Directions: Complete each problem. A complete problem has not only the answer, but the solution and reasoning behind that answer. All work must be submitted on separate pieces of paper.

1) Manatees are large, gentle sea creatures that live along the Florida coast. Many manatees are killed or injured by powerboats. Here are data on powerboat registration (in thousands) and the number of manatees killed by boats in Florida in the years 1977 to 1990.

Year	Powerboat Registration (1000)	Manatees killed	Year	Powerboat Registration (1000)	Manatees killed
1977	447	13	1984	559	34
1978	460	21	1985	585	33
1979	481	24	1986	614	33
1980	498	16	1987	645	39
1981	513	24	1988	675	43
1982	512	20	1989	711	50
1983	526	15	1990	719	47

a) We want to examine the relationship between number of powerboats and number of manatees killed by boats. Which is the explanatory variable?
b) Construct a scatterplot of these data (in context). What does the scatterplot show about the relationship between the variables?
c) Describe the direction of the relationship.
d) Describe the strength of the relationship. Can the number of manatees killed be predicted accurately from powerboat registrations? If powerboat registrations remained constant at 719,000, about how many manatees would be killed by boats each year?

2) Propelled by a stream of pressurized water, jet skis and other so-called we bikes carry from one to three people, retail for an average price of $5,700, and have become one of the most popular types of recreational vehicle sold today. But critics say that they're noisy, dangerous, and damaging to the environment. An article in the August 1997 issue of the *Journal of the American Medical Association* reported on a survey that tracked emergency room visits at randomly selected hospitals nationwide. Here are data on the number of jet skis in use, the number of accidents, and the number of fatalities for the years 1987-1996.

Year	Number in use	Accidents	Fatalities
1987	92,756	376	5
1988	126,881	650	20
1989	178,510	844	20
1990	241,376	1,162	28
1991	305,915	1,513	26
1992	372,283	1,650	34
1993	454,545	2,236	35
1994	600,000	3,002	56
1995	760,000	4,028	68
1996	900,000	4,010	55

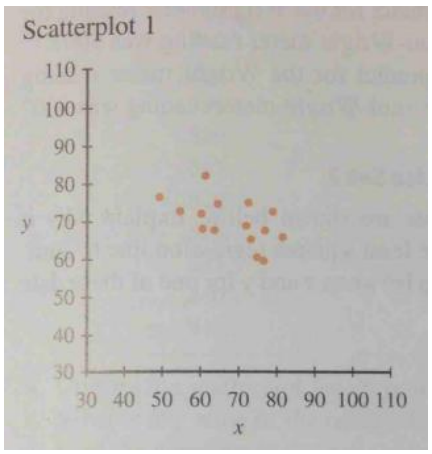
a) We want to examine the relationship between the number of jet skis in use and the number of accidents. Which is the explanatory variable?
b) Make a scatterplot of these data (in context). What does the scatterplot show about the relationship between these variables?
c) Describe the direction of the relationship.
d) Describe the form of the association.

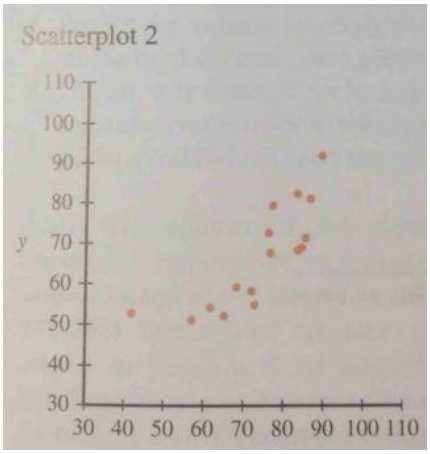
3) Metabolic rate, the rate at which the body consumes energy, is important in studies of weight gain, dieting, and exercise. Table 3.2 gives data on the lean body mass and resting metabolic rate for 12 women and 7 men who are subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all fat. Metabolic rate is measured in calories burned per 24 hours, the same calories used to describe the energy content of foods. The researchers believe that lean body mass is an important influence on metabolic rate.

