Unit 5: Boxplots

SUMMARY OF VIDEO

Hot dogs are an American icon – and we eat billions of them every year. It seems like there is a hot dog to fit just about every taste out there… all beef, some pork, turkey, skinless, even tofu for the vegetarian hot dog lover. Not all hot dogs are created equal though, at least in terms of calories. The calorie count varies quite a bit depending on the type of hot dog and also from brand to brand of a given type. The video gives us an inside view at Vallid Labs in Agawam, Massachusetts, and the calorie counting techniques they use to find the number of calories in particular hot dogs. After turning a hot dog into mush and then applying a series of treatments with acids and bases, distillations and titrations, the number of grams of fat, protein, and carbohydrates per hot dog is determined. Then using the information from the Nutrition Facts in Figure 5.1, we can determine a hot dog's calories.

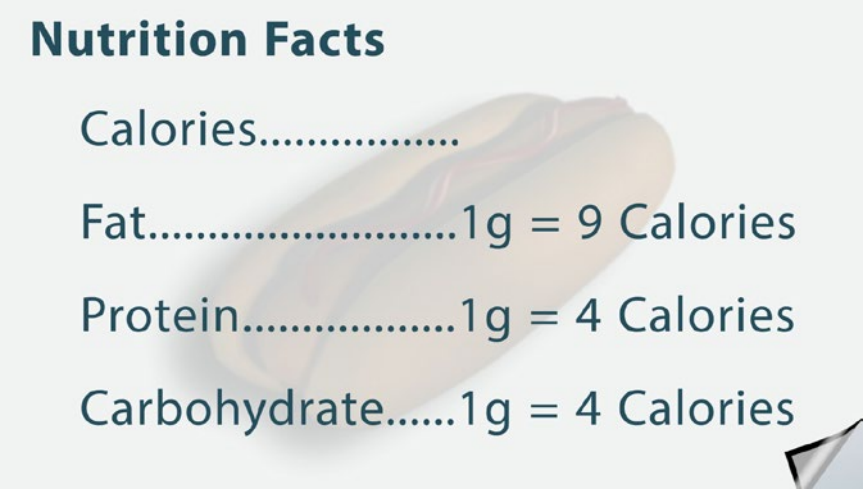

Figure 5.1. Determining the calories in a hot dog.

Despite similar appearances, hot dogs vary widely in nutritional content. For those calorie-conscious among us, statistics can help suggest the healthiest choice. We start with the calorie counts (listed in order from smallest to largest) of 20 different brands of all-beef hot dogs.

110	110	130	130	140	150	160	160	170	170
175	180	180	180	190	190	190	200	210	230

You can see that all-beef hot dogs range from 110 calories in the lowest brand to 230 calories in the highest. One way we can describe this distribution numerically is with the median – the number of calories in a typical beef hot dog. Since we have 20 brands, the location of the median is (10 + 1)/2, which is 10.5. So, we find the median by averaging the 10th and 11th calorie count in the ordered list:

$$\text{median} = \frac{170 + 175}{2} = 172.5 \text{ Calories}$$

So, your typical beef hot dog has 172.5 calories.

Next, we add some more numbers to our analysis of the spread of the calorie counts of beef hot dogs. We know the minimum is 110 calories and the maximum is 230 calories. However, we also need some information about the numbers in between. For that we can determine the quartiles. These are values one-quarter and three-quarters up the ordered list of calories. The first quartile – also known as Q_1 – has 25% of the ordered observations at or below it. It is the median of the lower half of the data:

110 110 130 130 140 150 160 160 170 170

The median of this 10 number set is between the fifth and sixth data value: $Q_1 = 145$. Now, we turn our attention to the third quartile – also known as Q_3 – which has 75% of the ordered observations at or below it. To find Q_3, take the median of the upper half of the data:

175 180 180 180 190 190 190 200 210 230

The median of these 10 upper numbers gives us $Q_3 = 190$.

So, with our minimum, first quartile, median, third quartile, and maximum, we have what is called the five-number summary, which we've summarized in Figure 5.2.

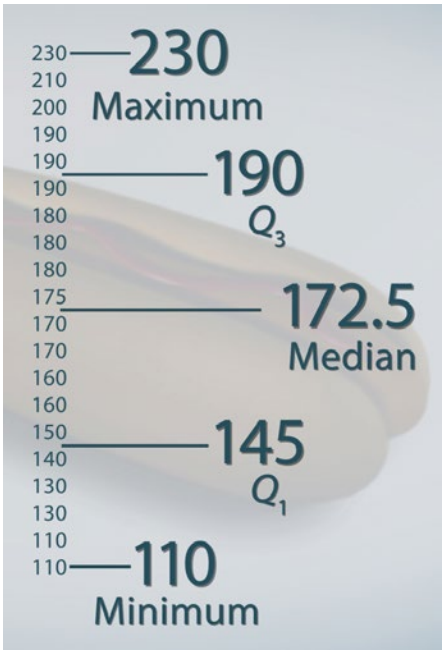

Figure 5.2. The five-number summary of all-beef hot dog calories.

