

CHECKING FOR NORMALITY

QUESTION:

I am knee deep into inferences with means and told my kids that they had to provide a plot (not just the verbal description). Here is the question, is there any reason that I should be persuading them to do a histogram over a NPP? I have some students that like to just do the NPP.

ANSWERS:

1) Either is fine.

2) Dave Bock outdid himself in the brevity of his response to this;-)Yes, probably either would be accepted, but I do not find NPPs especially easy to read. Can your students look at a NPP and say:

Are the data skewed? High or low? Outliers? High or low? Bimodal?

NPPs ARE more effective at spotting problems in the tails of the distribution, distinguishing between the mound shaped Cauchy and a normal, for example, but AP Stats. gives little attention to such problems.

Personally, I think you are not going to find out if the POPULATION is normally distributed unless you have a very large sample. In that case, a NPP might be in order. When you look at sample data, it is more important to detect problems WITH THIS SAMPLE. In a situation I looked at yesterday, the data were repeated measurements of something about seven feet long, and the data had a range of about four inches. First graders should be able to do better than that, so the data were discarded in favor of more work on measurement technique.

3) I, actually, am a fan of NPPs. With a NPP, unlike a histogram, you get to see ALL the data. Interpretation of these is something of an art -- i.e. I agree with Bob -- but in AP Statistics the only interpretation needed is about the credibility of a normal population. And in this, all one need do is say, "Yup, sufficiently straight."

4) Presumably the choice comes up when students are supplied raw data and have to choose what display to make, and presumably they would be making the display with a graphing calculator and typing in the data. (Personally, I would make a stem-and-leaf by hand and so would "see ALL the data," but then I'm weird;-) In this context it is extremely unlikely that they will be able to determine if the POPULATION is normally distributed because no AP exam question is going to ask them to type in a large enough sample to determine this. So I would disagree with Chris that checking the normality of the population is the only goal -- it may not even be a realistic goal.

AP Stats is more mathy than most college stats courses and puts a lot of emphasis on the mathematical assumptions, such as normality. In practice, though, you are not looking for normality, you are looking for trouble;-) For example, you may find outliers that are errors of data gathering and need to be corrected. (If the population actually includes such outliers, you need to leave them alone.) So "checking the normality assumption" is sort of AP code for "look at the data". For most problems you are likely to find there, more students are more likely to identify the problem from a stem and leaf, histogram or dotplot than from an NPP.

To give just one example, a NPP is better at detecting long or heavy tails than most alternate displays -- assuming you have a large enough sample to evaluate this. The NPP then looks S-shaped. But it curves one way for long tails (trouble) and the opposite way for short tails (minor/no problem) -- and which way is which depends on what you plot vs. what. These very different problems can lead to equal degrees of straightness in the NPP. Chris would know all this and make the right decision, but I think the typical AP Stats student could easily be led astray. I don't think such students would have trouble distinguishing between short tailed and long tailed distributions in a histogram.

5) Bob has the TEMERITY disagree with me...

“So I would disagree with Chris that checking the normality of the population is the only goal -- it may not even be a realistic goal.”

However, little does he realize that I actually agree with him! HA!

I must not have been sufficiently clear in my reply. Bob had stated that interpreting the NPP was pretty involved, with which I generally agree -- if the goal is to get all the information he wrote about in his posting. However, the particular use of the NPP in AP Statistics is checking for normality, which is (at least I THINK is) a pretty simple proposition -- is the NPP straight?

I agree with Bob that the typical garden variety sample sizes of data we give to cherubs is nowhere near what is necessary for an adequate check for an NPP -- but neither would a stem & leaf or box plot or (horrors!) a histogram. What is being honored by the check for the credibility of a presumption of normality (i.e. "checking for normality") is that assumptions are important in statistics and the checking should be done, not that with textbook sized data sets the check really works.