The standard deviation and the 95% summary interval

StatPREP Class Activity

Orientation

The length of the *95% summary interval* is a pretty good way to describe the amount of variation in a variable. Yet statisticians tend to prefer using the *standard deviation*. This makes sense, since statisticians are interested in confusing the poor student by using obscure names ("Deviation?". Really?) and formulas that can be hard to understand. . . . Not really!! Even though the standard deviation has an odd name and seems complicated, the actual reason it's used so much by statisticians is that it has some nice mathematical properties.

The purpose of this lesson is to help you understand the standard deviation in terms of the 95% summary interval. Often, there's a very close and relatively simple relationship between the two. This means that it's possible to make sense of the standard deviation without formulas. (There's still the odd name, but there's nothing we can do about that!)

Activity

Open up the Center and spread Little App (see footnote[1])., and select the `Births_2014` data frame. Set the sample size to n = 200.

1. Set the response variable to be `baby_wt`, the weight of the baby at birth. You'll see that there is variability, that is, not every baby weighs exactly the same.

2. Turn on the display of the mean and the standard deviation. Look closely at the annotation for the standard deviation. The standard deviation is a kind of measuring unit, like the inch marks on a ruler. The annotation is being drawn as a ruler. The mean is exactly in the middle and there are tick marks at ± 1 standard deviation and at ± 2 standard deviations.[2]

[2] Sometimes the ± 2 marks don't fit within the frame, so those aren't included on the ruler.

3. The numerical value of the standard deviation is the length between tick marks, just as an inch is the length between marks on a ruler.

   - Use the measuring stick to find the numerical value for the standard deviation for the data shown in your plot.
   - Open the "Statistics" tab underneath the graphic where you will find printed the standard deviation as calculated directly from the data. Compare that number to the number you found in (a). They should be just about the same.

- Use the measuring stick to find the length of the whole ruler being displayed. Compare that to the number you read in (b) from the Statistics tab.

Describing how long the ruler is by how many standard deviations it is long.
. . .

4. Turn on the display of the 95% summary interval. See where the ends of the 95% summary interval fall along the standard deviation ruler.

   How long is the standard deviation compared to the length of the 95% summary interval?   . . .

   - Make the sample size bigger, say, n = 2000.

   Does the relationship between the 95% summary interval and the standard deviation change?   . . .

   Fill in the blank in the following statement. (Feel free to round to a whole number.)   . . .

   The 95% summary interval is ____ times as long as the standard deviation.

5. The simple relationship between the 95% summary interval and the standard deviation often holds, but not always. Some variables consistently have the standard deviation ruler mis-aligned with the 95% summary interval.

   - Set the response variable to be APGAR score. Try several New Samples.

   Does the 95% summary interval align with the ruler?   . . .

   - Turn on the violin density display.

   Is the density shape symmetric (top to bottom)? Does it have a very long tail?
   . . .

   - Go back to `baby_wt` and look at the violin.

   How does the shape of this violin differ from that for APGAR. Is it symmetric? Does it have a very long tail?   . . .

   For response variables with a bell-shaped distribution, roughly how long is the 95% summary interval in terms of the standard deviation? Circle one of the following.   . . .

The 95% summary interval is ☐ the same length as ☐ twice as long as ☐ three times longer than ☐ the standard deviation.

*For response variables that don't have a bell-shaped distribution, e.g. they have a long tail on one side but not the other, does the answer you gave to the previous question still hold?   . . .

Version 0.1, 2019-05-29, Daniel Kaplan,