

Stat 134 Lec 40

Warmup: 11-11:10

Let $x, z \sim N(0,1)$, $-1 \leq \rho \leq 1$

Let $y = \rho x + \sqrt{1-\rho^2} z$

a) Find $E(y|x) = E(\rho x + \sqrt{1-\rho^2} z|x)$

b) Find $\text{Var}(y|x) = \text{Var}(\rho x + \sqrt{1-\rho^2} z|x)$

$$\begin{aligned} a) E(y|x) &= \rho E(x|x) + \sqrt{1-\rho^2} E(z|x) \\ &\quad \stackrel{x}{\text{''}} \quad \stackrel{z}{\text{''}} \\ &= \boxed{\rho x} \end{aligned}$$

Note
 $E(z|x) = \int z f_z(z) dz$

$$f_{z|x}(z) = \frac{f_{z,x}(z)}{f_x(x)} \stackrel{\text{indep}}{=} \frac{f_z(z) f_x(x)}{f_x(x)} = f_z(z)$$

$$\begin{aligned} b) \text{Var}(y|x) &= \text{Var}(\rho x|x) + \text{Var}(\sqrt{1-\rho^2} z|x) \\ &\quad \stackrel{x}{\text{''}} \quad \stackrel{z}{\text{''}} \quad \stackrel{y}{\text{indep}} \\ &= \rho^2 \text{Var}(x|x) + (1-\rho^2) \text{Var}(z|x) \\ &\quad \stackrel{x}{\text{''}} \quad \stackrel{z}{\text{''}} \\ &= \boxed{1-\rho^2} \end{aligned}$$

so $E(z|x) = \int z f_z(z) dz = E(z)$

$\text{Var}(z) = 1$

Last time

Sec 6.5 Bivariate Normal

Defn (Standard Bivariate Normal Distribution)

let $X, Y \sim \text{iid } N(0, 1)$, $-1 \leq \rho \leq 1$

$$Y = \rho X + \sqrt{1-\rho^2} Z \sim N(0, 1)$$

$$\text{Corr}(X, Y) = \rho$$

We call the joint distribution (X, Y) the
Standard bivariate normal with $\text{Corr}(X, Y) = \rho$

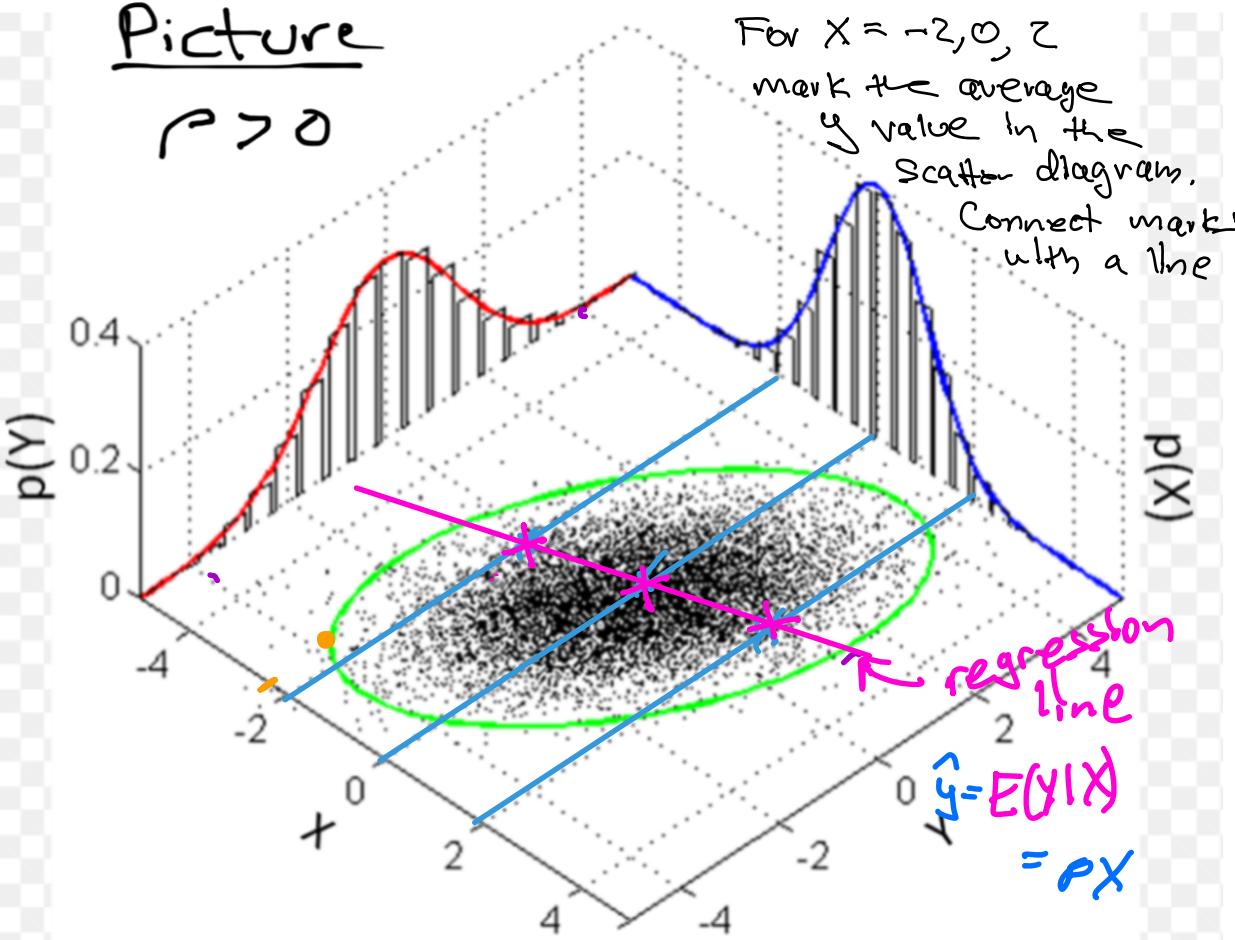
written $(X, Y) \sim \text{BV}(0, 0, 1, 1, \rho)$

