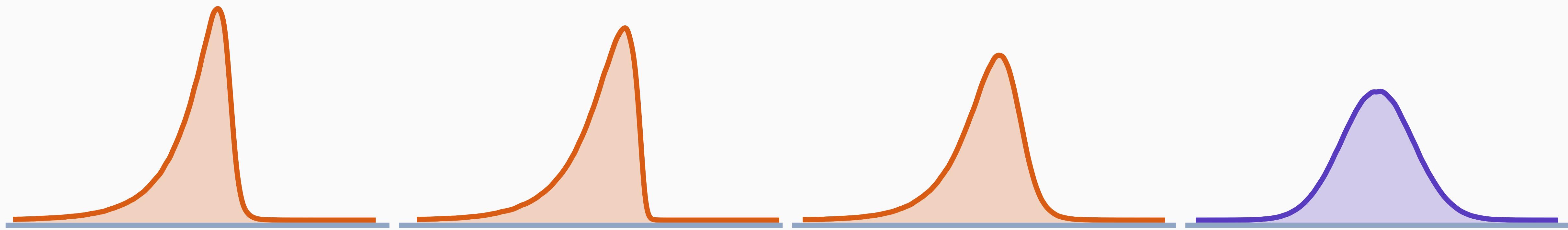
POSITIVE SKEWNESS

- As skewness increases, the distribution becomes increasingly asymmetric (kurtosis must increase too)



NEGATIVE SKEWNESS

- As skewness increases, the distribution becomes increasingly asymmetric (kurtosis must increase too)



Skewness = -2.00
Excess Kurtosis = 7.00

Skewness = -1.75
Excess Kurtosis = 5.00

Skewness = -1.25
Excess Kurtosis = 3.50

Skewness = 0
Excess Kurtosis = 0

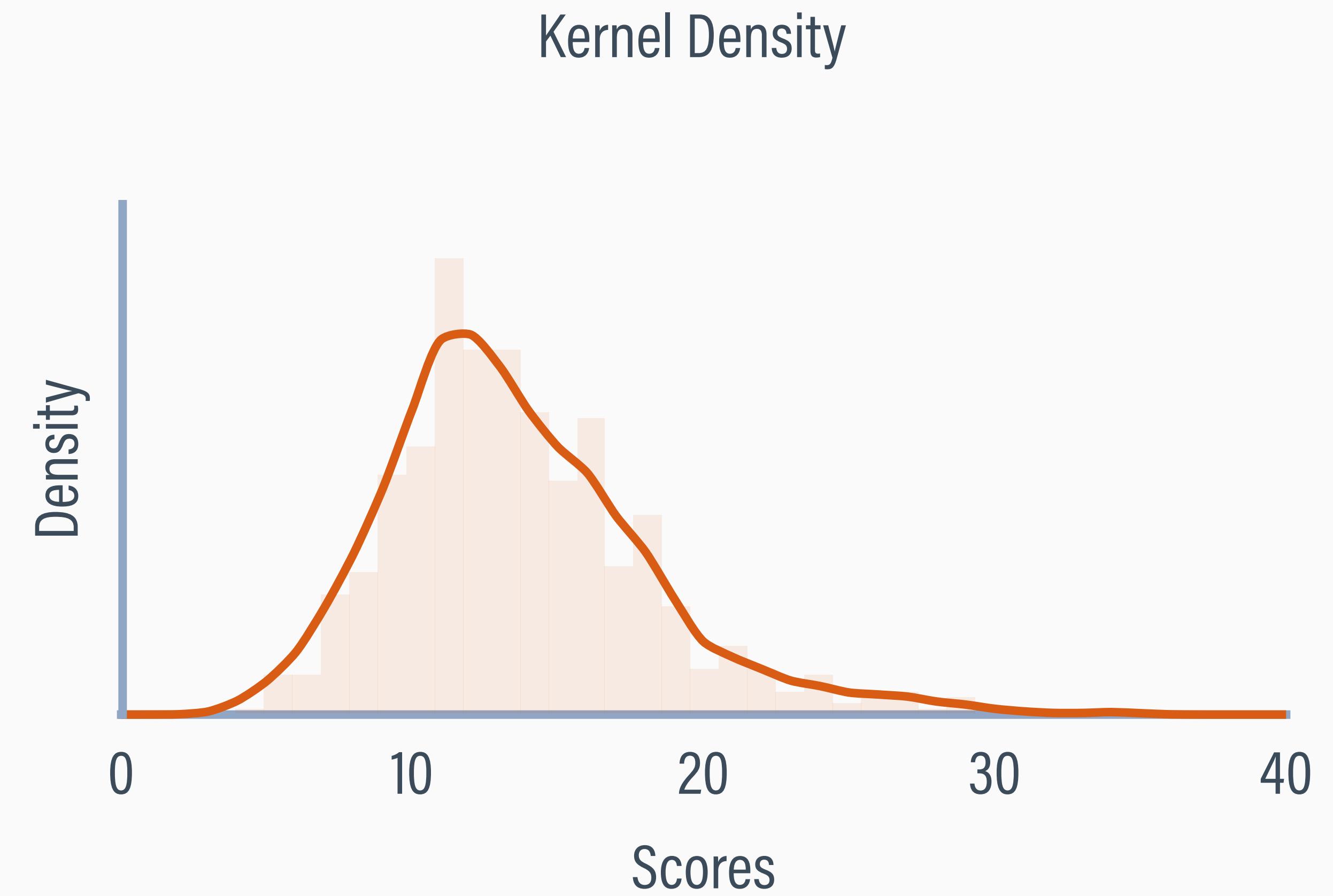
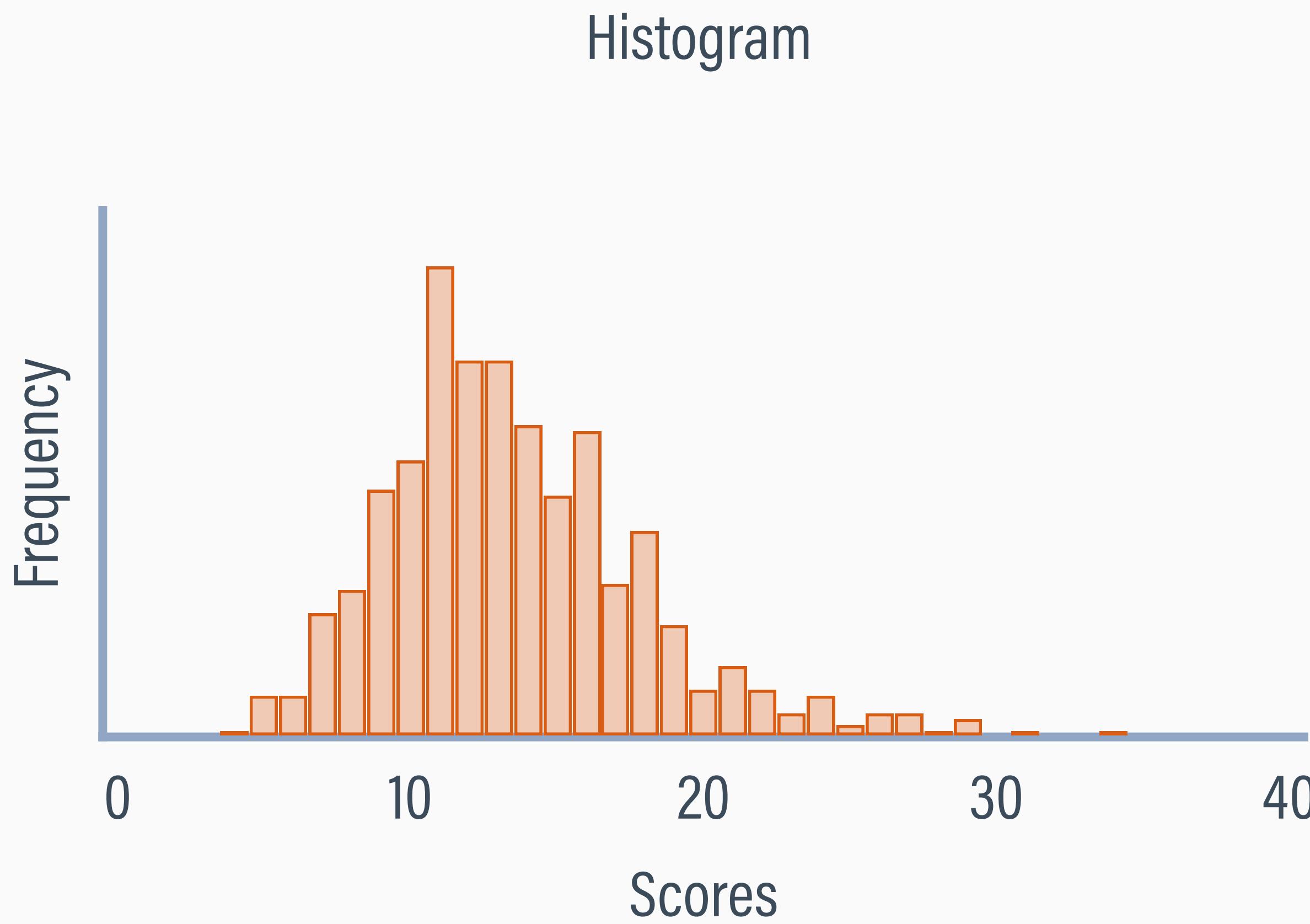
OUTLINE

- 1 Measuring psychological constructs
- 2 Distribution shapes for numeric variables
- 3 Graphical displays for numeric variables
- 4 Graphical and tabular displays for discrete and categorical variables
- 5 Study questions

HISTOGRAMS AND KERNEL DENSITY PLOTS

- A **histogram** is a bar plot for numeric variables (no breaks or gaps along the horizontal axis, which is a number line)
- A **kernel density plot** applies a smoothing algorithm that connects the histogram bars with a continuous curve
- Frequency = density = count = vertical elevation

HISTOGRAM AND KERNEL DENSITY GRAPHS



UVEAL MELANOMA and DEPRESSION

Uveal melanoma, a rare eye cancer, presents potential vision loss and life threat. This prospective, longitudinal study interrogated the predictive utility of visual impairment, as moderated by optimism/pessimism, on depressive symptoms in 299 adults undergoing diagnostic evaluation.

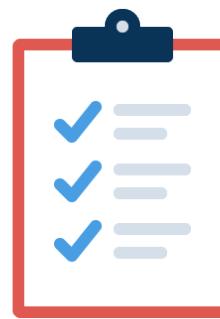


Annette
Stanton

James
MacDonald

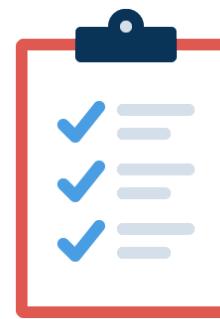
MacDonald, J.J., Jorge-Miller, A., Enders, C.K., McCannel, T., Beran, T., & Stanton, A.L. (2021). Perceived and objective visual impairment predicting depressive symptoms across one year in uveal melanoma diagnostic biopsy: Optimism and pessimism as moderators. *Health Psychology, 40*, 408-417.

KEY NUMERIC VARIABLES



Depressive Symptoms

The CES-D is a 20-item inventory that asks people to rate how often they experience depressive symptoms such as restless sleep, poor appetite, and feeling lonely.



Dispositional Optimism

The Life Orientation Test Revised (LOT-R) is a 3-item questionnaire that asks people to rate their tendency to expect positive outcomes in life.

