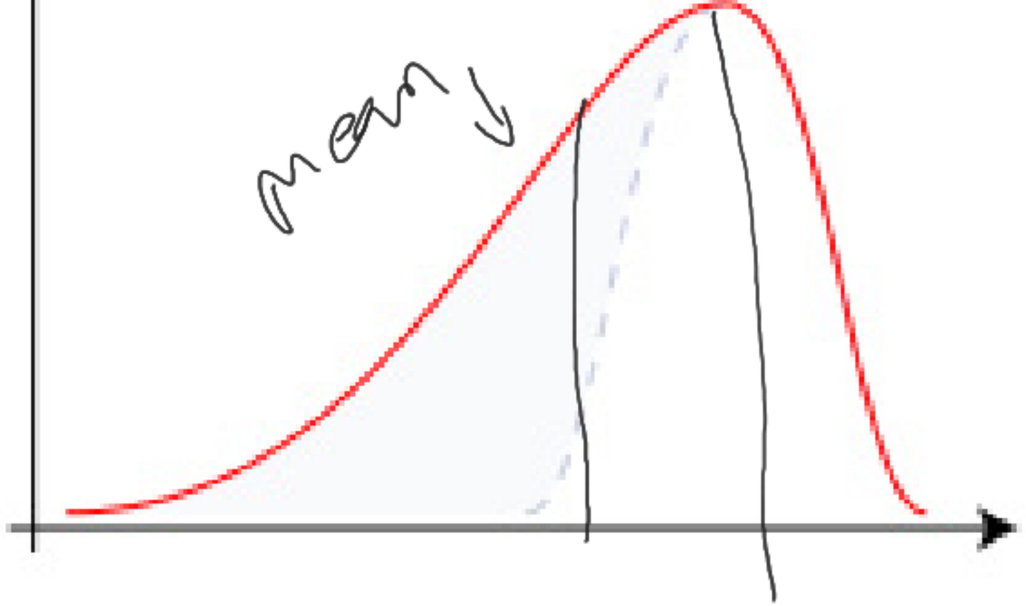
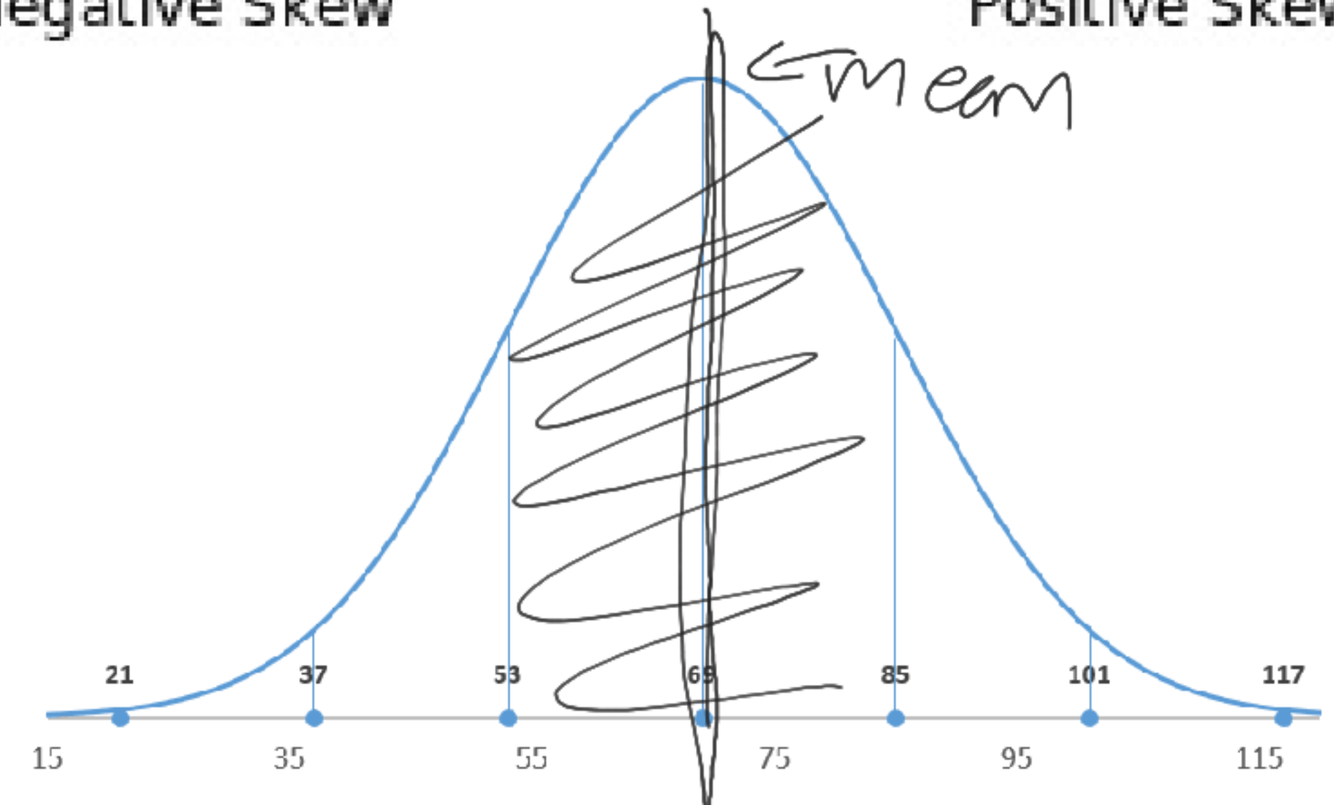