The five-number summary gives us a nice snapshot of both the center and spread of the data. The median marks the center. The first and third quartiles contain between them the middle half of the data. We can measure the spread of the inner 50% of the data by the interquartile range (IQR):

$$\text{IQR} = Q_3 - Q_1$$

The two extremes show how far out the data extends. We can measure that spread using the range:

range = maximum – minimum

In statistics, the best description of data often combines the precision of numbers with the clarity of pictures. A boxplot (or box-and-whisker plot) is a graphic display of the five-number summary. Figure 5.3 shows a boxplot of the all-beef hot dog calories. The box spans the first and third quartiles, the median is marked inside the box, and whiskers extend out to the extremes.

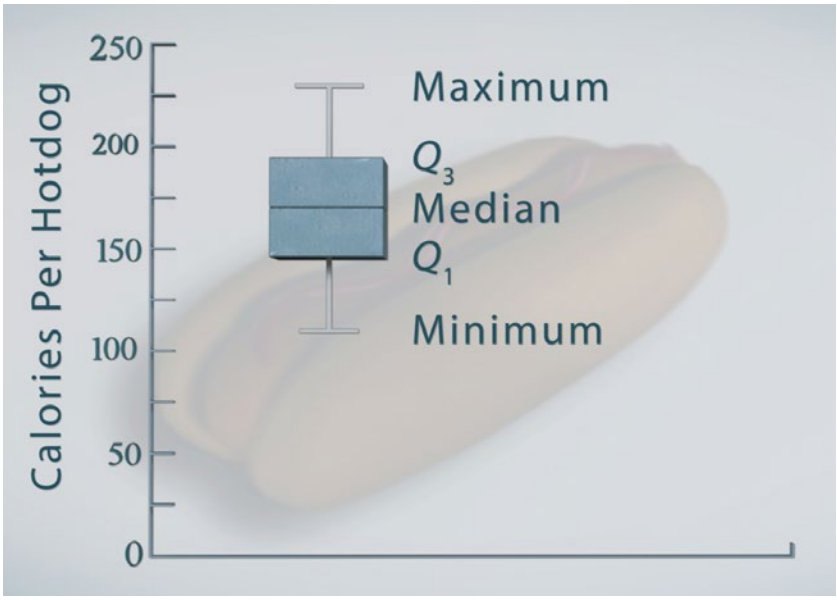

Figure 5.3. Boxplot of all-beef hot dog calories.

Boxplots don't show a distribution in detail the same way that a stemplot or histogram would, but boxplots can be a great way to make a quick side-to-side comparison of a few distributions. Take a look at Figure 5.4 comparing the calories of beef, poultry, and veggie hot dogs.

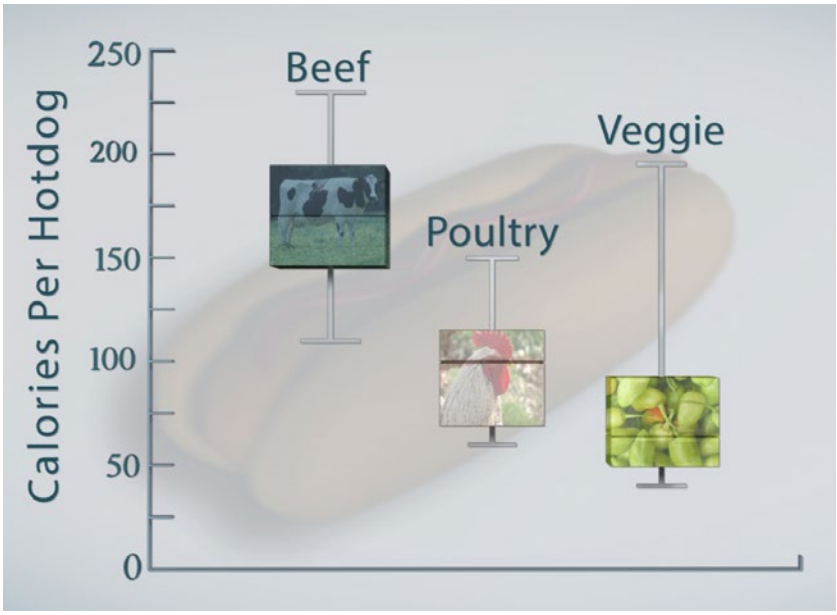

Figure 5.4. Comparing calories of beef, poultry, and veggie hot dogs.

Notice that the median of the poultry hot dogs is below the minimum for the beef hot dogs. So, half of the brands of poultry hot dogs have fewer calories than the lowest calorie brand of all-beef hot dogs. Now, check out the boxplot for the vegetarian hot dog. Notice that at least one

veggie brand has more calories than three quarters of the beef hot dogs!

So, now we can add boxplots to the list of ways we can graphically represent data – and, as shown by the hot dogs, this method allows for easy comparisons between groups.