Parameters of the normal distribution

StatPREP Class Lesson

Orientation

The bell-shaped distribution is often a good description of the distribution of a variable, so often that it's called the "normal" distribution.

Actually, there is not just one normal distribution but a whole family of them, all of which have the characteristic bell shape. To identify an individual member of this bell family, two numbers suffice:

- the mean: the location of the peak of the bell
- the standard deviation: the width of the bell

What's potentially confusing here is that you may know about the mean in terms of a calculation you do on a variable: add up the n values and divide by n. Similarly, the standard deviation is such a calculation. Amazingly enough, the numbers you get from these calculations identify the bell-shaped distribution that's the best match to the distribution of the variable. And if the variable has a bell-like distibution itself, you'll see that the distribution of the variable is often close to an exact match to the normal curve itself.

One consequence of this is that people often can make a pretty good estimate of the mean and standard deviation of a variable just by looking at its distribution, say with a histogram or a violin plot.

This lesson is about learning to eyeball the mean and standard deviation from a display of the distribution. Reasons to do this include having a better understanding of the mean and standard deviation and, importantly, being able to double-check calculations in order to avoid blunders. Even with a computer, it's easy to make mistakes like reading the wrong number from a table of calculations.

Activity

Open up the rare-and-common Little App. (See footnote[1]). Set the response variable to `height_adults` and leave the explanatory variable at `.none..`

[1] https://dtkaplan.shinyapps.io/LA_rare_and_common/

1. The graphic shows a traditional blot of the distribution of the response variable, called a *density plot*. If you're familiar with a histogram, you might like to think about a *density plot* as a kind of smoothed histogram without the jagged, abrupt changes from bar to var.

   - There are sliders in the app to let you define numerically what range of values are common, uncommon, or rare. We won't need these for this lesson, so arrange for the "common" slider to run the full span of the data.

2. Check the control box to overlay a theoretical normal distribution. The distribution will be shown as a white, transluscent, bell-shaped region. The location and width of a normal distribution are described by the mean and standard deviation. In picking the particular normal distribution to overlay, the mean and standard deviation have been set to those of the response variable.

   • Compare the theoretical distribution to the actual distribution. You'll see that the actual values of height occur less frequently near the center than they would be for a theoretical normal distribution. The tails, both right and left, line up pretty well with the theoretical normal distribution.
   • What ranges of height occur more frequently in the actual heights than in the theoretical distribution?

3. Set the explanatory variable to `sex`.

   • Compare the theoretical normal distribution for women to the actual distribution. Are they matched better than in step (2) when there was no explanatory variable? How about for men?

4. Several variables are listed below along with a description of the shape of the distribution.

   • diastolic blood pressure is slightly left skew
   • system blood pressure is slightly right skew
   • age is flat and truncated to the left
   • testosterone is bi-modal.
   • bmi_adults is right skew, truncated to the left.
   • Look at each of these variables and figure out what differences between the theoretical normal distribution and the actual data correspond to the various labels.
   • Use `sex` as an explanatory variable for each of the variables. Are there any variables where the distribution for the individual sexes has a different description than the distribution without an explanatory variable.

5. Look through the various data sets to find variables that are a good match to the normal distribution.

   • Comment on whether most variables have a "normal" shape.

   ───────────────────────────

Version 0.1, 2019-05-28, Thomas Kinzeler and Danny Kaplan, Word version