MATH 105-Test #3b.

1. A history instructor has given the same pretest and the same final examination each semester. He is interested in determining if there is a relationship between the scores of the two tests. He computes the linear correlation coefficient and notes that it is **1.9**. What does this correlation coefficient tells us?
a) There is no association between the tests.
b) There is a strong positive association between the tests.
c) There is a strong negative association between the tests.
d) The history instructor has made an error in calculating the correlation coefficient.

2. The least-squares regression line
a) Minimizes the sum of the predicted values
b) Maximizes the sum of the predicted values
c) Minimizes the sum of the residuals squared
d) Maximizes the sum of the residuals squared

3. A researcher determines that the linear correlation coefficient is .85 for paired data. This indicates that there is
a) No linear correlation but that there may be some other relationship
b) A strong positive linear correlation
c) Insufficient evidence to make any decision about the correlation of the data
d) A strong negative correlation

4. An instructor wishes to determine if there is a relationship between the number of absences from his class and a student's final grade in the course. What is the response variable?
a) Final Grade b) Absences c) Instructor

5. A traffic officer is compiling information about the relationship between the hour of the day and the speed over the speed limit at which the motorist is ticketed. He computes a correlation coefficient of .12. What does this tell the officer?
a) There is a weak positive linear correlation
b) There is a moderate negative linear correlation
c) There is a strong positive linear correlation

6. For a particular regression analysis, the following regression equation is obtained: $\hat{y} = 8.3x + 32$, where x represents the number of hours studied for a test and y represents the score on the test. True or False? If R^2 is 94%, the number of hours studied is very useful for predicting the test score.
a) True b) False

7. Classified ads in the *Ithaca Journal* offered several used Toyota Corollas for sale. A least squares regression was performed to predict the price of the Corolla based on the age of the car. We would expect that the association between age and price of a used Corolla to be (circle one):
Positive Negative
Explain:
As the age of a car increases we expect the price to decrease.

8. Coffee is a leading export from several developing countries. When coffee prices are high, farmers often clear forests to plant more coffee trees. Here are the data for five years on prices paid to coffee growers in Indonesia and the rate of deforestation in a national park that lies in a coffee-producing region.

Price (per lb)	29	40	54	55	72
Deforestation (%)	.49	1.59	1.69	1.82	3.10

The regression equation is $\hat{y} = -.976 + .054x$

a) Calculate the residual for the year that $55 dollars per pound was paid to coffee growers.

$\hat{y} = -.976 + .054 \times 55 = 1.994\%$ Resid $= y - \hat{y} = 1.82 - 1.994 = -0.174\%$

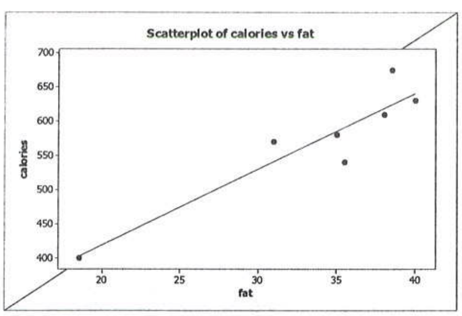
b) If you are an environmentalist and oppose deforestation, would you prefer for an observation to have a negative or positive residual. Explain.

Negative because there would be less deforestation than predicted by the line.

9. A researcher gathers data from a sample of households on the number of cars owned by the household and the total household income. The association appears to be linear and the least-squares line for predicting number of cars from household income turns out to be $\hat{y} = .4952 + .000042x$. The correlation coefficient (r) between cars and income is .75. What percent of the variation in the number of cars owned is explained by the least-squares regression line?

$.75^2 = .5625 = R^2$ 56.25% of the variation in # of cars owned (y) is explained by the regression line.

10. Fast food is often considered unhealthy because much of it is high in both fat and calories. The scatter diagram and regression equation below show the association between fat content (grams) and calories for several brand of burgers. The least-squares regression equation along with the scatter diagram appear below.



$\hat{y} = 197.7 + 11.08x$

a) What is the predicted number of calories for a sandwich that has 30 grams of fat?

$\hat{y} = 197.7 + 11.08 \times 30 = 530.1$ cal.

b) Give a one sentence interpretation of the slope in context of the problem.

For every additional gram of fat the predicted number of calories increase by 11.08.

