In order to receive extra credit you must do the following:
1. Complete the Part 2 Practice Test including all three FRQ.
2. Complete all or part of this set of FRQ. Without notes if possible. (You don't have to do all of them)
3. For any set of problems you complete, you must then correct those problems using a different color pen with the answers provided online.

Free Response Practice

2002 Form B Question 1

Animal-waste lagoons and spray fields near aquatic environments may significantly degrade water quality and endanger health. The National Atmospheric Deposition Program has monitored the atmospheric ammonia at swine farms since 1978. The data on the swine population size (in thousands) and atmospheric ammonia (in parts per million) for one decade are given below.

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Swine Population	0.38	0.50	0.60	0.75	0.95	1.20	1.40	1.65	1.80	1.85
Atmospheric Ammonia	0.13	0.21	0.29	0.22	0.19	0.26	0.36	0.37	0.33	0.38

(a) Construct a scatterplot of the data

(b) The value for the correlation coefficient for these data is 0.85. Interpret this value.

(c) Based on the scatterplot in part (a) and the value of the correlation coefficient in part (b), does it appear that the amount of atmospheric ammonia is linearly related to the swine population size?

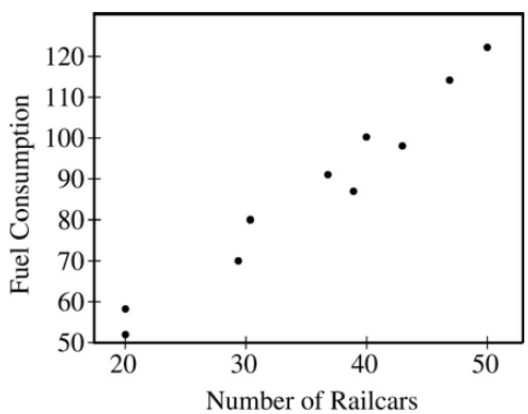
(d) What percent of the variability in atmospheric ammonia can be explained by swine population size?

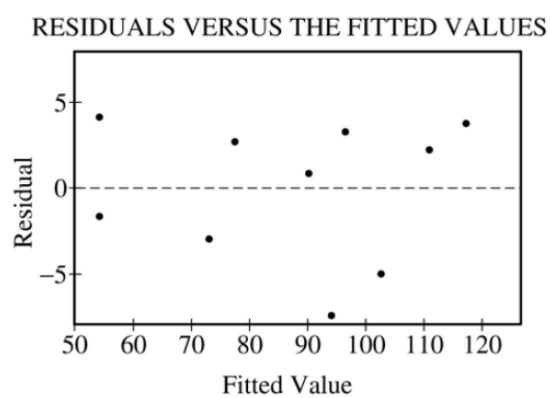
2005 Form A Question 3
The great Plains Railroad is interested in studying how fuel consumption is related to the number of railcars for its trains on a certain route between Oklahoma City and Omaha.

A random sample of 10 trains on this route has yielded the data in the table below.

Number of Railcars	Fuel Consumption (units/mile)
20	58
20	52
37	91
31	80
47	114
43	98
39	87
50	122
40	100
29	70

A scatterplot, a residual plot, and the output from the regression analysis for these data are shown below.





The regression equation is
Fuel Consumption = 10.7 + 2.15 Railcars

Predictor	Coef	StDev	T	P
Constant	10.677	5.157	2.07	0.072
Railcar	2.1495	0.1396	15.40	0.000

S = 4.361 R-Sq = 96.7% R-Sq(adj) = 96.3%

(a) Is a linear model appropriate for modeling these data? Clearly explain your reasoning.

(b) Suppose the fuel consumption cost is $25 per unit. Give a point estimate (single value) for the change in the average cost of fuel per mile for each additional railcar attached to a train. Show your work.

(c) Interpret the value of r^2 in the context of this problem.

(d) Would it be reasonable to use the fitted regression equation to predict the fuel consumption for a train on this route if the train had 65 railcars? Explain.

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.

