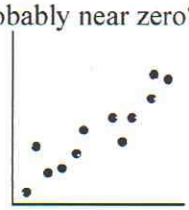
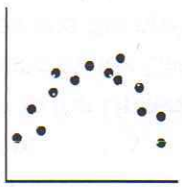
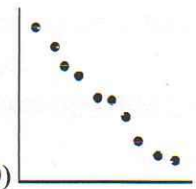
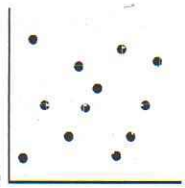
AP* Statistics Test A – Linear Regression ***Name*** _____

__ 1. Researchers studying growth patterns of children collect data on the heights of fathers and sons. The correlation between the fathers' heights and the heights of their 16 year-old sons is most likely to be . . .
A) near –1.0 B) near 0 C) near +0.7 D) exactly +1.0 E) somewhat greater than 1.0

__ 2. The auto insurance industry crashed some test vehicles into a cement barrier at speeds of 5 to 25 mph to investigate the amount of damage to the cars. They found a correlation of $r = 0.60$ between speed (MPH) and damage ($). If the speed at which a car hit the barrier is 1.5 standard deviations above the mean speed, we expect the damage to be __?__ the mean damage.
A) equal to B) 0.36 SD above C) 0.60 SD above D) 0.90 SD above E) 1.5 SD above

__ 3. Which scatterplot shows a strong association between two variables even though the correlation is probably near zero?

A)  B)  C)  D)  E) 

__ 4. The correlation between X and Y is $r = 0.35$. If we double each X value, decrease each Y by 0.20, and interchange the variables (put X on the Y-axis and vice versa), the new correlation
A) is 0.35 B) is 0.50 C) is 0.70 D) is 0.90 E) cannot be determined.

__ 5. A consumer group collected information on HDTVs . They created a linear model to estimate the cost of an HDTV (in $) based on the screen size (in inches). Which is the most likely value of the slope of the line of best fit?
A) 0.70 B) 7 C) 70 D) 700 E) 7000

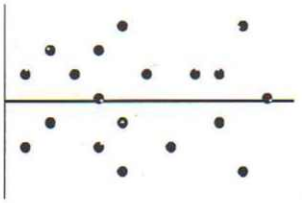
__ 6. The correlation between a family's weekly income and the amount they spend on restaurant meals is found to be $r = 0.30$. Which must be true?
I. Families tend to spend about 30% of their incomes in restaurants.
II. In general, the higher the income, the more the family spends in restaurants.
III. The line of best fit passes through 30% of the (*income*, *restaurant$*) data points.
A) I only B) II only C) III only D) II and III only E) I, II, and III

__ 7. A medical researcher finds that the more overweight a person is, the higher his pulse rate tends to be. In fact, the model suggests that 12-pound differences in weight are associated with differences in pulse rate of 4 beats per minute. Which is true?
I. The correlation between pulse rate and weight is 0.33
II. If you lose 6 pounds, your pulse rate will slow down 2 beats per minute.
III. A positive residual means a person's pulse rate is higher than the model predicts.
A) none B) I only C) II only D) III only E) II and III only

__ 8. Education research consistently shows that students from wealthier families tend to have higher SAT scores. The slope of the line that predicts *SAT score from family income* is 6.25 points per $1000, and the correlation between the variables is 0.48. Then the slope of the line that predicts *family income from SAT score* (in $1000 per point) …
A) is 0.037 B) is 0.16 C) is 3.00 D) is 6.25 E) is 13.02.

__ 9. A regression analysis of company profits and the amount of money the company spent on advertising found $r^2 = 0.72$. Which of these is true?
I. This model can correctly predict the profit for 72% of companies.
II. On average, about 72% of a company's profit results from advertising.
III. On average, companies spend about 72% of their profits on advertising.
A) none B) I only C) II only D) III only E) I and III

__ 10. A least squares line of regression has been fitted to a scatterplot; the model's residuals plot is shown. Which is true?



A) The linear model is appropriate.
B) The linear model is poor because some residuals are large.
C) The linear model is poor because the correlation is near 0.
D) A curved model would be better.
E) None of the above.

11. **Earning power** A college's job placement office collected data about students' GPAs and the salaries they earned in their first jobs after graduation. The mean GPA was 2.9 with a standard deviation of 0.4. Starting salaries had a mean of $47200 with a SD of $8500. The correlation between the two variables was $r = 0.72$. The association appeared to be linear in the scatterplot.

(*Show work*)

a. Write an equation of the model that can predict salary based on GPA.

b. Do you think these predictions will be reliable? Explain.

c. Your brother just graduated from that college with a GPA of 3.30. He tells you that based on this model the residual for his pay is -$1880. What salary is he earning?

12. **Assembly line** Your new job at *Panasony* is to do the final assembly of camcorders. As you learn how, you get faster. The company tells you that you will qualify for a raise if after 13 weeks your assembly time averages under 20 minutes. The data shows your average assembly time during each of your first 10 weeks.

