# Select Reading Question Responses (3/3)

> I really struggled with question 2.11. I did not know the equation/ what numbers to use to get the Q1, Q3, or IQR.

Q1 is the point which is above 25% of the data. The idea behind calculating Q1 in 2.11 is the start from the left, and start adding up the heights of the bars. When you hit roughly 0.25, you've found the first quartile. The heights of the first three bars starting from the left of the histogram are roughly: 0.025, 0.05, 0.2. The sum of those is 0.275, so we're past the Q1. This means that Q1 should be somewhere in that first tall bar, ie, somewhere between 15 and 20. Let's say Q1 is approximately 17.5.

To find the median, we keep going until we hit 0.5. The height of the next (short) bar is 0.05, and then we have 0.2. The sum of all the heights up till the second tall bar is 0.525, so we've overshot the median again and the median must be somewhere in that second tall bar. In other words, it's between 25 and 30. Let's say it's at 27.5.

To find Q3, we do the same thing, adding up till we hit 0.75. This happens somewhere between 40 and 45, so maybe we say Q3 is about 42.5.

Finally, IQR is Q3—Q2 $\approx 42.5 - 17.5 = 25$.

> In the reading there was discussion of modes as being the "peaks" of a distribution. Is there a specific term for the lower spikes or points in a distribution?

You could use the word "valley." But usually people just talk about "peaks" of distributions.

> In question 2.23, I am confused why it would be dependent, I would have answered it independent because both political ideologies seem to support and not support DREAM, wouldn't that mean it is independent of each other and the political ideologies do not affect it? or am I confusing the concept of independent and dependent?

> How do you tell if a relationship between variables is independent or dependent when looking at a mosaic plot?

"Independent" in this context would mean that all ideologies support DREAM *in equal percentages*. In other words, "independence" would mean that the percentage of conservatives who support DREAM is equal to the percentage of moderates who support DREAM is equal to the percentage of liberals who support DREAM. Similarly with percentages of "not support" and percentages of "not sure." This is not the case (as we can tell from the mosaic plot, since the three columns of the mosaic plot don't have boxes of the same height) so the variables are *dependent*.

We can generalize this as follows. If the "gaps between boxes" of the two-variable mosaic plot all line up, your two categorical variables are independent. In other words, if you can start at one end of your mosaic plot, and travel all the way through to the other end by going through the "white space" between boxes without ever making a turn, the variables are independent. On the other hand, if the gaps don't line up (ie, if you have to make a turn somewhere), the variables are dependent.

How do you determine when a difference is large enough to reject the independence model/ accept the alternative model?

This is really very subjective! There's a "standard answer" (reject $H_0$ when the "p-value" is less than 0.05, as we'll talk about), but people sometimes follow this "standard answer" without thinking about it and end up misunderstanding their results. There are two articles I'll ask you to read towards the end of this class about this issue. You might decide to read them sooner to get a better sense of this issue.

Can transformations of a data set distort the data being displayed in a harmful manner that prevents us from understanding something important about the data? For example, the textbook says that transforming data can 'reduce skew' for a data set. Isn't it possible that reducing the data set's skew harmfully alters our ability to understand the data set?

It can be harmful, but really only if you forget that you've done the transformation! On the other hand, if you do a transformation and see a pattern you didn't see before, you can use that to write down a model for the original "untransformed" data.

The answer to this question and the previous one are sort of related. The thing that I think is important to keep in mind is that statistics provides you with some very powerful tools for understanding the world around you, but *you should not apply these tools blindly*. Doing so will inevitably lead you astray. If, on the other hand, you think clearly about what you're doing and why it makes sense, you'll be okay.