$$\begin{matrix} & \uparrow & \uparrow & \uparrow & \uparrow \\ \uparrow & M_x & M_y & \sigma_x & \sigma_y \\ X & & Y & & \end{matrix}$$

Picture

$$\rho > 0$$

For $X = -2, 0, 2$
mark the average
 y value in the
Scatter diagram.
Connect marks
with a line



$$(Y|X=x) = \rho x + \sqrt{1-\rho^2} Z$$

Small x
 is a fixed constant

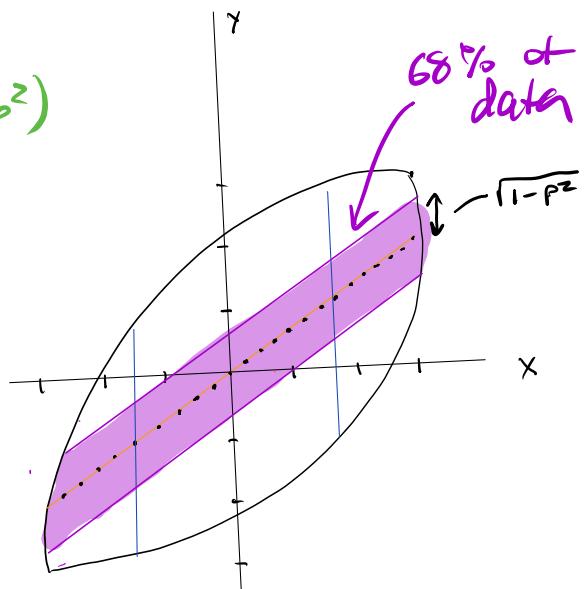
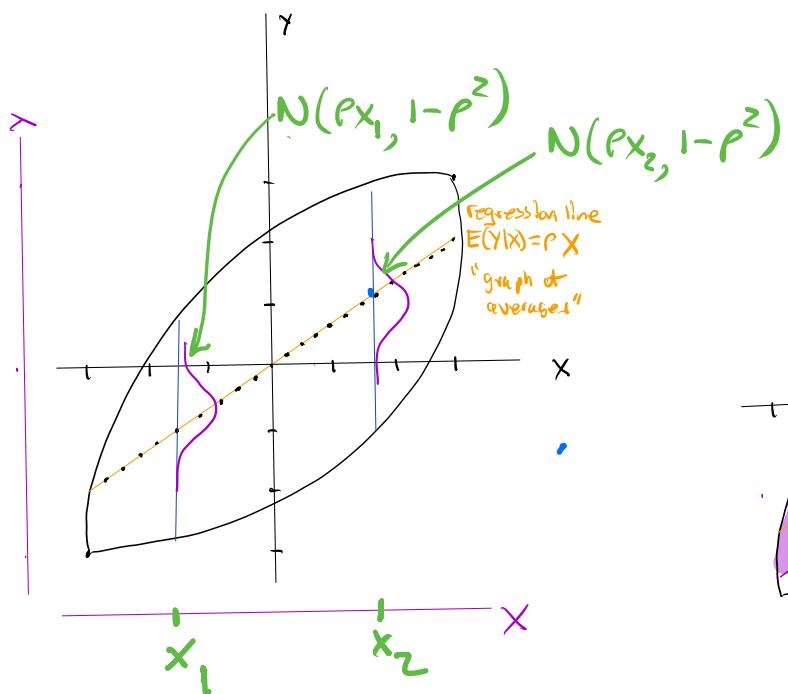
Since $Z \sim N(0, 1)$ and a linear combination of normal is normal, $Y|X=x$ is normal.

