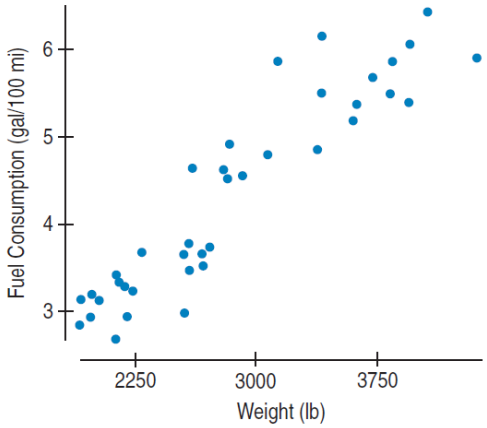
AP STATISTICS **Name: _____ Block: ____**
Review Chapters 26 – 27

1. Here is the scatterplot of weight (in pounds) versus fuel consumption (in gallons used per 100 miles driven) for a random sample of 38 automobiles. Describe the association between weight and fuel consumption for these automobiles. *(Remember: Describe strength, form, direction, and use context!)*



2. In the previous problem, suppose that the data for a hybrid automobile was added with a point at (4000, 3.1).

   a) How would the addition of this point affect the slope of the least squares regression line?

   b) How would the addition of this point affect the correlation coefficient?

3. The linear model for the association between fat grams and protein grams in Burger King's menu items is

$$\widehat{fat\ grams} = 6.8 + 0.97(protein\ grams)$$

   What is the estimated *increase* in fat grams that corresponds to an *increase* of 10 grams of protein?

4. Suppose that you were asked to analyze each of the situations described below. **(Note: Do not do these problems!)** For each, indicate which inference procedure you would use (from the list), the test statistic (*z, t,* or χ^2), and number of degrees of freedom (if appropriate).

	Type	*z, t,* or χ^2	df
a.			
b.			
c.			
d.			
e.			
f.			
g.			
h.			

1. proportions, 1 sample
2. proportions, 2 samples
3. means, 1 sample
4. means, 2 samples
5. means, matched pairs
6. goodness of fit
7. homogeneity
8. independence
9. regression, inference for β

a) "Parents should spank their children when they think it is necessary", said 52% of the adult respondents to a survey – though most child-development experts say spanking is not appropriate. The survey of 7225 adults was co-sponsored by *Working Mother* and Epcot Center at Disney World" (*Rochester Democrat and Chronicle*, 1990). Find the 95% confidence interval for the parameter representing the nation as a whole.

b) A random sample of 99 people was asked to indicate their favorite choice of sport among baseball, basketball, football, and soccer. The responses were also separated into three age groups – under 20, between 20 and 40, and over 40. Does this data provide evidence of an association between type of sport and age?

c) A marine biologist collects data on the median adult lengths and weights for 35 randomly selected marine species. Does the data provide evidence that the lengths of these marine species can be used to predict their weight?

d) The manager of a factory wants to compare the mean number of units assembled per employee in a week for two new assembly techniques. Two hundred employees from the factory are randomly selected and each is randomly assigned to one of the two techniques. After teaching 100 employees one technique and 100 employees the other technique, the manager records the number of units each of the employees assembles in one week. Estimate the difference in mean times for the two assembly techniques.

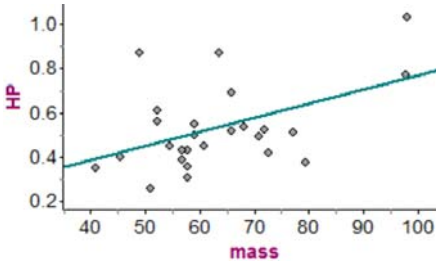
e) Suppose a coin purchaser took random samples of coins that were minted during 5 different decades. The physical condition of each of the coins is recorded as either an "A" (excellent), "B" (very good), or "C" (fair). Do the data present evidence of a difference in physical conditions of the coins between the decades?

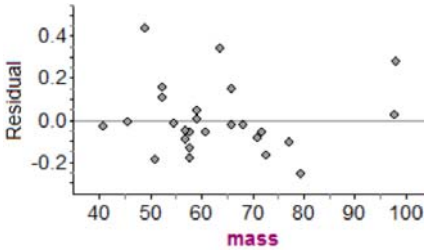
f) 22 people complaining of indigestion take an antacid. They report that their discomfort subsided in an average of 13 minutes; the standard deviation was 4 minutes. The manufacturer wants a 95% confidence interval for the "relief time."

