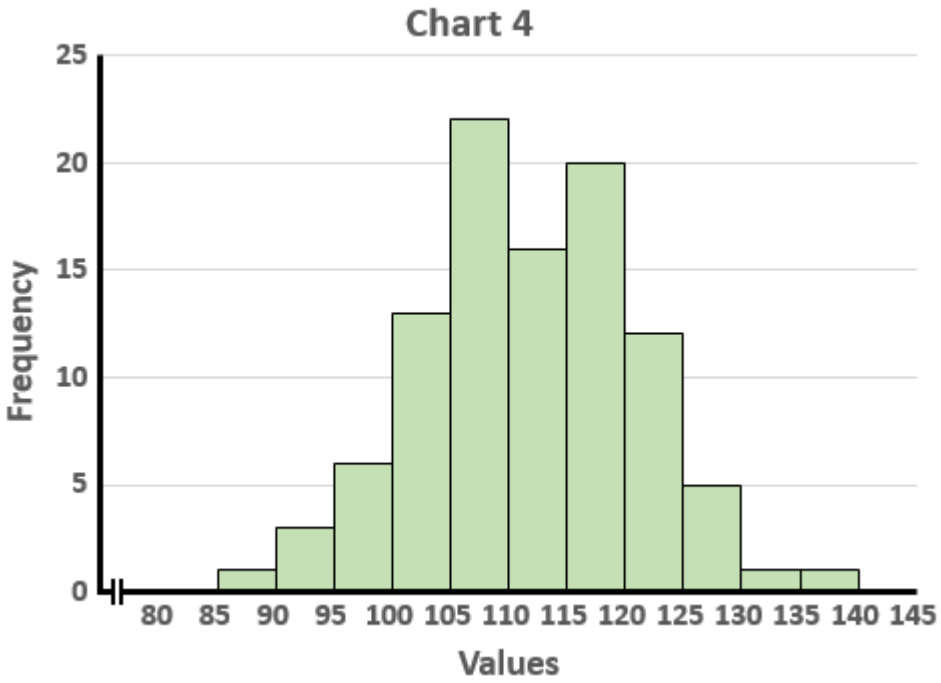
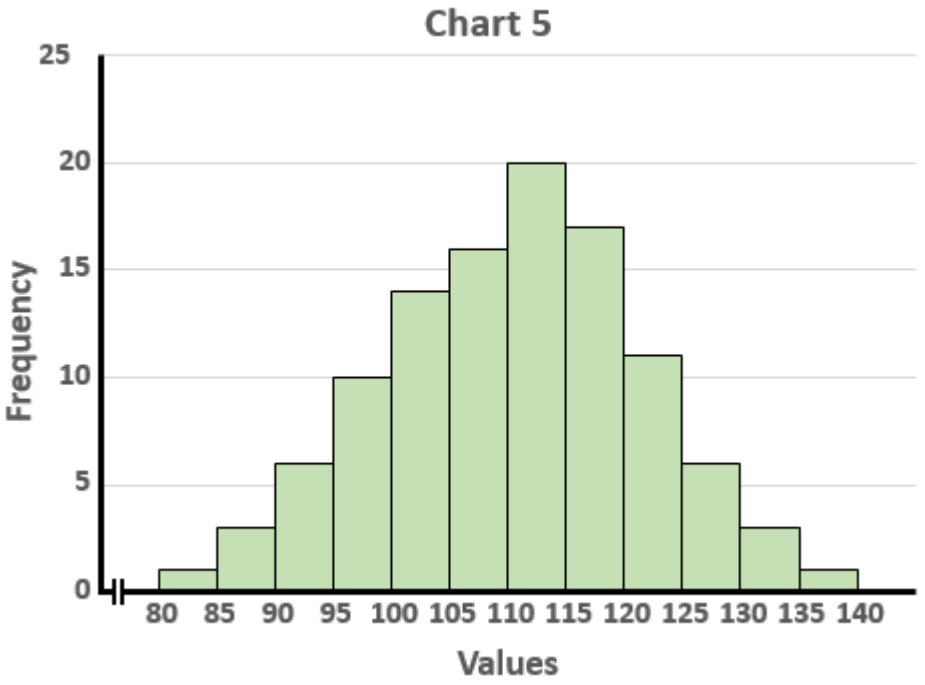
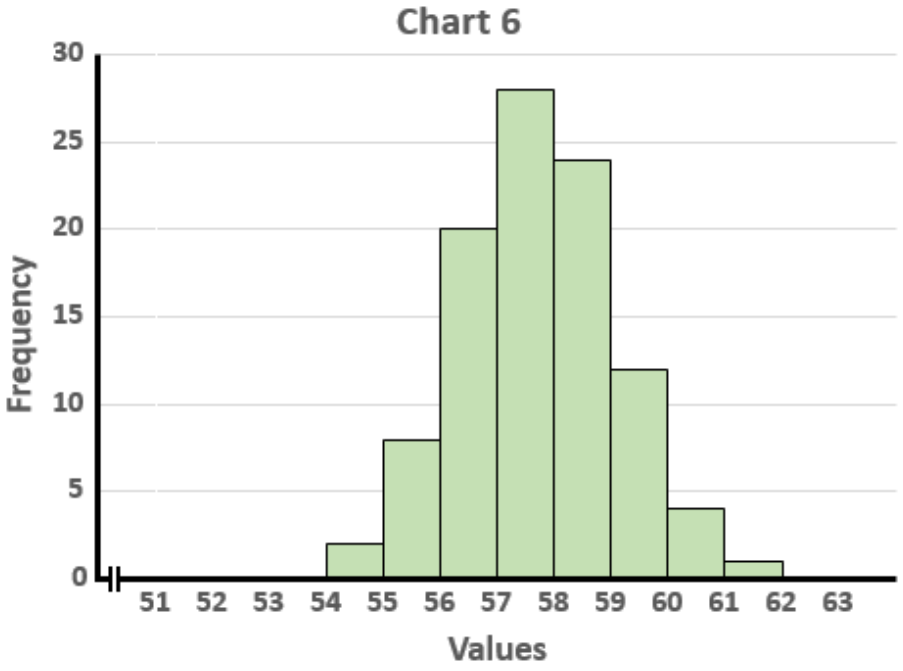
L2 Histograms Practice #2

