Competing Model Validation (2.6) Homework

1. Tycho Brahe was a nobleman from Denmark who spent his life taking diligent measurements of the positions of the stars and planets. Johannes Kepler was a poor German who spent his life staring at the data gathered by Brahe. He figured if he could find an equation that modelled Brahe's data well, he could unlock the secrets of the universe. He was somewhat successful in that pursuit. Below is the distance between Pluto and the Sun taken on January 1st every five years from 1965 to 2020. The values of t represent years since 1900. The values of "Distance" have the unit AU. The data did not all fit in one row. The second table is just a continuation of the first. (data source: https://www.heavens-above.com/planets.aspx)

t	60	65	70	75	80	85	90
Distance	33.8	32.7	31.7	30.8	30.2	29.8	29.7

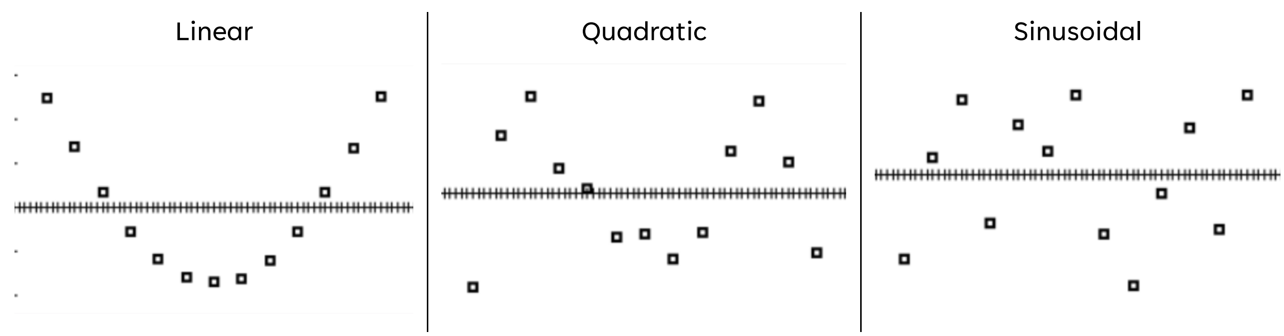
t	95	100	105	110	115	120
Distance	29.8	30.2	30.9	31.8	32.8	34.0

   a. Plot this data and make a prediction of what model type may be most appropriate.

   b. Use a quadratic regression model to predict the distance on January 1st of 2024 ($t = 124$).

   c. Use a sinusoidal regression model to predict the distance on January 1st of 2024 ($t = 124$).

   d. Below are the residual plots of the linear, quadratic, and sinusoidal regression models of this data. Based on these residual plots, which model is most appropriate for modelling the distance between Pluto and the Sun? Justify with a complete sentence or two.



   e. Based on the context of this data, is there a reason to believe one type of model may be more appropriate than another type?

2. When Tycho Brahe was not busy taking diligent measurements of the positions of the stars and planets, he spent his free time doing fun things like losing his nose in a duel, having an affair with the Queen of Denmark, falling in love with a commoner, and getting chased out of Copenhagen by a mob. On October 13th, 1601, Tycho Brahe fell ill. Eleven days later, he died. Some believe he was poisoned with mercury. Methylmercury has a biological half-life of 65 days. This means that every 65 days, the level of methylmercury in one's body can be expected to have decreased by 50%. If the concentration of methylmercury in Brahe's body was $100\,\frac{\text{ng}}{\text{mL}}$ on October 13th, what was the concentration in his body upon his death eleven days later?

3. Below is data reflecting the price of one dozen large grade A egg t years after 1980. (*Data Source: U.S. Bureau of Labor Statistics: Annual Average for Eggs, grade A, large, per doz. in U.S. city average, average price, not seasonally adjusted. CPI Average Price Data*). Use a graphing utility/Desmos to construct $P(t)$, an exponential regression model that fits this data.

t	0	5	10	15	20	25	30	35
Price	\$0.84	\$0.80	\$1.01	\$0.92	\$0.91	\$1.22	\$1.66	\$2.47

a. Based on the model, predict the price of one dozen large grade A eggs in 2025. Round to the nearest cent.
b. Based on this model, what value of t corresponds with a price of $\$1.50$?
c. What is the residual value of the data point for the year 1990 ($t = 10$)?
d. What is the residual value of the data point for 2000 ($t = 20$)?
e. Is an exponential model appropriate for this data set? Use the word “residual” in your justification for why or why not this model is appropriate.
f. In 2020, supply chain issues and an increased interest in homemade bread led to unusual egg prices. Based on this information, what domain restriction should be put on the exponential model?