Also $E(Y|X=x) = \rho x$ and $V_{\text{cov}}(Y|X=x) = 1 - \rho^2$

so $Y|X=x \sim N(\rho x, 1 - \rho^2)$

see warmup

Picture



efk

$$(Test 1, Test 2) \sim BV(0, 0, 1, 1, 0.6)$$

↙ standard bivariate normal

What is greater? ↗ mean ↘ variance

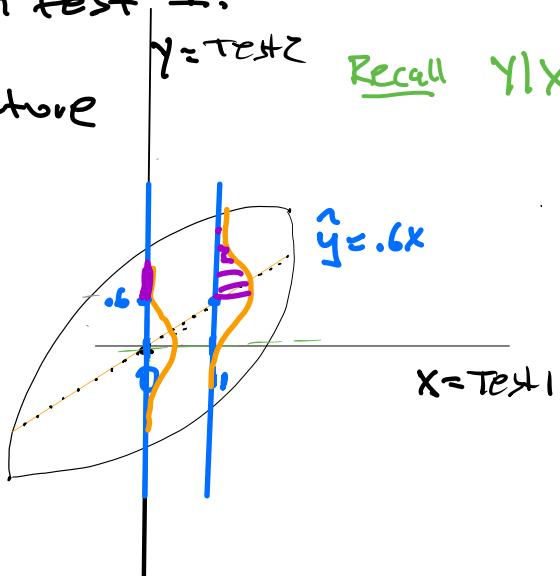
- a) The chance you get greater than .6 on test 2 among students who get 1 on test 1 → 50%

- b) The chance you get greater than .6 on test 2 among students who get 0 on test 1.

Picture

$$y = \text{Test 2}$$

Recall $y|x \sim N(0.6x, 1 - 0.6^2)$



Today ① Sec 6.5 Bivariate Normal

① Sec Grd Bivariate Normal

Defⁿ (Bivariate Normal Distribution)

Random variables U and V have bivariate normal distribution with parameters $\mu_U, \mu_V, \sigma_U^2, \sigma_V^2, \rho$ iff the standardized variables

$$X = \frac{U - \mu_U}{\sigma_U}$$

$$Y = \frac{V - \mu_V}{\sigma_V}$$

have std. bivariate normal distributions with corr ρ .

Then $\rho = \text{corr}(X, Y) = \text{corr}(U, V)$.

We write $(U, V) \sim BV(\mu_U, \mu_V, \sigma_U^2, \sigma_V^2, \rho)$

regression line of bivariate normal distribution

Let $(U, V) \sim BV(\mu_U, \mu_V, \sigma_U^2, \sigma_V^2, \rho)$

then $(X, Y) \sim BV(0, 0, 1, 1, \rho)$ where

$$\begin{aligned} X &= \frac{U - \mu_U}{\sigma_U} \\ Y &= \frac{V - \mu_V}{\sigma_V} \end{aligned}$$

$$\hat{Y} = E(Y|X) = E\left(\frac{V - \mu_V}{\sigma_V} \mid \frac{U - \mu_U}{\sigma_U}\right)$$

$$= E\left(\frac{V - \mu_V}{\sigma_V} \mid U\right)$$

$$= \frac{E(V|U) - \mu_V}{\sigma_V} = \frac{\hat{V} - \mu_V}{\sigma_V}$$

$\hat{Y} = \rho X$ is regression line in S.V.

$$\frac{\hat{V} - \mu_V}{\sigma_V} \sim N(0, 1)$$

$$\Leftrightarrow \hat{V} - \mu_V = \frac{\sigma_V}{\sigma_U} \rho (U - \mu_U)$$

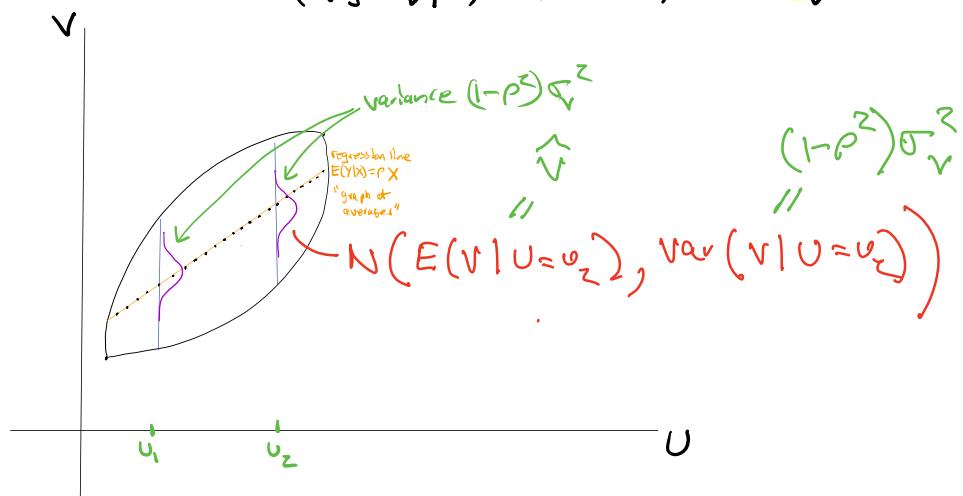
$$\Leftrightarrow \hat{V} = \left(\frac{\sigma_V}{\sigma_U} \rho \right) U + \mu_V - \frac{\sigma_V}{\sigma_U} \rho \mu_U$$

