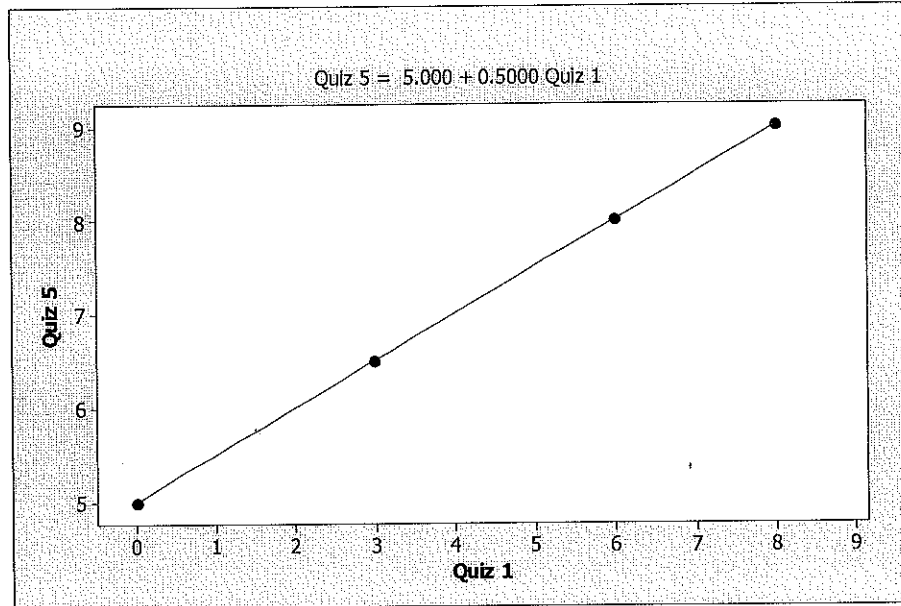


**Worksheet for Correlation and Regression (February 1, 2013)**

Part 1. Consider the following hypothetical data set. Here are data from four students on their Quiz 1 scores and their Quiz 5 scores.

(Fake) Data on Quiz scores:

x	y
Quiz 1	Quiz 5
0	5
3	6.5
6	8
8	9



Notice that this is an exact linear relationship, as we had in algebra classes. We need to review this before going on to learn about approximate linear relationships, as we will see in statistics classes.

- Fill in the following table of some predicted Quiz 5 scores.

x	y
Quiz 1	Quiz 5
0	5
1	5.5
2	6.0
3	6.5
4	7.0
6	8.0
8	9.0

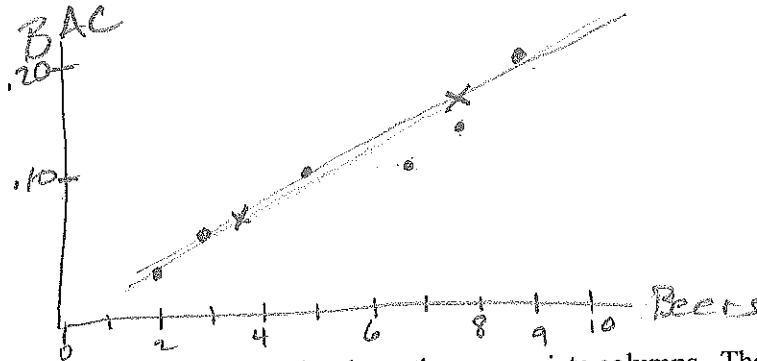
The equation of this line is  $y = 5 + 0.5x$ . The slope is 0.5 and the y-intercept is 5.