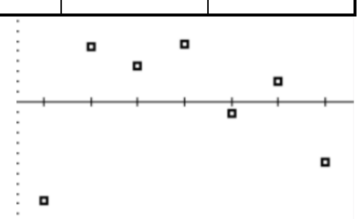
4. Bigger turkeys take longer than small turkeys to finish cooking. A turkey is done cooking once its internal temperature, checked at the thickest part of the turkey, reaches 165°F. Every year I take careful measurements of how long it takes my turkey to reach this temperature. The data are given below.

Weight of Turkey (lb)	12	13	13	18	18.5	19	19	20	21
Time Until Cooked (minutes)	165	180	170	240	246	253	250	255	280

a. I gather this data so that I can form a model that will input the weight of a turkey and output how long it will take for that turkey to reach 165°F. That way I know when I should open the oven to check the turkey’s temperature. Would it be better for such a model to underestimate or overestimate cooking time?
b. Uh-oh, I think my thermometer is broken. I’m just going to take the turkey out of the over based on how long the model predicts it will take to cook. Would it be better for such a model to underestimate or overestimate?
c. For which turkey weight should I have more confidence when using the model: 15 lbs or 30 lbs?

5. As America’s population grows, more telephone numbers are needed. On Jan 1, 1997, there were only 151 telephone area codes. Over the next six years, the number of area codes grew at a steady rate until Jan 1, 2003 when there were 274 area codes. The table below shows the number of area codes in the intervening years (Source: USA Today, NeuStar, Inc.). Using the linear regression feature in Desmos, construct a model for $N(t)$, the number of area codes in the year t.

Year	1997	1998	1999	2000	2001	2002	2003
Number of Area Codes	151	186	204	226	239	262	274



a. What in this scenario suggests that a linear model is appropriate? To the right is a plot of the residuals. Based on this plot, is a linear model appropriate? How do you know?
b. How many area codes does the model suggest would be needed in the year 2023? Round to the nearest whole number.
c. In what year does the model suggest 1000 area codes would be needed? Round to the nearest year.
d. Based on the model, how many years did it take to go from 200 area codes to 300 area codes?

6. Throughout the 1970s, scientists noted a growing area of depleted ozone over Antarctica. The mean area of the "ozone hole" for years in the 1980s is given below with unit "million km^2" (Source: https://ozonewatch.gsfc.nasa.gov). Using linear regression, construct a model for $A(t)$, the size of the ozone hole, in million km^2, in the year t.

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Area	1.4	0.6	4.8	7.9	10.1	14.2	11.3	19.3	10	18.7

   a. Do you believe a linear model is appropriate for this data? Justify.

   b. Based on this linear model, extrapolate the size of the ozone hole in the year 2023.

7. In 1987, an international treaty known as the Montreal Protocol banned certain ozone depleting substances. Ozone levels stabilized around the year 2000 and then began dropping. The mean area of the "ozone hole" for years in the 2010s is given below with unit "million km^2" (Source: https://ozonewatch.gsfc.nasa.gov). Using the linear regression feature in Demos/a graphing calculator, construct a model for $A(t)$, the size of the ozone hole, in million km^2, in the year t. Construct a model using only the 2010s data. Do not include the 1980s data from the previous question.

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Area	19.4	24.7	17.8	21	20.9	25.6	20.7	17.4	22.9	9.3

   a. Based on this model, extrapolate the year in which the ozone hole will reach a size of zero.

   b. Find the residual corresponding to year 2013.

8. Parris invents a new type of pickleball racket and begins selling it on Amazon. Below is data reflecting how many rackets she sells on certain days.

Days since sales began	0	50	100	150	200	250	300	350
Rackets sold on that day	100	376	562	658	665	584	415	157

   a. Parris has reason to believe her data may be modelled well by a quadratic function. Use quadratic regression to model Parris's sales. On what day does the model predict her sales will be greatest?

   b. On what day does the model predict Parris will make no sales?