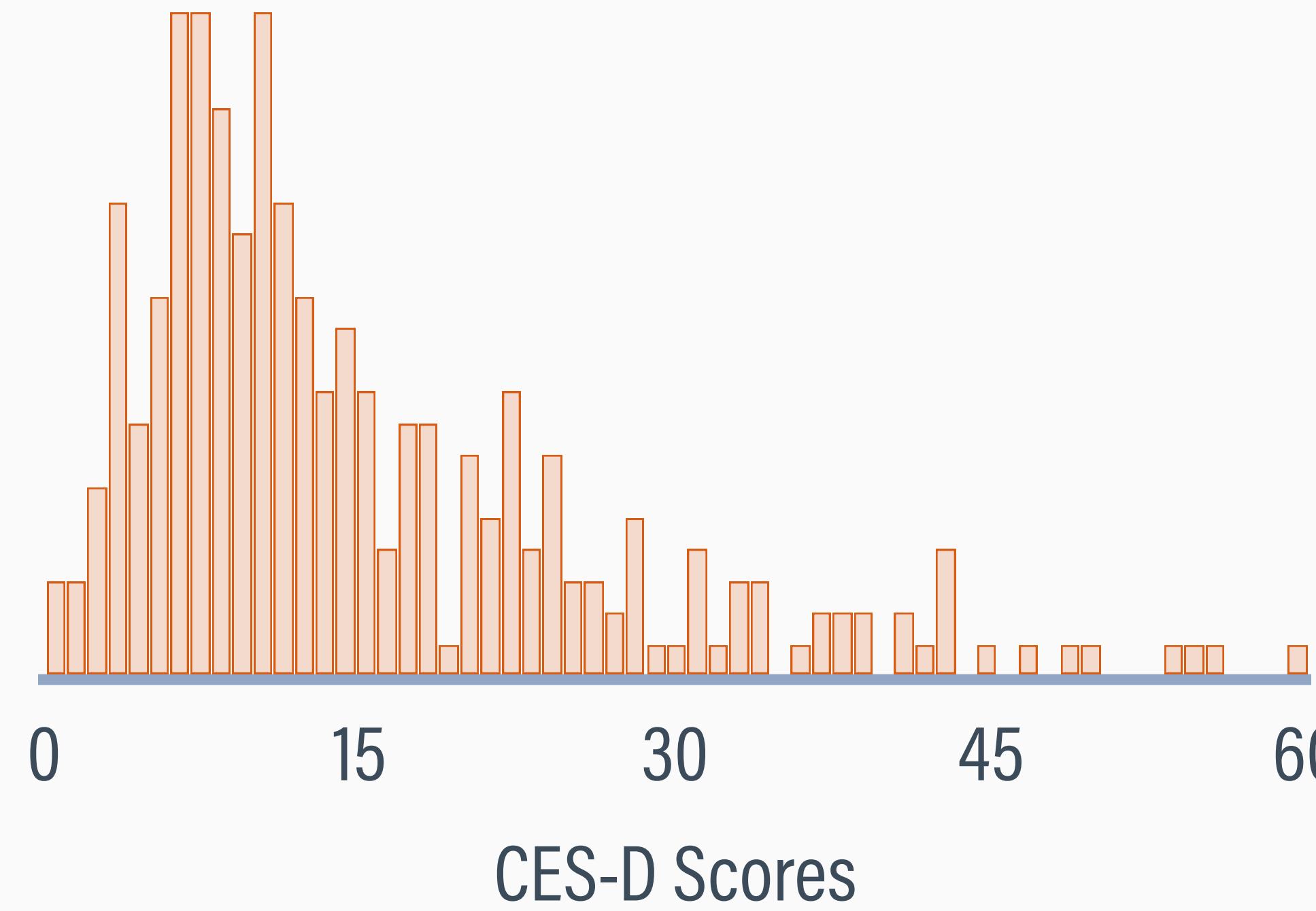


Visual Impairment

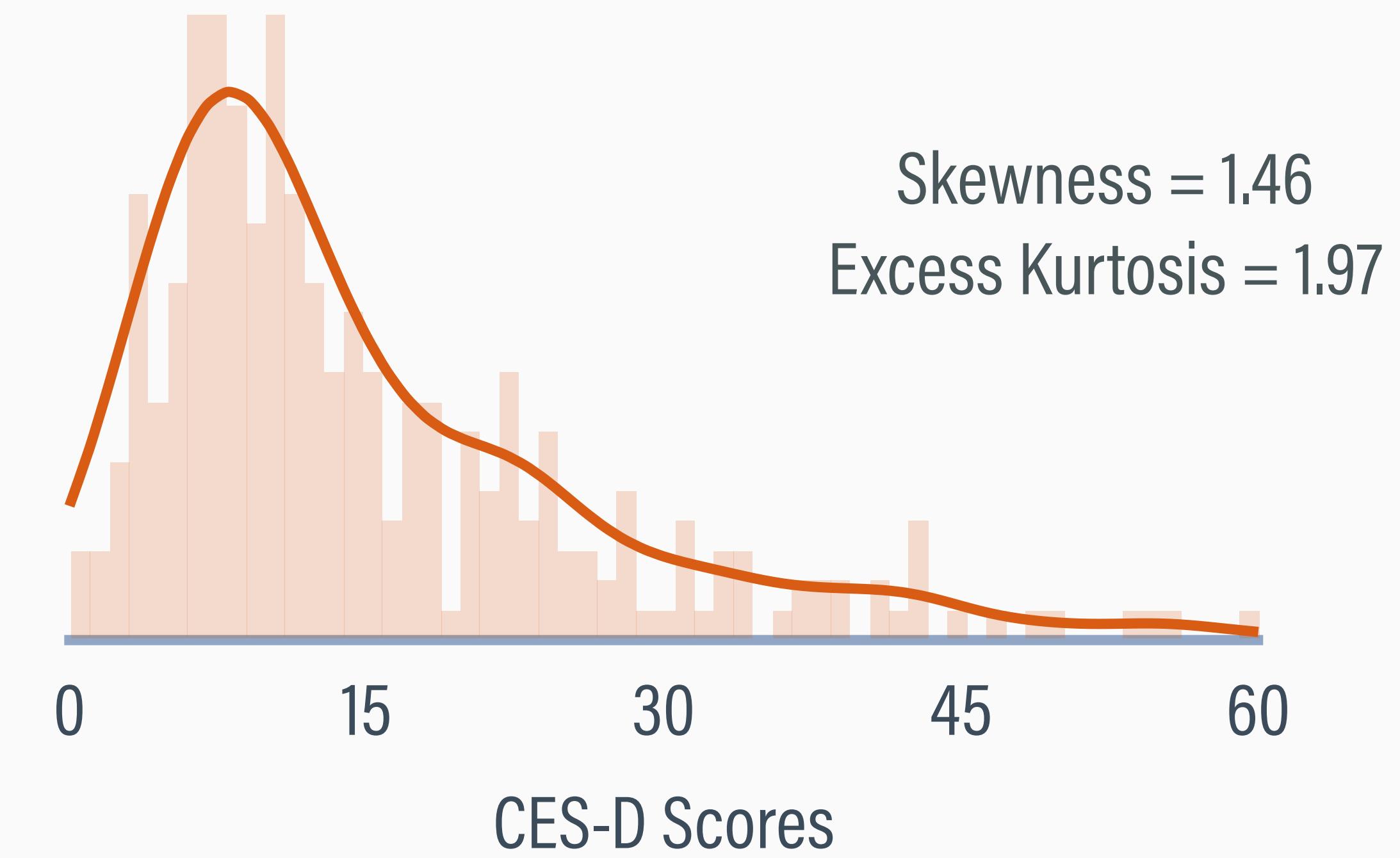
The Measure of Outcome in Ocular Disease vision subscale (MOOD-V) is a 13-item measure. Scores were summed and transformed to a 0-100 scale.

