

# Measures of Center in a Distribution

The **mean** is what we most commonly call the average value. It is defined as follows:


$$\text{mean} = \frac{\text{sum of all values}}{\text{total number of values}}$$

The **median** is the middle value in the sorted data set (or halfway between the two middle values if the number of values is even).

The **mode** is the most common value (or group of values) in a distribution.

# Middle Value for a Median

---

<u>6.72</u>	3.46	3.60	6.44	<u>26.70</u>	
3.46	3.60	6.44	6.72	26.70	(sorted list)
					(odd number of values)
<b>exact middle</b>			<b><u>median</u> is 6.44</b>		

# No Middle Value for a Median

<u>6.72</u>	3.46	3.60	<u>6.44</u>	
3.46	3.60	6.44	6.72	(sorted list)
				(even number of values)
				<u>median</u> is 5.02

<u>3.60 + 6.44</u>	
2	

# Mode Examples

---

a. 5 5 5 3 1 5 1 4 3 5



Mode is 5

b. 1 2 2 2 3 4 5 6 6 6 7 9



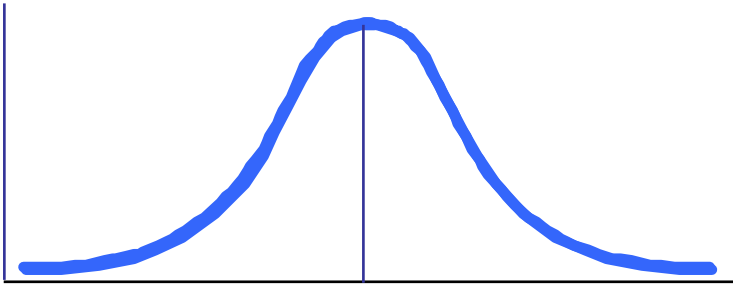
Bimodal (2 and 6)

