Thursday, February 1: Confidence Interval Example

Math 30853, George Gilbert

Heights of adults in many populations of one sex are approximately normally distributed with standard deviation 2.5 inches. The male students in a former class had heights

76	75	73	70	72	71	73	72	71	73
73	69	72	68	63	71	70	70	72	

If we take this as a random sample of some population (maybe male students in 30853?), find a 90% confidence interval for the mean height.

Two-sided is the default unless specifically told.

What is the first thing we should do?

Hint: What is the first thing you should do before applying any mathematical theorem?

Answer. Check that the assumptions are met. In statistics, that means plausible.

We enter the data into say, list 1, in our calculator (STAT: Edit) or spreadsheet and find $\bar{x} \approx 71.26315789$ and $s \approx 2.79$. These values are stored under VARS:Statistics:XY, so we can always retrieve them and avoid rounding in intermediate calculations. This value is pretty consistent with the assumption that $\sigma = 2.5$. We will be able to quantify this when we look at the distribution of s under the assumption of a normal population distribution, which will allow us to compute confidence intervals.

What about normality?

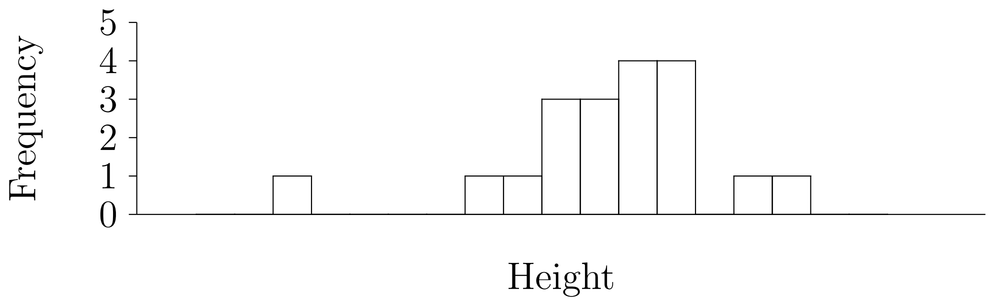
If the sample size were greater than 30, we could invoke the Central Limit Theorem.

Check #1. Draw a histogram. On the TI-83/84.

1. Go to STAT PLOT (2nd Y=). Move down to one of the three plots, turn it on, choose Type Histogram (obvious, last entry of 1st row), tell the Xlist where your data is, L_1 in my case, and Freq where the frequency is, another list or 1. Make sure all other plots, both Y= and STAT PLOT are off. If you get get a bizarre graph, they probably aren't.

2. Go to WINDOW and set up the ranges. Looking at the data, I chose XMin=61.5 and XMax=78.5 with Xscl=class width=1. Then Ymin=0 and YMax=5$\geq$ the mode.

3. GRAPH



Keep in mind the differences from the theoretical that we saw in even in large samples in the Central Limit Theorem demo when we were simulating various distributions.

Check #2. Do a rough check of the shape. One standard deviation from the mean would be those values between about $71.26315789 - 2.5$ and $71.26315789 + 2.5$. For our data this is $16/19 \approx 84\%$ versus the theoretical 68%. All but one value is within two standard deviations, so we compare the $18/19 \approx 95\%$ to the theoretical 95%.

Check #3. Normal Probability Plot (Section 4.6 in the text).

There is nothing magical about 1 or 2 standard deviations, so we avoid making such arbitrary choices. The (approximate) z-score $z = \dfrac{x - \bar{x}}{s}$ is a linear function of x. If the distribution is normal, we should also be able to get the z-score by computing the percentile of x and fining the corresponding z-score using invNorm. If we plot x versus this latter value, it should be roughly linear.

Recall that the median of 2, 3, 5, 7, 11 is 5, which is the 50th percentile. Thus, the 3rd number on the list is the 2.5/5 percentile. In general, the ith sorted(!!) value is assigned percentile, expressed as a probability, $(i - 1)/n$.

"By hand," i.e. with some help from your calculator or spreadsheet,

1. Sort your data. If you want to keep the orginal data in order, save your ist into a different on. 2nd STAT: SortA(L_1) will sort L_1 in ascending order and overwrite L_1.

2. Assign "percentile" $(i-1)/n$ to each value. Easy on a spreadheet, but I don't know an easy way on a TI. You can avoid a little effort by entering $i - .5$ intot he TI and later dividing the whole list by n. If you were really using you TI a lot to do this by hand, you could write a short TI program to do it. I put it in L_3.

3. Use invNorm to compute a z-score for each percentile.

4. Under STAT PLOT, choose Type Scatterplot (non-obvious, first entry of 1st row), the Xlist is your x values and the YList your z-scores.

5. Set up your WINDOW and GRAPH .

This is much easier in Excel. (See accompanying spreadsheet.)

In fact, this is built into the TI. Nonobviously, the 2nd row, 3rd entry of the STAT PLOT is a normal probability plot. Simply check it, adjust your WINDOW, and GRAPH. The one unusally short male makes distorts the graph a bit.

The class also had 6 female students with heights

$$65 \quad 61 \quad 69 \quad 64 \quad 63 \quad 67$$

So let's suppose we are sufficiently satisfied the assumptions are met. We compute $z = \text{inv}(.95) = 1.64485\ldots$, store to avoid rounding. Then the interal is $\bar{x} - z \cdot 2.5/\sqrt{19} \approx 70.3$ inches to $\bar{x} + z \cdot 2.5/\sqrt{19} \approx 72.23$ inches (one more decimal place than the data). Alternatively, you could compute $\text{invNorm}(.05, \bar{x}, 2.5)$ and $\text{invNorm}(.95, \bar{x}, 2.5)$.