

Transcript: Live Describing Distributions

These histograms are all symmetric and by symmetric we mean basically if I split it in the middle it looks the same above and below the middle - approximately.

Certainly I don't expect the bars to fall exactly on the other bars if I folded over the middle but close.

I think about symmetry I have to have an idea in my mind about where center is so I can mentally fold it. Here the idea is, this is my center, and if I folded across the center those bars are going to land pretty much on top of each other.

And then here we have are two skewed distributions, skewed right - the way that I remember it is whichever way the tail is pointing - skewed right to the right and skewed left on the bottom pointing to the left.

The other way you might think about it is - which end is messing me up? well if I could ignore most of the right side it would be almost symmetric right? If I could ignore most the left side it would be almost symmetric. Sometimes that doesn't even apply though - it can be so skewed that no matter what you do it's always going to be skewed.

But that's the idea, skewed to the left, skewed to the right, the right is what's messing me up.

Now here this is where we get difficult right? Where is the center of those two pictures, how do you measure center. Well we could all come up with different answers and we might all be right!

I can go max to min - halfway, that's a center a bad one for this distribution.

(Bimodal Distributions)

If I graphed your heights, heights of students, I would get that kind of picture because there's males and females and the heights of those bars would be related to how many males and how many females I surveyed so if I surveyed more males I'm going to get a higher peak for the males and if I surveyed more females I'm going to get a higher peak for the females.

The peaks generally tell me that I have multiple groups that I may or may not know about - there's something going on there - can you figure out what it is? Heights - really easy - but something random that you just measured, you might not even have measured the other variable that exists that explains that - it could happen - you don't know what that variable is.

This is one to show you that you can have a skewed distribution with two modes this is the little mode you know it's not as pronounced as the initial one. Sometimes these are even so not pronounced that I ignore them - I really wouldn't call it bimodal - it's just a quirk potentially in

the data - but if you see a clear pattern there could be something going on there you could have.

This - almost looks like another little hump and another little hump and by the way after bimodal is multimodal - no trimodal or quadrimodal or anything like that.

And there's an example of one that has an outlier - so here we have a skewed right distribution but it doesn't really have any outliers there's nobody outside of my current pattern right? my current pattern says skewed right - so these guys are extreme but they're within the pattern that the data is presenting.

Now if this continued nothing nothing over here that's an outlier something that's so far out its outside the current skewed right pattern so that's kinda like that right one - although this is almost symmetric with an outlier right? almost

(student question - how do outliers enter into the shape)

Its part of the spread but we don't usually take it into account when we think about shape, you know, one value, I'm not going to call it skewed right because there's an outlier, I'm going to call it skewed right because this is a little longer than that - if I believe that visually. Again I think it's pretty close to symmetric.

But an outlier is going to be outside the pattern of the data and it might even be an error, you know, again, if we see this, it might be a data error and we need to delete it fix it or something like that.