Week	Time(min)
1	43
2	39
3	35
4	33
5	32
6	30
7	30
8	28
9	26
10	25

a. Which is the explanatory variable? _____

b. What is the correlation between these variables? _____

c. You want to predict whether or not you will qualify for that raise. Would it be appropriate to use a linear model? Explain.

13. **Associations** For each pair of variables, indicate what association you expect: positive(+), negative(-), curved(C), or none(N).

___	power level setting of a microwave	number of minutes it takes to boil water
___	number of days it rained in a month (during the summer)	number of times you mowed your lawn that month
___	number of hours a person has been up past a normal bedtime	number of minutes it takes the person to do a crossword puzzle
___	number of hockey games played in Minnesota during a week	sales of suntan lotion in Minnesota during that week
___	length of a student's hair	number of credits the student earned last year

14. **Music and grades** (True Story) A couple of years ago, a local newspaper published research results claiming a positive association between the number of years high school children had taken instrumental music lessons and their performances in school (GPA).

a. What does "positive association" mean in this context?

b. A group of parents then went to the School Board demanding more funding for music programs as a way to improve student chances for academic success in high school. As a statistician, do you agree or disagree with their reasoning? Explain briefly.

15. **Gas mileage again** In the *Data Desk* lab last week you analyzed the association between a car's fuel economy and its weight. Another important factor in the amount of gasoline a car uses is the size of the engine. Called "displacement", engine size measures the volume of the cylinders in cubic inches. The regression analysis is shown.

Dependent variable is: MPG
89 total cases of which 0 are missing
R squared = 60.9% R squared (adjusted) = 60.0%
s = 3.056 with 89 - 2 = 87 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	696.744	1	696.744	74.6
Residual	448.236	48	9.33826	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	34.9799	1.231	28.4	< 0.0001
Eng. Displcmt	-0.066196	0.0077	-8.64	< 0.0001

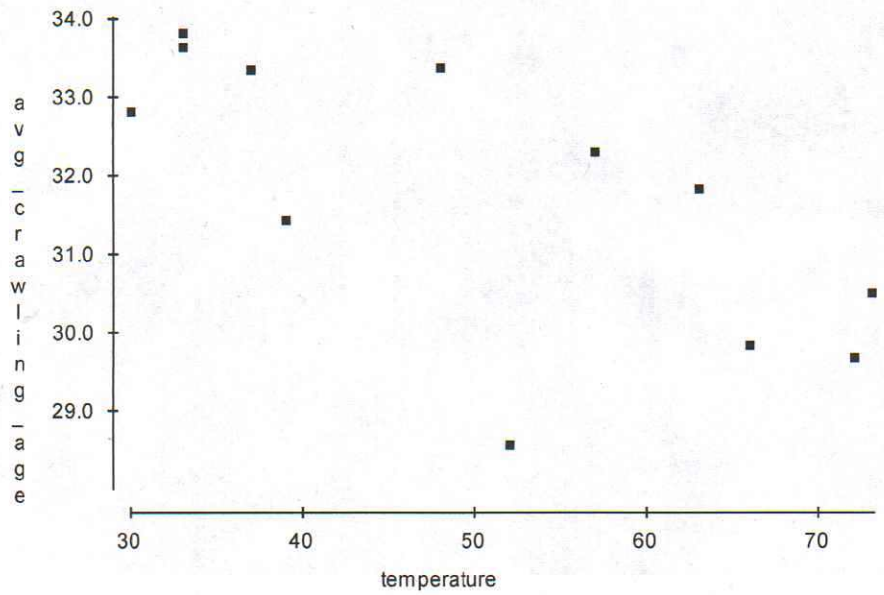
a. How many cars were included in this analysis?

b. What is the correlation between engine size and fuel economy?

c. A car you are thinking of buying is available with two different size engines, 190 cubic inches or 240 cubic inches. How much difference might this make in your gas mileage? (*Show work*)

16. **Crawling** Researchers at the University of Denver Infant Study Center investigated whether babies take longer to learn to crawl in cold months (when they are often bundled in clothes that restrict their movement) than in warmer months. The study sought an association between babies' first crawling age (in weeks) and the average temperature during the month they first try to crawl (about 6 months after birth). Between 1988 and 1991 parents reported the birth month and age at which their child was first able to creep or crawl a distance of four feet in one minute. Data were collected on 208 boys and 206 girls. The graph below plots average crawling ages (in weeks) against the mean temperatures when the babies were 6 months old. The researchers found a correlation of $r = -0.70$ and their line of best fit was

$$AvAge = 36 - 0.08AvTemp.$$



a. Draw the line of best fit on the graph. (Show your method clearly.)

b. Describe the association in context.

c. Explain (in context) what the slope of the line means.

d. Explain (in context) what the y-intercept of the line means.

e. Explain (in context) what R^2 means.

f. In this context, what does a negative residual indicate?