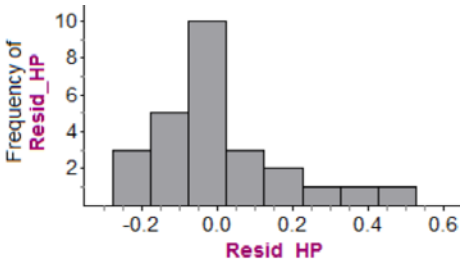
g) Twelve automobiles were selected at random to test two new mixtures of unleaded gasoline. Each car was given a measured allotment of the first mixture, *x*, and driven; then the distance traveled was recorded. The second mixture, *y*, was tested in the same manner. The order in which the *x* and *y* mixtures were tested was also randomly assigned. Can you conclude that there is a difference in mileage obtained by these two gasoline mixtures?

h) An SRS of 101 postal employees found that the average time these employees had worked for the postal service was 7 years with standard deviation 2 years. Assume the distribution of the time the population of employees has worked for the postal service is approximately normal with mean μ. Is this data evidence that μ has changed from the value of 7.5 years of 20 years ago?

5. During a "horsepower lab" activity, a group of twenty six AP Statistics students ran up a flight of stairs and measured their estimated horsepower, based on their mass (and also the time required to make it up the stairs). The information from the regression analysis is shown below. The response variable in this analysis is "horsepower".







Predictor	**Coefficient**	**S.E. of Coeff**	**T-ratio**	**Probability**
Mass (kg)	0.0064	0.0024	2.658	0.0138
Constant	0.1224	0.1551	0.789	0.4377
S = 0.167	R-sq = 22.7%	R-sq adj = 19.5%		

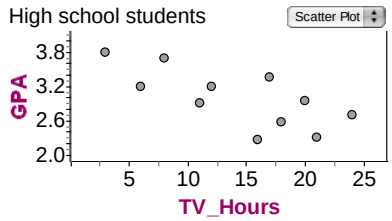
a) In the computer output, the coefficient for "Constant" is 0.1224. Interpret this value in context.

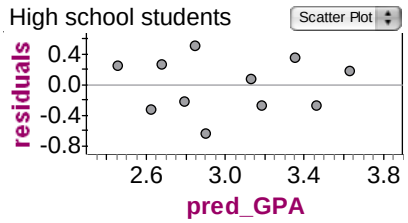
b) At the 5% significance level, is there evidence of a linear association between mass and horsepower? Be sure to check the necessary conditions for inference.

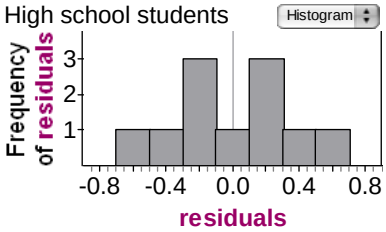
c) What is the proportion of the variation in horsepower that can be explained by the person's mass?

d) Interpret the meaning of the value "s = 0.167".

6. A random sample of 11 high school students produced the following results for number of hours of television watched per week and GPA.







Predictor	**Coef**	**StdDev**	**T**	**P**
Constant	3.8	2.0426	1.86	0.05
Hours	-0.0558	0.01769	-3.154	0.012

R-sq = 53% s = 0.355

a) State the least squares regression equation in context.

b) What is the correlation coefficient for this regression analysis?

c) Are the conditions for inference met?

d) Is there evidence of a relationship between the number of hours of TV a high schooler watches and his/her GPA? Test an appropriate hypothesis and state your conclusion in proper context. At this point, assume that all conditions for inference have been met.

e) Construct a 95% confidence interval to estimate the slope of the least squares regression line.

7. A food delivery company in a large city charges customers a flat fee to for food delivery plus an additional fee per mile. There is also a charge for the amount of time that the courier spends in traffic. The company wants to develop a new method for determining fares based on mileage and a flat fee only, not on time spent in traffic. A random sample of 10 recent delivery fees was selected, and the distance (in miles) and the total fee (in dollars) were recorded. A regression model was fit to the data, and the output is given below.

Variable	**Coefficient**	**Std. Err.**	**T-stat**	**Prob.**
Intercept	4.290	0.298	14.400	<0.0001
Mileage	1.229	0.166	7.418	<0.0001

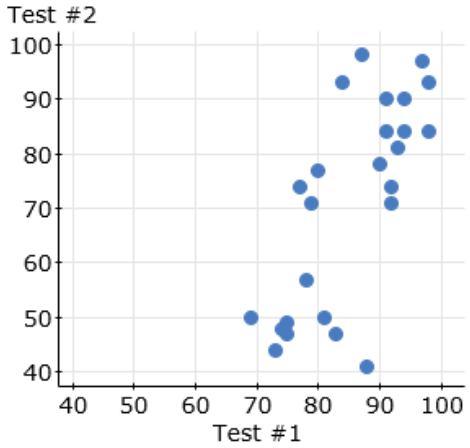
a) Construct and interpret a 95% confidence interval for the slope of the least-squares regression line. Assume that the conditions for inference are met.

b) A recent social media post claims that the cost of food delivery this large city comes out to about $1.50 per mile travelled. Based on your interval in part (a), is this claim of $1.50 per mile a reasonable value? Explain.

c) The company is considering setting their per mile delivery fee at $1.87. Based on your interval in part (a), would this be enough for the company to increase their average revenue per mile? Explain.

d) One of the recent deliveries used in the regression analysis had a distance of 2.13 miles, and the residual for the delivery fee was –$1.45. Calculate the actual fee for this delivery.

