Applied Math Modeling Exam 1 (Spring 2020)

Name:

Directions: Show your work! Answers without justification will likely result in few points. Your written work also allows me the option of giving you partial credit in the event of an incorrect final answer (but good reasoning). Indicate clearly your answer to each problem (e.g., put a box around it). **Good luck!**

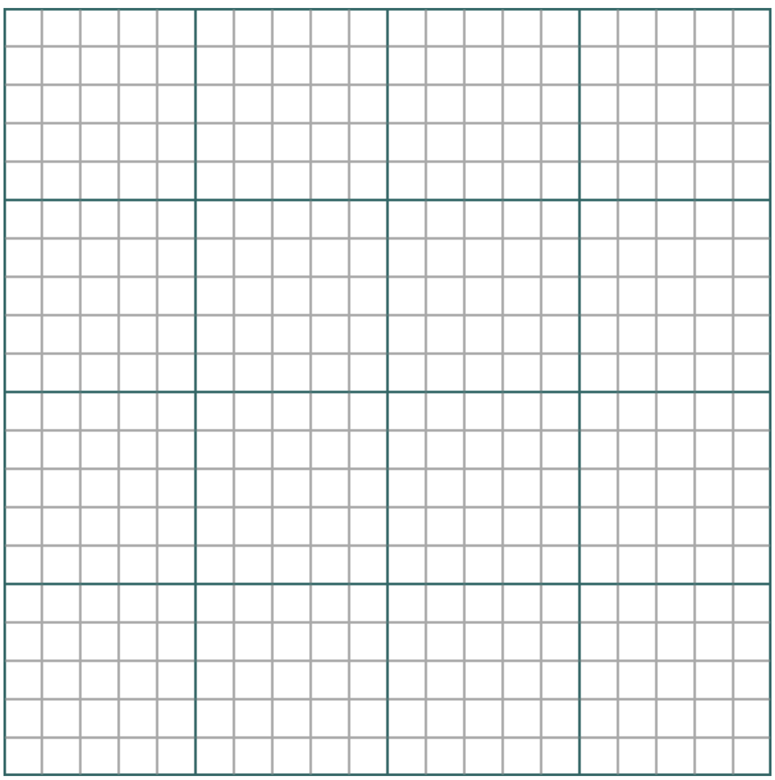
Problem 1. (13 pts) Use the following data to fill in the tables at right (3 pts), and then

x_i	0	1	2	3	4
y_i	-1	2	1	0	3

x_i^2					
$x_i y_i$					

$\overline{x}$	$\overline{y}$	$\overline{x^2}$	$\overline{xy}$

a. (6 pts) construct the simple linear regression model of best fit (find $y(x) = a + bx$). Show all work. Sketch the data and your line in the grid provided.



b. (4 pts) Compute R^2 ($SSE = 6.40$; $SS_{Total} = 10$). How do we interpret R^2? Illustrate in your graph of the regression line which "errors" are used in its calculation.

Problem 2. (10 pts) I have issues with some regressions: I hope that you do, too!

a. (4 pts) Comment on the linear regression model featured in this graphic, and sketch in a model you might prefer (justify). Note: one data point is on the corner of the box.

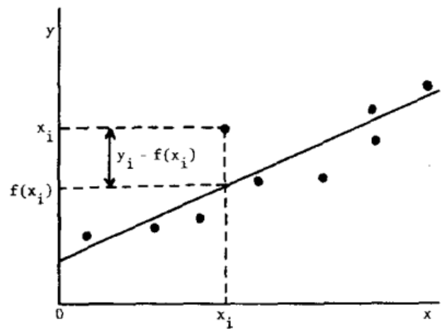
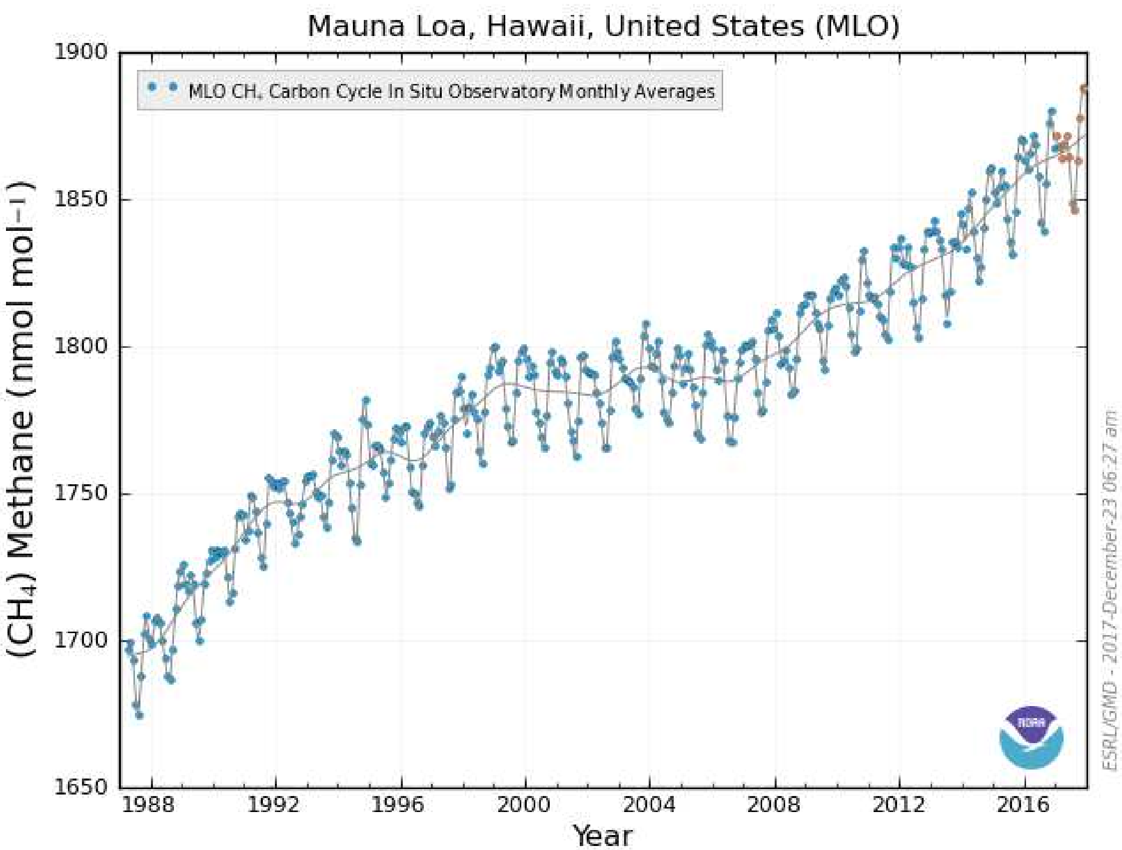