c. 1 2 3 6 7 8 9 10

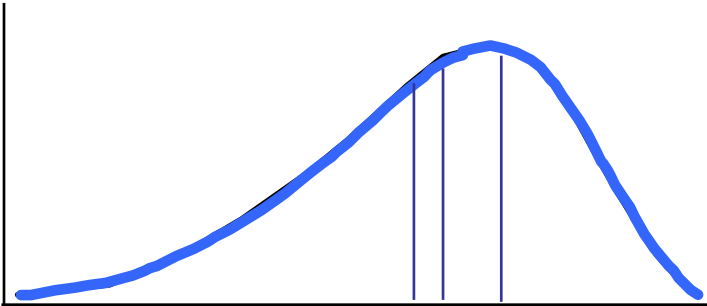


No Mode

# Symmetric and Skewed Distributions

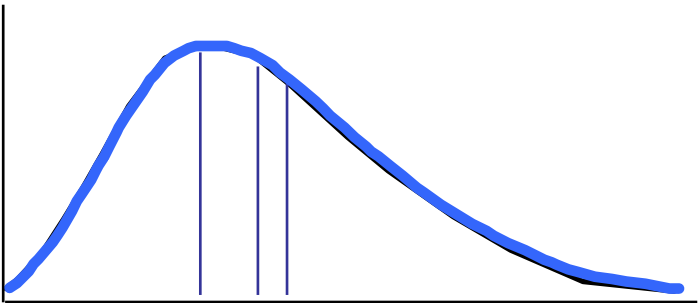


Mode = Mean = Median  
**SYMMETRIC**



Mean ——— | ——— Mode  
          ↑    ↑    ↑  
          Median

**SKEWED LEFT  
(negatively)**

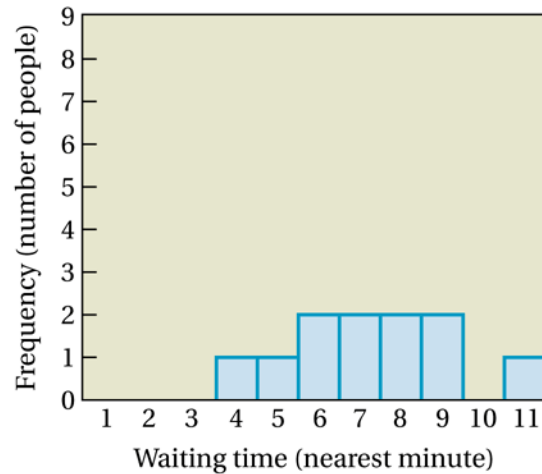


Mode ——— | ——— Mean  
          ↑    ↑    ↑  
          Median

**SKEWED RIGHT  
(positively)**