Negative Skew

Positive Skew



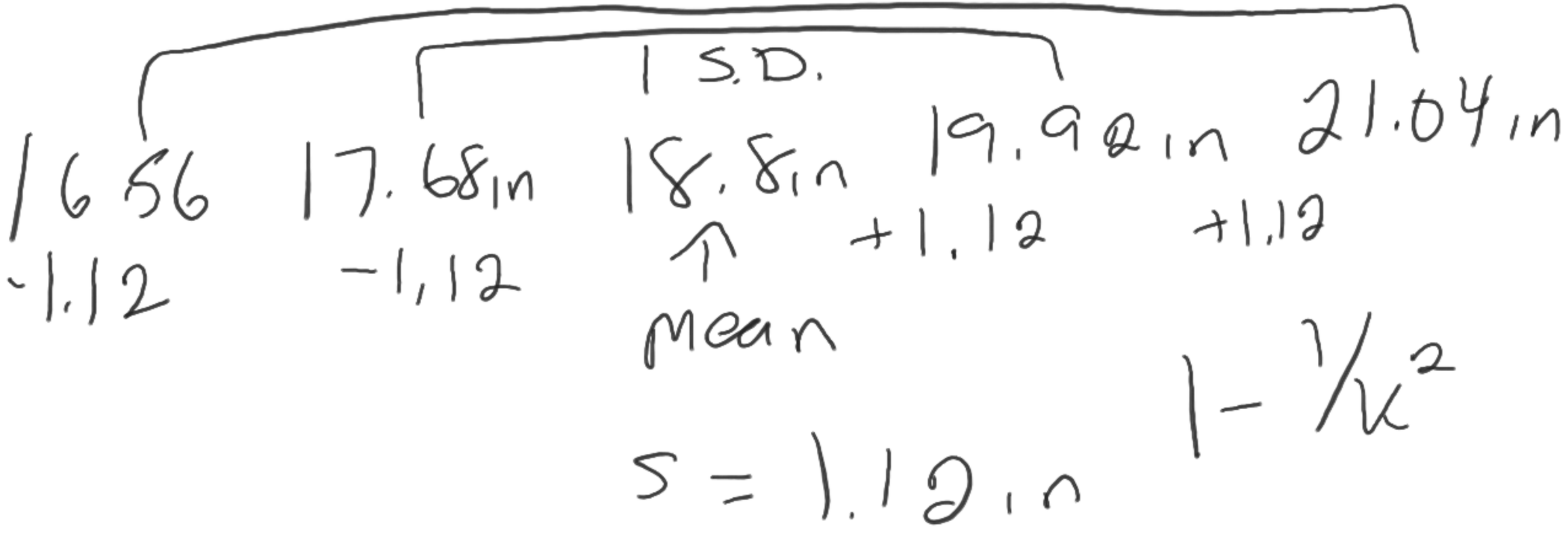
Key Fact 3.3: Chebyshev's Rule

For any quantitative data set and any real number k greater than or equal to 1, at least $1 - 1/k^2$ of the observations lie within k standard deviations to either side of the mean, that is, between $\bar{x} - k \cdot s$ and $\bar{x} + k \cdot s$.

Forearm Length In 1903, K. Pearson and A. Lee published the paper "On the Laws of Inheritance in Man. I. Inheritance of Physical Characters" (*Biometrika*, Vol. 2, pp. 357–462). The article examined data on forearm