In this tutorial, you're going to learn mainly some cautions about using best-fit lines to make predictions. So let's take a look. The data here is the airfares for different city destinations for Minneapolis-Saint Paul Airport and how many miles away from Minneapolis-Saint Paul they are. We figure this is the explanatory variable because we think that things that are further away should cost more to get to, based on gasoline and things like that.

So the regression equation is predicted airfare is equal to 113.11 plus 0.137 times miles. So it's a pretty straightforward question asking what the predicted airfare would be for a flight from Atlanta from Minneapolis. The distance is 1,064 miles. Simply put this into the regression equation and you get about $259. The biggest question here, though, is how confident are we in that prediction. How are we confident that that prediction is close to what it actually cost to get to Atlanta?

Well, let's look back at the data. Our data, or our line, our linear model, is based on data that had distances that were less than the distance to Atlanta and also more than the distance to Atlanta. So it seems to make sense to use that model to predict what the cost would be for a place that is 1,064 miles from Minneapolis.

But what about the predicted airfare for a flight to Anchorage at a distance of 3,163 miles from Minneapolis? Well, in this case, the prediction differs largely from what the actual airfare ends up being. We can't really use the prediction equation to predict what the airfare to Anchorage would be because it's so far out of the bounds of the data that we use to actually create the model.

So what I'm getting at is this-- so the range of miles that we use to actually create our model was from about 400 miles away from Minneapolis to about 1,300 miles away from Minneapolis. This was the longest distance away, Charleston. This was the shortest distance away, Chicago. What we're saying is that within this window, from about 400 to about 1,300, the line gives reasonable predictions for airfare.

Outside of that window, though, it might not. It might, but it might not. So we have to be very cautious about using this prediction line to predict the airfare to, say, Milwaukee, which is closer than 400 miles to Minneapolis, or to a place like Anchorage. It might not give accurate predictions outside of this particular window.

The whole idea of that is called extrapolation. And it's using the linear model to make predictions outside the range of values for which the estimate was intended, which is this window here. Now it's not always bad to extrapolate. Sometimes linear trends do continue outside of the window from the data that made them. But they don't always.

And so we need to proceed with caution if we do, in fact, end up extrapolating data. Extrapolation is pretty risky.

You're sort of trusting the linear model to continue outside of the bounds that created the model itself. So using the linear model to try and predict outside those bounds might be an unwise decision.

So let's look at a nonsense example. Men's Olympic gold medal 100 meter dash times have decreased at a rate of about 100th of a second per year for the last 60 years. Now, someone who extrapolates might say, well, if this trend continues, then about in 1,000 years, there will be a person whose gold medal sprint time is zero seconds.

Well, we know that that's nonsense. We can't use this line to predict what might happen even 100 years down the road, much less 1,000 years down the road. So, extrapolation might not be a good idea, especially with this particular data set. Extrapolation can lead to nonsense results.

And so to recap. A linear model is a reasonable predictor of response values. And in our example, it was airfare values within the range of values of the explanatory variable, within the range that created it. So, in the 400 to 1,300 mile range that created our airfare graph.

And using that model to predict responses for values outside that range is called extrapolation. And it might not be a good idea. It's not always a bad idea, but a lot of the times, it's risky. And you should be aware that it's risky if you go down that road.

Sometimes, it gives you values that don't make any practical sense. Extrapolation is the reason why sometimes the y-intercept of a least-squares line doesn't have a meaningful interpretation. In our example, the y-intercept was about $113, which means that if you go nowhere from Minneapolis, you pay $113, or at least you're predicted to.

That makes no sense because zero isn't near the values that created the line. So using zero miles to predict an airfare would be extrapolation. And so it doesn't have any meaningful interpretation in that particular context.

So we talked about extrapolation and how it can be risky. Good luck. And we'll see you next time.