Name _____ **Statistics Exam Topics 7 - 11**

St Algebra College has a problem with attrition. A sizable number of non-graduating students do not return to classes each semester. Is attrition importantly related to race? There is a table below illustrating the attrition/retention of college students based on race. – Answers are Rounded

	Race			
Attrition	**White**	**Black**	**Other**	
Returned in Spring	260 / 256	80 / 79	55 / 60	***395***
Did not Return	85 / 89	26 / 27	26 / 21	***137***
	345	***106***	***81***	***532***

(a) Which variable would consider to be the explanatory variable, and which the response?

Race: Explanatory;
Attrition: Response

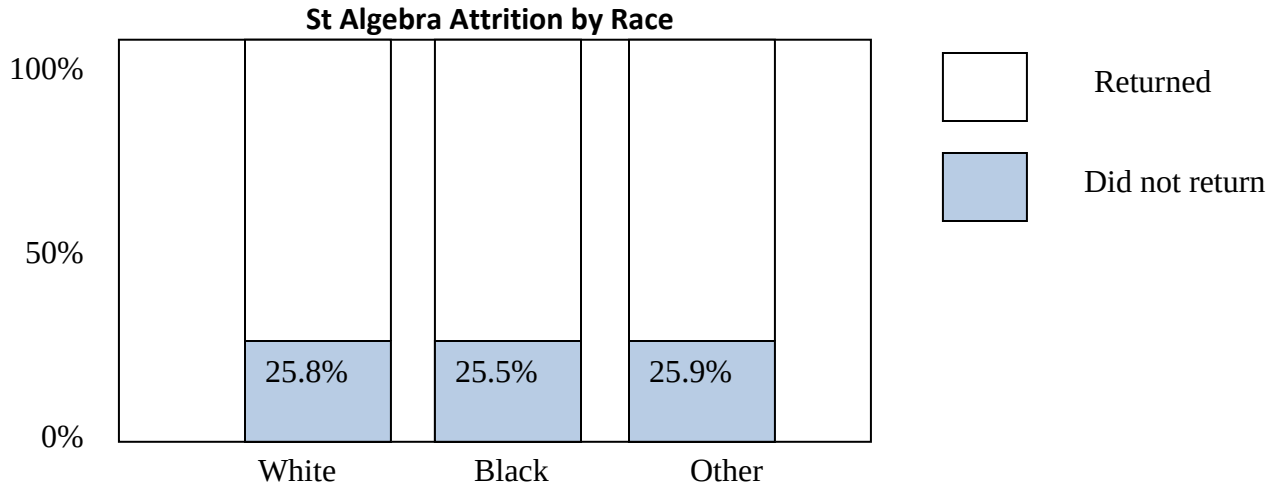
(b) In the table above fill in the marginals and expected values, if the variables are independent

(c) What proportion of Blacks did not returned in the spring? $^{26}/_{106} = .245$

(d) What proportion of those that returned in the spring were white? $^{260}/_{395} = .658$

(d) What proportion of students did not return in the spring? : $^{137}/_{532} = .258$

(e) Make a Segmented Bar graph of the table.



(f) Does St. Algebra College need to develop a plan to retain minority students (Explain using information from a and b to back your explanation)?

It appears that percentages are even across the board, therefore the attrition rate is independent of Race so targeting minorities would be inappropriate, overall attrition should be addressed.

In the following situations, identify the variables, observational units and whether the variable is the explanatory or response variable.

(a) How much income tax a person pays and their years of education.

Obs Unit: Person; Exp: Yrs of Ed; Resp: Income Tax

(b) A person's age and their blood pressure.

Obs Unit: Person; Exp: Age; Resp: Blood Pressure

(c) Annual per capita income for a different nations and the infant mortality rate for that nation.