- Interpret the y-intercept. (Write a sentence about Quiz 1 and Quiz 5 score relationship using the y-intercept.)  
*a = 5. When Quiz 1 score is 0, we predict Quiz 5 score is 5.*

- Interpret the slope. (Write a sentence about the Quiz 1 and Quiz 5 score relationship using the slope.)  
*b = 0.5. When the Quiz 1 score increases by 1 point, we predict the Quiz 5 score will increase by 0.5 points.*

Part 2 Data: In EESEE, an experiment is described in which they want to predict Blood Alcohol Content from the number of beers. Volunteer college students (some men and some women) are assigned a certain number of beers to drink and then, after a half an hour, their Blood Alcohol (BAC) level is measured. (In our text, the full dataset is in Ch. 24 beers and is the example on making predictions of BAC from Beers. For this handout, for ease of calculation, we will use only the first six of these data points)

x	y
Beers	BAC
5	0.1
2	0.03
9	0.19
8	0.12
3	0.04
7	0.095



- In the blank space above, put labels  $x$  and  $y$  above the appropriate columns. Then quickly sketch a scatterplot appropriate to investigate whether how BAC depends on the number of beers consumed.
- Guess what  $r$  is. (Is it positive or negative? Is it close to zero or close to positive or negative one?)  
*positive, close to +1*
- For each variable ( $x$  and  $y$ ) find the mean and standard deviation. **On homework and quizzes, use software (Minitab or CrunchIt) to compute the standard deviations. On tests, you won't have to compute the standard deviations.**

$$\bar{x} = 5.667 \quad \bar{y} = 0.0958$$

$$s_x = 2.80 \quad s_y = 0.0582$$

- On homework and quizzes, use Minitab or CrunchIt to compute the correlation coefficient. On tests, you won't have to compute the correlation coefficient.** For these data, we have  $r = 0.927$ .  
Is this consistent with your guess? *Yes*

5. On your scatterplot, guess where you would draw a line that comes close to describing the data. Do a very light sketch of it. You will compare your guessed line with the one you actually compute below.

6. **On homework, quizzes and tests, be able to compute the equation of the regression line by hand, using a scientific calculator, using these formulas.** On homework and quizzes, use this often enough to learn to do it for the test. You may use the computer to find the equation of the regression line on most homework and quiz problems.

Compute the slope and intercept of the regression line.

$$b = r \frac{s_y}{s_x} = .927 \frac{0.0582}{2.80} = 0.0192684$$

$$a = \bar{y} - b\bar{x} = 0.0958 - (0.0192684)(5.667) = -0.01339$$

- Write the equation of the regression line: ( $\hat{y} = a + bx$ . You write this, but put in your computed values for  $a$  and  $b$ .)

$$\hat{y} = -0.01339 + 0.01927x$$

8. Use the equation of the regression line to predict BAC twice, for 4 beers and 8 beers:

$$x = 4 \quad \hat{y} = -0.01339 + 0.01927(4) = .0637$$

$$x = 8 \quad \hat{y} = -0.01339 + 0.01927(8) = .1408$$

9. Using a different colored pencil, put these two points you just computed on your graph and then draw a line through them. This is the graph of the regression line that you just computed. Does it look pretty close to the line you guessed? *Yes it looks very close*

10. Compute the residual for  $x = 8$ . (The residual is the difference between the observed  $y$  value and the predicted  $y$  value.)

$$\text{residual} = y - \hat{y} = .12 - .1408 = -.0208$$

11. Use the line you drew to predict  $y$  when  $x = 6$ . Write your prediction here.

*about BAC = 0.11*

12. After reading the example in the text about interpreting the regression coefficients, write your interpretations about ~~quiz grades~~ *BAC & Beers* here:

Intercept: (If  $x = 0$ , then the predicted value of  $y$  is  $a$ .)

*If the person drinks 0 beers, we predict the BAC is approximately -0.01339.*

*(Since this is negative, it is unrealistic. But it is very close to 0 and we believe it should be 0.)*

Slope: (If  $x$  increases by 1, then we predict that  $y$  will increase by  $b$ .)

*For each additional one beer, we predict BAC will increase by about 0.19.*

Variable	N	N*	Mean	StDev
Beers_6	6	0	5.67	2.80
BAC_6	6	0	0.0958	0.0582

Pearson correlation of Beers\_6 and BAC\_6 = 0.927  
 The regression equation is  
 $BAC_6 = -0.0132 + 0.0192 \text{ Beers}_6$

Variable	N	N*	Mean	StDev
Beers	16	0	4.813	2.198
BAC	16	0	0.0738	0.0441

Pearson correlation of Beers and BAC = 0.894  
 The regression equation is  
 $BAC = -0.0127 + 0.0180 \text{ Beers}$