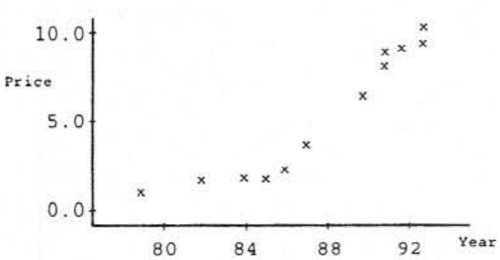
Percent of Section II grade—25

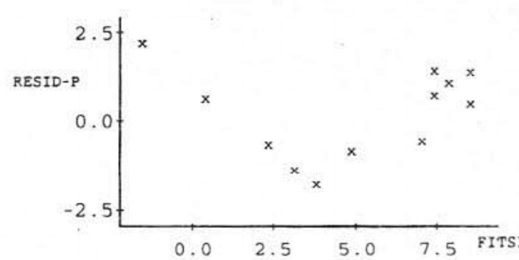
6. You are planning to sell a used 1988 automobile and want to establish an asking price that is competitive with that of other cars of the same make and model that are on the market. A review of newspaper advertisements for used cars yields the following data for 12 different cars of this make and model. You want to fit a least squares regression model to these data for use as a model in establishing the asking price for your car.

Production Year	1990	1991	1992	1987	1993	1991	1993	1985	1984	1982	1986	1979
Asking Price (in thousands of dollars)	6.0	7.7	8.8	3.4	9.8	8.4	8.9	1.5	1.6	1.4	2.0	1.0

The computer printouts for three different linear regression models are shown below. Model 1 fits the asking price as a function of the production year, Model 2 fits the natural logarithm of the asking price as a function of the production year, and Model 3 fits the square root of the asking price as a function of the production year. Each printout also includes a plot of the residuals from the linear model *versus* the fitted values, as well as additional descriptive data produced from the least squares procedure.

Model 1





The regression equation is
Price = -58.1 + 0.719 Year.

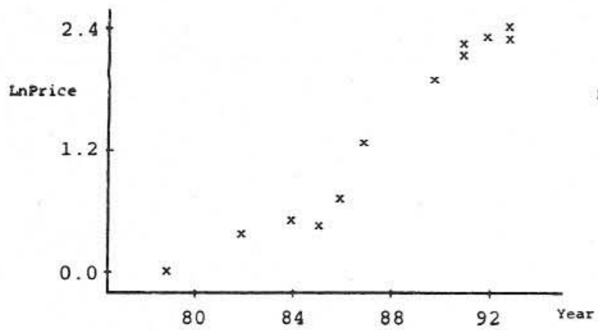
Predictor	Coef	Stdev	t-ratio	p
Constant	-58.050	7.205	-8.06	0.000
Year	0.71900	0.08200	8.77	0.000

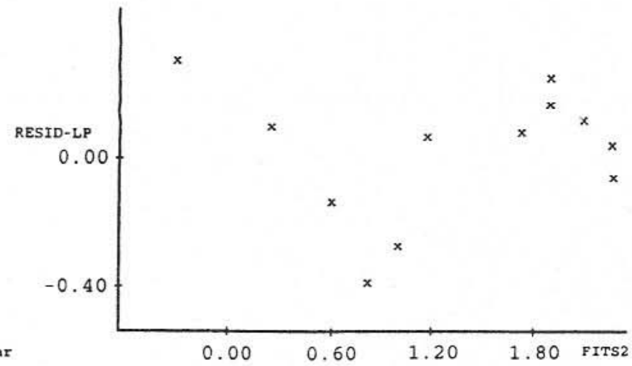
s = 1.255   R-sq = 88.5%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	121.10	121.10	76.88	0.000
Error	10	15.75	1.58		
Total	11	136.85			

Model 2





The regression equation is
LnPrice = -14.9 + 0.185 Year.

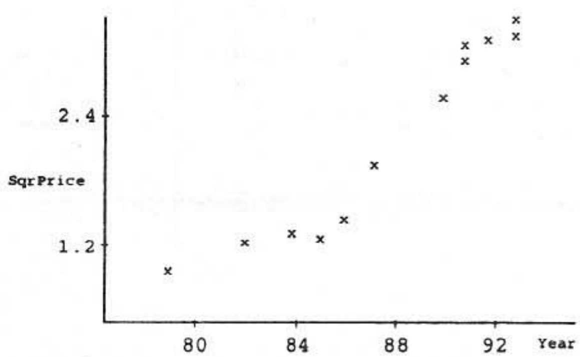
Predictor	Coef	Stdev	t-ratio	p
Constant	-14.924	1.223	-12.21	0.000
Year	0.18502	0.01392	13.30	0.000

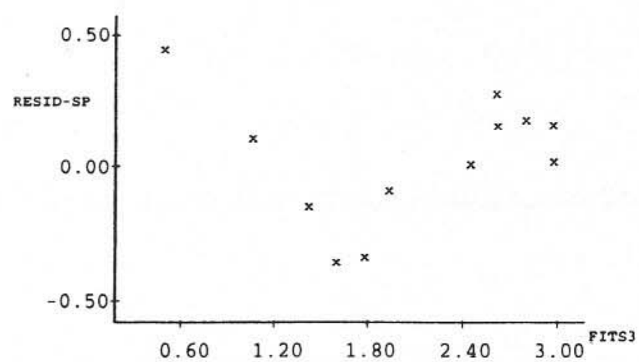
s = 0.2130 R-sq = 94.6%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	8.0190	8.0190	176.77	0.000
Error	10	0.4536	0.0454		
Total	11	8.4726			

Model 3





The regression equation is
SqrPrice = -13.3 + 0.176 Year.

