

## REGRESSION IN BIVARIATE NORMAL POPULATIONS

X, Y bivariate normal.

$\mu_X$  = mean of X,  $\sigma_X$  = standard deviation of X

$\mu_Y$  = mean of Y,  $\sigma_Y$  = standard deviation of Y

$\rho$  = correlation coefficient

What do the conditional distributions Y|X look like?

Their pdf's can be calculated from the joint density using the formula

$$\begin{aligned}
 f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left[ -\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{2(1-\rho^2)} \right] \\
 &\quad \div \frac{1}{\sqrt{2\pi}\sigma_X} \exp \left[ -\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 \right] \\
 &= \text{(Details left to the interested student; completing the square should be useful.) ...} \\
 &= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp \left[ -\frac{1}{2}\left(\frac{y-\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)}{\sigma_Y(1-\rho^2)}\right)^2 \right]
 \end{aligned}$$

Result: Y|X is normal, with mean and variance:

- $E(Y|X = x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X)$
- $\text{Var}(Y|X) = \sigma_Y^2(1 - \rho^2)$

**Consequences:** (X and Y are bivariate normal)

1. "Constant variance":  $\text{Var}(Y|X)$  does *not* depend on X.
2. "Linear mean function":  $E(Y|X)$  is a *linear function* of X, with slope  $\rho \frac{\sigma_Y}{\sigma_X}$

(Note how *the slope depends on all three of the parameters*  $\rho$ ,  $\sigma_X$ , and  $\sigma_Y$ .)

So the pipe cleaner model fits!

**Alternate perspectives:**

1. Rearranging the mean function,

$$\frac{E(Y|X = x) - \mu_Y}{\sigma_Y} = \rho \frac{x - \mu_X}{\sigma_X}$$

Recall:  $E(E(Y|X)) = \underline{\hspace{2cm}}$ .

So:

Left side: Like  $E(Y|X)$  standardized

Right side =  $x$  standardized.

Thus: If  $X$  and  $Y$  are bivariate normal, then for every increase of 1 in standardized  $x$ ,  $E(Y|X)$  "standardized" increases  $\rho$  units. (If you've seen least squares regression, you have seen the analogue for the least squares regression line, using  $sd(x)$ ,  $sd(y)$  and  $r$ .)

2. Rearranging as

$$E(Y|X = x) = \mu_Y + \rho \sigma_Y \frac{x - \mu_X}{\sigma_X},$$

we see: For every increase of  $\sigma_X$  in  $X$ ,  $E(Y|X)$  increases  $\rho\sigma_Y$ .

**Similarly for  $X|Y$ :**

$$E(X|Y = y) = \mu_X + \rho \frac{\sigma_X}{\sigma_Y} (y - \mu_Y) \quad \text{Var}(X|Y) = \sigma_X^2(1 - \rho^2)$$

**Example:** If  $X$  and  $Y$  have a standard bivariate normal distribution with  $\rho = 0.5$ , then

$$E(Y|X = x) = \rho x = x/2 \text{ (which gives graph } y = x/2\text{)}$$

$$E(X|Y = y) = \rho y \text{ (which gives graph } x = y/2 \text{ -- i.e., } y = 2x\text{)}$$

These are different! (More on homework.)

**Note:** The mean lines are *not* the same as the axes of the ellipses forming the level curves of the bivariate normal pdf. Here is a picture of a sample of 200 from a standard bivariate normal distribution with  $\rho = 0.5$ . Also shown in the picture are:

- Some level curves for the pdf
- The axes of the ellipse
- The line showing  $E(Y|X=x)$  as a function of  $x$ .
- The line showing  $E(X|Y = y)$  as a function of  $y$ .

Which of these four lines is which?

