Select Reading Question Responses (9/16/2021)

> I'm confused by why you would ever use the R value when the R^2 value seems to give us more information. Can you explain?

In fact, you have it backwards. The R^2 value only tells you about the strength of the relationship. The R value tells you both about the strength and the direction of the relationship.

That being said, note that the information contained in the R value is completely equivalent to the information contained in the R^2 value and the *sign* of the slope of the line of best fit. In other words, there isn't *that* much more information in the R value than the R^2 value, and by the time you've computed R^2 you've likely also computed a line of best fit, so you anyway have the additional information that would be found in R.

> In example 8.10, I understood how the point-slope formula came together with the values from the gift aid data set as $y - 19940 = -0.0431(x - 101,780)$, but I did not understand how that equation simplified to $y = 24327 - 0.0431x$.

First note that $-0.0431 \times -101780 = 4386.718$, so when we distribute $-0.0431(x - 101,780)$, we get

$$y - 19940 = -0.0431x + 4386.718 = 4386.718 - 0.0431x.$$

Then we move the 19940 from the left-hand side to the right-hand side by adding 19940 to both sides and we get

$$y = 24326.718 - 0.0431x.$$

At this point, we just remind ourselves that x represents `family_income` and y is predicted `aid`!

> How do you know if a linear regression is the best fit? Do you just try different types of regressions functions to see which one has the higher R^2 value?

"Line of best fit" is a vague concept. Before you can "know if a linear regression is the best fit," you have to formalize what "best fit" means. The usual way of doing this is to say that "best fit" means "minimizes the sum of squared residuals," and the resulting line is called a least squares regression. But there are other formal definitions of "best fit" (including "minimizes the sum of residual magnitudes," as discussed below).

Note that R (and hence R^2) is a number that's computed using the *data*, not a proposed line (take a look at the footnote on p. 310, or the formula on Wikipedia). It turns out that R^2 has an interpretation in terms of the the residuals of the least squares regression (top of p. 323, or the more detailed discussion on Wikipedia).

Now, you could turn the discussion on the top of p. 323 on its head: you could say that, by "best fit" line, you formally mean the line that maximizes the quantity

$$\frac{s^2_{\text{response}} - s^2_{\text{residual}}}{s^2_{\text{response}}}.$$

I think this is what you probably had in mind when you proposed your question, and this would definitely be an interesting thing to do! But I have no idea if there's an explicit algorithm for computing the equation of the line that maximizes the above quantity. You could just try many different lines and compute the above quantity for each of them and see which one "comes out on top," but that would be quite tedious. It *might* be that the line that maximizes the above quantity is the same as the least squares line, but I don't know for sure. . .

> In what types of applications would we want to minimize the sum of residual magnitudes instead of the sum of squared residual magnitudes?

I have never seen people constructing best-fit lines that minimize the sum of residual magnitudes in practice. There are some theoretical benefits to such a line (it's less impacted by outliers than the usual least squares regression), but those theoretical benefits are usually far outweighed by the computational cost of minimizing the sum of residual magnitudes. But I am also not a statistician, so you should *not* interpret any of this to mean that people never minimize the sum of residual magnitudes in practice ☺

If you'd like, you can read more about minimizing the sum of residual magnitudes on Wikipedia under "least absolute deviations."

> Why are we so concerned with a pure linear regression? x to the first seems pretty arbitrary when we have a tool that can more accurately give us a correlation.

People can (and do!) fit non-linear models to data as well. The thing to bear in mind, though, is *Occam's razor*. If a linear model does a good enough job explaining the data, constructing a nonlinear model might be unnecessarily complicated. What's more, there's actually a practical (not just philosophical) trade-off being made here. When you start fitting complicted nonlinear models, you run the risk of overfitting, ie, of writing down a model which very accurately describes the data you used to construct the model but which doesn't have good predictive power for new data that might arise in the future.