DEPRESSIVE SYMPTOMS DISTRIBUTION

Histogram



Kernel Density

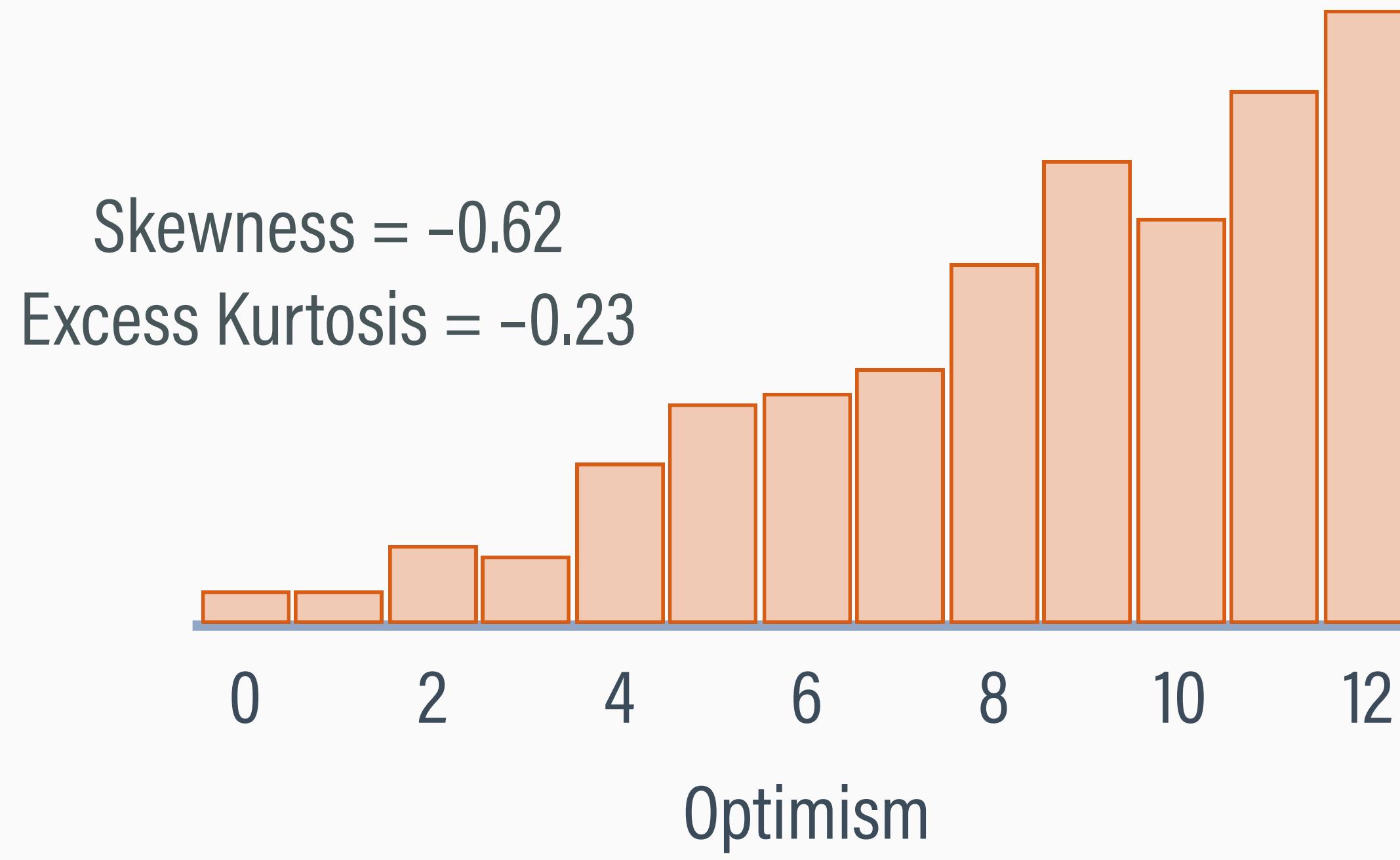




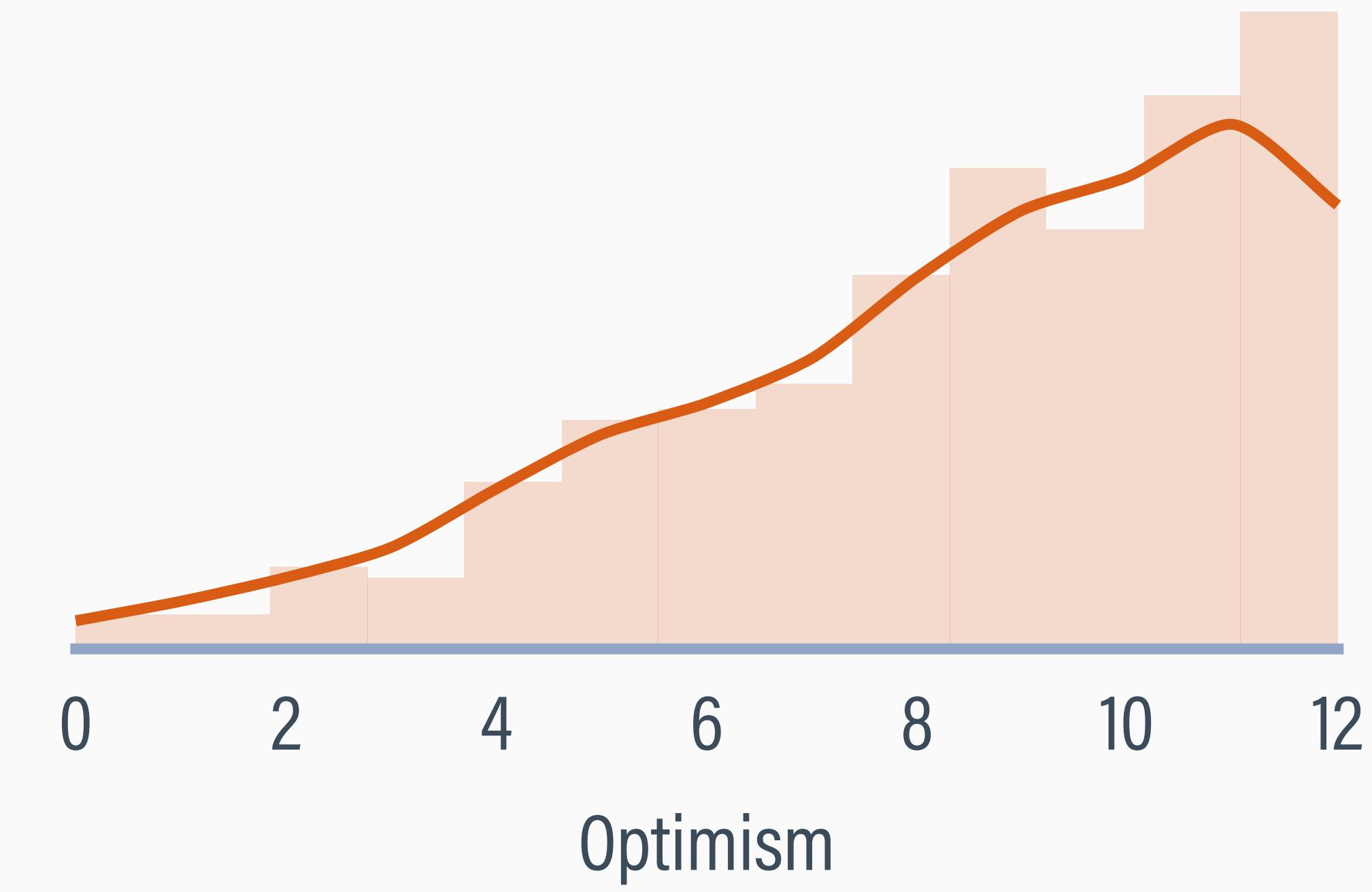
In small groups of two or three, discuss the skewness and kurtosis statistics for the depression variable. Interpret each statistic in practical terms (e.g., what it says about tail heaviness or asymmetry). How do these statistics align with your visual impression from the plot? Based on the distribution shape, how would you characterize the overall magnitude of depressive symptoms in this sample?