Subject	Sex	Mass(kg)	Rate(cal)	Subject	Sex	Mass(kg)	Rate(cal)
1	M	62.0	1792	11	F	40.3	1189
2	M	62.9	1666	12	F	33.1	913
3	F	36.1	995	13	M	51.9	1460
4	F	54.6	1425	14	F	42.4	1124
5	F	48.5	1396	15	F	34.5	1052
6	F	42.0	1418	16	F	51.1	1347
7	M	47.4	1362	17	F	41.2	1204
8	F	50.6	1502	18	M	51.9	1867
9	F	42.0	1256	19	M	46.9	1439
10	M	48.7	1614				

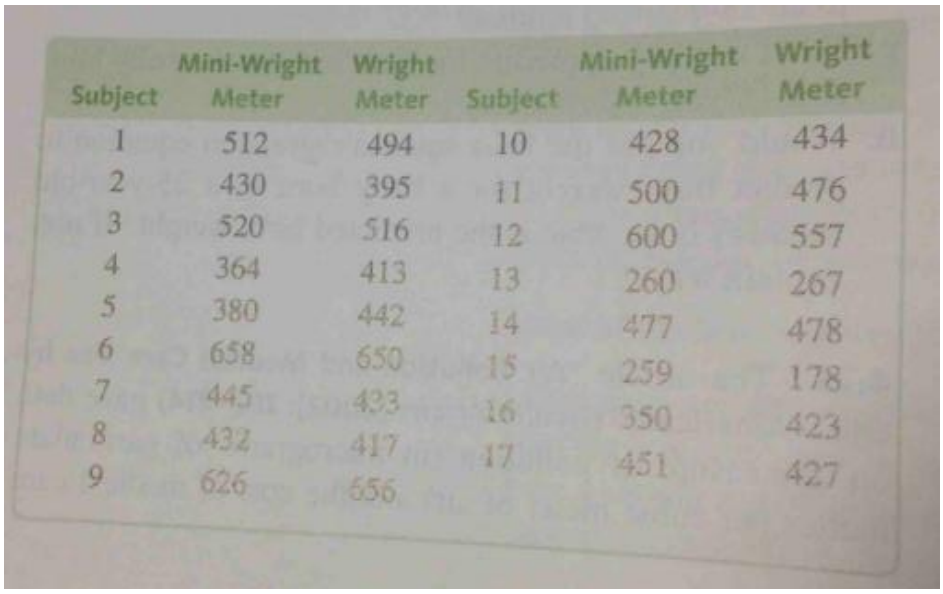
a) Consider only the female subjects from problems a) and b). Which is the explanatory variable?
b) Is the association between these variables positive or negative? What is the form of the relationship? How strong is the relationship?
c) Now add the data for the male subjects to your graph, using a different color or a different plotting symbol. Does the pattern of relationship that you observed in b) hold for men also? How do the male subjects as a group differ from the female subjects as a group?

4) Two scatterplots are shown below. Explain why it makes sense to use the least squares regression line to summarize the relationship between x and y for one of these sets but not the other.





5) The authors of the paper "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement" (*International Journal of Nursing Studies* [2010]; 931-936) compared two different instruments for measuring a person's ability to breathe out air. This measurement is helpful in diagnosing various lung disorders. The two instruments considered were a Write peakf low meter and a mini-Wright peak flow meter. Seventeen people participated in the study, and for each person air flow as measured once using the Write meter and once using the mini-Write meter.



Subject	Mini-Wright Meter	Wright Meter	Subject	Mini-Wright Meter	Wright Meter
1	512	494	10	428	434
2	430	395	11	500	476
3	520	516	12	600	557
4	364	413	13	260	267
5	380	442	14	477	478
6	658	650	15	259	178
7	445	433	16	350	423
8	432	417	17	451	427
9	626	656			

a) Suppose that the Write meter provides a better measure of air flow, but the mini-Write meter is easier to transport and to use. If the two types of meters produce different readings, but there is a strong relationship between the readings, it would be possible to use a reading from the mini-Wright meter to predict the reading that the larger Write meter would have given. Use the given data to find an equation to predict Write meter reading using a reading from the mini-Wright meter.
b) What would you predict for the Wright meter reading for a person whose mini-Wright meter reading was 500?
c) Describe the strength and direction of the actual data. What conclusions can you draw about the connection between the mini-Wright and Wright meters based on the r^2 value?

6) The data in the accompanying table are from the paper "Six-Minute Walk Test in Children and Adolescents" (*The Journal of Pediatrics* [2007]: 395-399). Boys completed a test that measures the distance that the subject can walk on a flat, hard surface in 6 minutes. For each age group shown in the table, the median distance walked by the boys in that age group is given.

