Interquartile Range

11.1 - Using Normal Distributions

Spread and Range

- Always report a measure of spread along with a measure of center when describing a distribution numerically.
- The range of the data is the difference between the maximum and minimum values.

Range = max - min

- A disadvantage of the range is that a single extreme value can make it very large and, thus, not representative of the data overall.

11.1 - Using Normal Distributions

Spread: The Interquartile Range

- The interquartile range (IQR) lets us ignore the extreme data values and concentrate on the middle of the data.

- To find the IQR, we first need to know what quartiles are …

11.1 - Using Normal Distributions

Spread: The Interquartile Range (cont)

- Quartiles divide the data into four equal sections.
  - The lower quartile is the median of the half of the data below the median. (Q1)
  - The upper quartile is the median of the half of the data above the median. (Q3)

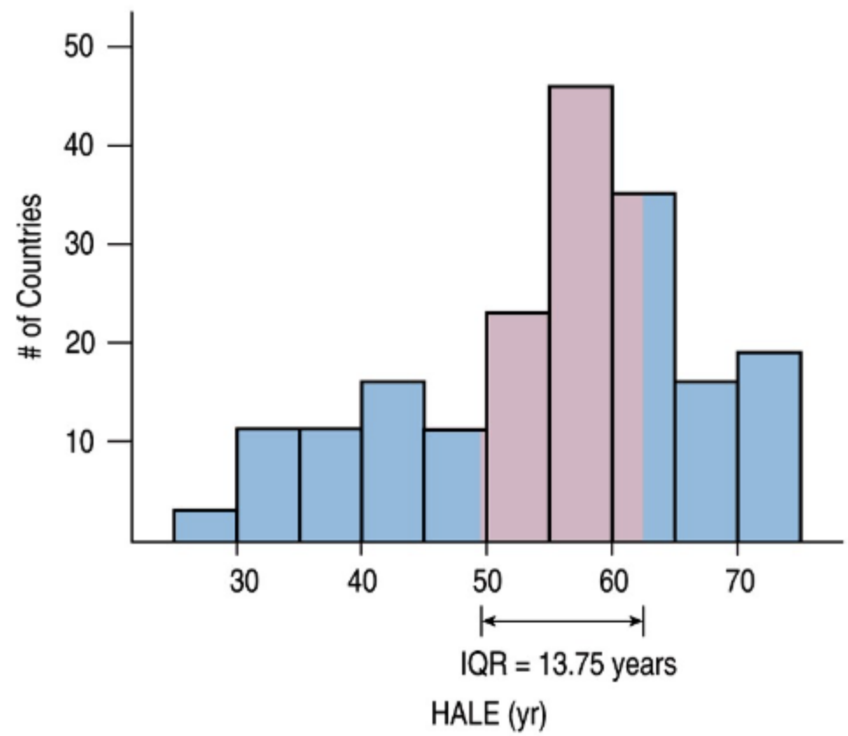
- The difference between quartiles is the IQR, so

IQR = upper quartile - lower quartile

11.1 - Using Normal Distributions

Spread: The Interquartile Range (cont)

- The lower and upper quartiles are the 25th and 75th percentiles of the data, so …
- The IQR contains the middle 50% of the values of the distribution, as shown in the figure:



11.1 - Using Normal Distributions

The Five Number Summary

- The five-number summary of the distribution reports its median, quartiles, and extremes (maximum and minimum).
- Example: The five-number summary for the ages at death for rock concert goers who died from being crushed is

Max	47 years
Q3	22
Median	19
Q1	17
Min	13

11.1 - Using Normal Distributions

Shape, Center, and Spread

- When telling about a quantitative variable, always report the *shape* of its distribution, along with a *center* and a *spread*.
  - If the shape is skewed, report the median and IQR.
  - If the shape is symmetric, report the mean and standard deviation and possibly the median and IQR as well.