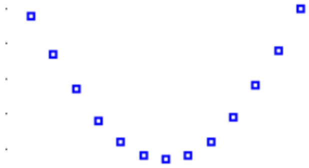
   c. What is an appropriate domain restriction for this model?

   d. Based on the model, how much time passes between the first day that Parris makes 500 sales and the last day that Parris makes 500 sales.

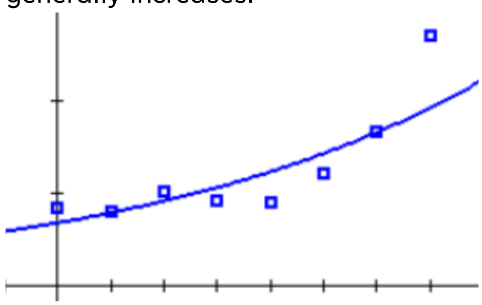
   e. Let $S(t)$ be this model for the number of sales on day t. Find the average rate of change of S on the interval $0 \leq t \leq 7$.

Competing Model Validation HW Answers

1.
    a. Below is a plot of the data. Based off this plot alone, a quadratic or sinusoidal model may be appropriate. The context of the problem and the residual plots can help us realize one of these is far more appropriate than the other (see part d and part e).

    b. 35.290 AU
    c. 34.927 AU
    d. The sinusoidal model is most appropriate as there is no clear pattern in the graph of the residuals.
    e. The sinusoidal model is most appropriate as Pluto's distance from the Sun increases and decreases periodically as it orbits the Sun. The end behavior of a quadratic model would suggest that as time goes on, Pluto drifts further and further from the Sun and never returns.
2. 88.932
3.
    a. $\$2.46$
    b. $t = 27.321$
    c. 0.086
    d. -0.312
    e. An exponential model is not the most appropriate model as there is a clear pattern to the residuals. Even without making a residual plot, we can see it in the below plot of the data and the exponential regression. For $t = 0$ the residual is positive. Then, the residual generally decreases. Then the residual generally increases.

    f. Because t represents years since 1980 and the model should only apply for years less than 2020, the model should only allow input values $t < 40$.
4.
    a. It would be better in this context to underestimate cooking time. An underestimate would lead to checking the turkey too soon, but no harm comes of that. An overestimate would lead to checking the turkey took late, risking it burning or doing any of the other weird things meat may do when cooked for too long.
    b. On the other hand, if you do not have a thermometer, and are using only the model to decide when to remove the turkey from the oven, then it would be best if the model overestimates cooking time. One would rather eat a turkey that was cooked too long and is a little dry than eat a turkey that was not cooked long enough and is a little food poison-y.
    c. One should have more faith in the model's ability to predict the cooking time of a 15 lbs turkey than a 30 lbs turkey. The maximum weight provided in the data is 21 lbs. It is unclear if the trend captured by the model can be extended to 30 lbs. On the other hand, 15 lbs falls within the range of input values

of the given data. In other words, we trust interpolation (making predictions for 15 lbs) more than we trust extrapolation (making predictions for 30 lbs).

5.
   a. The phrase “grew at a steady rate” seems to be suggesting a relatively constant rate of change. Linear functions have a constant rate of change. The residuals have no clear pattern, so a linear model seems appropriate.
   b. 677
   c. 2039.266. In 2039 the model predicts there will be 1000 area codes.
   d. 5.036 years

6.
   a. Yes, a linear model is appropriate. The data appears linear, and there is no clear pattern to the residuals.
   b. $81.767 \text{ million km}^2$

7.
   a. $t = 2046.9$, so 2046 or 2047
   b. 0.105

8.
   a. Day 179.469
   b. Day 374.112
   c. For $t > 374.112$, the model has negative output values. In reality, negative sales are not possible. So we must exclude these t-values from the domain. For $t < 0$, the model has non-zero output values. But this corresponds with a time before the racket was released. In reality, no sales were made before the racket was released. So, we must also exclude these t-values from the domain. This model should only be used for input values in the interval $0 \leq t \leq 374.112$. Outside of this interval, the model gives output values without real meaning.
   d. 197.456 days
   e. $\frac{S(7)-S(0)}{7-0} = 6.254 \frac{\text{sales}}{\text{day}}$