Lecture 16

(1)

Topics: Bivariate Normal continued

There is an easier way to Remember Master formula 3.

$$\left(\frac{E(Y|X=x) - E(Y)}{sd(Y)}\right) = \rho\left(\frac{x - E(X)}{sd(X)}\right)$$

This is the Z-score for Y evaluated at the prediction $E(Y|X=x)$.

This is the Z-score for X evaluated at x.

Simplified Master formula 3:

$$(\underbrace{\text{predicted Z-score}}_{\text{for } Y}) = \rho\,(\underbrace{\text{observed Z-score}}_{\text{for } X}).$$

Important fact: Since Z-scores quantify rarity & correlation is always btwn -1 & 1 the above formula implies that no matter how rare the observed X is among all possible X's, the associated Y value will be predicted to be less rare among all possible Y's. (this is called the regression effect)

(2)

e.g. Suppose (X,Y) are Bivariate Normal with parameters

$E(X)=10, \quad sd(X)=2$

$E(Y)=5, \quad sd(Y)=100$

Suppose you observe $X=16.5$ which corresponds to a Z-score (for X) of

$$\frac{16.5-10}{2} = 3.25$$

This is fairly rare.

How rare... well $P(Z \geq 3.25) = 0.00058$ so getting a Z-score ≥ 3.25 only happens 0.058% of the time (which is about 1 in 2000).

Predict the Z-score for Y in the following three cases

Case 1: $\rho = 0.01$

Case 2: $\rho = 0.5$

Case 3: $\rho = 0.99$.

Answer:

Case 1: $\rho = 0.01$.

By MF3

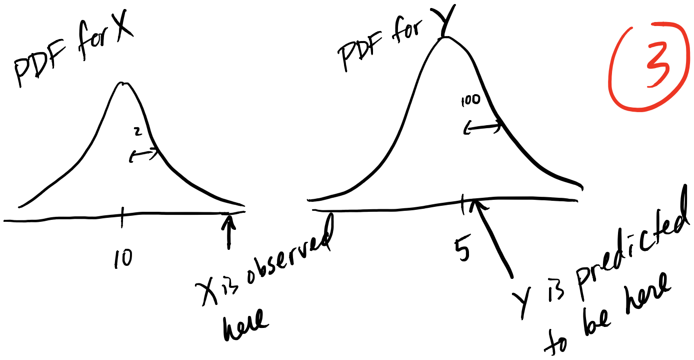
predicted Z-score for Y

$= \rho\,(\text{obs Z-score for } X)$

$= 0.01\,(3.25)$

$= \underbrace{0.0325}$

This Z-score for Y is not at all rare. $P(Z > 0.0325) = 0.48$.



Case 1: $\rho = 0.5$

By MF3

predicted z-score for Y

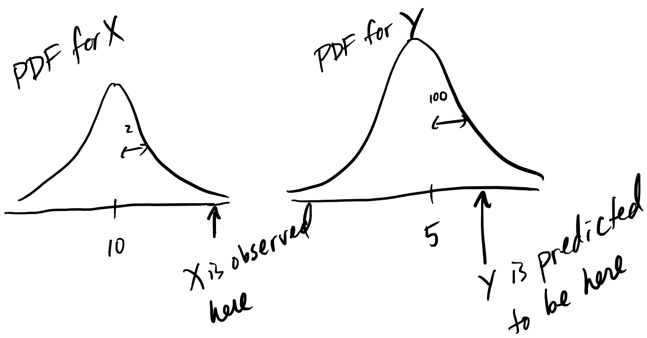
$= \rho\,(\text{obs z-score for } X)$

$= 0.5\,(3.25)$

$= \underbrace{1.625}$

This z-score for Y is not at all rare. $P(Z > 1.625) = 0.052$

... so about 5 out of 20.



Case 1: $\rho = 0.99$

By MF3

predicted z-score for Y

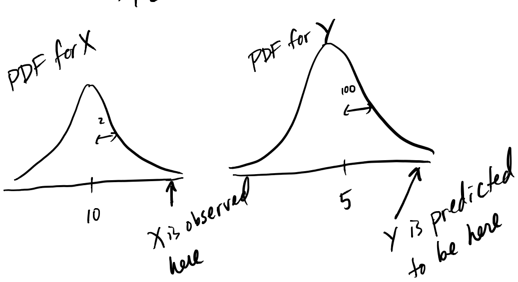
$= \rho\,(\text{obs z-score for } X)$

$= 0.99\,(3.25)$

$= \underbrace{3.2175}$

This z-score for Y is rare but still not as rare as X...

$P(Z > 3.2175) = 0.00065$



We observe the regression effect all the time but don't realize it.

e.g. It seems that whenever my son has a high fever in the middle of the night, when I bring him to the doctor the next day his fever isn't as bad.