11.1 - Using Normal Distributions

Practice

On Monday a class of students took a big test, and the highest score was 92. The next day a student who had been absent made up the test, scoring 100. Indicate whether adding that student's score to the rest of the data made each of these summary statistics increase, decrease, or stay about the same:

a. mean ____increase____

b. median ____same____

c. range ____increase____

d. IQR ____same____

e. standard deviation ____increase____

The Standard Deviation as a Ruler and the Normal Model

11.1 - Using Normal Distributions

Standard Deviation as a Ruler

- The trick in comparing very different-looking values is to use standard deviations as our rulers.
- The standard deviation tells us how the whole collection of values varies, so it's a natural ruler for comparing an individual to a group.
- As the most common measure of variation, the standard deviation plays a crucial role in how we look at data.

11.1 - Using Normal Distributions

Standardizing with z-scores

- We compare individual data values to their mean, relative to their standard deviation using the following formula:

$$z = \frac{x - \mu}{\sigma}$$

- We call the resulting values standardized values, denoted as z. They can also be called z-scores.

11.1 - Using Normal Distributions

Standardizing with z-scores (cont)

$$z = \frac{x - \mu}{\sigma}$$

- Standardized values have no units.
- z-scores measure the distance of each data value from the mean in standard deviations.
- A negative z-score tells us that the data value is below the mean, while a positive z-score tells us that the data value is above the mean.

11.1 - Using Normal Distributions

Benefits of Standardizing

- Standardized values have been converted from their original units to the standard statistical unit of *standard deviations from the mean*.

- Thus, we can compare values that are measured on different scales, with different units, or from different populations.