OPTIMISM DISTRIBUTION

Histogram

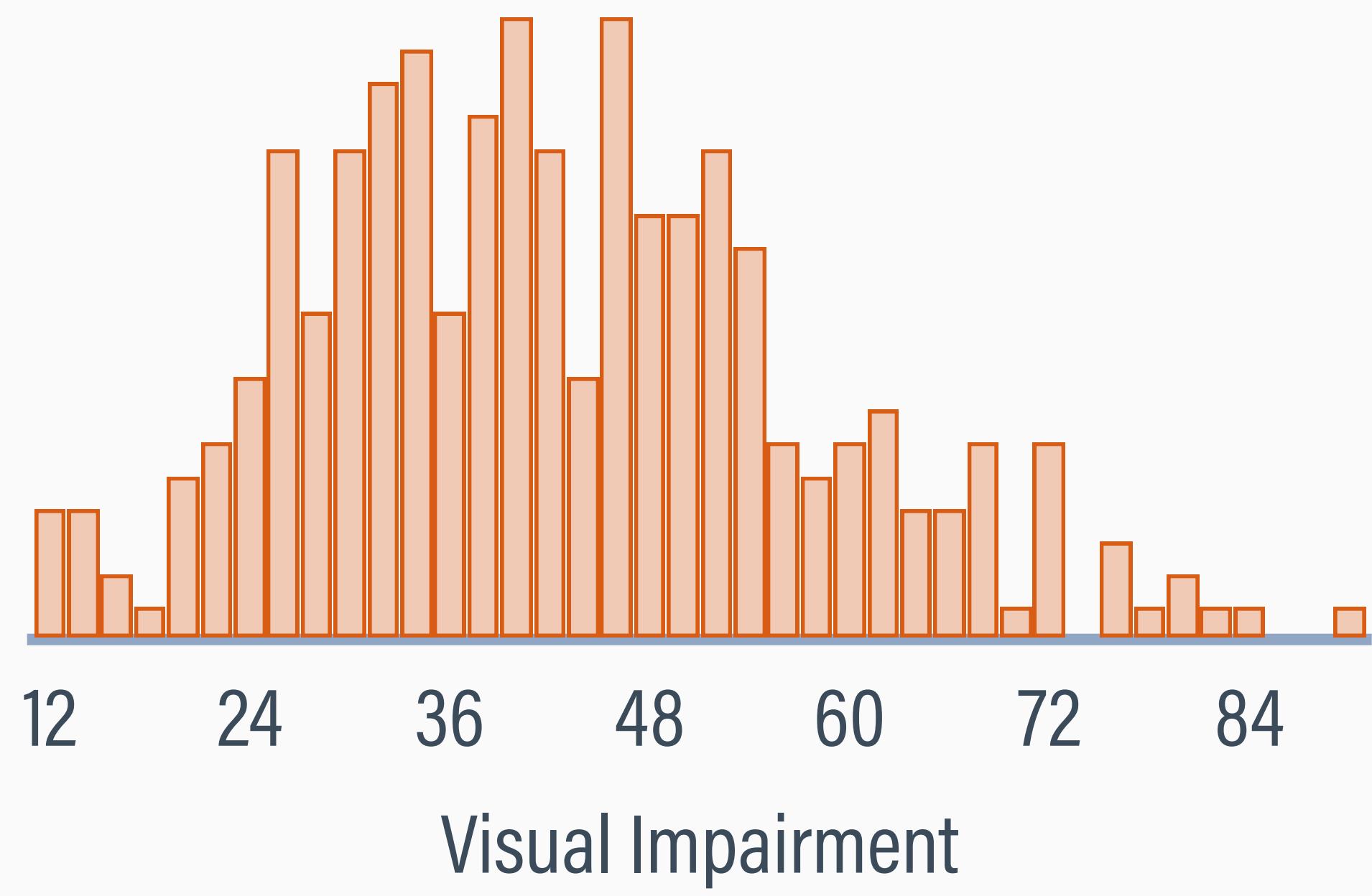


Kernel Density

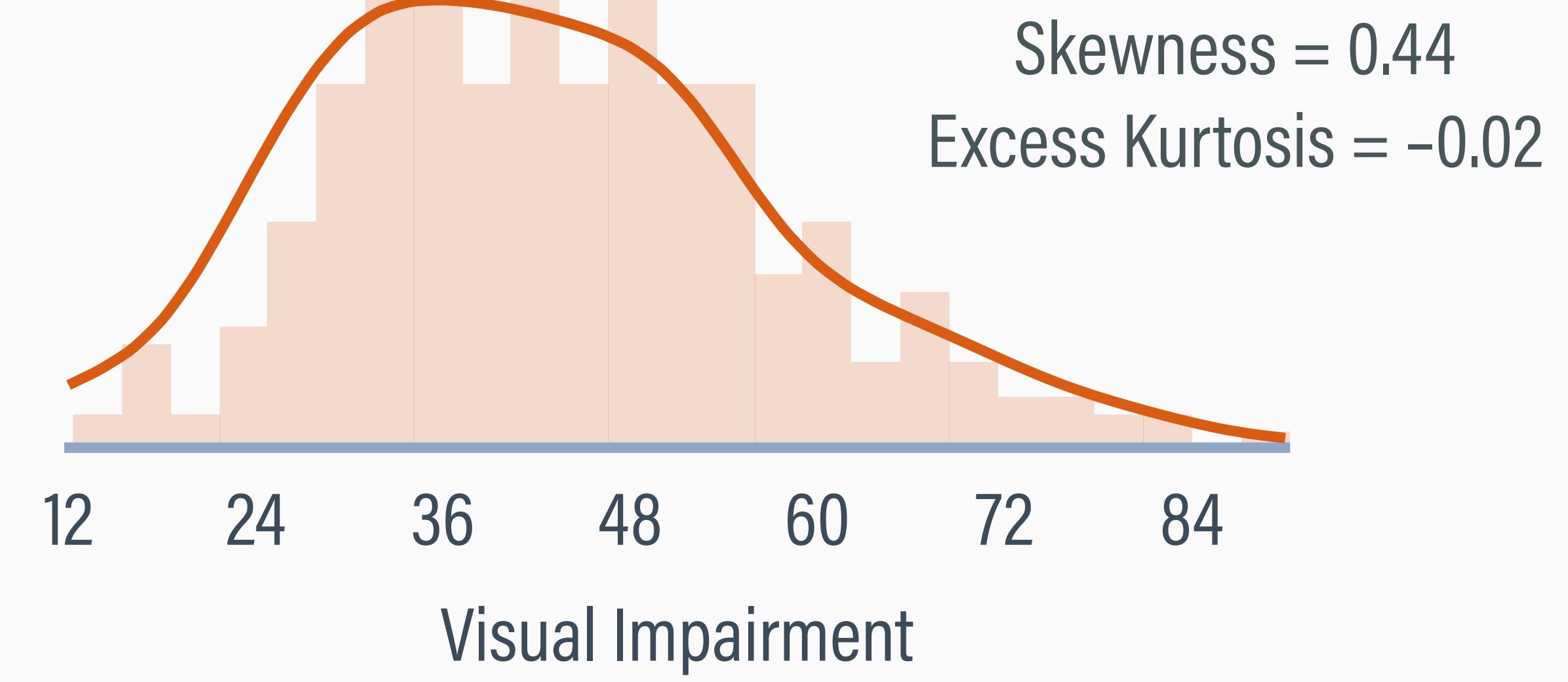


VISUAL IMPAIRMENT DISTRIBUTION

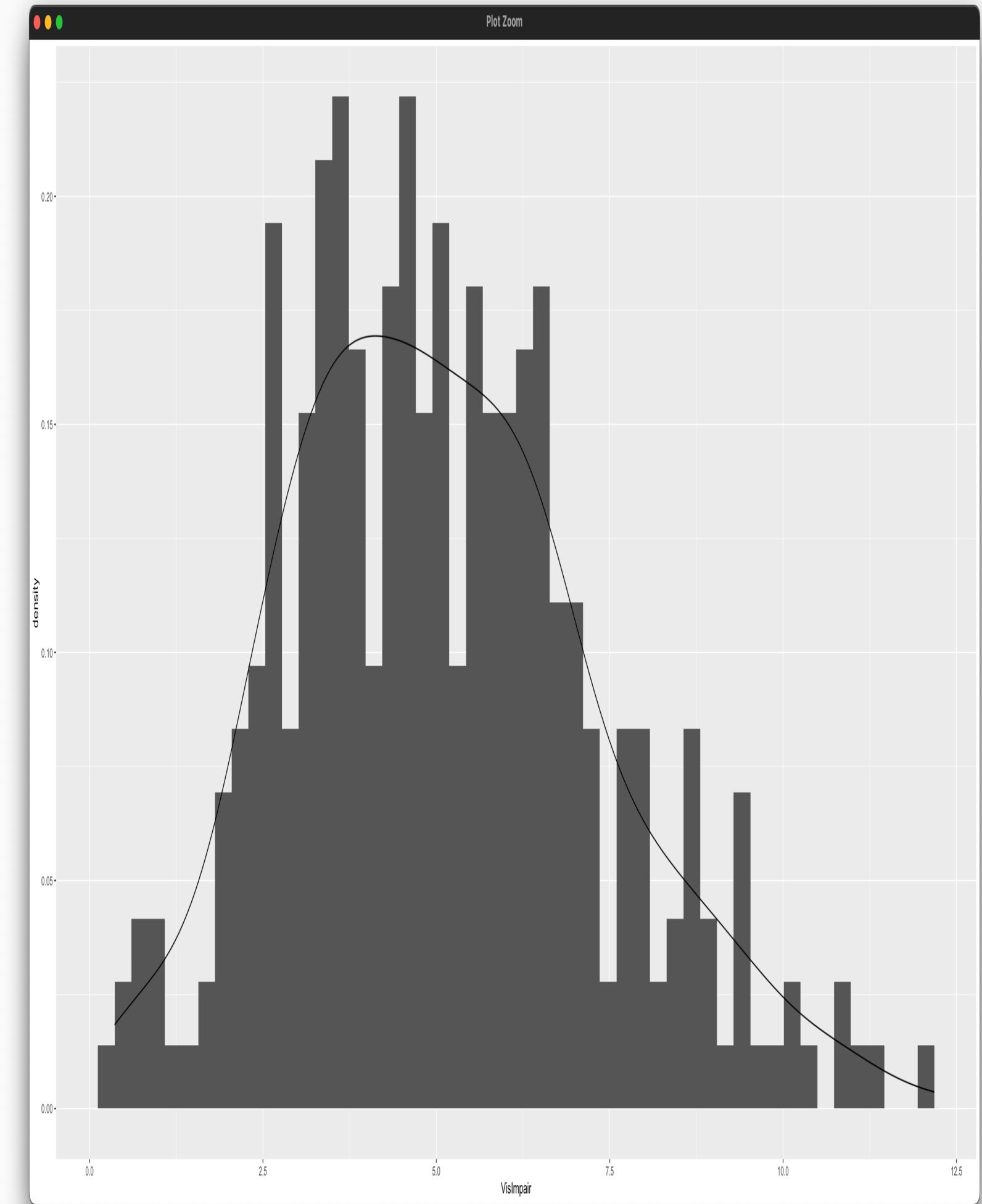
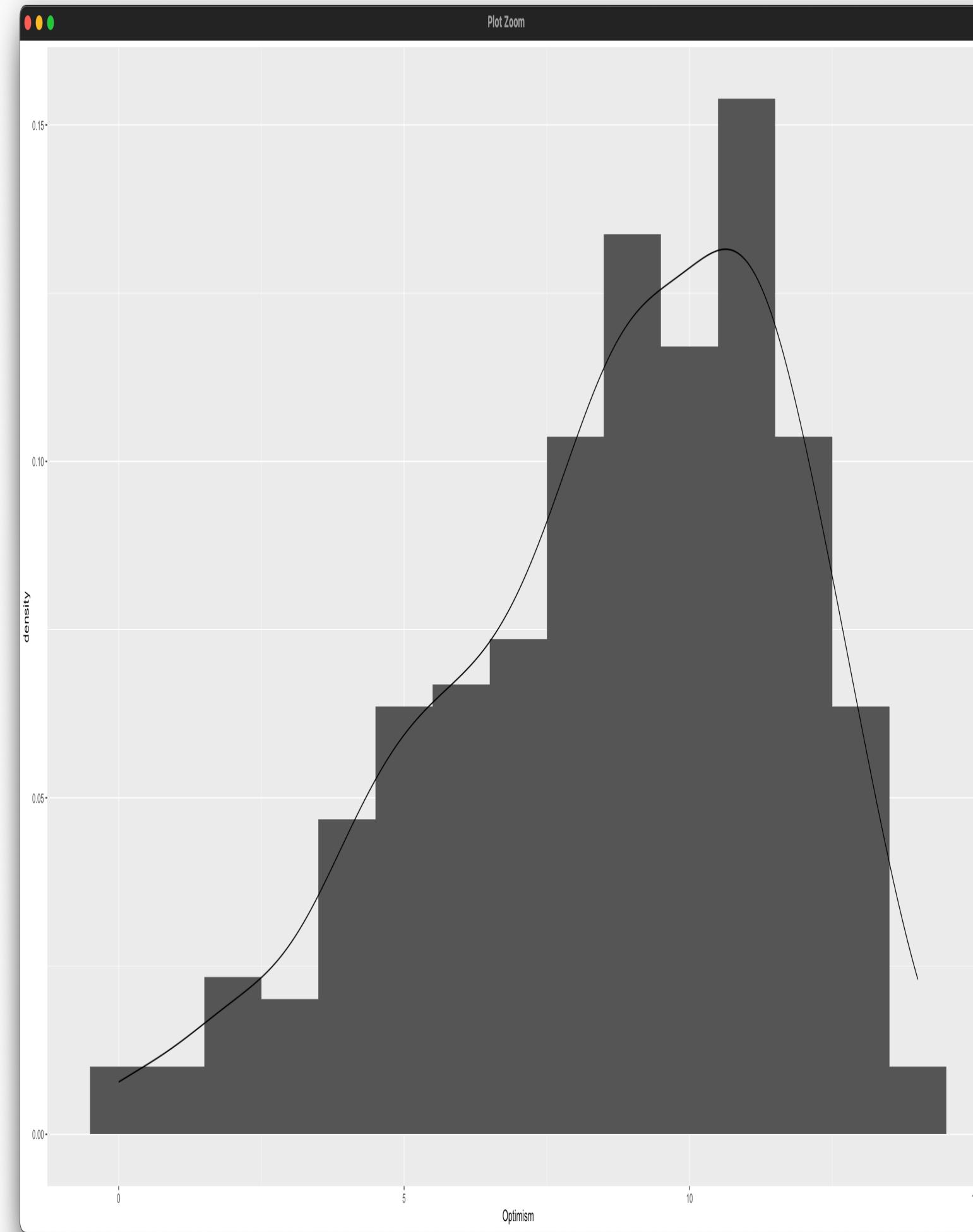
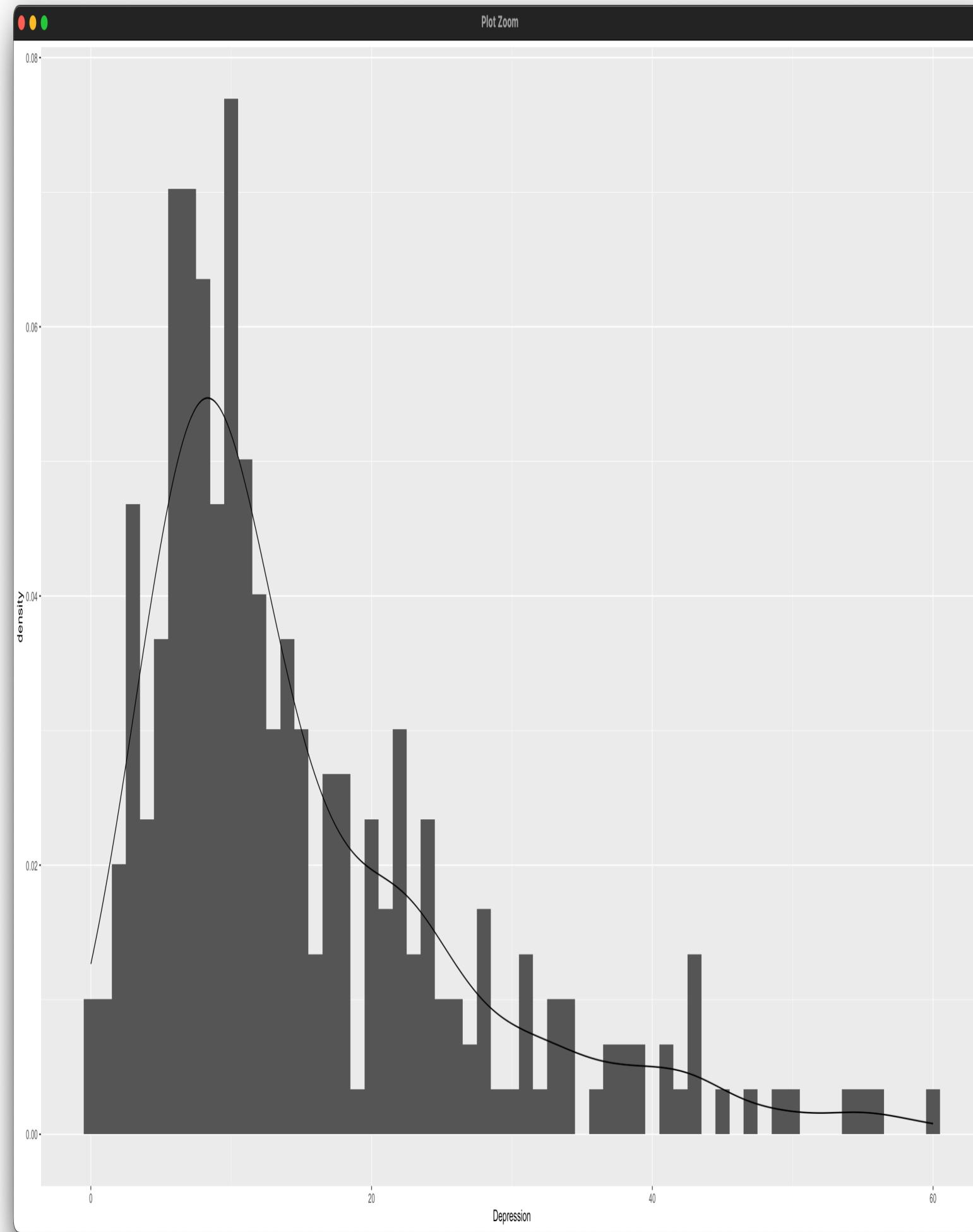
Histogram