11. Recall that we collected data in the beginning of the semester on travel time to campus and travel mode (bike, walk, car, etc). Suppose that I want to compare the average travel time for students who bike and students who drive to campus. I decide to calculate a 95 % confidence interval for the difference between bike travel time and car travel time ($\mu_{bike}-\mu_{car}$). The StatCrunch output appears below.

95% confidence interval results:
μ_1 : mean of population 1
μ_2 : mean of population 2
$\mu_1 - \mu_2$: mean difference
(with pooled variances)

Difference	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-9.85	0.80509675	1151	-11.429622	-8.270378

a) In the context of the problem, interpret the 95% confidence interval (be sure to specify the direction of the difference). We can be 95% confident that the average commute time for students who bike to campus is between 8.3 & 11.5 minutes shorter than students who drive.

b) If I were to conduct the following hypothesis test: H_0: $\mu_{bike}=\mu_{car}$ versus H_1: $\mu_{bike}\neq\mu_{car}$. Would the p-value be larger or smaller than 0.05. Explain.

We would reject H_0 since the confidence interval shows a difference between commute times.

12. Political pundits talk about the "bounce" that a presidential candidate gets after his party's convention. In the last forty years, it has averaged about 6 percentage points. Just before the 2004 Democratic convention, Rasmussen Reports polled 1500 likely voters at random and found that 705 favored John Kerry. Just afterward they took another random sample of 1500 likely voters and found that 735 favored Kerry. Use a .05 significance level to test the claim that there was a bounce after the convention.

a) State the null and alternative hypotheses. $H_0: P_b = P_a$ $H_1: P_b < P_a$

b) The p-value for the hypothesis test was 0.1365. State a conclusion in context of the problem.

We can not reject H_0 therefore there is no evidence that John Kerry got a "bounce" after the convention.

c) A confidence interval for the difference between before and after was (-0.056, 0.0158). Does your confidence interval agree with the results of the hypothesis test? Explain.

Yes, since the confidence interval includes 0 we have no evidence that the proportions are different which is what we included in our hypothesis test.

13. The USStates data file contains information from data collected on each state. Two of the variables recorded were average IQ score and the percent of residents with a college education. I decided to run a regression to predict state's average IQ score from the percentage of college educated residents. The regression output appears below:

Simple linear regression results:
Dependent Variable: IQ
Independent Variable: College
IQ = 93.0025 + 0.2356217 College
Sample size: 50
R (correlation coefficient) = 0.4635
R-sq = 0.21484086
Estimate of error standard deviation: 2.424381

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	93.0025	2.0545542	$\neq 0$	48	45.26651	<0.0001
Slope	0.2356217	0.06501523	$\neq 0$	48	3.6241002	0.0007

a) Which variable is the explanatory variable and which variable is the response variable?

% college residents

IQ

b) Report the regression equation.

$\hat{y} = 93.0 + 0.236x$ (or $\hat{IQ} = 93.0 + 0.236$ college)

c) California has an average IQ score of 96 and 31.1% of its residents are college educated. Compute the residual.

$\hat{y} = 93 + 0.236 \times 31.1 = 100.34$ Resid $= 96 - 100.34 = -4.34$

d) In context of the problem, interpret the residual you computed in c).

The predicted average IQ for CA is 4.34 points more than the actual IQ score for CA.

e) Explain why we can't conclude that having more college educated residents **causes** higher average IQ scores for a state.

We can not conclude a cause & effect association between 2 variables because there may be lurking variables.

14. The two-way table below shows survey results which asked students if they had the choice of winning an Academy Award, a Nobel Prize, or an Olympic gold medal which they would choose. The StatCrunch output appear below. The expected values appears in the parentheses.

Contingency table results:

	Academy	Nobel	Olympic	Total
Female	20 (14.47)	76 (69.56)	73 (84.97)	169
Male	11 (16.53)	73 (79.44)	109 (97.03)	193
Total	31	149	182	362

Chi-Square test for independence:

Measure	DF	Value	P-value
Chi-Square	2	8.23924	0.0163

a) State the null and alternative hypotheses.

H_0: Gender & Award are independent
H_1: " " " dependent

b) Using a significance level of 0.05, give a conclusion to the hypothesis test.

We can reject Ho & conclude that a students choice of award depends on gender.

c) In the table below, calculate the residuals, observed minus expected values (O-E).

	Academy	Nobel	Olympic
Female	5.53	6.44	-11.97
Male	-5.53	-6.44	11.97

11-16.53=-5.53

d) Write a few sentences summarizing the results in c).

More males than expected chose an olympic medal while more females than expected picked an Academy award or nobel prize.