11.1 - Using Normal Distributions

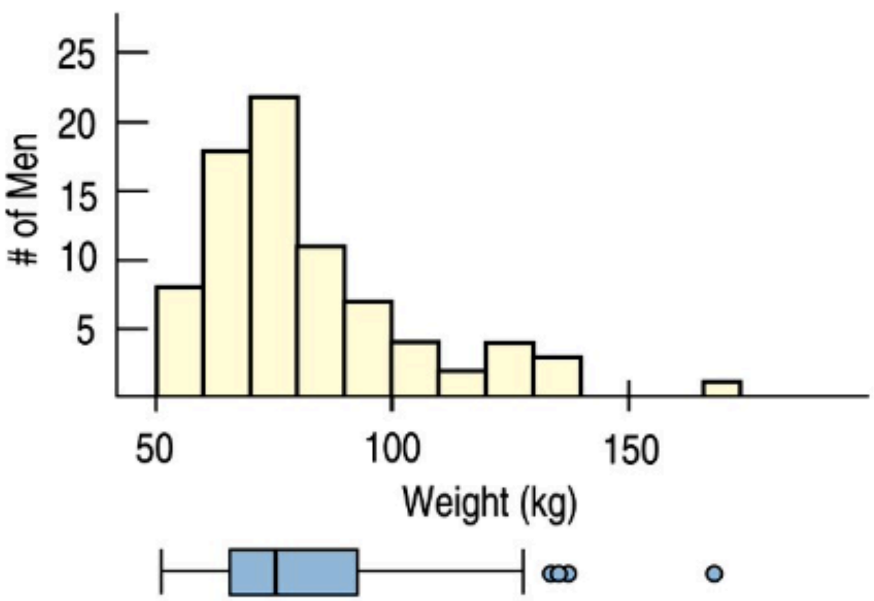
Shifting Data

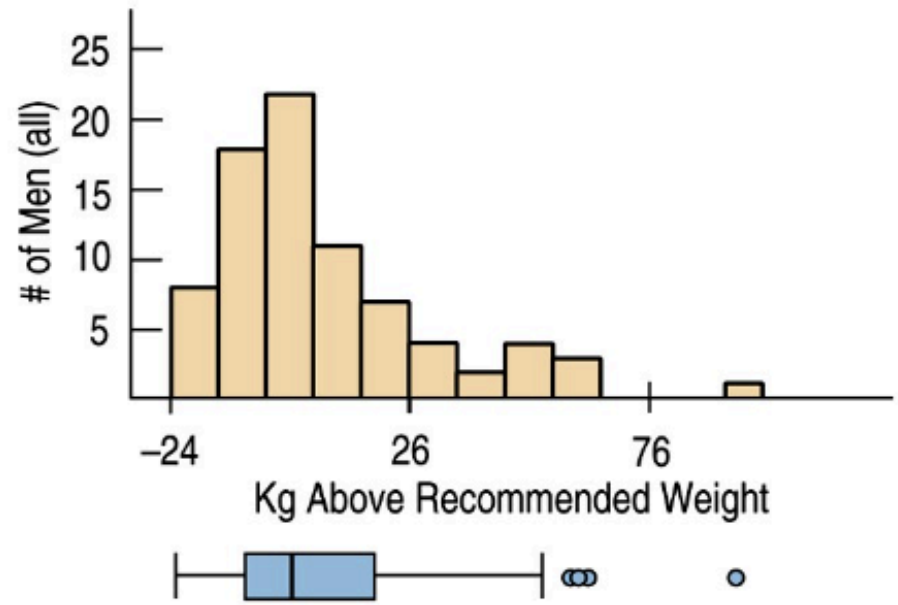
- Adding (or subtracting) a constant amount to each value just adds (or subtracts) the same constant to (from) the mean. This is true for the median and other measures of position too.
- In general, adding a constant to every data value adds the same constant to measures of center and percentiles, but leaves measures of spread unchanged.

11.1 - Using Normal Distributions

Shifting Data (cont)

- The following histograms show a shift from men's actual weights to kilograms above recommended weight:





11.1 - Using Normal Distributions

Back to z-scores

- Standardizing data into z-scores shifts the data by subtracting the mean and rescales the values by dividing by their standard deviation.
    - Standardizing into z-scores does not change the shape of the distribution.
    - Standardizing into z-scores changes the center by making the mean 0.
    - Standardizing into z-scores changes the spread by making the standard deviation 1.

11.1 - Using Normal Distributions

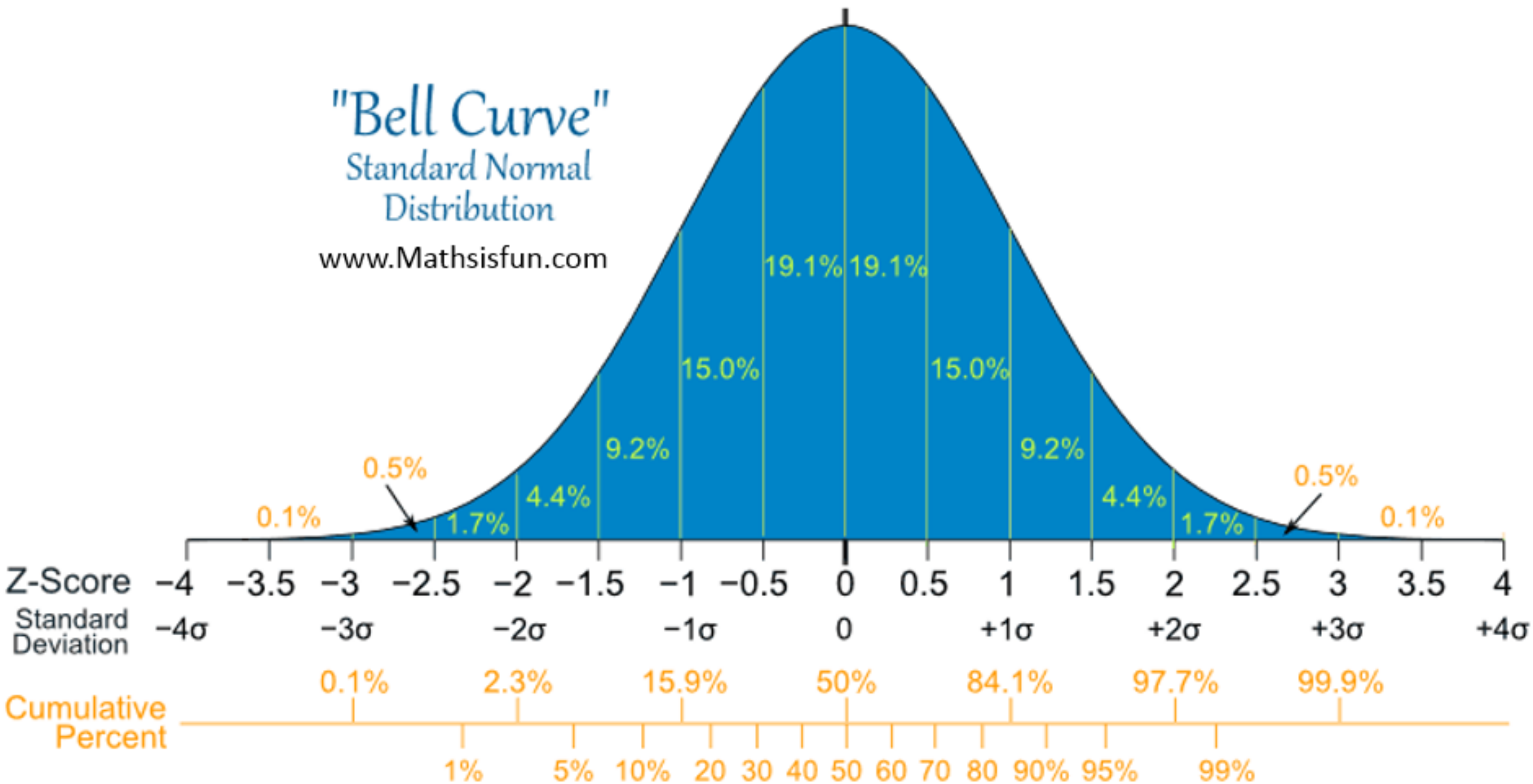
When is a z-score Big?