Kernel Density



R OUTPUT



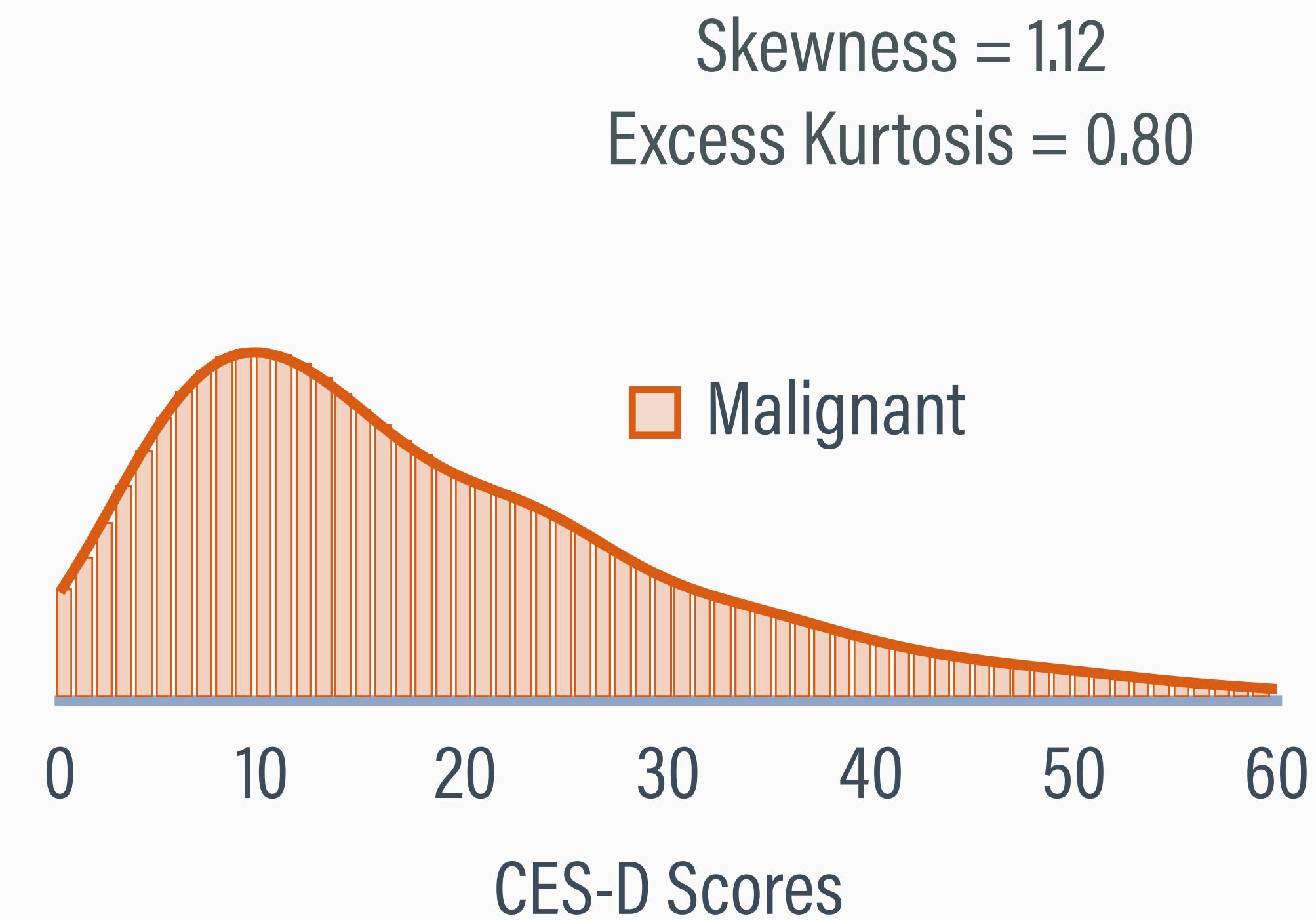
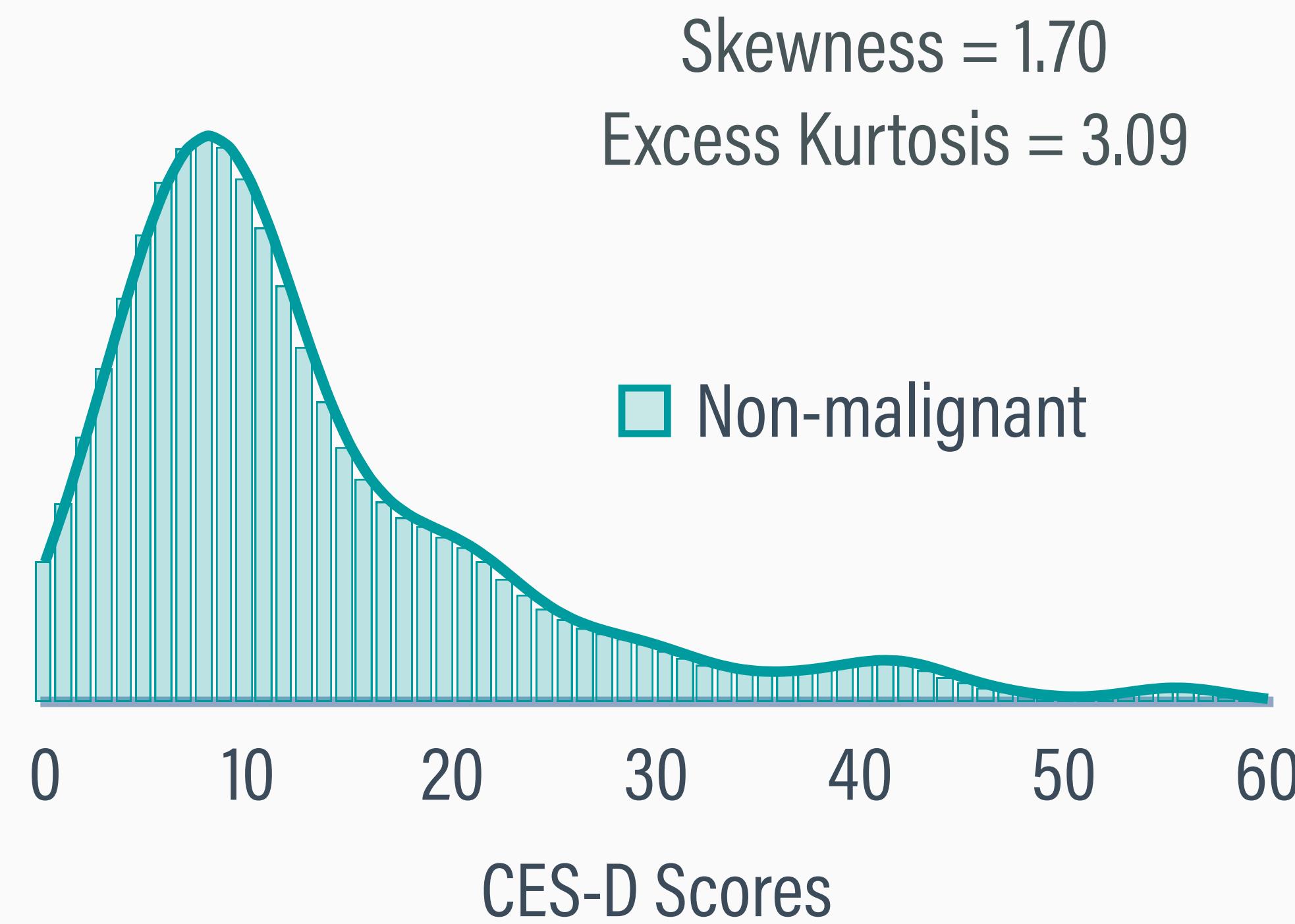
R OUTPUT

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Participant	1	299	150.00	86.46	150.00	150.00	111.19	1.00	299.00	298.00	0.00	-1.21	5.00
Diagnosis*	2	299	1.64	0.48	2.00	1.68	0.00	1.00	2.00	1.00	-0.59	-1.66	0.03
Age	3	299	58.99	13.94	60.00	59.47	11.86	19.00	91.00	72.00	-0.33	-0.15	0.81
Gender*	4	299	1.54	0.50	2.00	1.55	0.00	1.00	2.00	1.00	-0.15	-1.98	0.03
Comorbrids	5	299	0.94	0.74	1.00	0.93	1.48	0.00	2.00	2.00	0.10	-1.17	0.04
Optimism	6	299	8.65	3.06	9.00	8.87	2.97	0.00	14.00	14.00	-0.62	-0.23	0.18
Depression	7	299	14.85	11.44	11.00	13.12	7.41	0.00	60.00	60.00	1.46	1.97	0.66
VisImpair	8	299	5.09	2.23	4.85	4.98	2.21	0.36	12.18	11.82	0.43	-0.03	0.13

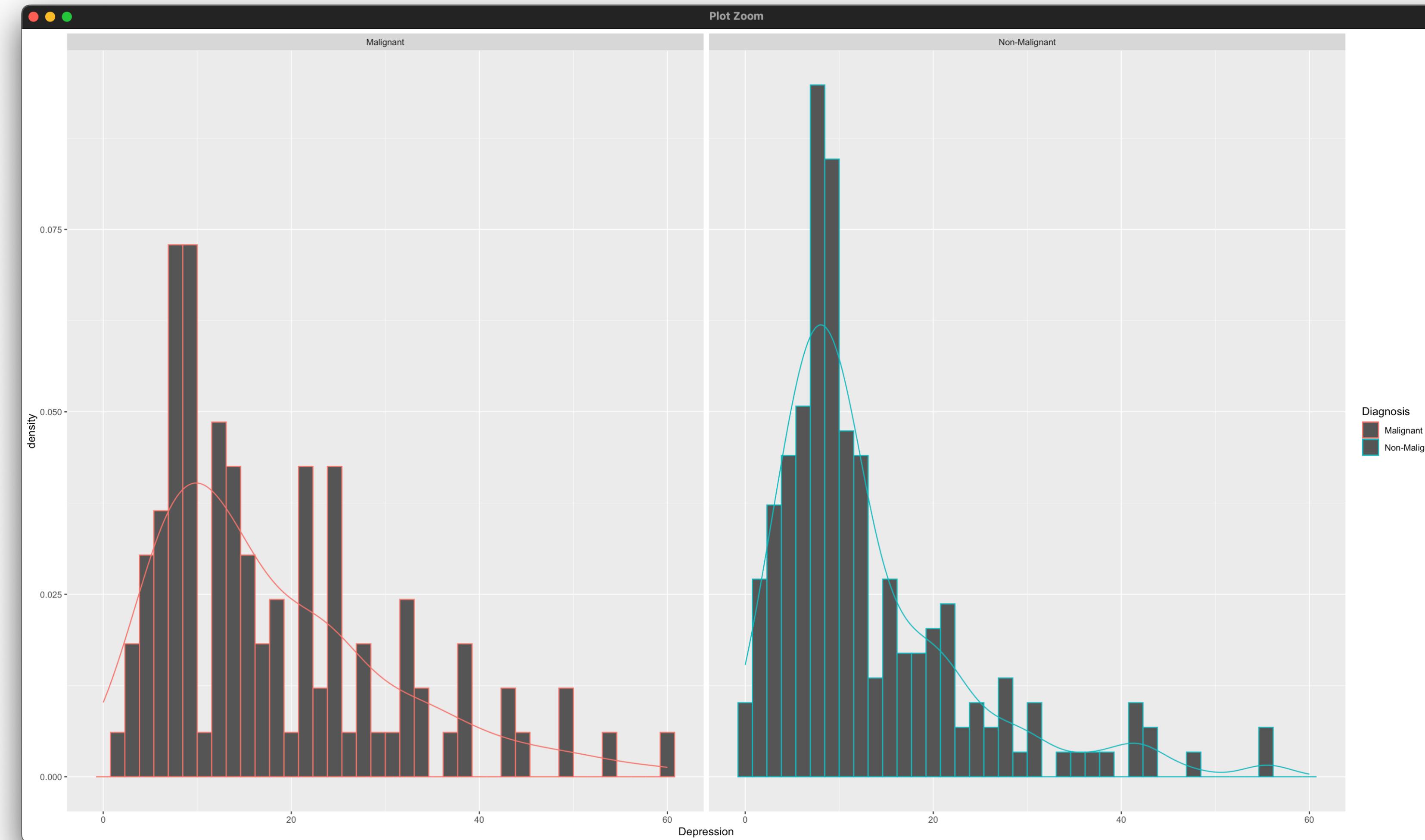
COMPARATIVE RESEARCH QUESTIONS

- Comparative research questions ask whether two or more groups (or occasions) differ from one another
- Question: Do participants diagnosed with cancer differ in depressive symptoms compared to those receiving a negative (no cancer) diagnosis?
- We can begin to answer this question by comparing the distributions of the two groups

GROUP DISTRIBUTIONS



R OUTPUT





One of the main goals of the research study is to determine whether the diagnostic groups differ. In small groups of two or three, examine and discuss the two distributions. Identify specific features of the distributions (e.g., location, spread, shape) that suggest differences between groups, as well as features that suggest similarities. Conclude with a brief statement about whether the visual evidence alone suggests that a cancer diagnosis affects depressive symptoms.

OUTLINE

- 1 Measuring psychological constructs
- 2 Distribution shapes for numeric variables
- 3 Graphical displays for numeric variables
- 4 Graphical and tabular displays for discrete and categorical variables
- 5 Study questions

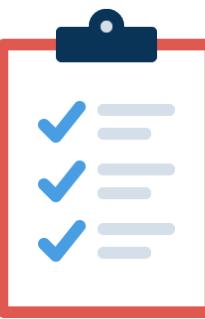
CATEGORICAL VARIABLES

- Nominal variables group features into mutually exclusive, unordered categories with no inherent rank order (e.g., gender identity, racial ethnic identity, clinical diagnosis).
- Ordinal variables group features into ordered categories that can be ranked from low to high (e.g., five-point Likert scales, educational attainment groupings)

BAR PLOTS AND FREQUENCY DISTRIBUTIONS

- A **bar plot** is a graphical display for categorical variables that shows the number (or percent) of responses in each bin
- A **frequency distribution** shows the same information (the range of scores, number and % in each bin) in a table

KEY VARIABLES



Cancer Diagnosis

Based on a diagnostic clinical evaluation for a possible intraocular malignancy, participants were classified as having malignant or nonmalignant diagnoses.