Obs Unit: Nation; Exp: per Capita Inc; Resp: Infant Mortality

_____1) All of the following are possible values for ***r*** except
a) 1.0
b) -.4341
c) 1.001
d) 0

_____2) Which of the following shows weakest linear association:
a) -.998
b) -.512
c) .023
d) .401

_____3) A pair of variables is found to have a correlation coefficient of .456, based on this number you could say:
a) There is a strong positive linear association between these two variables
b) The association is non-existent since r is so low.
c) The association appears to be weak, but a scatterplot may show a strong non-linear association.
d) There could only be a linear association between these two variables since they it a bivariate set.

_____4) When two variables are said to be related in some way, there is a(n) _____ between the them.
a) variability
b) causation
c) confidence
d) association

_____5) The measure of the degree and direction in which two variables is related is called the _____.
a) mean
b) standard deviation
c) proportion of variability
d) correlation coefficient

_____6) The symbol used to represent the measure indicated in correlation coefficient is_____.
a) x-bar
b) r^2
c) r
d) S_x

_____7) If we square correlation coefficient, we obtain a measure of _____ in the amount the independent variable affects the dependent variable.
a) Confidence
b) Deviation
c) Direction
d) Association

_____8) If you can successfully predict the value of the response variable by knowing the measure of the explanatory variable then there must be causation involved.
a) True
b) False

_____9) If there is a significant statistical tendency between two variable, then we know all pairs of data in the data set follow this tendency.
a) True
b) False

_____10) If a bivariate set of data has an r = -.4, all of the following are true except
a) The slope of the regression line goes up going from left to right.
b) The association is fairly weak.
c) The proportion of variability is .16.
d) We cannot claim causation exists between these two variables

_____11) If a bivariate set of data has an r^2 = .95, which of the following must be true
a) The regression line has a positive slope
b) There exists a strong association between these two variables
c) r is approximately .9025
d) the scatterplot is randomly scattered over a large area.

_____12) If the correlation coefficient between two variables is close to zero then there can be no relationship whatsoever between the two variables.
a) True
b) False

Answers 1 – 12 MC
1) C 2) C 3) C 4) D 5) D 6) C 7)A 8) b 9) B 10) A 11) B 12) B

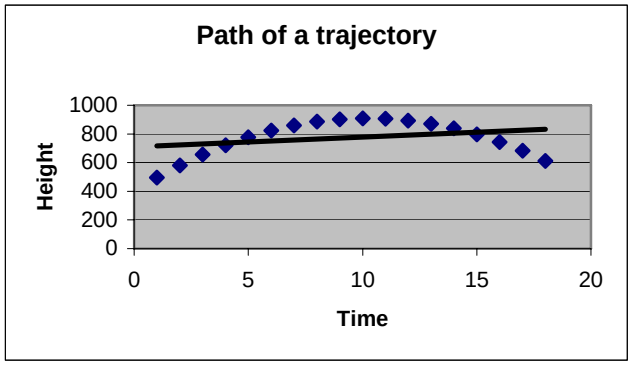
_____13) The following Scatterplot relates height to time in the path of a trajectory. For this pair of data r = .288.

a) Looking strictly at the correlation coefficient, what would you say about the strength and direction of this association?

Weak Positive Linear Association

b) Looking at the graph, what would you say about the association of these two variables, and what should be done?

A transformation should be done to make the variables form a straight line so a Linear Regression could be done.



The *Archaeopteryx* is an extinct beast having feathers like a bird but also teeth and a long bony tail like a reptile. They have only have five fossil specimens to study. The following are femur lengths and Humerus lengths of the five fossils:

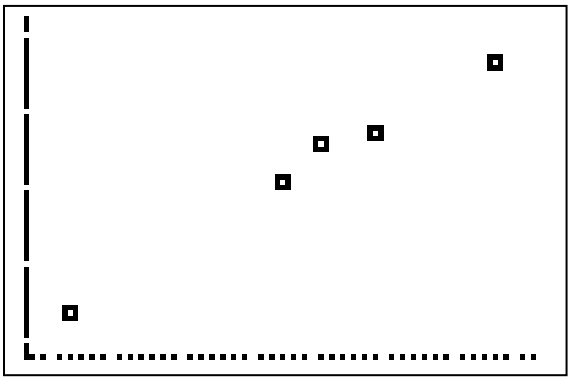
Specimen	**1**	**2**	**3**	**4**	**5**
Femur	38	56	59	64	74
Humerus	41	63	70	72	84

Use your calculator to find the following values [3 points]:

	Mean	**St Dev**	**r**
Femur	58.2	13.2	.994
Humerus	66	15.9	

a. Create a scatterplot of the data and comment on the plot in the box to the right.

