MATH-1230 (DUPRÉ) FALL 2010 TEST 3 ANSWERS

FIRST: PRINT YOUR LAST NAME IN LARGE CAPITAL LETTERS ON THE UPPER RIGHT CORNER OF THIS SHEET.

SECOND: PRINT YOUR FIRST NAME IN CAPITAL LETTERS DIRECTLY UNDERNEATH YOUR LAST NAME.

THIRD: WRITE YOUR FALL 2010 MATH-1230 LAB DAY DIRECTLY UNDERNEATH YOU FIRST NAME.

Suppose that X and Y are unknowns with

$$\mu_X = 20,\ \mu_Y = 100,\ \sigma_X = 2,\ and\ \sigma_Y = 10.$$

Suppose that the coefficient of correlation between X and Y is

$$\rho = .8.$$

Let

$$W = \beta_0 + \beta_1 X$$

be the regression equation for the regression of Y on X.

1. What is the covariance of X with Y or $Cov(X, Y)$?

ANSWER: By definition of the correlation coefficient we have

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y},$$

so therefore when given the standard deviations and correlation coefficient, we find the covariance using

$$Cov(X,Y) = \rho \cdot \sigma_X \cdot \sigma_Y.$$

This means here we have

$$Cov(X,Y) = (.8)(2)(10) = 8 \cdot 2 = 16.$$

FINAL ANSWER: 16

2. What is the value of the regression slope β_1?

ANSWER: The regression slope β_1 is alsways given by the simple formula

$$\beta_1 = \frac{\rho \cdot \sigma_Y}{\sigma_X}.$$

This means here we have

$$\beta_1 = \frac{(.8)(10)}{2} = \frac{8}{2} = 4.$$

FINAL ANSWER: 4

3. What is the value of the regression intercept β_0?

ANSWER: The regression intercept is always given by the simple formula

$$\beta_0 = \mu_Y - \beta_1 \cdot \mu_X.$$

This means here we have

$$\beta_0 = 100 - 4 \cdot 20 = 100 - 80 = 20.$$

Alternately, we can recall that the regression line always goes through the point $(\mu_X,\ \mu_Y)$, so we can easily write down the regression line equation in point-slope form:

$$W = \beta_1(X - \mu_X) + \mu_Y.$$

Here this means that

$$W = 4(X - 20) + 100,$$

so when we simplify the expression on the right hand side we get the equation

$$W = 4X - 80 + 100 = 4X + 20,$$

or simply

$$W = 4X + 20,$$

from which we can easily see that the verticle intercept is $\beta_0 = 20$.

FINAL ANSWER: 20

4. What is $E(Y|X = 24)$?

ANSWER: Since the regression equation is $W = \beta_0 + \beta_1 \cdot X$, this means that

$$E(Y|X = x) = \beta_0 + \beta_1 \cdot x,$$

no matter what value x we are given for X. Here this means that

$$E(Y|X = x) = 20 + 4x$$

for any numerical value x. In particular then,

$$E(Y|X = 24) = 20 + 4 \cdot (24) = 20 + 96 = 116.$$

Alternately, we can notice that the value $X = 24$ is 2 standard deviations above the mean, since $\mu_X = 20$ and $\sigma_X = 2$. Therefore, as the correlation coefficient is $\rho = .8$, we should guess that a value for Y which is $(.8) \cdot 2$ standard deviations above the mean for Y. For Y we have $\sigma_Y = 10$, therefore two standard deviations above the mean is 20 and 80 percent of 20 is 16. As the mean of Y is 100, we should increase our guess for the value of Y from 100 up to 116 when we are given that $X = 24$. Another way to look at this is with the deviations from the mean. The regression equation in point-slope form says

$$W = \beta_1 \cdot (X - \mu_X) + \mu_Y,$$

and we have $\mu_W = E(W) = \mu_Y$. (Why?) This means that using the deviations, we can write the regression equation as

$$D_W = \beta_1 \cdot D_X.$$

The idea of linear regression is to guess that whatever the deviation of X is from its mean, the corresponding deviation of W from its mean (which is also the mean of Y) should be used as your guess for the deviation of Y from its mean.

FINAL ANSWER: 116

5. When we use $E(Y|X = 24)$ as our guess for the value of Y, what do we expect for our squared error?

ANSWER: In general, if we use the linear expression

$$W = b_0 + b_1 \cdot X$$

as our equation for guessing values of Y from values for X, then we have some error

$$R = W - Y$$

and the expected squared error is then given by the equation

$$E(R^2) = \mu_R^2 + (\rho \cdot \sigma_Y - b_1 \cdot \sigma_X)^2 + (1 - \rho^2) \cdot \sigma_Y^2.$$

The optimal values for b_0 and b_1, denoted respectively, β_0 and β_1, are chosen so as to minimize the expected squared error. The second term on the right hand side of the equation for the squared error is made zero by choosing

$$\beta_1 = \rho \frac{\sigma_Y}{\sigma_X}$$

and the first term is made zero obviously by choosing $\mu_W = \mu_Y$ which is the same as defining β_0 to be given by

$$\beta_0 = \mu_Y - \beta_1 \cdot \mu_X.$$

This means that using the optimal linear regression equation we have an expected squared error given by

$$E(R^2) = (1 - \rho^2) \cdot \sigma_Y^2.$$

Here this means simply

$$E(R^2) = (1 - (.8)^2) \cdot (100) = (1 - .64) \cdot (100) = (.36)(100) = 36.$$

Notice that if you do not use information about X to help with your guess for Y, then the optimal guess is simply $\mu_Y = 100$ and your expected squared error is then

$$E(D_Y^2) = \sigma_Y^2 = (10)^2 = 100,$$

so by using information about the value of X we get a substantial reduction in our expected squared error.

FINAL ANSWER: 36