... Regression effect!

e.g. why does it always seem that after winning an oscar a directors follow up movie isn't as good...

Regression effect!

e.g. If I look at the top scorers on midterm 1, why are many of them no longer top scorers on midterm 2?

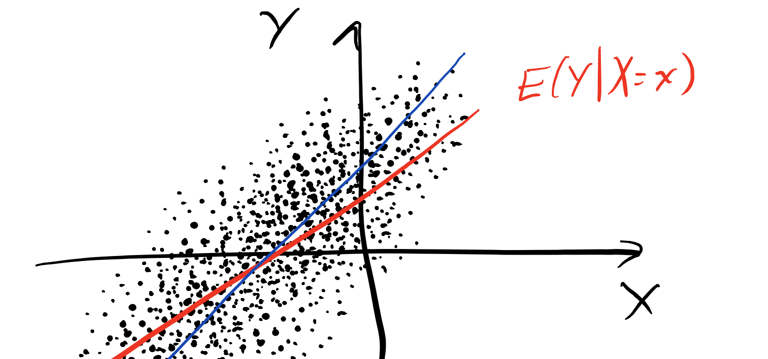
... Regression effect!

$\boxed{\text{Show Julia Notebook.}}$

e.g. If you plot the prediction line

i.e. plot $E(Y|X=x) = sd(Y)\left(\dfrac{x - E(X)}{sd(X)}\right) + E(Y)$

as a function of x ... the slope "look" too small



... Regression effect

5

Prediction error

Fact 1: when X, Y are Bivariate Normal the typical prediction error for Y is always smaller when incorporating information on X.

i.e. $sd(Y|X=x) \overset{MF4}{=} \underbrace{\sqrt{(1-\rho^2)}}_{\leq 1}\, sd(Y)$

only equal when $\rho=0$ $\longrightarrow \leq sd(Y)$

Fact 2: when X, Y are Bivariate Normal the prediction error gets smaller as $\rho \to 1$ or $\rho \to -1$.

$sd(Y|X=x) = \underbrace{\sqrt{(1-\rho^2)}}_{\to 0 \text{ as } \rho \to 1 \text{ or } -1}\, sd(Y)$

e.g. Midterm 1 scores: X

$E(X) = 10.62$, $sd(X) = 1.46$

Midterm 2 scores: Y

$E(Y) = 8.54$, $sd(Y) = 2.08$

Correlation btwn X & Y

$\rho = 0.155$.

Find $E(Y|X=11)$ & $sd(Y|X=11)$

For $E(Y|X=11)$ use MF3

$$(\text{predicted z-score}) = 0.155(\text{obs z-score})$$
$$= 0.155\Bigg(\underbrace{\frac{11-10.62}{1.46}}_{0.26}\Bigg)$$
$$= 0.04$$

6

$$\therefore E(Y|X=11) = sd(Y)(\text{predicted z-score}) + E(Y)$$
$$= 2.08(0.04) + 8.54$$
$$= \underbrace{8.623}_{\text{best prediction for } Y \text{ given } X=11}$$

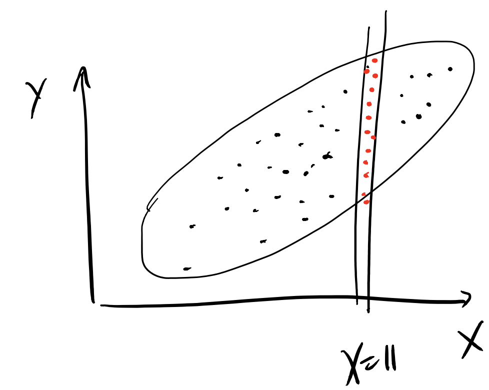
slightly above average.

The typical prediction error is

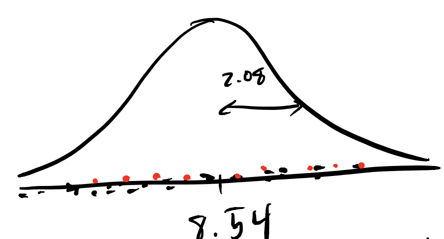
$$sd(Y|X=11) = \sqrt{(1-\rho^2)}\, sd(Y)$$
$$= \sqrt{1-0.155^2}\; 2.08$$
$$\approx 2.05$$

smaller than $sd(Y)$.

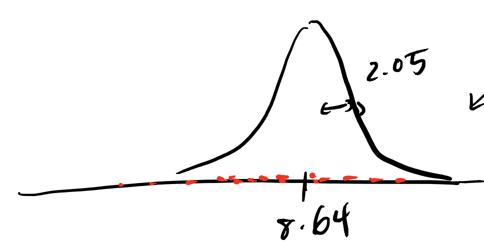
Here is the picture



All the Y's



The Y's for the subgroup that got $X=11$



given $X=11$ this is the subgroup you use to do prediction.