Scale and Label properly →



b. ___R^2= 98.8%_____What is the proportion of variability between these two variables?

c. Using your values for mean, standard deviation, and Correlation coefficient and the formulas find the equation for predicting the humerus length from the femur length. You must show your work for credit! (You may use your calculator to check your work but I need to see the work!)

Humerus-***hat*** = -3.7 +1.2 (femur)

1.2 = r * $S_{humerus}/S_{femur}$ -3.7 = y-bar – 1.2 (x-bar) From data table above.

d. _____Suppose a fossil was recently found that had a femur length of 42cm. Based on your equation, predict the length of the recently found fossils humerus.

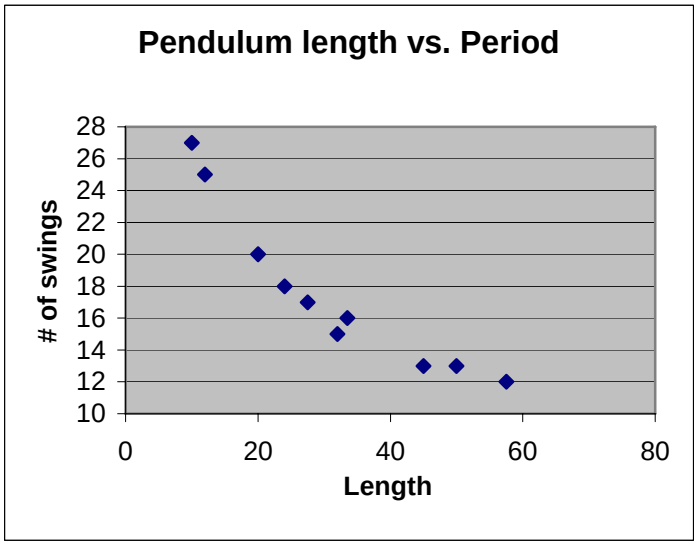
46.7 = -3.7 + 1.2(42)

e. _____ Suppose the actual humerus length is 56cm. What is the value of the residual?

Residual = 56 – 46.7 = 9.3

f. **True/False** If we reverse our perspective, that is, we want to predict the femur length based on the length of the humerus we can simply use the same regression equation as was used above (algebraically solved for femur). False

The following Scatterplot and data came from one of your classmates for the pendulum experiment. The linear regression for this shown with the r^2.



Pendulum	
Length	# of Swings
33.5	16
20	20
57.5	12
12	25
32	15
45	13
24	18
27.5	17
10	27
50	13

$y = -0.2986x + 26.9$
$r^2 = 0.8632$

a) Why is a transformation necessary and what type of transformation would you use?

Since the data asymptotically approaches the x and y axis the power regression model seems to fit. Thus I will take the log of both sides of x an y

b) Perform the transformation you indicated– you may use the other regressions to check your work but not as an answer.

PwrReg y=a*x^b a=80.33172786 b=-.4703409992 r²=.9918591524 r=-.9959212582	LinReg y=a+bx a=1.904887108 b=-.4703409992 r²=.9918591524 r=-.9959212582
This Regression was done with L1 and L2 original data	This Regression was done on with the transformation: X = log(L1) and Y = log(L2)

c) Did the transformation improve the linearity, (*explain*). The R^2 is very close to 1, it appears to have straightened the data quite a bit.

NOTE:in MINITAB a log10 on both variables would have accomplished the same thing or transforming the columns of data as I showed in the LinReg example above.