- A z-score gives us an indication of how unusual a value is because it tells us how far it is from the mean.

- Remember that a negative z-score tells us that the data value is *below* the mean, while a positive z-score tells us that the data value is *above* the mean.

- The larger a z-score is (negative or positive) the more unusual it is.

11.1 - Using Normal Distributions



11.1 - Using Normal Distributions

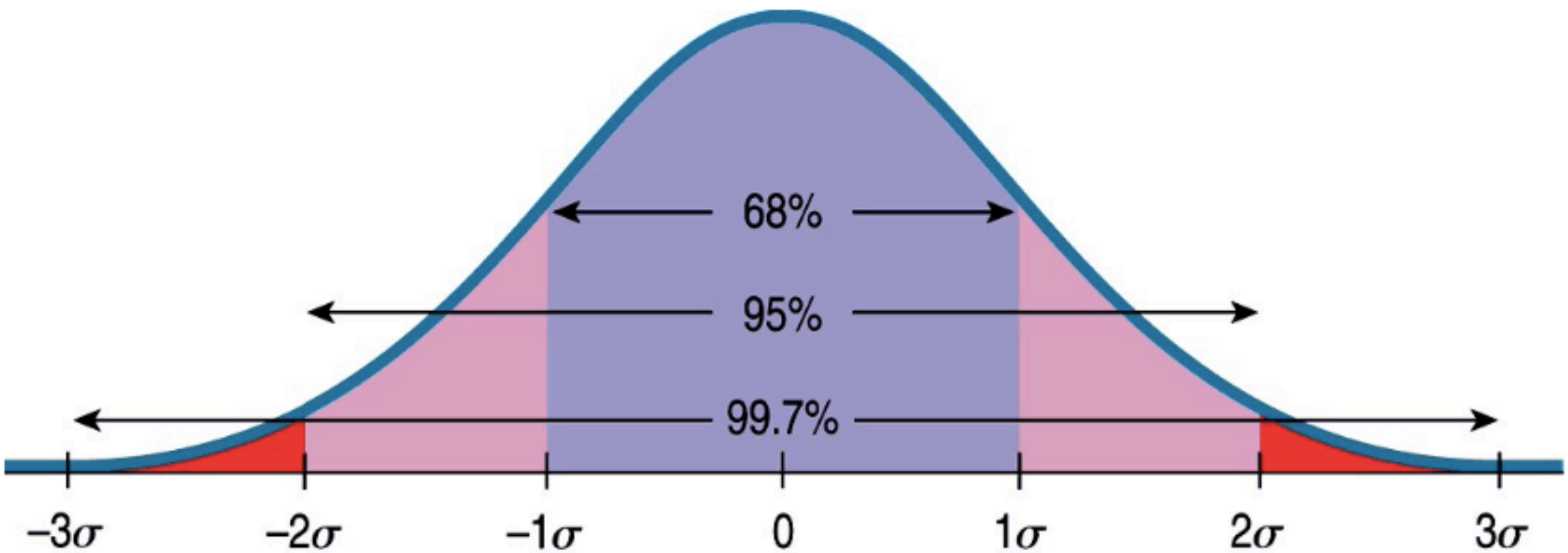
The 68-95-99.7 Rule

- Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean.
- We can find these numbers precisely, but until then we will use a simple rule that tells us a lot about the Normal model…

11.1 - Using Normal Distributions

The 68-95-99.7 Rule (cont)

- The following shows what the 68-95-99.7 Rule tells us:



11.1 - Using Normal Distributions

When is a z-score Big? (cont)

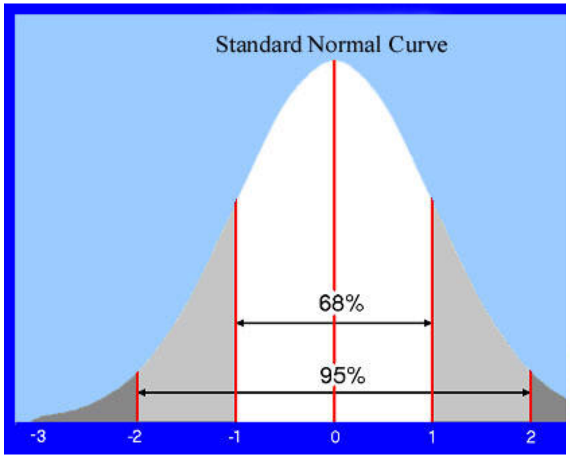
- There is a Normal model for every possible combination of mean and standard deviation.
  - We write $N(\mu, \sigma)$ to represent a Normal model with a mean of μ and a standard deviation of σ.
- Once we have standardized, we need only one model:
  - The N(0,1) model is called the standard Normal model (or the standard Normal distribution).

11.1 - Using Normal Distributions

When is a z-score Big? (cont)

- Summaries of data, like the sample mean and standard deviation, are written with Latin letters. Such summaries of data are called statistics.
- When we standardize Normal data, we still called the standardized value a z-score, and we write

$$z = \frac{x - \mu}{\sigma}$$



z-score Table

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

11.1 - Using Normal Distributions

Practice

Assuming normal find the following

1) Percent greater than 2.23 st dev

1.29%

2) Find z-score if area = 0.73 on right side of z

Z = -0.61

3) Find z-score if area = 0.61 on left side of z

Z = 0.28

11.1 - Using Normal Distributions

Practice

Adult female Dalmatians weigh an average of 50 pounds with a standard deviation of 3.3 pounds. Adult female Boxers weigh an average of 57.5 pounds with a standard deviation of 1.7 pounds. One statistics teacher owns an underweight Dalmatian and an underweight Boxer. The Dalmatian weighs 45 pounds, and the Boxer weighs 52 pounds. Which dog is more underweight? Explain.

use z-scores

Dalmatian: $z_D = \dfrac{45 - 50}{3.3} = -1.52$

Boxer: $z_B = \dfrac{52 - 57.5}{1.7} = -3.24$

The Dalmatian is 1.52 standard deviations underweight, while the Boxer is 3.24 standard deviations underweight. So, the Boxer is more underweight.

★