MTH 325, Homework #2, Spring 2023 Name ____

Instructions: Show work or attach R code used to perform calculations (or any other technology used). Be sure to answer all parts of each problem as completely as possible, and attach work to this cover sheet with a staple.

1. No tortilla chip aficionado likes soggy chips, so it is important to find characteristics of the production process that produce chips with an appealing texture. The following data on x=frying time (sec) and y=moisture content (%) appeared in an article on the subject.

x	5	10	15	20	25	30	45	60
y	16.3	9.7	8.1	4.2	3.4	2.9	1.9	1.3

   a. Construct a scatterplot of y vs x and comment.
   b. Construct a scatterplot of the $(\ln(x), \ln(y))$ pairs and comment.
   c. What probabilistic relationship between x and y is suggested by the linear pattern in the plot of part (b)?
   d. Predict the value of moisture content when frying time is 20, in a way that conveys information about reliability and precision.
   e. Analyze the residuals from fitting the simple linear regression model to the transformed data and comment.

2. In each of the following cases, decide whether the given function is intrinsically linear. If so, identify $x' \, and \, y'$, and then explain how a random error term ϵ can be introduced to yield an intrinsically linear probabilistic model.
   a. $y = \frac{1}{\alpha+\beta x}$
   b. $y = \frac{1}{1+e^{\alpha+\beta x}}$
   c. $y = e^{e^{\alpha+\beta x}}$
   d. $y = \alpha + \beta e^{\lambda x}$

3. An article reported on an experiment to investigate how the behavior of mozzarella cheese varied with temperature. Consider the accompanying data on x (temperature) and y (elongation %) at failure of the cheese.

x	59	63	68	72	74	78	83
y	118	182	247	208	197	135	132

   a. Construct a scatterplot.
   b. Model the data with a nonlinear model: try quadratic, exponential, log and power models. Which model appears to fit the best?

4. Continuous recording of heart rate can be used to obtain information about the level of exercise intensity or physical strain during sports participation, work or other daily activities. An article reported on a study to investigate using heart rate response (x as a percentage of the maximum heart rate) to predict oxygen uptake (y, as a percentage of maximum uptake).

HR	43.5	44.0	44.0	44.5	44.0	45.0	48.0	49.0
VO_2	22.0	21.0	22.0	21.5	25.5	24.5	30.0	28.0
HR	49.5	51.0	54.5	57.5	57.7	61.0	63.0	72.0
VO_2	32.0	29.0	38.5	30.5	57.0	40.0	58.0	72.0

Perform a simple linear regression analysis, paying particular attention to the presence of any unusual or influential observations.

5. The following data on mass rate of burning x and flame length y is representative of that which appeared in an article on the subject.

x	1.7	2.2	2.3	2.6	2.7	3.0	3.2
y	1.3	1.8	1.6	2.0	2.1	2.2	3.0
x	3.3	4.1	4.3	4.6	5.7	6.1	
y	2.6	4.1	3.7	5.0	5.8	5.3	

   a. Estimate the parameters of a power function model.
   b. Construct diagnostic plots to check whether a power function is an appropriate model choice.
   c. Test $H_0: \beta = \frac{4}{3}$ vs. $H_a: \beta < \frac{4}{3}$ using a level 0.05 test.
   d. Test the null hypothesis that states that the median flame length when burning rate is 5.0 is twice the median flame length when burning rate is 2.5 against the alternative that this is not the case.

6. The accompanying data on y is energy output (W) and x is temperature difference (K) was provided in an article. The article's authors fit a cubic model to the data.

x	23.20	23.50	23.52	24.30	25.10	26.20	27.40	28.10	29.30	30.60	31.50	32.01
y	3.78	4.12	4.24	5.35	5.87	6.02	6.12	6.41	6.62	6.43	6.13	5.92
x	32.63	33.23	33.62	34.18	35.43	35.62	36.16	36.23	36.89	37.90	39.10	41.66
y	5.64	5.45	5.21	4.98	4.65	4.50	4.34	4.03	3.92	3.65	3.02	2.89

   a. Fit a cubic model to the data.
   b. What proportion of the observed variation in energy output can be attributed to the model relationship?
   c. Fit a quadratic model to the data.
   d. Calculate the adjusted R^2 for the quadratic model and the ordinary R^2. Compare these values to the R^2 and adjusted R^2 for the cubic model.
   e. Does the cubic predictor appear to provide useful information about y over and above provided by the linear and quadratic models? State and test the appropriate hypotheses.
   f. Obtain a 95% confidence interval for the true average energy output in this case, when $x = 30$, and also a 95% prediction interval for a single energy output to be observed when temperature difference is 30.
   g. Conduct a hypothesis test of the mean prediction being 5 or different from 5 when $x = 35$ using a 0.05 significance level.

7. A trucking company considered a multiple regression model for relating the dependent variable y (total daily travel time for one of its drivers (hours)) to the predictors x_1 (distance traveled, miles) and x_2 (the number of deliveries made). Suppose that the model equation is

$$Y = -0.800 + 0.060x_1 + 0.900x_2 + \epsilon$$

   a. What is the mean value of travel time when distance traveled is 50 miles and three deliveries are made?
   b. How would you interpret $\beta_1 = 0.060$, the coefficient of the predictor x_1? What is the interpretation of $\beta_2 = 0.900$?

c. If $\sigma = 0.5$ hours, what is the probability that travel time will be at most 6 hours when three deliveries are made, and the distance traveled is 50 miles?

8. An investigation of a die-casting process resulted in the accompanying data on x_1 (furnace temperature), x_2 (die close time) and y (temperature difference on the die surface).

x_1	1250	1300	1350	1250	1300	1250	1300	1350	1350
x_2	6	7	6	7	6	8	8	7	8
y	80	95	101	85	92	87	96	106	108

a. Construct a multiple regression model of this data.
b. Carry out the model utility test.
c. Calculate and interpret a 95% confidence interval for β_2, the population regression coefficient of x_2.
d. When $x_1 = 1300$ and $x_2 = 7$, the estimated standard deviation of $\hat{y}$ is 0.353. Calculate a 95% confidence interval for the true average temperature difference when furnace temperature is 1300 and die close time is 7.
e. Calculate a 95% prediction interval for the temperature difference resulting from a single experimental run with a furnace temperature of 1300 and a die close time of 7.