# Why Variation Matters

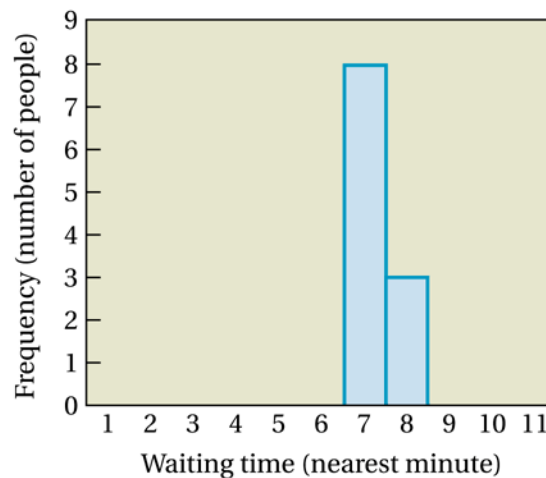
**Big Bank**



**Big Bank** (three line wait times):

4.1	5.2	5.6	6.2	6.7	7.2
7.7	7.7	8.5	9.3	11.0	

**Best Bank**

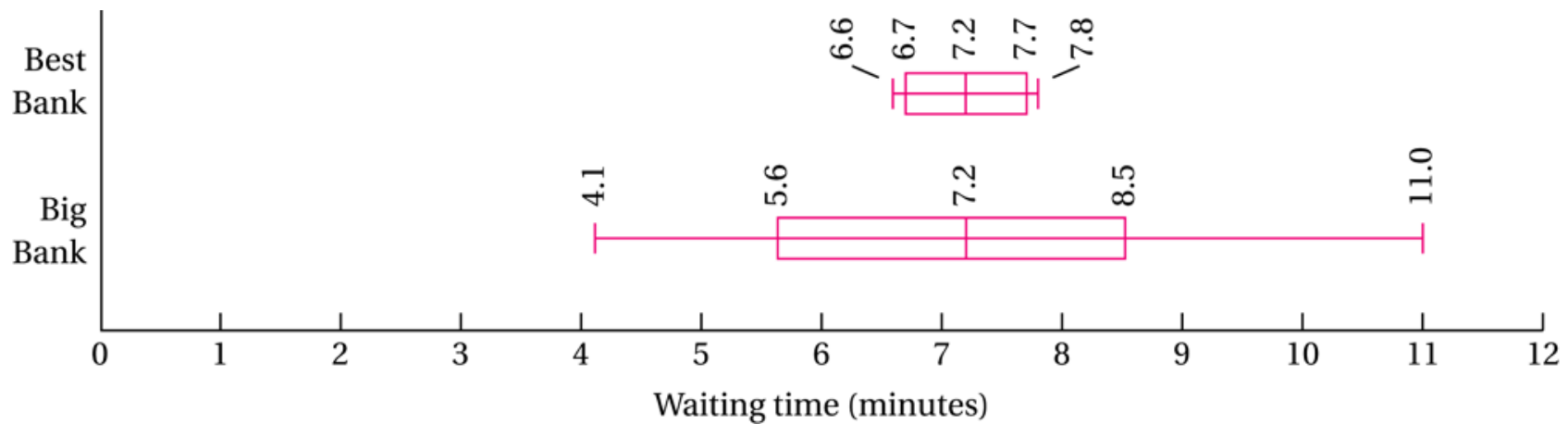


**Best Bank** (one line wait times):

6.6	6.7	6.7	6.9	7.1	7.2
7.3	7.4	7.7	7.8	7.8	

# Five Number Summaries & Box Plots

6-B



## Big Bank

low value (min) = 4.1  
lower quartile = 5.6  
median = 7.2  
upper quartile = 8.5  
high value (max) = 11.0