length, in inches, for a sample of 140 men. The mean and standard deviation of the forearm lengths are 18.8 in. and 1.12 in., respectively.

a. Apply Chebyshev's rule with $k = 2$ to make pertinent statements about the forearm lengths of the men in the sample.

b. Repeat part (a) with $k = 3$.



$s = 1.12$ in

$1 - 1/k^2$

2SD → 16.56 to 21.04 $\geq 75\%$

3SD → 15.44 to 22.16 $\geq 88.9\%$

Key Fact 3.4: Empirical Rule

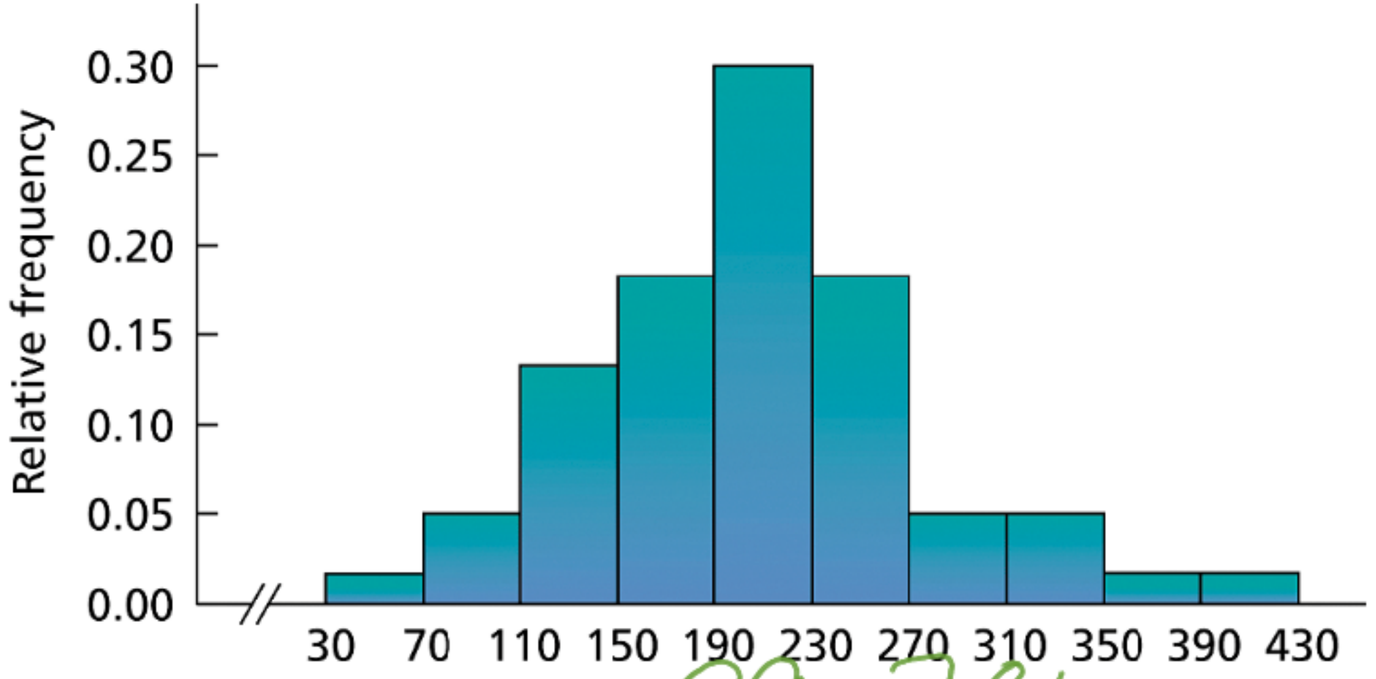
For any quantitative data set with roughly a bell-shaped distribution, the following properties hold.