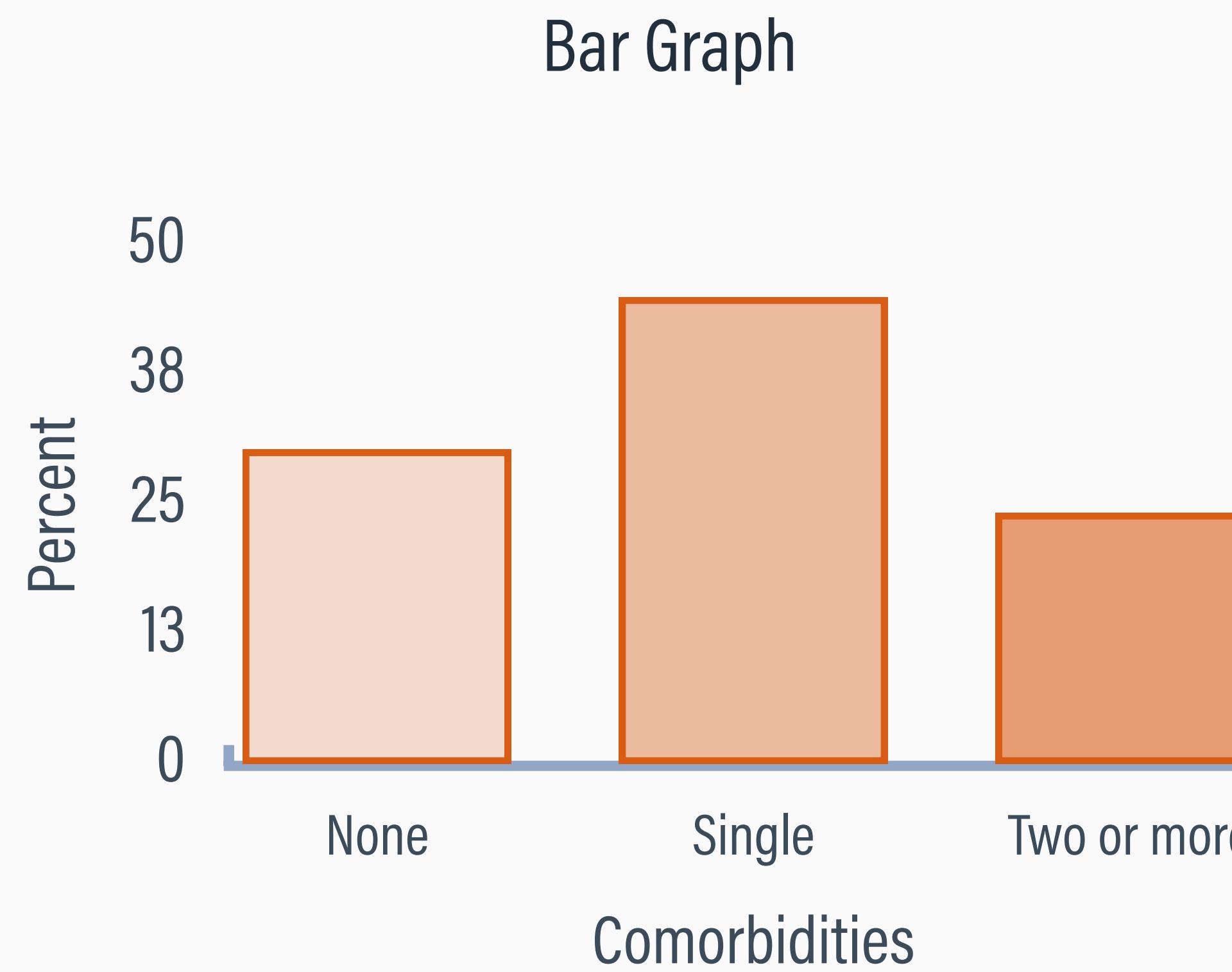


Comorbidities

Participants reported their number of medical comorbidities. Comorbidities were coded 0 (no comorbidities), 1 (a single comorbidity), or 2 (two or more comorbidities).

BAR GRAPH VS. FREQUENCY DISTRIBUTION

- Frequency distributions provide the same information in tabular format

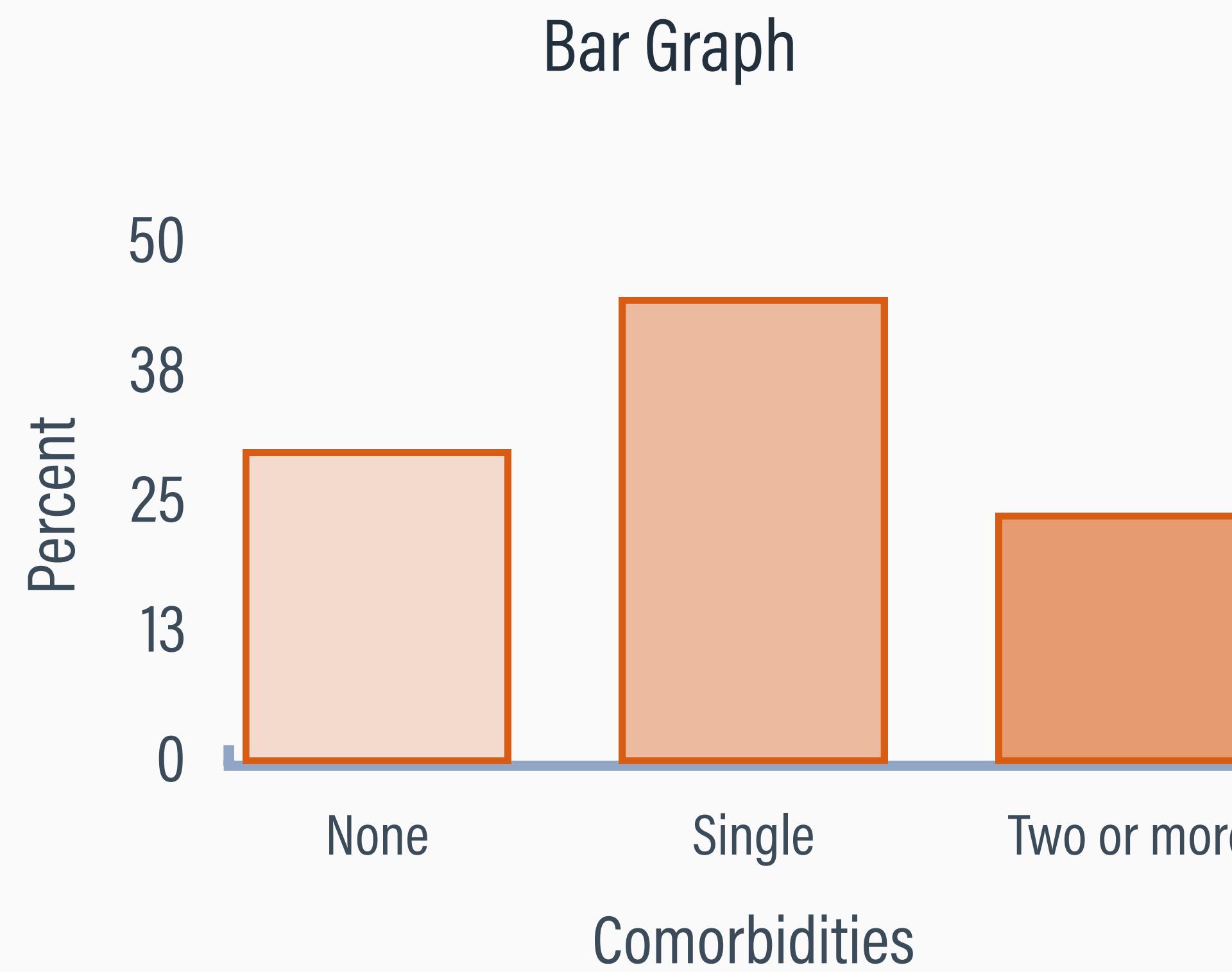


Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

BAR GRAPH VS. FREQUENCY DISTRIBUTION

- Frequency distributions provide the same information in tabular format



Frequency Distribution

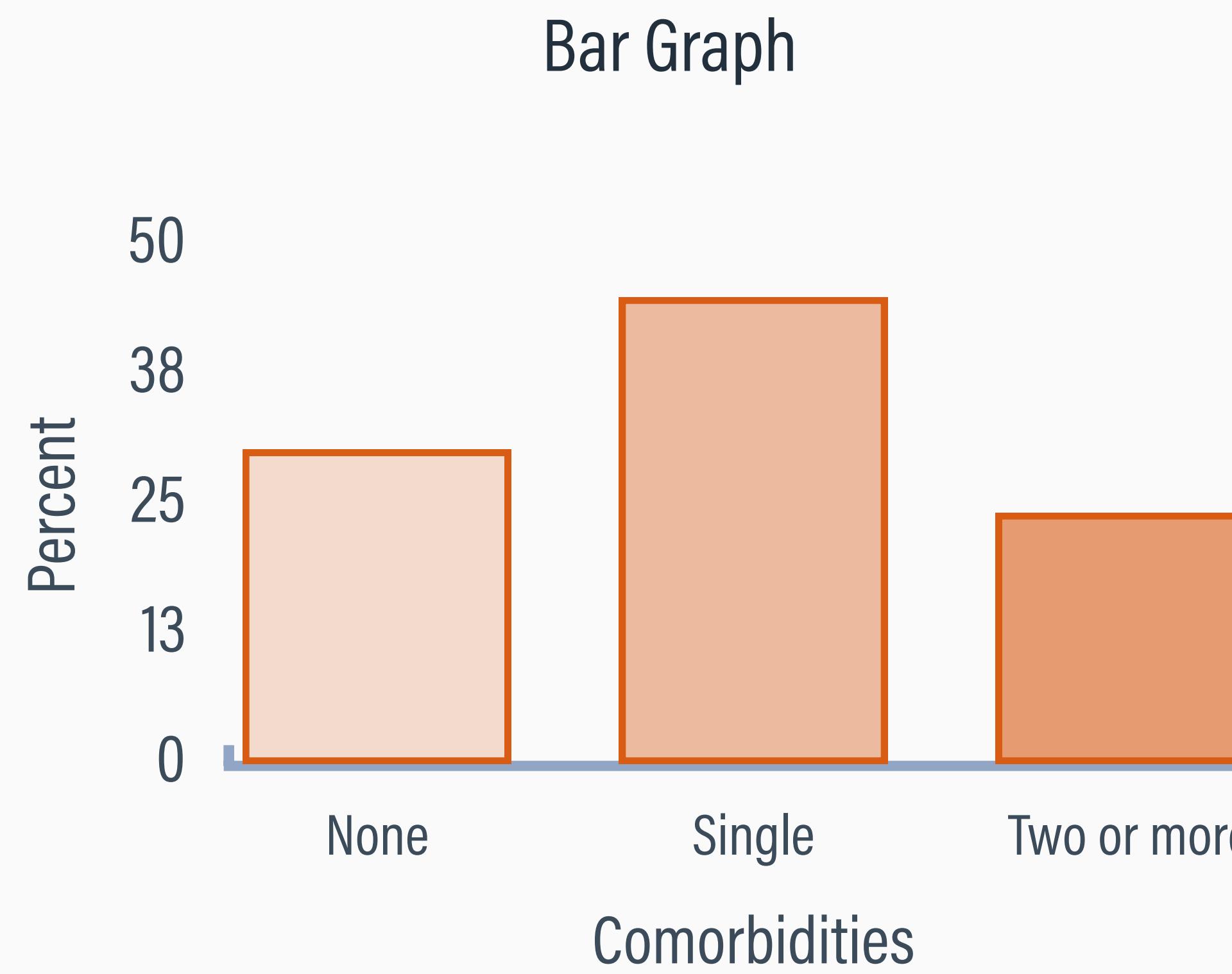
Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Horizontal axis values