## Best Bank

low value (min) = 6.6  
lower quartile = 6.7  
median = 7.2  
upper quartile = 7.7  
high value (max) = 7.8

# Standard Deviation

Let  $A = \{2, 8, 9, 12, 19\}$  with a mean of 10. Use the data set  $A$  above to find the sample standard deviation.

$x$ (data value)	$x - \text{mean}$ (deviation)	(deviation) <sup>2</sup>
2	$2 - 10 = -8$	$(-8)^2 = 64$
8	$8 - 10 = -2$	$(-2)^2 = 4$
9	$9 - 10 = -1$	$(-1)^2 = 1$
12	$12 - 10 = 2$	$(2)^2 = 4$
19	$19 - 10 = 9$	$(9)^2 = 81$
	Total	<b>154</b>

$$\text{standard deviation} = \sqrt{\frac{154}{5-1}} = 6.2$$

$$\text{standard deviation} = \sqrt{\frac{\text{sum of (deviations from the mean)}^2}{\text{total number of data values} - 1}}$$

# Range Rule of Thumb for Standard Deviation

$$\text{standard deviation} \approx \frac{\text{range}}{4}$$

We can estimate standard deviation by taking an approximate range, (usual high – usual low), and dividing by 4. The reason for the 4 is due to the idea that the usual high is approximately 2 standard deviations above the mean and the usual low is approximately 2 standard deviations below.

Going in the opposite direction, on the other hand, if we know the standard deviation we can estimate:

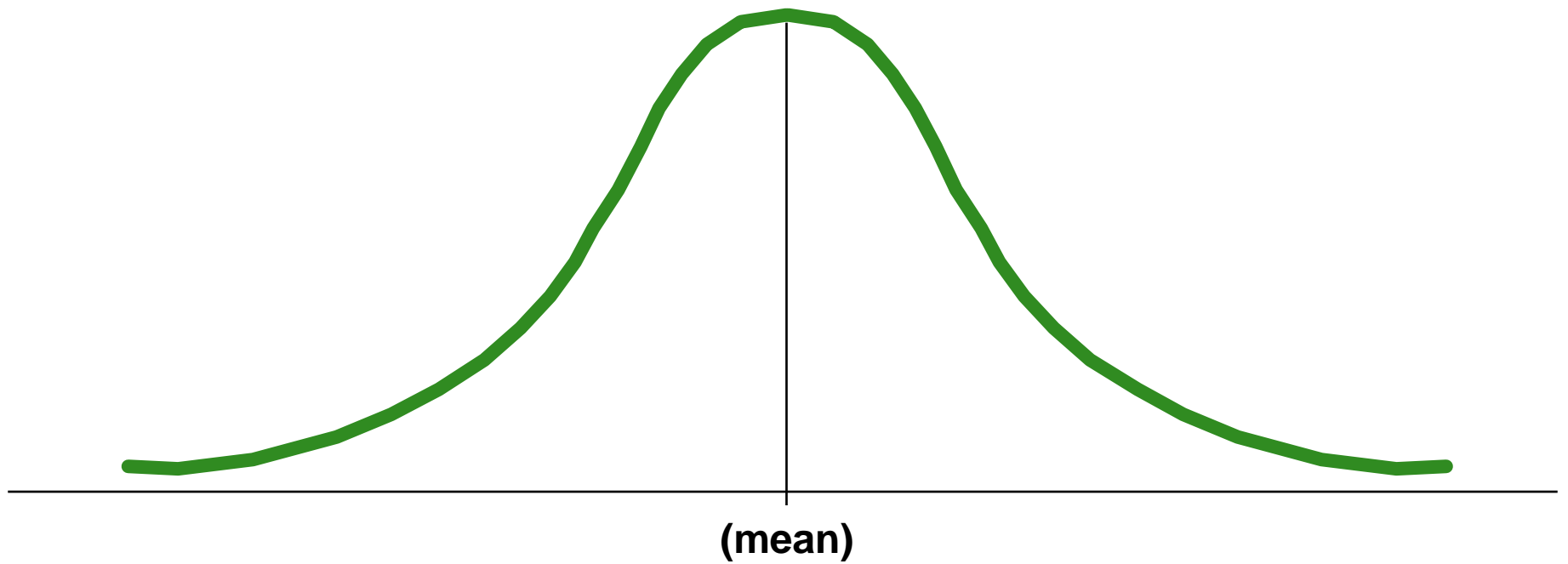
$$\text{low value} \approx \text{mean} - 2 \times \text{standard deviation}$$