Age Group	Representative Age (midpoint of age group)	Median Six-minute Walk Distance (meters)
3–5	4.0	544.3
6–8	7.0	584.0
9–11	10.0	667.3
12–15	13.5	701.1
16–18	17.0	727.6

a) Determine the least squares regression line, as well as the meaning of it's slope.
b) Compute the residuals, and construct a residual plot. Are there any unusual features in the residual plot?

7) Briefly explain why a large value of r^2 is desirable in a regression setting.

8) The following data on x = frying time (in seconds) and y = moisture content (%) appear in the paper "Thermal and Physical Properties of Tortilla Chips as a Function of Frying Time." (*Journal of Food Processing and Preservation* [1995]: 175-189)

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) Base on the accompanying MINITAB output, does the least squares regression line effectively summarize the relationship between y and x?

The regression equation is
log(moisture) = 2.02 – 1.05 log(time)

Predictor	Coef	SE Coef	T	P
Constant	2.01780	0.94978	219.99	0.000
log(time)	-1.05171	0.07091	-14.83	0.000

S=0.0657067 R-Sq=97.3% R-Sq(adj)=96.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.94978	0.94978	219.99	0.000
Residual Error	6	0.02590	0.00432		
Total	7	0.97569			

b) Use the MINITAB output to predict moisture content when frying time is 35 seconds.
c) Based on the MINITAB output, what type of function transformation occurred? Verify this by sketching a scatterplot of the original data.

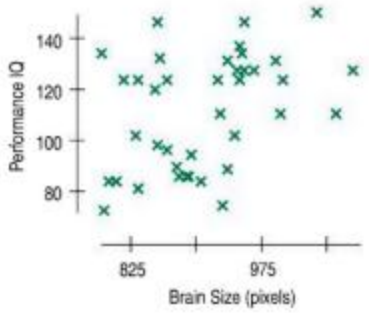
9) Most baseball hitters perform differently against right-handed and left-handed pitching. Consider two players, Joe and Moe, both of whom bat right-handed. The table below records their performance against right-handed and left-handed pitchers.

Player	Pitcher	Hits	At bats
Joe	Right	40	100
	Left	80	400
Moe	Right	120	400
	Left	10	100

a) Make a two-way table by player (Joe vs. Moe) versus outcome (hit vs. no hit) regardless of pitcher.
b) Determine the overall batting average (hits divided by total times at bat) for each player. Who has the higher batting average?

c) Make a separate two-way table of player versus outcome for each kind of pitcher. From these tables, find the batting average of Joe and Moe against right-handed pitching. Who does better? Do the same for left-handed pitching. Who does better?
d) The manager doesn't believe that one player can hit better against both left-handers and right-handers yet have a lower overall batting average. Explain in simple language why this happens to Joe and Moe.

10) A study examined brain size (measured as pixels counted in a digitized magnetic resonance image [MRI] of a cross section of the brain) and IQ (4 Performance scales of the Weschler IQ test) for college students. The scatterplot shows the Performance IQ scores vs the brain size. Comment on the association between brain size and IQ.



11) The correlation between a car's horsepower and its fuel economy (in mpg) is r= -0.869. What fraction of the variability in fuel economy is accounted for by the horsepower?

12) To the right are data collected on the prices and capacity of hard drives. Below is the regression analysis of Price vs Capacity.

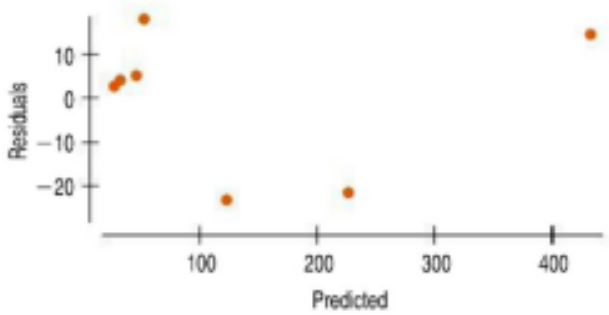
Dependent variable is Price
R-squared = 98.8%
s = 17.95

Variable	Coefficient
Intercept	18.617
Capacity	103.929

Capacity (in TB)	Price (in $)
0.080	29.95
0.120	35.00
0.250	49.95
0.320	69.95
1.0	99.00
2.0	205.00
4.0	449.00