For the three histograms here:

- discuss whether the distribution is Normal or not;
- estimate mean and median, and discuss any difference;
- estimate standard deviation.







Answers – L2 Histograms Practice #2

Marks will depend as much on discussion of evidence as values given. Key terminology should be used rather than general terms.

The key features for a distribution being Normal should be discussed in each answer:

- "bell curve" shape fits the middle of the columns;
- symmetrical, allowing for some random variation in practice;
- drops off to effectively zero at 3 standard deviations.

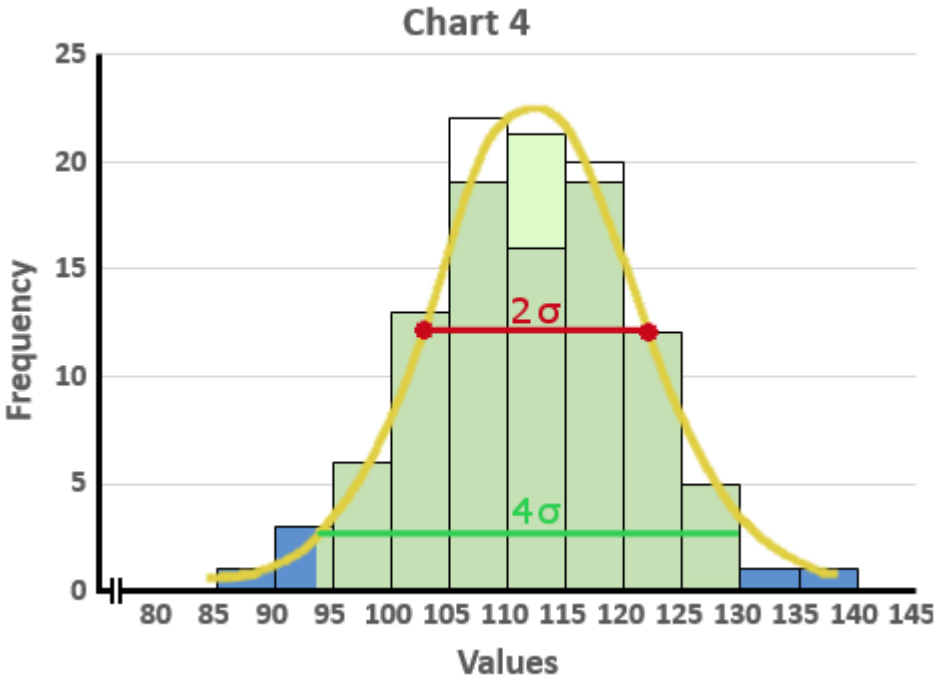
The median can be found by adding up the total frequency and counting in from either end.