$$\text{high value} \approx \text{mean} + 2 \times \text{standard deviation}$$

# The 68-95-99.7 Rule for a Normal Distribution

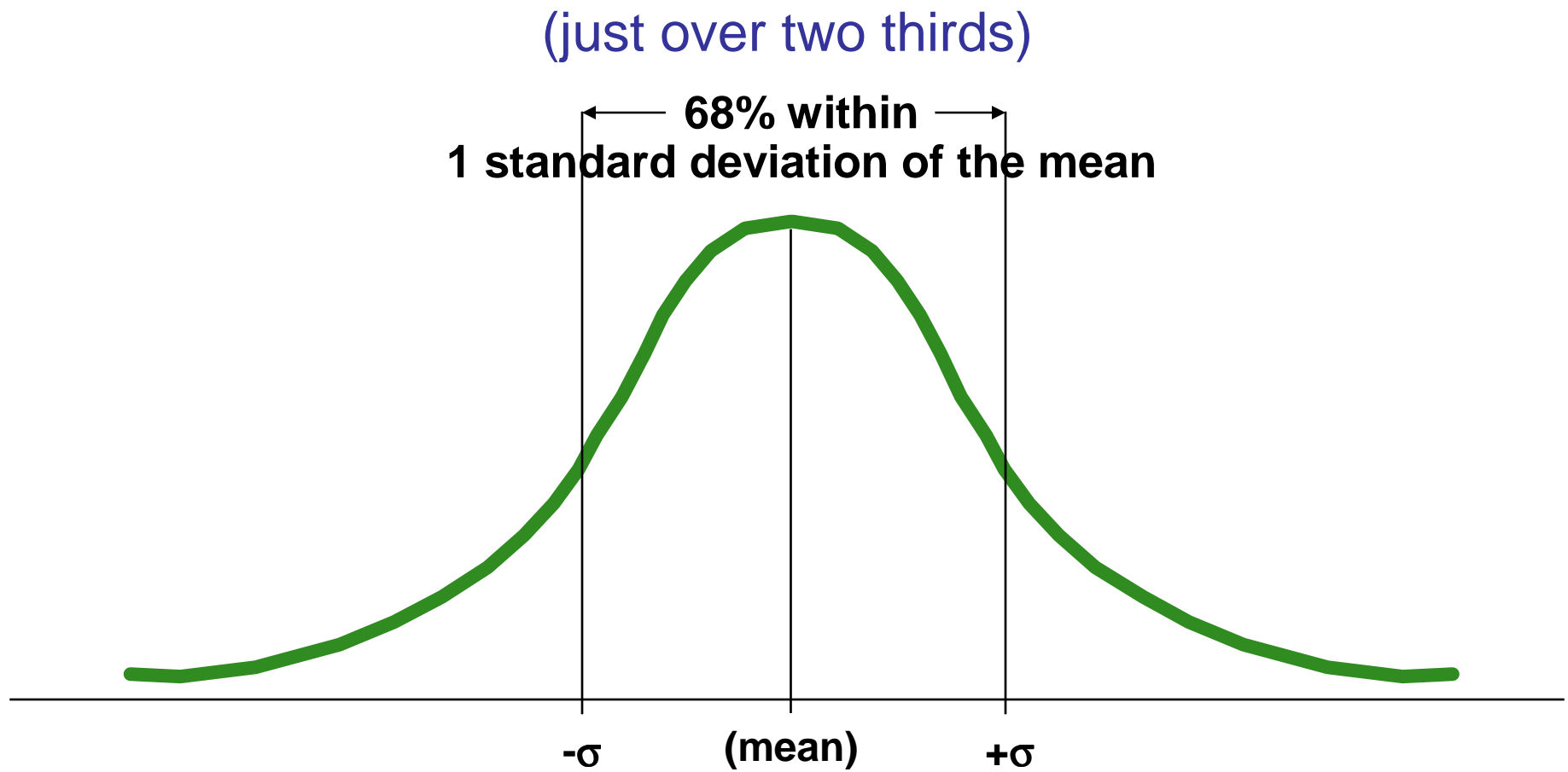
---

6-C



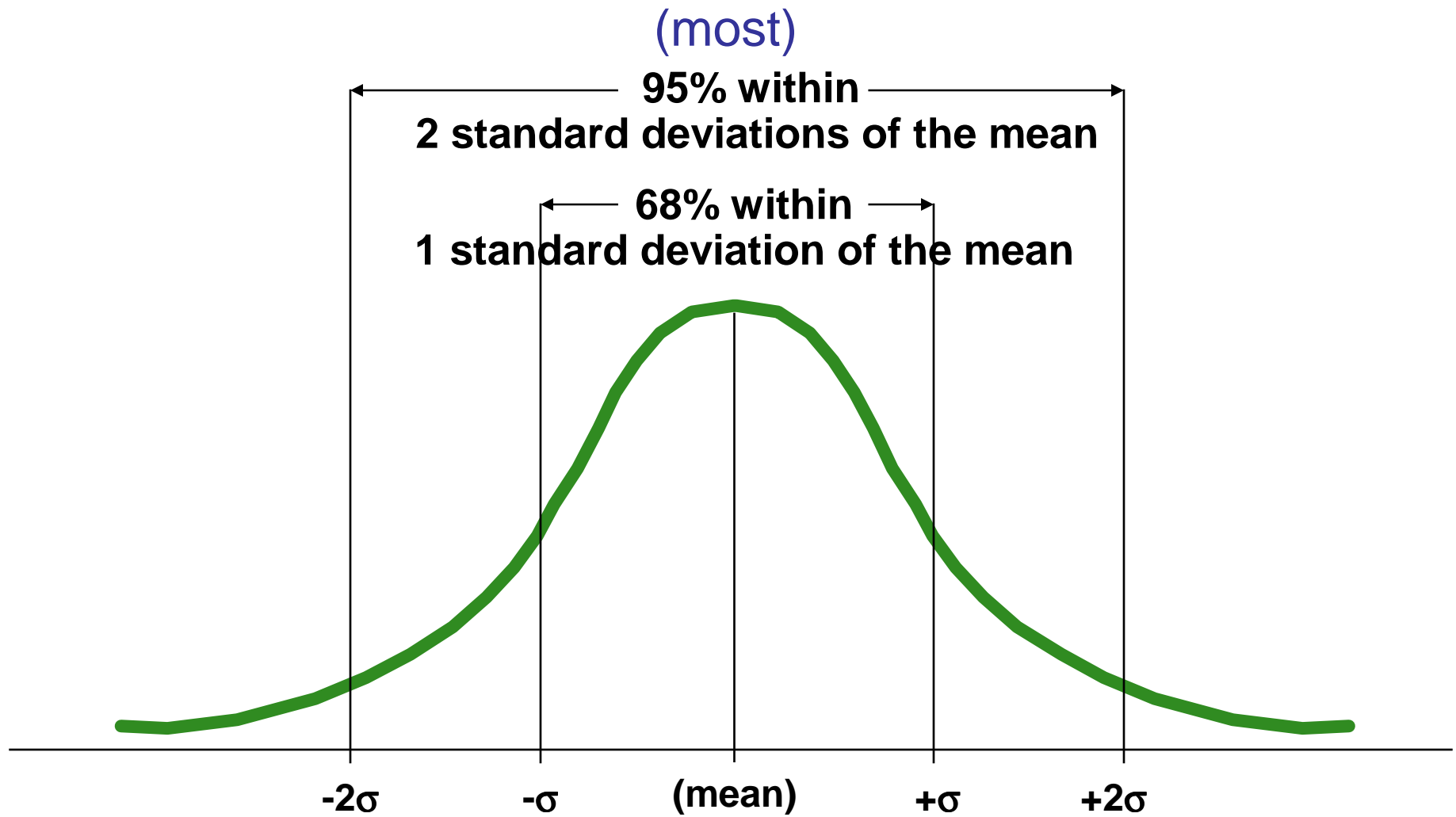
# The 68-95-99.7 Rule for a Normal Distribution

