MTH 325, Homework #1, 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. A February 2, 2008 article in the Columbus Dispatch reported a study on the distance students lived from campus and average GPA. Here is a summary of the results.

Residence	*Avg. GPA*
Residence Hall	3.33
Walking Distance	3.16
Near campus (long walk or short drive)	3.12
Within the county (not near campus)	2.97
Outside the county	2.94

Based on this data, the association between the distance a student lives from campus and GPA is negative. Many universities require freshman to live on campus, but this data has prompted some to suggest that sophomores should also be required to live on campus in order to improve grades. Does this data imply that living closer to campus actually does improve grades or might there be another factor that should be considered?

2. The gas mileage of an automobile first increases as the speed increases and then decreases as the speed increases. Suppose that this relationship is very regular. Make a scatterplot of the mileage vs. speed data shown below and find the correlation (show that r=0). Explain why the correlation is zero even though there is a strong relationship?

Speed	**25**	**35**	**45**	**55**	**65**
Mileage	20	24	26	24	20

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. What does the plot suggest about the relationship between the two variables?

4. Suppose that a certain chemical process the reaction time y (hr) is related to temperature (F) in the chamber in which the reaction takes place according to the simple linear regression model with equation $y = 5.00 - 0.01x$ and $\sigma = 0.075$.
   a. What is the expected change in reaction time for 1-degree Fahrenheit increase in temperature? For a 10-degree Fahrenheit increase in temperature?
   b. What is the expected reaction time when temperature is 200-degrees Fahrenheit? When 250?
   c. Suppose five observations are made independently on reaction time, each one for a temperature of 250 degrees. What is the probability that all five times are between 2.4 and 2.6 hrs?

d. What is the probability that two independently observed reaction times for temperatures 1-degree apart are such that the time at the higher temperature exceeds the time at the lower temperature?

5. The efficiency ratio for a steel specimen immersed in a phosphating tank is the weight of the phosphate coating divided by the metal loss (both in mg/ft^2). The accompanying data was provided on tank temperature (x) and efficiency ratio (y).

Temp	170	172	173	174	174	175	176	177	180
Ratio	0.84	1.31	1.42	1.03	1.07	1.08	1.04	1.80	1.45
Temp	180	180	180	181	181	182	182	182	182
Ratio	1.60	1.61	2.13	0.84	1.43	0.90	1.81	1.94	2.68
Temp	184	184	185	186	188				
Ratio	1.49	2.52	3.00	1.87	3.08				

a. Construct a stem-and-leaf display of both temperature and efficiency ratio. Comment on any interesting features.
b. Is the value of efficiency ratio completely and uniquely determined by tank temperature?
c. Construct a scatterplot of the data. Does it appear that efficiency could be well-predicted by temperature? Explain.
d. Determine the equation of the estimated regression line.
e. Calculate a point estimate for the true average efficiency ratio when tank temperature is 182.
f. Calculate the residuals for which temperature is 182. Why do they not all have the same sign?
g. What proportion of the observed variation in efficiency ratio can be attributed to the simple linear regression relationship between the two variables?

6. The accompanying data describes the relationship between the independent variable SO_2 deposition rate ($\frac{mg}{m^2}/d$), and the dependent variable is steel weight loss (g/m^2).

x	**14**	**18**	**40**	**43**	**45**	**112**
y	280	350	470	500	500	1200

a. Construct a scatterplot. Does the simple linear regression model appear to be reasonable in this situation?
b. Calculate the equation of the regression line.
c. What percentage of observed variation in steel weight loss can be attributed to the model relationship in combination with variation in deposition rate?
d. Because the largest x value in this sample greatly exceeds the others, this observation may have been very influential in determining the equation of the estimated line. Delete this observation and recalculate the equation. Does the new equation appear to differ substantially from the original one (you might consider predicted values, or compare the new slope of the regression line to the confidence interval for the slope of the initial line)?
e. Conduct a hypothesis test on the slope of the regression line. Is it statistically significantly different from the initial slope?

7. The catch basin in a storm-sewer system is the interface between surface runoff and the sewer. The catch-basin insert is a device for retrofitting catch basins to improve pollutant-removal

properties. An article reported on tests of various inserts under controlled conditions for which inflow is close to what can be expected in the field. Consider the following data for one particular type of insert where x is the amount filtered (kiloliters) and y is percent total suspended solids removed.

x	23	45	68	91	114	136	159	182	205	228
y	53.3	26.9	54.8	33.8	29.9	8.2	17.2	12.2	3.2	11.1

a. Does a scatterplot support the choice of the simple linear regression model? Explain.
b. Obtain the equation of the least squares line.
c. What proportion of the observed variation in percentage removed can be attributed to the model relationship?
d. Does the simple linear regression model specify a useful relationship? Carry out an appropriate test of hypotheses using a significance level of 0.05.
e. Is there strong evidence for concluding that there is at least a 2% decrease in true average suspended solid removal associated with a 10 kiloliter increase in the amount filtered? Test appropriate hypotheses using the 0.05 level.
f. Calculate and interpret a 95% confidence interval for true percentage removed when amount filtered is 100 kiloliters. How does this interval compare in width to a confidence interval when amount filtered in 200 kiloliters?
g. Calculate and interpret a 95% prediction interval for percent removed when amount filtered is 100 kiloliters. How does this interval compare in width to the confidence interval calculated above, and to a prediction interval when amount filtered is 200 kiloliters?

8. Physical properties of six flame-retardant fabric samples were investigated in an article. Use the accompanying data and a 0.05 significance level to determine whether a linear relationship exists between stiffness x (mg-cm) and thickness y (mm). Is the result of the test surprising in light of the value of r?

x	**7.98**	**24.52**	**12.47**	**6.92**	**24.11**	**35.71**
y	0.28	0.65	0.32	0.27	0.81	0.57