Module 1 Discussion Reply

1. Interpreting $E(y)$ and $E(y|x)$ for when x is either fixed or random.
    a. I think that your reply to how to understand $E(y)$ is interesting. My approach was to plug in the formula directly into the expectation and try to see the result before making any statements. My simple math gave me that $E(y) = \beta_0 + \beta_1 x$, it does not show anywhere that a summation exists within the result. Looking at your reply, I think I have an idea of what you are trying to say, but I do not entirely agree with it. It is possible that you mean something else and are correct, but I am not sure how to think about the problem in that manner. I see $E(y)$ in this case sort of like the average value for y, given that it is a linear function of x. I think a more accurate explanation that I saw in another classmate's post is that it is the expected value over all possible x's, rather than the summation of them.

       In the part about the conditional expectation, I did something similar. Instead of using random variable X and its realization x, I used $x = x_0$ as the part within the conditional section. My interpretation though is the same as yours, in that it is the expectation given a certain point x_0.

       I think however that the understanding of the difference between $E(y)$ and $E(y|x)$ are about the same, in that one is based over all possible x's, while the other is based on a single x.
    b. I like how you mention here that it's a joint distribution between two random variables. However, it is not stated whether they are bivariate normal for example, and so the answer I think is less straightforward. I like also how you mention that x is some independent random variable. In part a, I plugged in the formula to see what the equation would come out to be. My result is that its $\beta_0 + \beta_1 E(x)$, which is slightly different from what's seen in part a.

       I looked at your description of $E(y|x)$, and I think it is an interesting response. I didn't say the same thing, so I can't immediately say whether I agree or not. My thinking is that it couldn't be directly evaluated without more knowledge of the underlying distribution, but I am not sure that this is correct. I went on to mention the result given that they are bivariate normal, but this is not assumed in the problem so it doesn't quite answer the question either. I think your response is interesting, and I am not sure whether it's correct. However, I need to think about it more to try and understand it.
2. I agree with your response to this question. I noticed in the textbook that there is an example of β_1 having a negative slope, and so I feel like this is a trick question. That is, if x increases by Δ, then the change in y can be either positive or negative, depending on the slope being positive or negative. The problem didn't specify any case, so I just added in this distinction. However, I do think that the increase in x will result in a $\beta_1 \Delta$ change in the overall equation.
3. For this problem I had some similar idea. It seems rather unnatural to have data that falls exactly on the straight line, unless it is simulated data that has been generated artificially. You mention also that it would imply that there is zero noise and error. This also is not something that we would typically see in the real-world. I suppose that this situation is

possible, but it just isn't likely. I think it's interesting that you mention it being a true relationship, however I am not sure what that means. It reminds me of the textbook saying things like the true regression line, which I understand to be the regression line generated by the population parameters which are generally unknown. Here, with true relationship, I suppose that could mean that we have insight then into the actual population relationship between y and x.