↑

BAR GRAPH VS. FREQUENCY DISTRIBUTION

- Frequency distributions provide the same information in tabular format



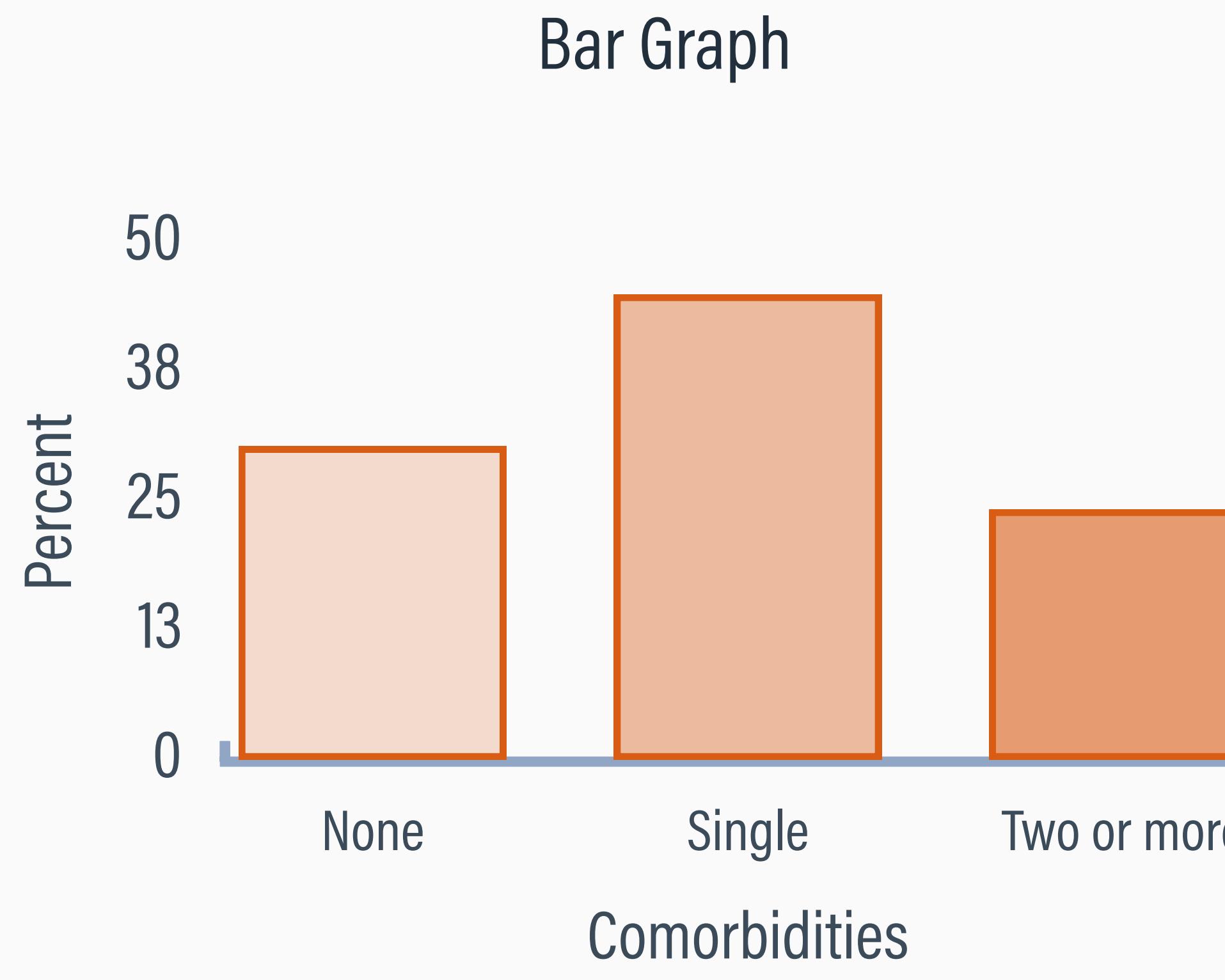
Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Bar height (raw counts)

BAR GRAPH VS. FREQUENCY DISTRIBUTION

- Frequency distributions provide the same information in tabular format



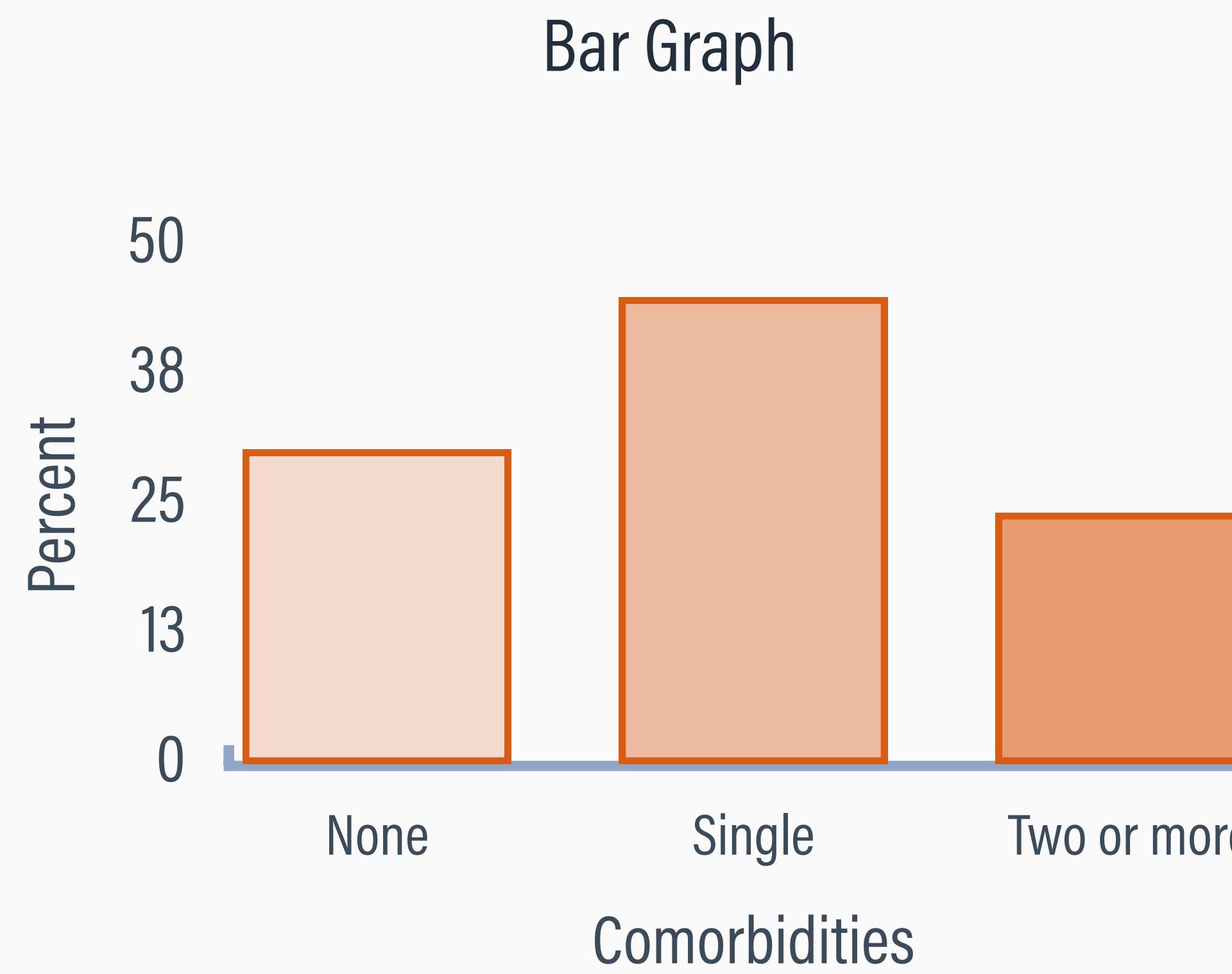
Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Bar height (as %)

BAR GRAPH VS. FREQUENCY DISTRIBUTION

- Frequency distributions provide the same information in tabular format



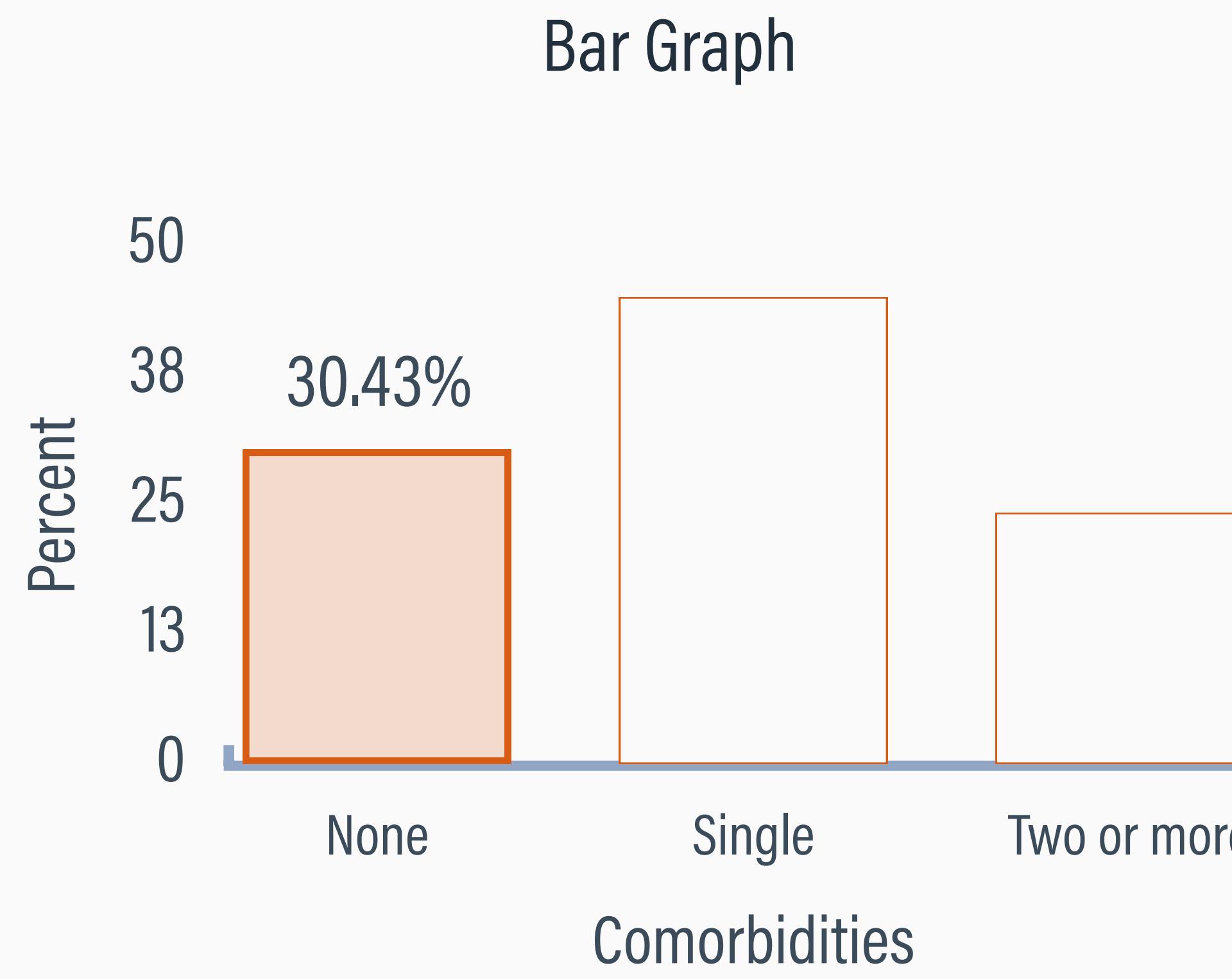
Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Accumulate %s

CUMULATIVE (ACCUMULATING) PERCENTAGES

- Frequency distributions provide the same information in tabular format



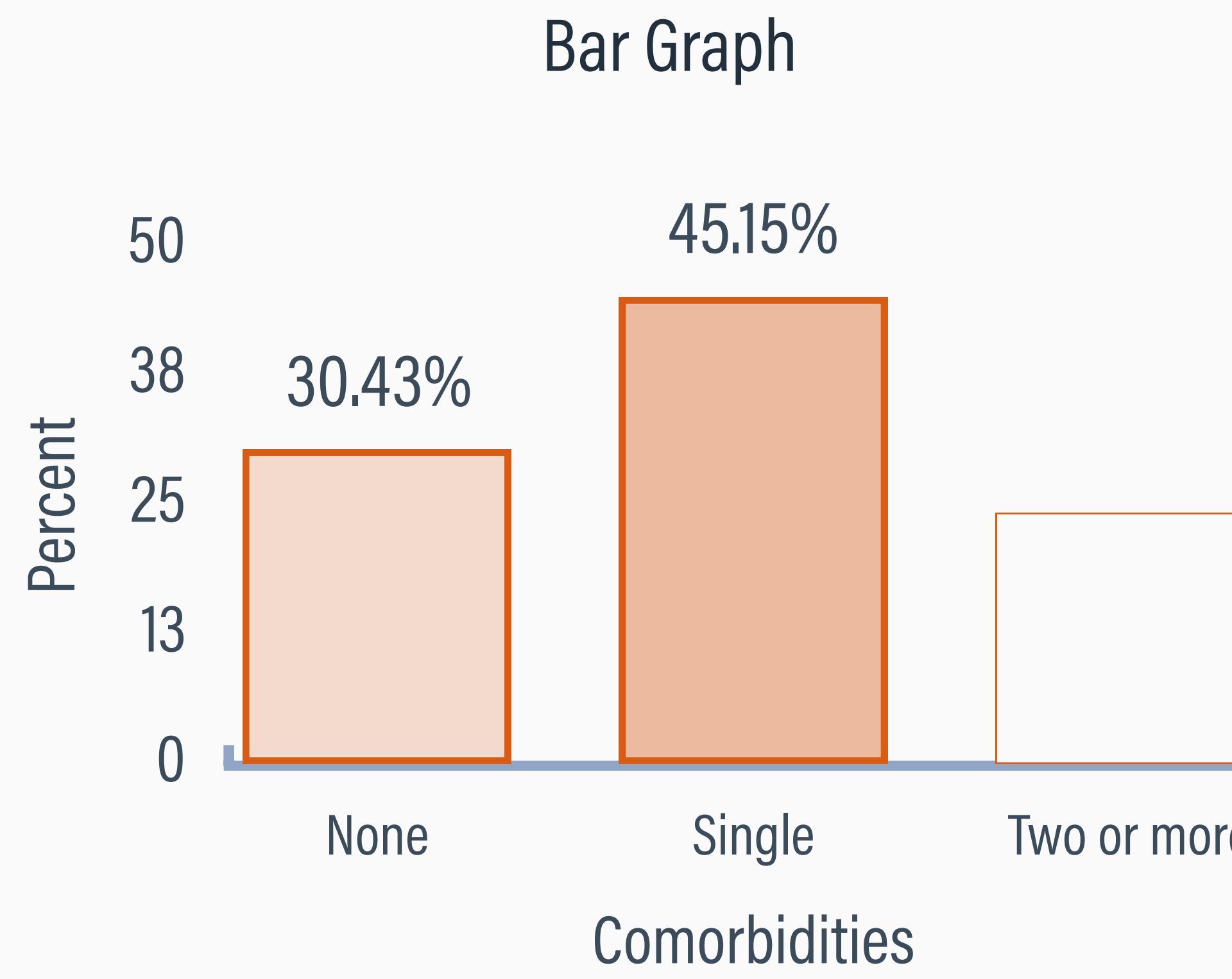
Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Accumulate %s