Figure 5. Deviations

b. (6 pts) Atmospheric methane (CH_4) is also measured at Mauna Loa; what model might you suggest for this data? (The curve running through the middle of the scatter is a smoothed version of the data.) Propose a model that will capture the trend and oscillatory behavior (**be as specific as you can**). What aspects of the trend are important to capture?



Problem 3. (13 pts) For the Keeling data, with data given monthly (time t in years), oscillations in the CO_2 levels suggest a model containing a sinusoidal component – $A \sin\left(2\pi \frac{t - t_0}{T}\right)$.

a. (3 pts) Why did we need to use **a pair** of trigonometric functions in the **linear regression** model? And how did we get rid of T?

b. (8 pts) Suppose that a linear regression involving $c\cos(2\pi x) + d\sin(2\pi x)$ gives 95% confidence intervals of $c \in [2.5, 3.5]$ and $d \in [-5, -3]$.

   i. (4 pts) What would be the best **point** estimates (single values) for the oscillation's amplitude A and its phase t_0?

   ii. (4 pts) What are approximate 95% confidence intervals for A and t_0?

c. (2 pts) In the log ruler problem, linear regression helped us discover the quadratic model of the Ontario rule. What was the only source of error in the regression? That is, why wasn't the linear regression a perfect fit, since the markings for the number of board feet of lumber were computed from a quadratic function?

Problem 4. (12 pts) These questions pertain to the cross country problem.

a. (8 pts) The linear model at right is a regression model for log-transformed data (original data shown at left, with an inappropriate linear regression). We can use the slope and intercept from it to estimate the Riegel-like power model $P(m) = am^b$ for the period(P) as a function of mass(m). Find that model. (Note: the logarithms used are base 10 logs, $\log_{10}(x)$.)

Figure 1: "...the period of an object oscillating at the end of a spring depends on the object's mass."

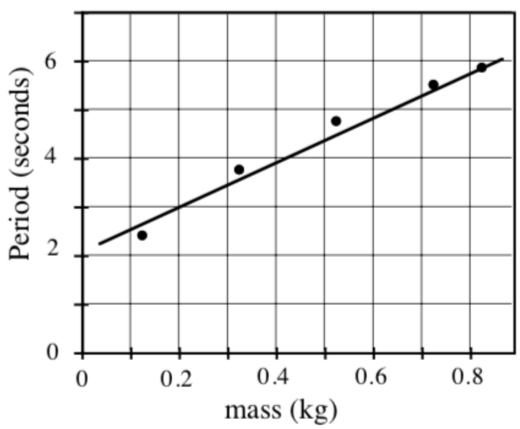


Figure 5.1: Graph of the oscillation period as a function of mass.

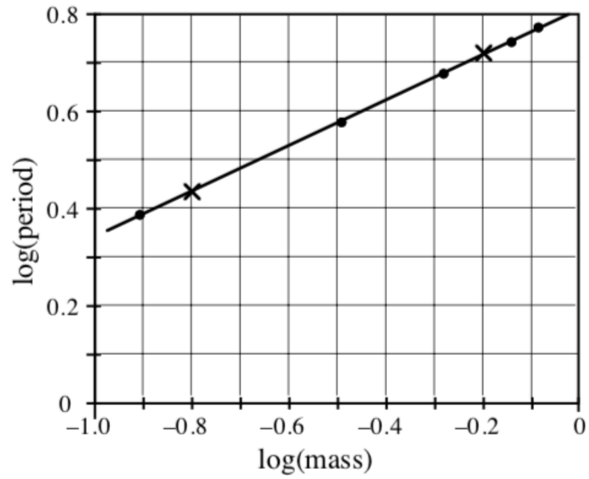


Figure 5.2: Graph of the log of the oscillation period as a function of the log of the mass.

b. (4 pts) One strategy to control for the different eras in which Thad and Thomas ran was to compare each to the distribution of winning times of their era's champions. Explain how that worked. What assumptions were made about the distributions, and how did we use those to measure Thad against Thomas?