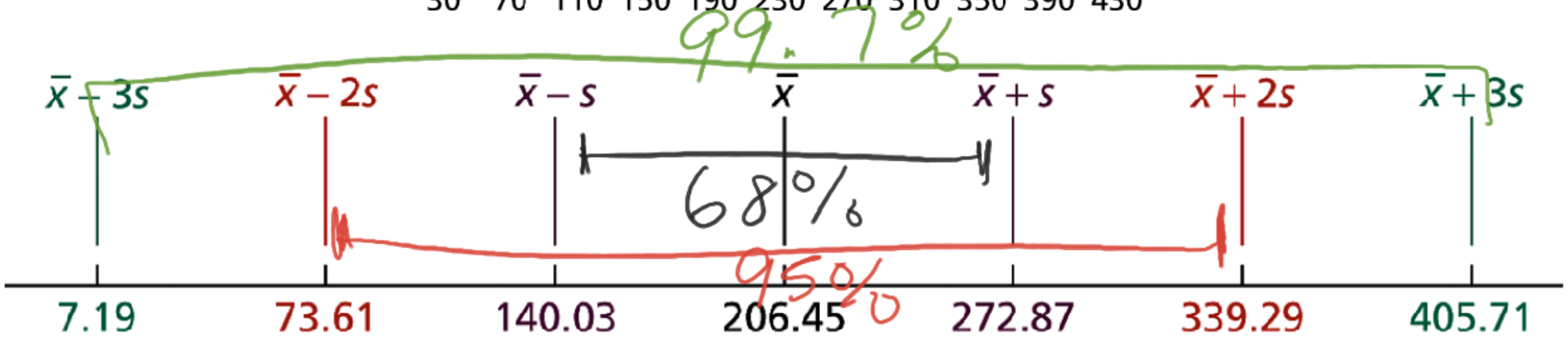
Property 1: Approximately 68% of the observations lie within one standard deviation to either side of the mean, that is, between $\bar{x} - s$ and $\bar{x} + s$.