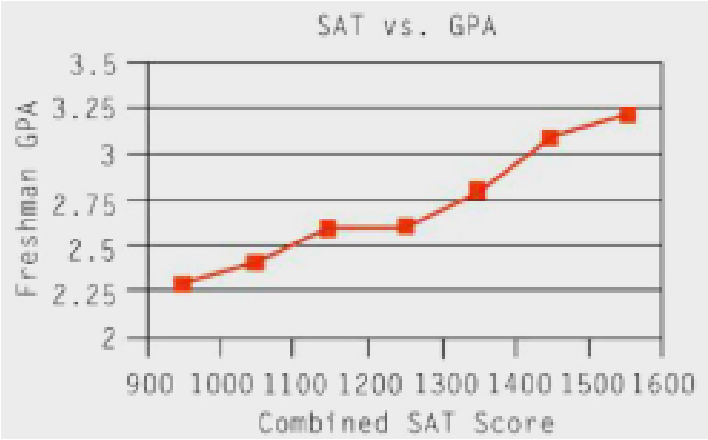
a) Without the use of a calculator, write the regression equation. Define the variables used in your equation. (You may use a calculator to check your equation for accuracy)
b) Explain the meaning of the slope and y-intercept in context.
c) Predict the price of a 3.0 TB drive.
d) Shahin found a 3 TB hard drive for $300. According to your regression equation, determine why this either is or is not a good buy.
e) Determine the average price for a 3.0 TB hard drive today (provide your data, where you found it, and any necessary citations as evidence of statistical legitimacy). How does the price you found compare to the price listed in part d), and what does that suggest about the context of this problem as a whole?

13) Below is a scatterplot of the residuals from the regression of the hard drive prices in exercise 12.

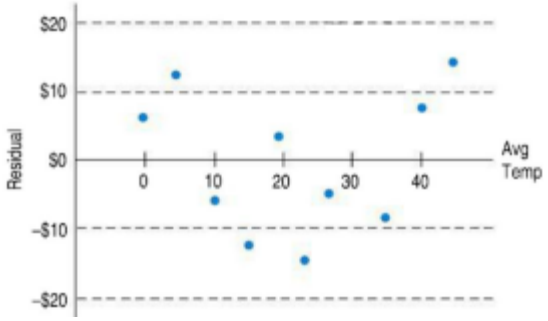


a) Are any assumptions or conditions violated? If so, explain in what ways.
b) What would you recommend about this regression?

14) A college admissions officer, defending the college's use of SAT scores in the admissions process, produced the graph below. It shows the mean GPAs for last year's freshmen, grouped by SAT scores. How strong is the evidence that *SAT Score* is a good predictor of *GPA*? What concerns you about the graph, the statistical methodology, or the conclusions reached?



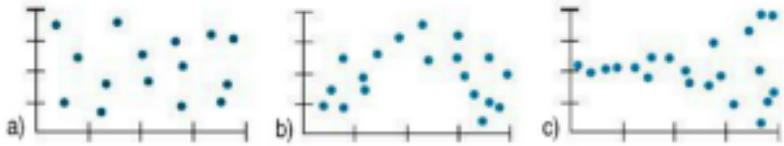
15) After keeping track of his heating expenses for several winters, a homeowner believe she can estimate the monthly cost (c) from the average daily Fahrenheit temperature (F) with the model $\hat{c} = 133 - 2.13F$. Here is the residuals plot for her data:



a) Interpret the slope and y-intercept of the line in this context.
b) During months when the temperature stays around freezing, would you expect cost predictions based on this model to be accurate, too low, or too high? Explain.

c) What heating cost does the model predict for a month that averages $10°$? How does this compare to the actual cost of a month averaging $10°$?
d) Should the homeowner use this model? Explain your reasoning.
e) Would this model be more successful if the temperature were expressed in degrees Celsius? Explain your reasoning.

16) Suppose you have fit a linear model to some data and now take a look at the residuals. For each of the following possible residual plots, tell whether you would try a re-expression and, if so, why.



17) Scientist Robert Boyle examined the relationship between the volume in which a gas is contained and the pressure in its container. He used a cylindrical container with a moveable top that could be raised or lowered to change the volume. He measured the *Height* in inches by counting equally paced marks on the cylinder, and measured the *Pressure* in inches of mercury (as in a barometer). Some of his data are listed in the table. Create an appropriate model, including statistical evidence of how it fits the data.

Height	48	44	40	36	32	28
Pressure	29.1	31.9	35.3	39.3	44.2	50.3
Height	24	20	18	16	14	12
Pressure	58.8	70.7	77.9	87.9	100.4	117.6