Predictor	Coef	Stdev	t-ratio	p
Constant	-13.313	1.447	-9.20	0.000
Year	0.17559	0.01647	10.66	0.000

s = 0.2520 R-sq = 91.9%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	7.2221	7.2221	113.72	0.000
Error	10	0.6351	0.0635		
Total	11	7.8572			

(a) Use Model 1 to establish an asking price for your 1988 automobile.

(b) Use Model 2 to establish an asking price for your 1988 automobile.

(c) Use Model 3 to establish an asking price for your 1988 automobile.

(d) Describe any shortcomings you see in these three models.

(e) Use some or all of the given data to find a better method for establishing an asking price for your 1988 automobile. Explain why your method is better.

If you need more room for your work, use the space below.

END OF EXAMINATION

4.

The 1970 draft lottery involved matching birthdates with a number from 1 to 366. The lower the number, the more likely the individual with the matching birthday was to be drafted to fight in Vietnam. The average selection numbers by month are given in the following table.

MONTH	AVERAGE NUMBER
January	201.2
February	203.0
March	225.8
April	203.7
May	208.0
June	195.7
July	181.5
August	173.5
September	157.3
October	182.5
November	148.7
December	121.5

The following is part of the computer output for the least-squares regression line for predicting draft number from birth month (January = 1, February = 2, etc.).

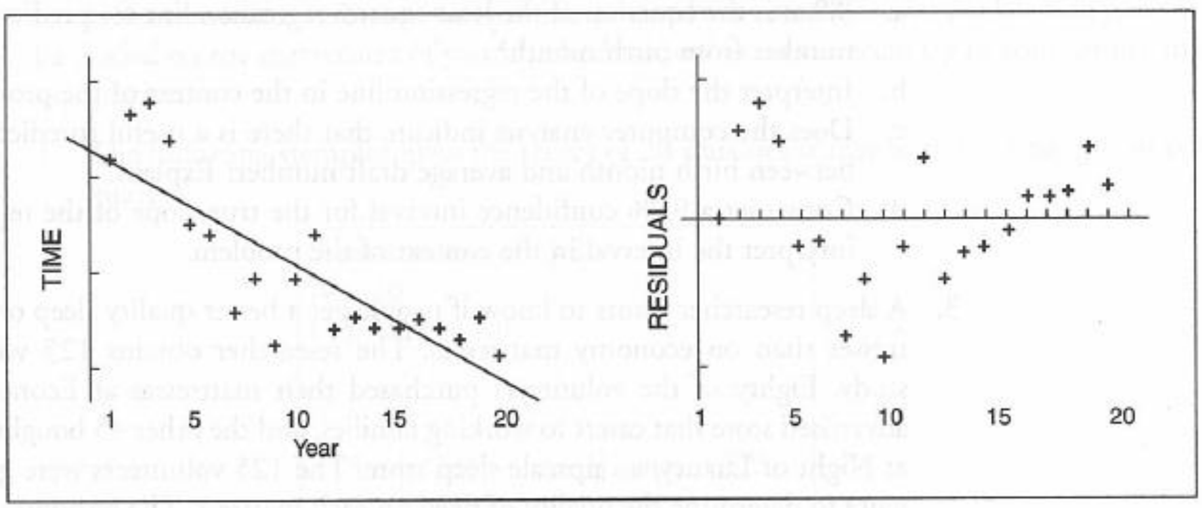
Predictor	Coef	St Dev	*t* ratio	*P*
Constant	229.402	9.466	24.23	.000
Month	−7.057	1.286	−5.49	.000

a. What is the equation of the least-squares regression line for predicting average draft number from birth month?

b. Interpret the slope of the regression line in the context of the problem.

5.