The mean will be the median, moved towards any skew away from symmetrical.

The standard deviation can be estimated by two methods:

- shown in red, the distance between the inflection points (where the curve stops getting steeper) is 2 σ;
- shown in green, the middle 95% of values covers 4 σ, so the outermost 5% of values can be counted off and the distance found ÷ 4.

Note, while all distributions have a standard deviation that we cannot make probability calculations for non-Normal ones. If a distribution is wildly non-Normal then σ cannot be calculated as if it is Normal.



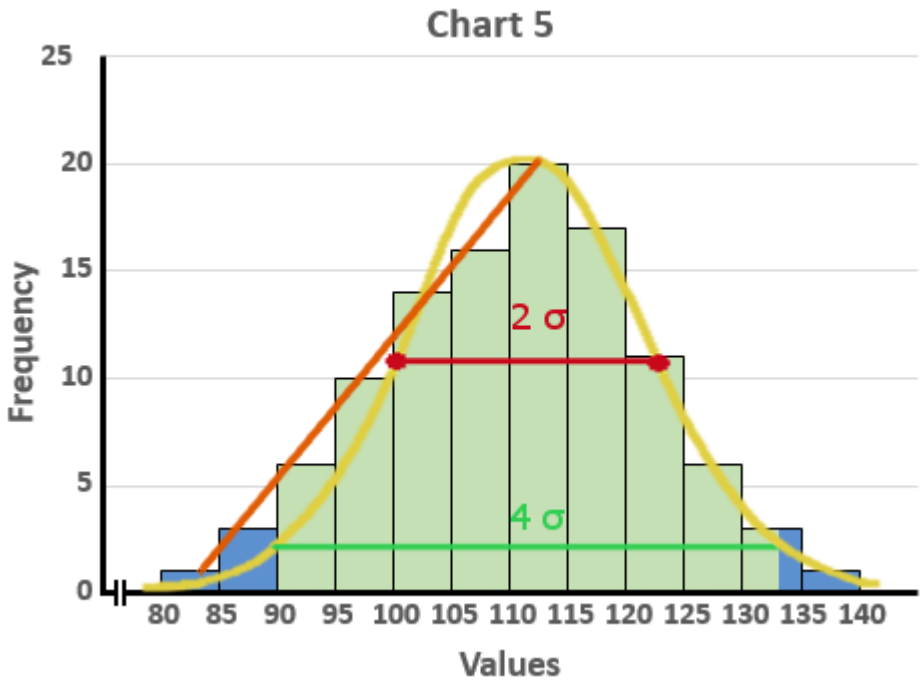
This may well be Normal. We only have to move five or six values – shown from empty to paler green – to make it fit a bell curve perfectly. With a sample size of only 100 there is likely to quite a lot of deviation from exact. ,

