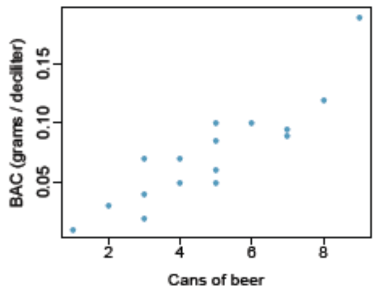
Math 183 HW17, Name: _____________________________, PID: ____________

7.36 Beer and blood alcohol content. Many people believe that gender, weight, drinking habits, and many other factors are much more important in predicting blood alcohol content (BAC) than simply considering the number of drinks a person consumed. Here we examine data from sixteen student volunteers at Ohio State University who each drank a randomly assigned number of cans of beer. These students were evenly divided between men and women, and they differed in weight and drinking habits. Thirty minutes later, a police officer measured their blood alcohol content (BAC) in grams of alcohol per deciliter of blood.[23] The scatterplot and regression table summarize the findings.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0127	0.0126	-1.00	0.3320
beers	0.0180	0.0024	7.48	0.0000

(a) Describe the relationship between the number of cans of beer and BAC.

(b) Write the equation of the regression line. Interpret the slope and intercept in context.

(c) Do the data provide strong evidence that drinking more cans of beer is associated with an increase in blood alcohol? State the null and alternative hypotheses, report the p-value, and state your conclusion.

(d) The correlation coefficient for number of cans of beer and BAC is 0.89. Calculate R^2 and interpret it in context.

(e) Suppose we visit a bar, ask people how many drinks they have had, and also take their BAC. Do you think the relationship between number of drinks and BAC would be as strong as the relationship found in the Ohio State study?

a) The relationship between the number of cans of beer and BAC is a moderate/strong positive linear relationship, with no real outliers.

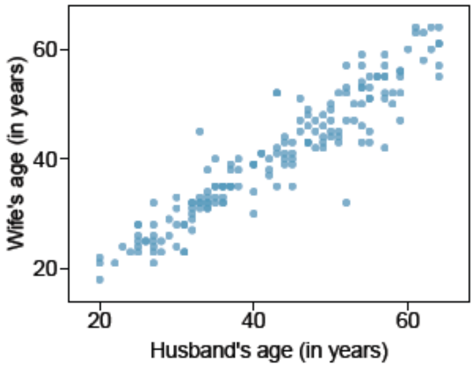
b) y_bar = -0.0127 + 0.0180 * x. In this context y_bar is the predicted value of the BAC of the linear regression model. X in this context, represents the number of beers consumed. Thus, for every can of beer the BAC of that person is expected to increase by 0.018.

c)
H0: B1 = 0
HA: B1 ≠ 0
The significance level (alpha) = 0.05. As given in the table, the p-value of B1 is 0.0000. Since, 0.0000 < 0.0500 we can reject the null hypothesis, thus drinking more cans of beer is correlated to your BAC.

d) Since, the correlation coefficient given, R = 0.89; R2 = 0.7921. Thus, the BAC of a participant is 79.21% dependent on the number of cans that participant has consumed.

e) Yes, I do believe the relationship would be just as strong.

7.38 Husbands and wives, Part III. Exercise 7.37 presents a scatterplot displaying the relationship between husbands' and wives' ages in a random sample of 170 married couples in Britain, where both partners' ages are below 65 years. Given below is summary output of the least squares fit for predicting wife's age from husband's age.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.5740	1.1501	1.37	0.1730
age_husband	0.9112	0.0259	35.25	0.0000
				$df = 168$

(a) We might wonder, is the age difference between husbands and wives consistent across ages? If this were the case, then the slope parameter would be $\beta_1 = 1$. Use the information above to evaluate if there is strong evidence that the difference in husband and wife ages differs for different ages.

(b) Write the equation of the regression line for predicting wife's age from husband's age.

(c) Interpret the slope and intercept in context.

(d) Given that $R^2 = 0.88$, what is the correlation of ages in this data set?

(e) You meet a married man from Britain who is 55 years old. What would you predict his wife's age to be? How reliable is this prediction?

(f) You meet another married man from Britain who is 85 years old. Would it be wise to use the same linear model to predict his wife's age? Explain.

a) H0: ß1 = 1,
HA: ß1 ≠ 1
ageW = 1.5740 + 0.9112 x (1)
ageW = 2.4852
t = (2.4852 – 1) / 0.0259 = 57.34362934362934, df = 168
The significance level, alpha = 0.05. Since, the p-value 0.000001 < alpha, we can reject the null hypothesis and state that there is strong evidence that the difference in husband and wife ages differs for different ages.

b) ageW = 1.5740 + 0.9112 x ageH

c) When the husband is being born the wife is predicted to be about 1.5740 years old. For every year older the husband is, the wife is 0.9112 years older than him.

d) The age of the wife has 88% to do with the age of the husband.

e)
ageW = 1.5740 + 0.9112 x ageH
ageW = 1.5740 + 0.9112 x (55)
ageW = 1.5740 + 50.116
ageW = 51.69 years old. This prediction is 88% reliable.

f) No, it would not be wise to use this model to predict his wife's age, because his age group was not accounted for in our data. We have trained our predictor on possibly unrelated data, so attempting to predict in his age range would probably not be very reliable. For instance, what if his generation tended to marry women with much higher ages than people in the generations we trained our data on? If we zoomed out and gathered statistics to include his age range, we might see a curved graph, thus rendering our linear regression unreliable.

7.42 Babies. Is the gestational age (time between conception and birth) of a low birth-weight baby useful in predicting head circumference at birth? Twenty-five low birth-weight babies were studied at a Harvard teaching hospital; the investigators calculated the regression of head circumference (measured in centimeters) against gestational age (measured in weeks). The estimated regression line is

$$\widehat{head\_circumference} = 3.91 + 0.78 \times gestational\_age$$

(a) What is the predicted head circumference for a baby whose gestational age is 28 weeks?

(b) The standard error for the coefficient of gestational age is 0.35, which is associated with $df = 23$. Does the model provide strong evidence that gestational age is significantly associated with head circumference?

a) head_circumference = 3.91 + 0.78 x (28 weeks)
head_circumference = 3.91 + 21.84
head_circumference = 25.75 cm

b) H0: ß1 = 0,
HA: ß1 ≠ 0
s = 0.35
The significance level, alpha = 0.05.
t = (0.78 - 0) / 0.35
t = 2.2285714285714286, df = 23
p-value = 0.017954. p-value < alpha, thus we can reject the null hypothesis. There is strong evidence to suggest that gestational age is significantly associated with head circumference at birth.

1. For you, personally, what was the most interesting statistics example (or two examples) we discussed in this class? I ask this so that I can see what examples resonate with students and change those that no one enjoyed. [example, why you enjoyed it so much]

The most interesting things for me were the linear regression models for the ages of married couples, and the regression model for the correlation of height vs. weight. I enjoyed it, because I was concurrently taking CSE 158, which a data mining and recommender systems class. This class focused heavily on predictor models and accuracy statistics which were very heavily discussed in the latter portion of this class.

2. I always want to make my teaching practice and classes better. What is one thing you would change about this class to make it better? (You will get credit for giving some honest, **constructive** feedback. Do not write “Everything was great – change nothing!” Note that you can use the CAPE system to give anonymous feedback.)

I would definitely say that giving less homework would be my biggest critique. For people who really want to learn the subject, but also have programming assignments for three to four other classes and extra curricular activities, such as sports, it is nearly impossible to keep up with two assignments per week. Maybe one longer one per week would suffice for learning the material and helping students to keep up.

(No R Questions this week.)