The graphs below give the times of the winner of the women's race at the Boston Marathon from 1975 to 1994. The graph on the left is the scatterplot of the year (numbered from 1 so that 1975 = 1, 1976 = 2, etc.) versus the times (which range from 2 hours 48 minutes down to 2 hours, 22 minutes). The graph on the right is a plot of the residuals versus the year. The equation of the regression line is *Time* = 163.1 − 1.23(*Year*), where the year is the number of years after 1975.



a. What would be the predicted time for the winner in 1999 (that would be year 25)?

b. The winner in 1999 actually ran the race in 143 minutes. What is the residual for this prediction?

c. Does a line appear to be a good model for the data? Explain.

d. If your goal was to predict the time for the winner in 1999, suggest an improvement to the situation described above but using the same data, or part of the same data.

6.

Scientists have suspected that animals, when deficient in certain chemicals, tend to ingest natural resources that have high concentrations of those chemicals to offset the deficiency. In a study, scientists drained saliva from the parotid gland of sheep in order to make them sodium deficient. Then they offered these sheep a solution of sodium bicarbonate and measured the sheep's sodium intake. The sodium deficiency and the sodium intake, both measured in millimoles, are recorded as follows:

Sodium Deficit (in millimoles)	Sodium Intake (in milimoles)
100	110
200	180
570	610
850	790
700	750
425	390
375	420
325	380
450	300
850	790

The summary statistics and the regression output for this data are as follows:

Variable	N	Mean	Median	StDev	Q1	Q3
Deficit	10	484.5	437.5	256.4	293.7	737.5
Intake	10	472.0	405.0	250.0	270.0	760.0

Predictor	Coef	StDev	T	P
Constant	15.55	47.94	0.32	0.754
Deficit	0.94211	0.08843	10.65	0.000

(a) Does a line appear to be a reasonable model for this data? Explain your answer.

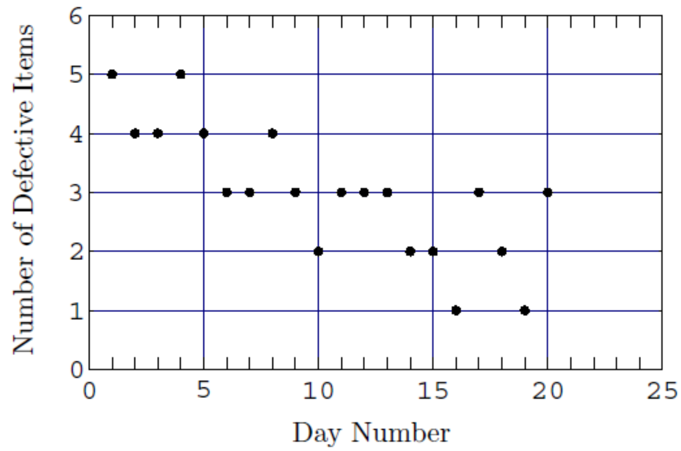
(b) State and interpret the slope in terms of the problem.

(c) Estimate the correlation between sodium deficit and sodium intake. Interpret your answer in the context of this problem.

(d) Estimate the amount of a sheep's sodium intake if the sheep is found to be deficient in sodium by 800 millimoles.

7.

A plot of the number of defective items produced during 20 consecutive days at a factory is shown below.



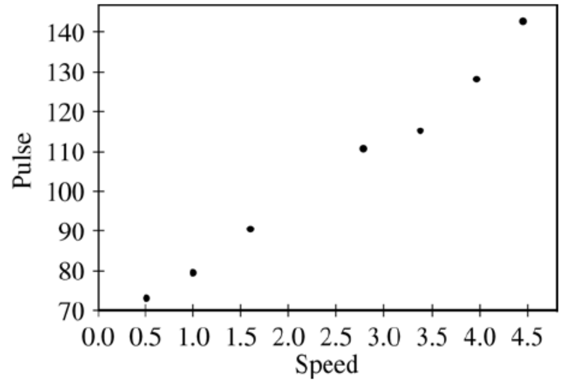
(a) Draw a histogram that shows the frequencies of the number of defective items.

(b) Give one fact that is obvious from the histogram but is not obvious from the scatterplot.

(c) Give one fact that is obvious from the scatterplot but is not obvious from the histogram.

8.

John believes that as he increases his walking speed, his pulse rate will increase. He wants to model this relationship. John records his pulse rate, in beats per minute (bpm), while walking at each of seven different speeds, in miles per hour (mph). A scatterplot and regression output are shown below.