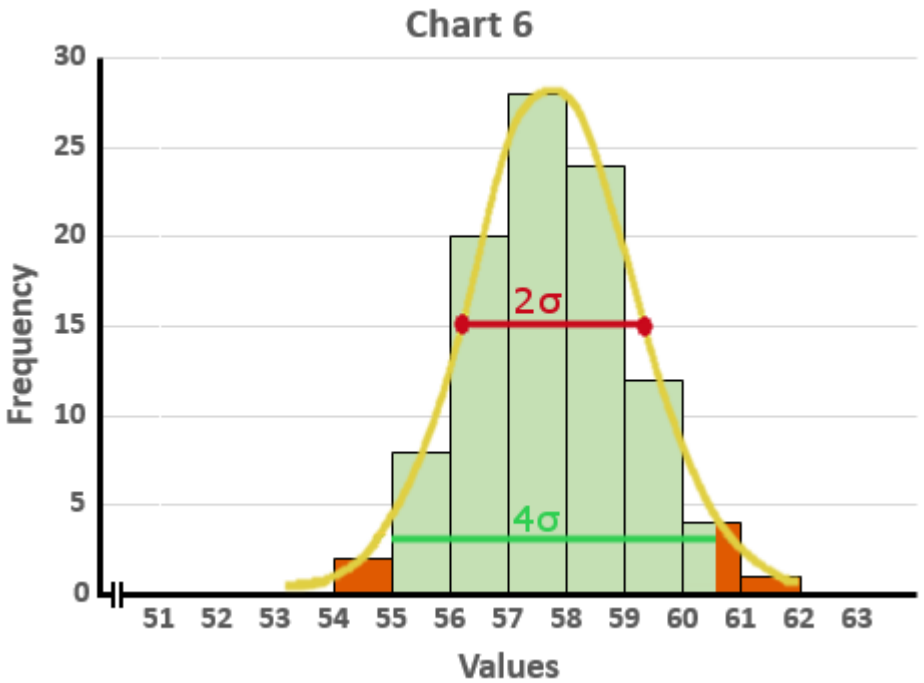
Other than the lack of central peak the rest of the data is symmetrical, tails off quickly at the ends and looks very Normal.

However, the two spikes on either side may represent an actually bimodal distribution. We'd need more data to tell the difference – if a second set of 100 points showed the same pattern, we would assume it was not Normal.

Based on counting in 50 values, and also where the bell curve peaks, we can estimate the mean and media is about 113, definitely in the range 111–114.

Standard deviation is between 9 and 10. This can be seen from the red 2σ distance between the two points of inflection which is just under 20 apart, indicating σ = 10 or a bit less. Alternatively the green 4σ distance containing the middle 95% of values is 38, indicating σ = 9.5.





This appears to not be Normal.

- The graph is asymmetric, with a lot more values below the peak than above.
- While a bell curve approximates the right hand side, the left hand side fits a line much better than a curve.

However that there are only 108 data values, so it may just be a particularly unlikely sample of a Normal distribution. We can't tell for absolute sure.

Counting 54 data points in, the median is around 111-112.

The mean will be lower than the median, as the tail values to the left are more distant (although not a lot further away as the skew is not great) so perhaps 109—110.

Standard deviation is in the range 11–12, although assuming it is non-Normal the s.d. is not useful for calculations of probability.

This is perfect Normal – too perfect to be real in fact for such a small number of data points. That the curve is more compressed horizontally than the others shown is a choice of the x axis scale only.

A bell curve sketch goes exactly through the centres of each column, is completely symmetric and tails to zero at 4σ. (Note, although the curve is symmetric, that does not require the actual columns to be. They will only be symmetric if the value of the median is exactly in the middle of its column.)

There are 99 values, so the median is the 50th, most of the way into the 57–58 column, giving a value of about 57.8, which is also where our sketch peak is. The mean will be the same as the median because it is symmetric.

Standard deviation is about 1.5. This can be seen from the red 2σ distance between the two points of inflection which are 3 apart, indicating σ = 1.5. Alternatively the green 4σ distance containing the middle 95% of values is 6.5, indicating σ = 1.6.