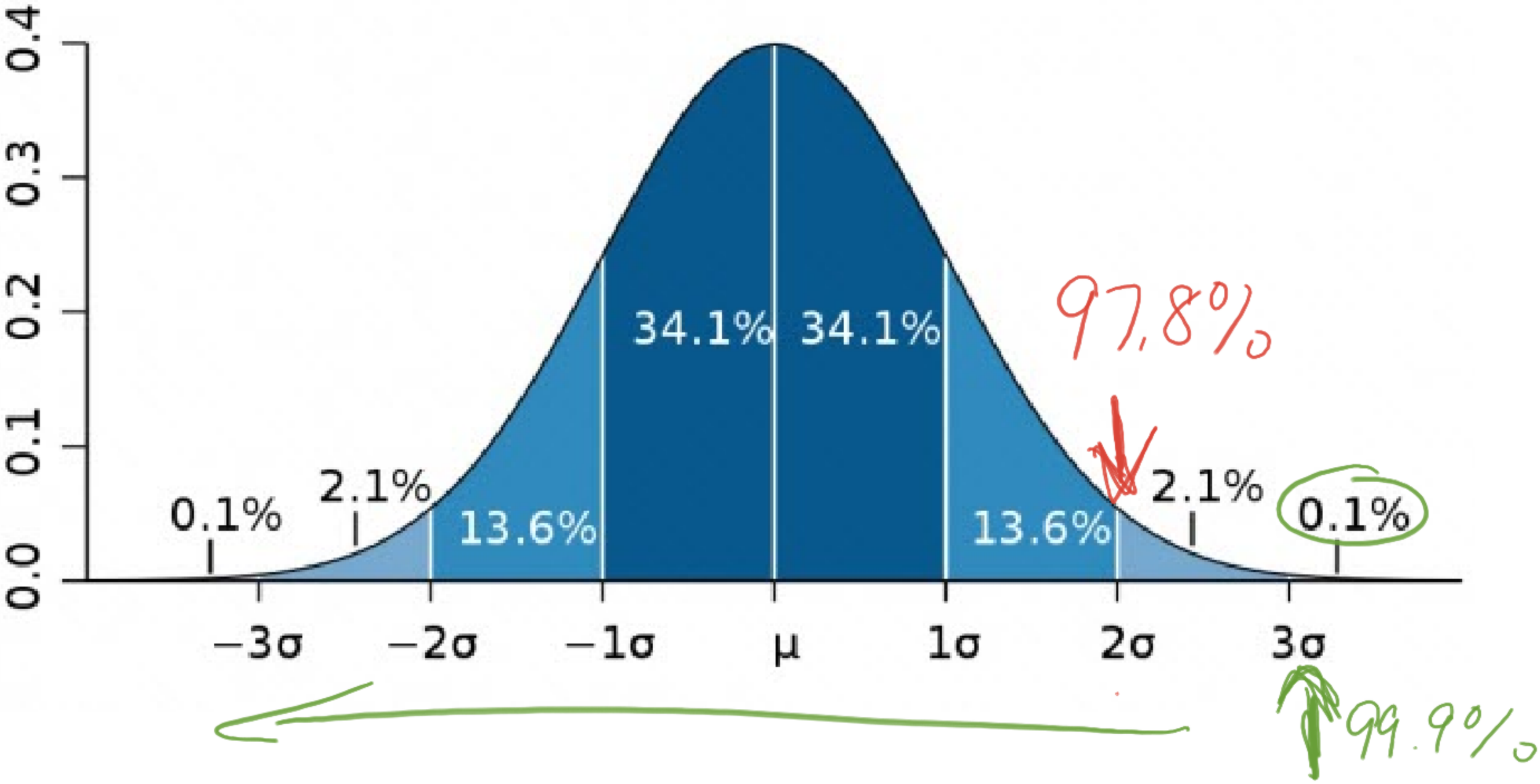
Property 2: Approximately 95% of the observations lie within two standard deviations to either side of the mean, that is, between $\bar{x} - 2s$ and $\bar{x} + 2s$.

Property 3: Approximately 99.7% of the observations lie within three standard deviations to either side of the mean, that is, between $\bar{x} - 3s$ and $\bar{x} + 3s$.

$$\bar{x} = 206.45 \text{ ppm and } s = 66.42 \text{ ppm.}$$






0.0
0.1
0.2
0.3
0.4
0.1%
2.1%
13.6%
34.1%
34.1%
13.6%
2.1%
0.1%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ
97.8%
99.9%

Definition 3.15: *z*-Score

For an observed value of a variable x, the corresponding value of the standardized variable z is called the ***z*-score** of the observation. The term **standard score** is often used instead of *z-score*.

$$Z = \frac{X - \mu}{\sigma}$$

Augusta National Golf Course The lengths of the holes on the Augusta National Golf Course are presented in the third column of Table 3.17. We determined earlier that the mean and standard deviation of the lengths are 413.1 yd and 132.0 yd, respectively. So, in this case, the standardized variable is

a. Find and interpret the z-score of Juniper's length of 180 yd.

b. Find and interpret the z-score of Camellia's length of 495 yd.

a) $z = \frac{180 - 413.1}{132}$

$= -1.766$

b) $z = \frac{495 - 413.1}{132}$

$= 0.62$

Definition 3.7: Quartiles

First, arrange the data in increasing order. Next, determine the median. Then, divide the (ordered) data set into two halves, a bottom half and a top half; if the number of observations is odd, include the median in both halves.