CUMULATIVE (ACCUMULATING) PERCENTAGES

- Frequency distributions provide the same information in tabular format



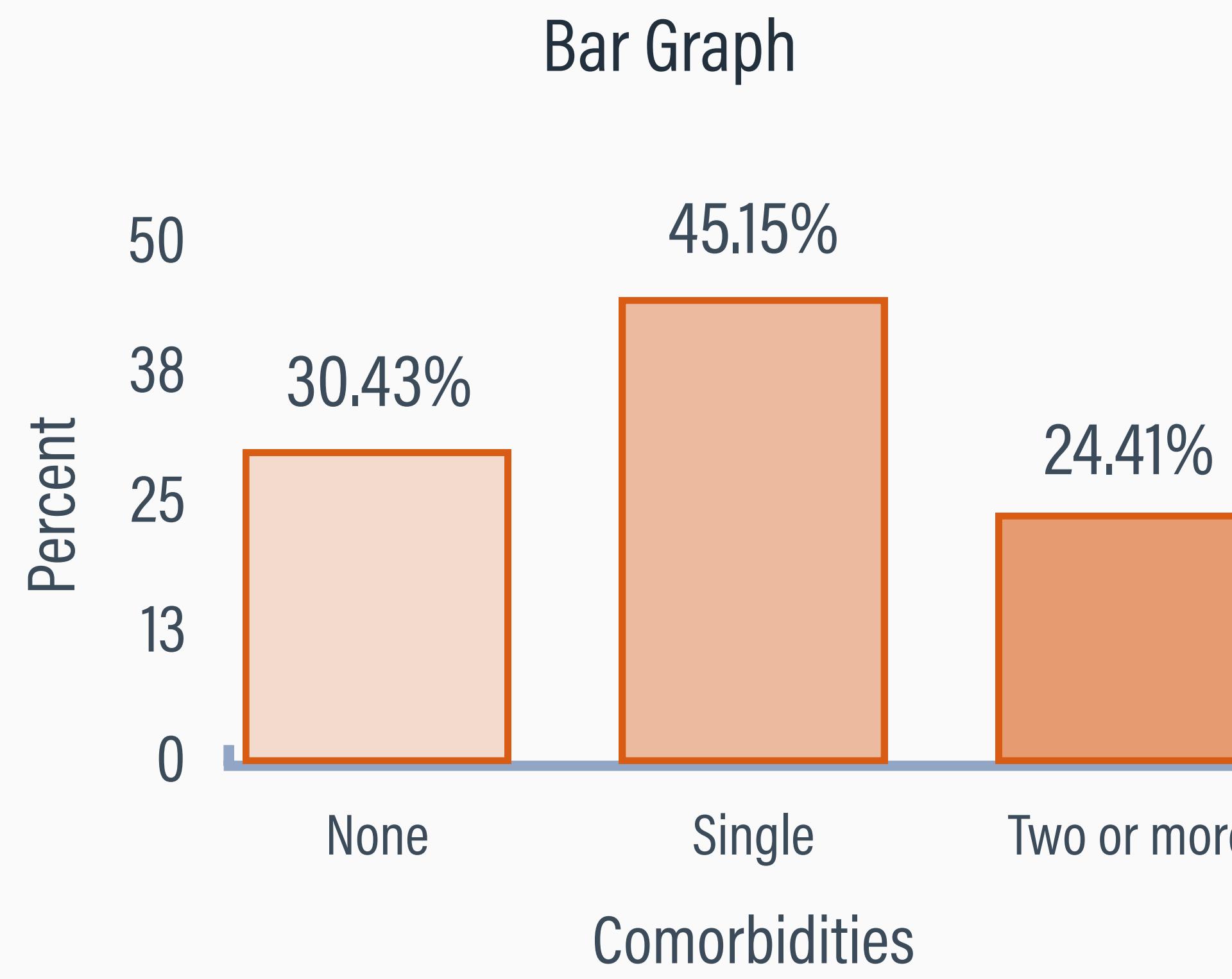
Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Accumulate %s

CUMULATIVE (ACCUMULATING) PERCENTAGES

- Frequency distributions provide the same information in tabular format



Frequency Distribution

Score	Count	%	Cum. %
0 = None	91	30.43	30.43
1 = Single comorbidity	135	45.15	75.59
2 = Two or more	73	24.41	100.0

Accumulate %s

R OUTPUT

Frequencies

Cancer\$Comorbid

Type: Integer

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
0	91	30.43	30.43	30.43	30.43
1	135	45.15	75.59	45.15	75.59
2	73	24.41	100.00	24.41	100.00
<NA>	0			0.00	100.00
Total	299	100.00	100.00	100.00	100.00

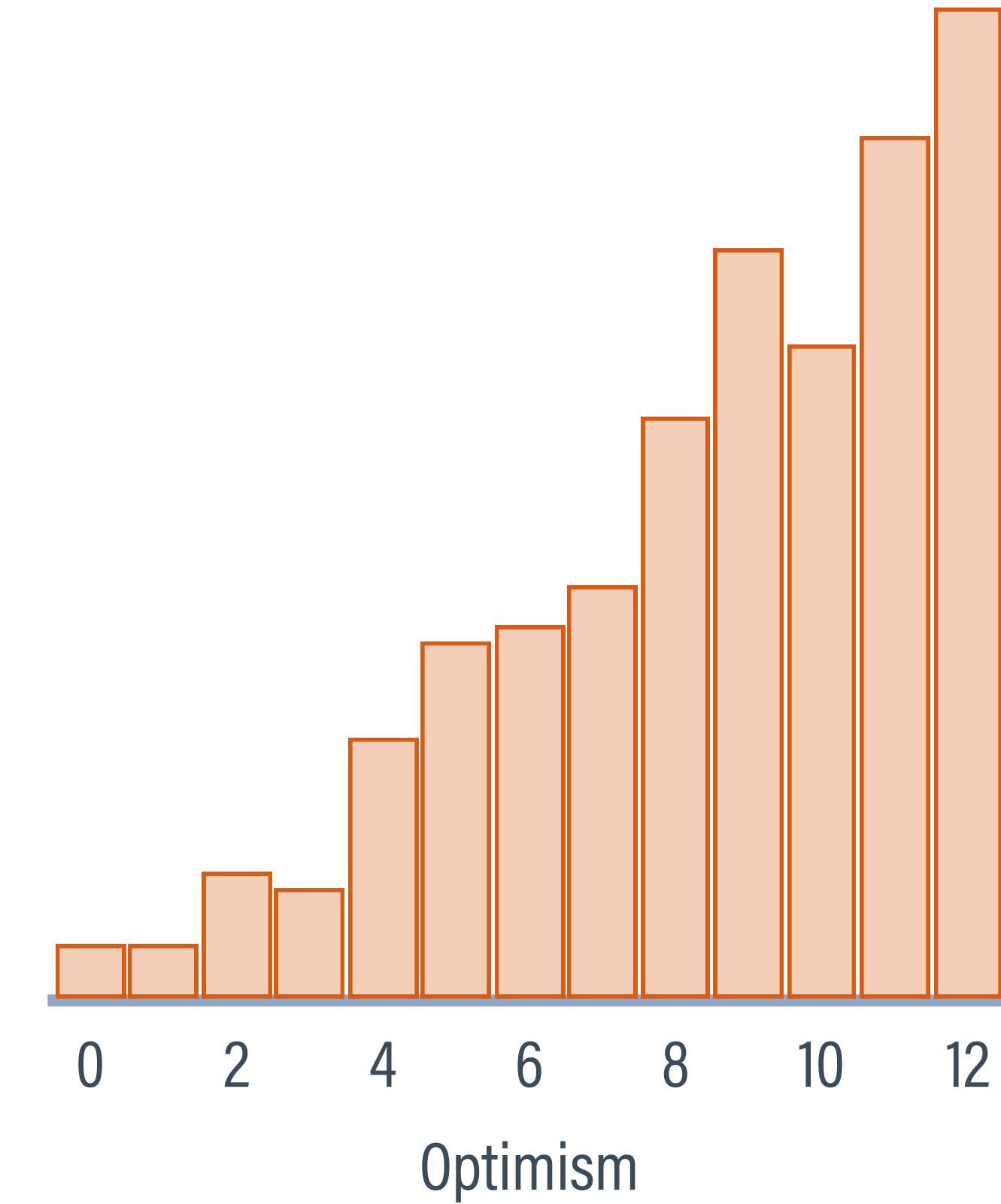


In small groups of two or three, discuss the optimism frequency distribution. What shape best describes the distribution? What features of the table convey the variable's shape?

Score	Count	%	Cum. %
0	3	1.0	1.0
1	3	1.0	2.0
2	7	2.3	4.4
3	6	2.0	6.4
4	14	4.7	11.0
5	19	6.4	17.4
6	20	6.7	24.1
7	22	7.4	31.4
8	31	10.4	41.8
9	40	13.4	55.2
10	35	11.7	66.9
11	46	15.4	82.3
12	53	17.7	100.0



Score	Count	%	Cum. %
0	3	1.0	1.0
1	3	1.0	2.0
2	7	2.3	4.4
3	6	2.0	6.4
4	14	4.7	11.0
5	19	6.4	17.4
6	20	6.7	24.1
7	22	7.4	31.4
8	31	10.4	41.8
9	40	13.4	55.2
10	35	11.7	66.9
11	46	15.4	82.3
12	53	17.7	100.0



In small groups of two or three, examine the frequency distribution and corresponding graph of the optimism scores. Evaluate the strengths and limitations of each representation for understanding the data. In your view, which provides more meaningful insight for summarizing this variable. Why?

OUTLINE

- 1 Measuring psychological constructs
- 2 Distribution shapes for numeric variables
- 3 Graphical displays for numeric variables
- 4 Graphical and tabular displays for discrete and categorical variables
- 5 Study questions

STUDY QUESTIONS (1)

1. Participants in a health psychology study rate their health using four categories: poor, good, very good, and excellent. Sketch a histogram that depicts a sample with relatively poor overall health.

2. What distribution shape would be consistent with a sample where most respondents were in excellent or very good health?

STUDY QUESTIONS (2)

3. A large hospital system evaluates patient satisfaction by asking a large sample of former patients to rate the hospital on a 0 (worst) to 10 (best) scale. Sketch a histogram that depicts a sample with high satisfaction.

4. What distribution shape would be consistent with a sample where most respondents were very satisfied?

STUDY QUESTIONS (3)

5. Compute the missing values from the frequency distribution table of education below.

Response	Count	Percent
Graduate School	50	?
Graduated college	100	20.0
Some college	?	20.0
Graduated high school	250	50.0
Total	500	100

STUDY QUESTIONS (4)

6. Referring to the previous frequency distribution table, compute the cumulative percentage for the “some college” bin and interpret its value.

7. What distribution characterizes the education variable depicted in the frequency table?