6-C



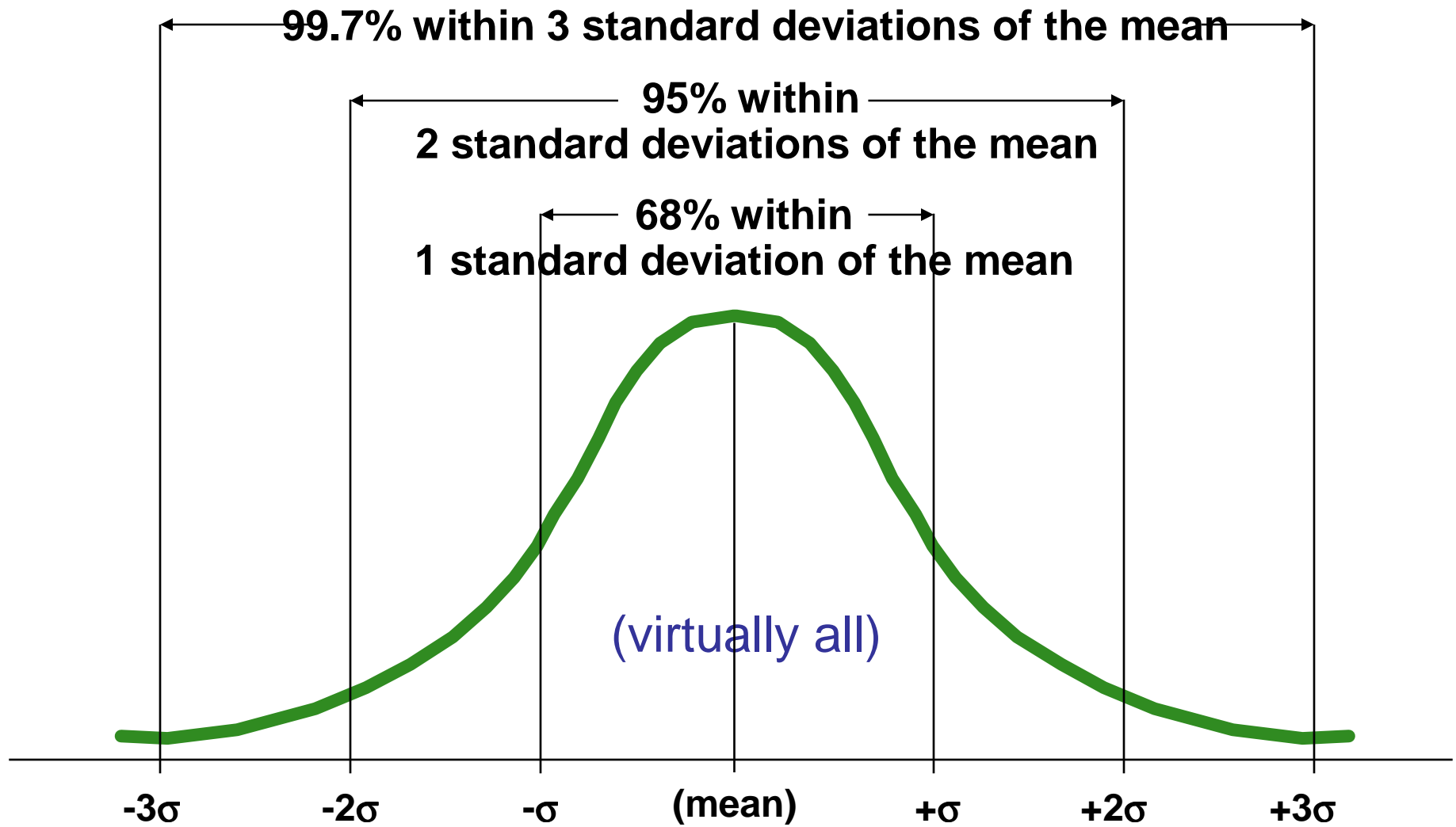
# The 68-95-99.7 Rule for a Normal Distribution