8. The graph below shows the test scores for a class of Statistics students on two different exams. Which of the following statements is/are true?



I. Every student performed better on test #1 than test #2.

II. Every student performed better on test #2 than test #1.

III. Most students – but not all – performed better on test #1 than on test #2.

IV. Most students – but not all – performed better on test #2 than on test #1.

V. The association between the scores on the two exams is negative.

VI. There is more variability in the distribution of scores for test #2 than for test #1.

9. Based on data collected from a random sample of 27 students at a large university, a 95 percent confidence interval for the slope of the least squares regression line is determined to be (–1.417, 1.213). Which of the following statements is/are true?

I. The slope of the sample data is negative.

II. The correlation coefficient of the data is negative.

III. There is a strong association between the x and y variables in the scatterplot.

IV. The r-squared value of the data is negative.

V. When conducting a t-test to determine if there is a linear association between x and y, the p-value of the test will be greater than 5%.

10. _____ An honors business program at the University of North Podunk admits twenty students each semester, and requires each student to purchase 2 items: a smartphone and a laptop computer. The students are required to buy a separate smartphone and laptop for this program (even if they already have their own), but are given up to $2000 to shop for whatever type of smartphone and laptop that they want. For example, if a student chooses to spend $700 on a smartphone, then they have up to $1300 to spend on a laptop computer. However, students are <u>not</u> allowed to keep any unused funds, and therefore most students will spend very close to (but never more than) the full $2000. For these 20 students, let "S" represent the amount of money spent on a smartphone and "L" represent the amount of money spent on a laptop. Which of the following statements is true regarding the relationship between these two variables?

A) "S" and "L" will have a correlation of 0

B) "S" and "L" will have a negative correlation

C) "S" and "L" will have a positive correlation that is less than 1

D) "S" and "L" will have a correlation of 1

E) The relationship between "S" and "L" cannot be determined from the information given

AP Statistics
Unit X Review – Inference with Regression Slope

1. The association between weight and fuel consumption for these automobiles is...
   - strong (or moderately strong),
   - linear (or fairly linear), and
   - positive – as weight of the automobile increases, so does fuel consumption.

2. a) Slope would decrease (more specifically, it become less positive)
   b) Correlation would weaken (decrease, since it would become less positive)

3. 9.7 grams

4.

	Type	*z, t,* or χ^2	df
a.	1	z	
b.	8	chi-sq	6
c.	9	t	33
d.	4	t	calc
e.	7	chi-sq	8
f.	3	t	21
g.	5	t	11
h.	3	t	100

5. a) A person with zero mass is predicted to generate 0.1224 horsepower.
   b) t-test for regression slope, df = 26 – 2 = 24
   Ho: There is NO linear association between mass and horsepower
   Ha: There IS a linear association between mass and horsepower

   Conditions:
   - Random sample condition: We must treat this group of students as a representative sample of students
   - The scatterplot for HP vs mass is *fairly* linear, and there is no clear pattern in the plot of residuals vs mass
   - The scatterplot of HP vs mass maintains roughly equal variance in HP.
   - The histogram of residuals is slightly skewed right, but there are no outliers. Normality for the population of residuals should still be reasonable.

   $t = \frac{b_1 - 0}{s_{b_1}} = \frac{0.0064}{0.0024} \approx 2.67$ (according to the printout, $t = 2.658$),

   p-value = 0.0138 (according to the printout)
   Since the p-value < alpha, we reject Ho.
   We have evidence of a linear association between mass and horsepower.
   c) r-squared = 0.227
   d) 0.167 is the typical difference between observed and predicted horsepower.

6. a) $\widehat{GPA} = 3.8 - 0.0558(hours)$
   b) r = -0.728 (remember: if slope is negative, so must the correlation coefficient be)
   c) Conditions are: 1) Random sample 2) Plot is linear 3) Equal variance in "y" 4) Near normality in the histogram of residuals
   d) t-test for regression slope (df = 11 – 2 = 9).
   [Check #5b for how to write the hypotheses]
   t = -3.1543, p-value = 0.01166…
   e) (-0.0958, -0.0158)

7. a) (0.846, 1.612)
   Interpretation of the interval: We are 95% confident that the <u>mean</u> increase in fee for each increase of 1 mile in distance is between \$0.846 and \$1.612

   b) Yes. \$1.50 is a value that is contained in our confidence interval.
   c) Yes, as \$1.87 is ABOVE the interval for the average per-mile fee.

8. Statements III and VI are true.
9. Statements I, II, and V are true.
10. B