Regression Analysis: Pulse Versus Speed

Predictor	Coef	SE Coef	T	P
Constant	63.457	2.387	26.58	0.000
Speed	16.2809	0.8192	19.88	0.000

S = 3.087 R-Sq = 98.7% R-Sq (adj) = 98.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	3763.2	3763.2	396.13	0.000
Residual	5	47.6	9.5		
Total	6	3810.9			

(a) Using the regression output, write the equation of the fitted regression line.

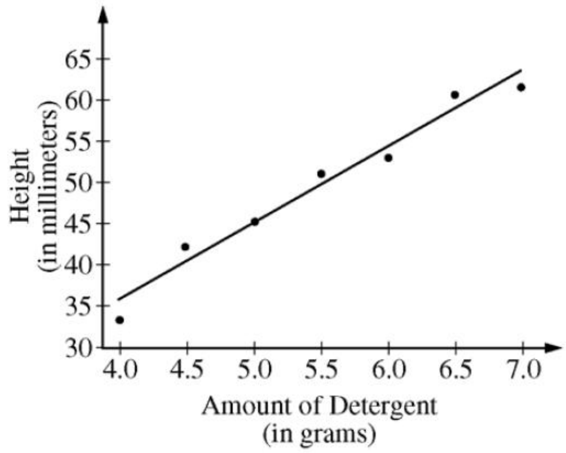
(b) Do your estimates of the slope and intercept parameters have meaningful interpretations in the context of this question? If so, provide interpretations in this context. If not, explain why not.

(c) John wants to provide a 98 percent confidence interval for the slope parameter in his final report. Compute the margin of error that John should use. Assume that conditions for inference are satisfied.

9.

A manufacturer of dish detergent believes the height of soapsuds in the dishpan depends on the amount of detergent used. A study of the suds' heights for a new dish detergent was conducted. Seven pans of water were prepared. All pans were of the same size and type and contained the same amount of water. The temperature of the water was the same for each pan. An amount of dish detergent was assigned at random to each pan, and that amount of detergent was added to the pan. Then the water in the dishpan was agitated for a set amount of time, and the height of the resulting suds was measured.

A plot of the data and the computer output from fitting a least squares regression line to the data are shown below.



Predictor	Coef	SE Coef	T	P
Constant	−2.679	4.222	−0.63	0.554
Amount	9.5000	0.7553	12.58	0.000

S = 1.99821 R–Sq = 96.9% R–Sq(adj) = 96.3%

(a) Write the equation of the fitted regression line. Define any variables used in this equation.

(b) Note that $s = 1.99821$ in the computer output. Interpret this value in the context of this study.

(c) Identify and interpret the standard error of the slope.

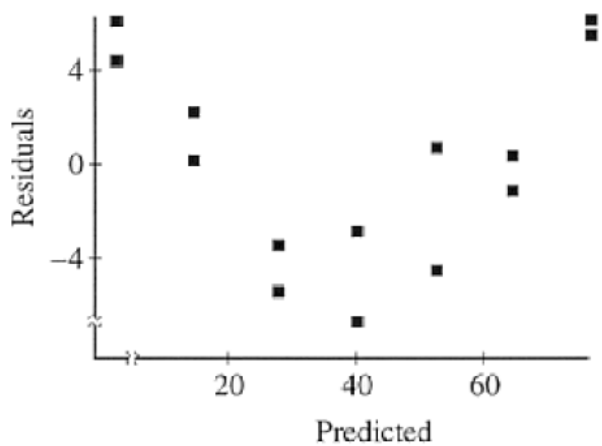
10.

In a study of the application of a certain type of weed killer, 14 fields containing large numbers of weeds were treated. The weed killer was prepared at seven different strengths by adding 1, 1.5, 2, 2.5, 3, 3.5, or 4 teaspoons to a gallon of water. Two randomly selected fields were treated with each strength of weed killer. After a few days, the percentage of weeds killed on each filed was measured. The computer output obtained from fitting a least squares regression line to the data is shown below. A plot of the residuals is provided as well.

Dependent variable is: percent killed
R squared = 97.2% R squared (adjusted) = 96.9%
$s = 4.505$ with $14 - 2 = 12$ degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	8330.16	1	8330.16	410
Residual	243.589	12	20.2990	

Variable	Coefficient	s.e. of Coeff	t-ratio	Prob
Constant	-20.5893	3.242	-6.35	≤ 0.0001
No. Teaspoons	24.3929	1.204	20.3	≤ 0.0001



(a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation.

(b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of weed killer are used, which of the following would you expect?

- ❍ The prediction will be too large.
- ❍ The prediction will be too small.
- ❍ A prediction cannot be made based on the information given on the computer output.

Explain your reasoning.