- The **first quartile** (Q_1) is the median of the bottom half of the data set.
- The **second quartile** (Q_2) is the median of the entire data set.
- The **third quartile** (Q_3) is the median of the top half of the data set.

Definition 3.8: Interquartile Range

The **interquartile range**, or **IQR**, is the difference between the first and third quartiles; that is, $\text{IQR} = Q_3 - Q_1$.

Weekly TV-Viewing Times The A. C. Nielsen Company publishes information on the TV-viewing habits of Americans in *Nielsen Report on Television*. A sample of 20 people yielded the weekly viewing times, in hours, displayed in Table 3.13. Determine and interpret the quartiles for these data.

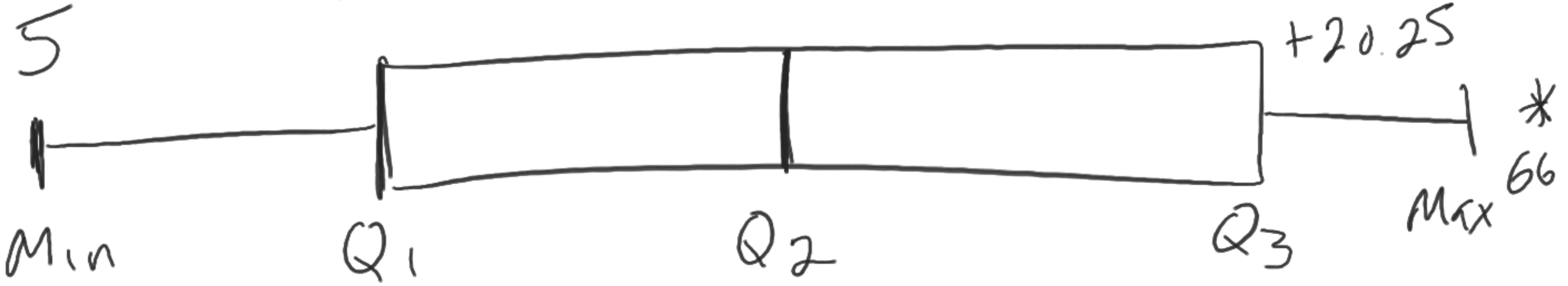
Table 3.13 Weekly TV-viewing times

25	41	27	32	43
66	35	31	15	5
34	26	32	38	16
30	38	30	20	21

5 15 16 20 21 25 26 27 30 30 31 32 32 34 35 38 38 41 43 66

Q_1 Q_2 Q_3

23 30.5 36.5



IQR $Q_3 - Q_1 = 13.5$

Outlier $\rightarrow 1.5(IQR) = 20.25$

Definition 3.10: Lower and Upper Limits

The **lower limit** and **upper limit** of a data set are

$$\text{Lower limit} = Q_1 - 1.5 \cdot \text{IQR};$$
$$\text{Upper limit} = Q_3 + 1.5 \cdot \text{IQR}.$$

Definition 3.9: Five-Number Summary

The **five-number summary** of a data set is Min, Q_1, Q_2, Q_3, Max.

To Construct a Boxplot

Step 1 Determine the quartiles.

Step 2 Determine potential outliers and the adjacent values.

Step 3 Draw a horizontal axis on which the numbers obtained in Steps 1 and 2 can be located. Above this axis, mark the quartiles and the adjacent values with vertical lines.

Step 4 Connect the quartiles to make a box, and then connect the box to the adjacent values with lines.

Step 5 Plot each potential outlier with an asterisk.