Modeling HW 4: Romeo the Kitten

This homework will be a follow-up to the lab on foster kitten weights; in particular, we'll be looking at Romeo's weight, which was very interesting for the first few days he was fostered. We'll use his weight changes as a good warning about the misuse of best-fit lines. Go ahead and open the spreadsheet for this homework now.

In that first tab, you'll see the same "First 7 days" data that we started the lab with (but only for Romeo this time). Now, it's hard to believe, but there are certain folks who work with numbers who would blindly add a trendline to that very limited, not-linear-looking-at-***all*** data set and pretend it made sense. Let's see why it doesn't.

1. **(2 points)** Add the trendline, and give the slope of that trendline here.

2. **(2 points)** In a sentence or two, tell me why that slope makes ***no*** sense in the context of a young, healthy foster kitten's weight[1].

Now, if you recall, we removed a data point or two from the set before actually arriving at a trendline that appeared to make sense. Let's do that again.

3. **(1 point)** Remove the "Day 1" weight. What's the slope now?

Now, it's positive...but just barely (in case you forget what that "E-03" means, watch this!). At this growth rate, Romeo would need about 115 days (or roughly 4 months) until he hit 2 pounds—and, from someone who understands how healthy kittens grow, that makes no sense. In your lab, you figured that it would take fewer than ***20*** days, for all three of the kittens (which agreed not only with the two-week data, but also with what the Humane Society shares with kitten foster families).

OK! Now remove the" Day 2" weight like we did in the lab!

4. **(3 points) (w)** According to your new slope (of the trendline with both Day 1 and Day 2 data removed) how many days would it take for Romeo to get to 2 pounds?

Well, that's closer to the answer you got in your lab, but still about twice as large as it should be! So again we have to ask...what's going on?

Well, we know ***now*** (after doing the lab)…we didn't have enough data here! This HW is ***only*** using week 1's data, and not ***both*** week 1's and week 2's. If you try to analyze just that first week's data, you might get results that aren't indicative of what should be going on, because they were pretty incomplete.

You might have realized something as you've made your way through MTH 098: just because you ***can*** do a certain math process on data doesn't mean you ***should***. If you give a computer program 2 or more points, it'll be able to put a best-fit line through them...whether or not that line makes ***sense***.

So your job as a consumer of data is to ask: am I looking at a reasonably representative set of data, enough to draw a believable conclusion about the behavior of this data? The answer is often "I'm not 100% sure" ...and that's OK! Statistics is rooted in probability, and that word "probability" shares the same start as "probably" ...not "certainly".

Let's do one more exercise to drive that point home. Open up this sheet next! In that sheet, there's a hidden dataset of points (you'll see 5 of them when you get there). Your job is simple: using the check boxes, add points, one at a time, until you're fairly confident that you can describe the ***shape*** of the data. You do not have to add them in any particular order. Go ahead and try it out!

5. **(2 points)** What shape is the data? And how many points did it take you until you saw it?

In a statistical study, you'd be in charge of uncovering those data points, one at a time, through sampling. And then, as the data began to show itself, you'd begin to understand the underlying behavior more and more. However, one difference: in this exercise, you could get to "see" all the data points if you wanted/needed to, but in an actual statistical study, you would likely never get to see every possible data point. Instead, you'd have to draw your conclusions based on a smaller subset. But, as you'll learn if you go on to statistics courses, those subsets are good enough!

---

[1] In this HW, you'll often hear me say things like, "in light of _____, that makes no sense." A big part of understanding the quality of data is being familiar with how it's collected, and the typical behavior of similar data. IN this case, since I've owned cats of all ages for most of my life, I'm familiar with how they grow. However, were I not a cat owner, I'd have no idea! Math, by itself, won't help you be an expert on an applied topic (dedicated to roughly 98% of people who use Twitter). 😊