6-C

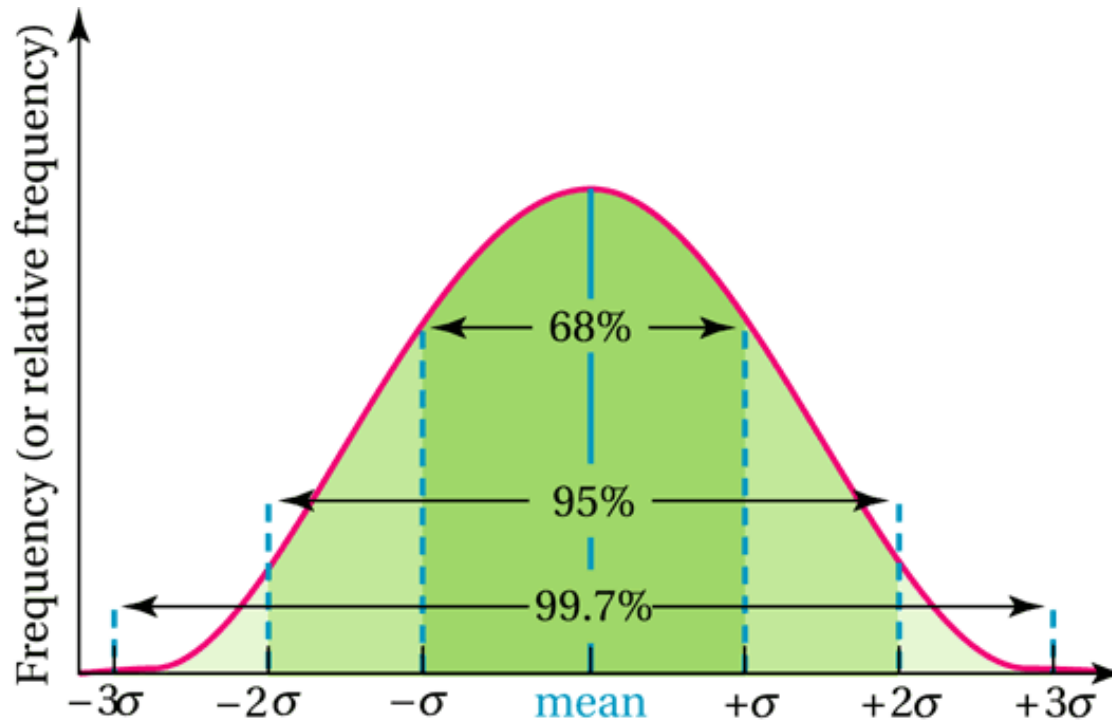


# The 68-95-99.7 Rule for a Normal Distribution

6-C



# Also known as the Empirical Rule



# Z-Score Formula

---

$$\text{standard score} = z = \frac{\text{data value} - \text{mean}}{\text{standard deviation}}$$

Example:

If the nationwide ACT mean were 21 with a standard deviation of 4.7, find the z-score for a 30.

What does this mean?

$$z = \frac{30 - 21}{4.7} = 1.91$$

This means that an ACT score of 30 would be about 1.91 standard deviations above the mean of 21.

# Standard Scores and Percentiles

**TABLE 6.3** Standard Scores and Percentiles

Z-SCORE	PERCENTILE	Z-SCORE	PERCENTILE	Z-SCORE	PERCENTILE	Z-SCORE	PERCENTILE
-3.5	0.02	-1.0	15.87	0.0	50.00	1.1	86.43
-3.0	0.13	-0.95	17.11	0.05	51.99	1.2	88.49
-2.9	0.19	-0.90	18.41	0.10	53.98	1.3	90.32
-2.8	0.26	-0.85	19.77	0.15	55.96	1.4	91.92
-2.7	0.35	-0.80	21.19	0.20	57.93	1.5	93.32
-2.6	0.47	-0.75	22.66	0.25	59.87	1.6	94.52
-2.5	0.62	-0.70	24.20	0.30	61.79	1.7	95.54
-2.4	0.82	-0.65	25.78	0.35	63.68	1.8	96.41
-2.3	1.07	-0.60	27.43	0.40	65.54	1.9	97.13
-2.2	1.39	-0.55	29.12	0.45	67.36	2.0	97.72
-2.1	1.79	-0.50	30.85	0.50	69.15	2.1	98.21
-2.0	2.28	-0.45	32.64	0.55	70.88	2.2	98.61
-1.9	2.87	-0.40	34.46	0.60	72.57	2.3	98.93
-1.8	3.59	-0.35	36.32	0.65	74.22	2.4	99.18
-1.7	4.46	-0.30	38.21	0.70	75.80	2.5	99.38
-1.6	5.48	-0.25	40.13	0.75	77.34	2.6	99.53
-1.5	6.68	-0.20	42.07	0.80	78.81	2.7	99.65
-1.4	8.08	-0.15	44.04	0.85	80.23	2.8	99.74
-1.3	9.68	-0.10	46.02	0.90	81.59	2.9	99.81
-1.2	11.51	-0.05	48.01	0.95	82.89	3.0	99.87
-1.1	13.57	0.0	50.00	1.0	84.13	3.5	99.98

# Inferential Statistics Definitions

---

- A set of measurements or observations in a statistical study is said to be **statistically significant** if it is unlikely to have occurred by chance.
- 0.05 (1 out of 20) or 0.01 (1 out of 100) are two very common **levels of significance** that indicate the probability that an observed difference is simply due to chance.
- At the level of 95% confidence we can estimate the **margin of error** that a sample of size  $n$  is different from the actual population measurement (parameter) by calculating:

$$\text{margin of error} \approx \frac{1}{\sqrt{n}}$$

- The **null hypothesis** claims a specific value for a population parameter null hypothesis: population parameter = claimed value
- The **alternative hypothesis** is the claim that is accepted if the null hypothesis is rejected.

# Outcomes of a Hypothesis Test

---

- Rejecting the null hypothesis, in which case we have evidence that supports the alternative hypothesis.
- Not rejecting the null hypothesis, in which case we lack sufficient evidence to support the alternative hypothesis.