regression line.

$E(V|U)$

furthermore,

$$\text{and } \text{var}(V|U) = \text{var}(\sigma_V Y + \mu_V | X) = \sigma_V^2 \text{var}(Y|X) = (1-\rho^2) \sigma_V^2$$



$$\begin{array}{l}
 \text{Test 1 is } \mu_U = 60 \\
 \sigma_U = 20 \\
 \text{Test 2 is } \mu_V = 60 \\
 \sigma_V = 20
 \end{array}
 \quad \left. \begin{array}{l} \mu_V = 60 \\ \sigma_V = 20 \end{array} \right\} \rho = .6$$

a) Find the regression line \hat{V}

$$\frac{\hat{V} - \mu_V}{\sigma_V} = \rho \frac{U - \mu_U}{\sigma_U}$$

$$\hat{V} = \left(\frac{\sigma_V \rho}{\sigma_U} \right) U + \mu_V - \frac{\sigma_V \rho \mu_U}{\sigma_U}$$

regression line.

$$\begin{aligned}
 \hat{V} &= \frac{20}{20} (.6) U + 60 - \frac{20}{20} (.6)(60) \\
 &= .6U + 24
 \end{aligned}$$

b) If you get a 70 on Test 1 what score do you predict to get on Test 2?

$$E(V|U=70) = .6(70) + 24 = 66$$

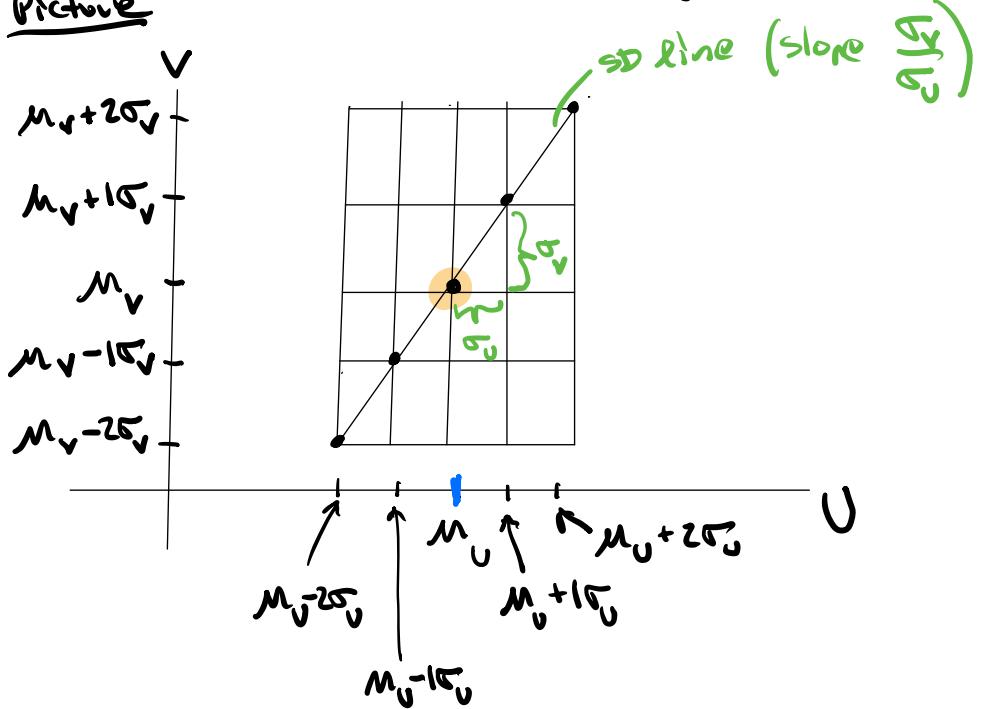
We can see the "regression effect" here. 70 is 1 SD above the mean and 66 is $6/10$ SD above the mean. On Test 2 your predicted score is smaller (regressed towards the mean of 60).

We discuss "regression effect" below:

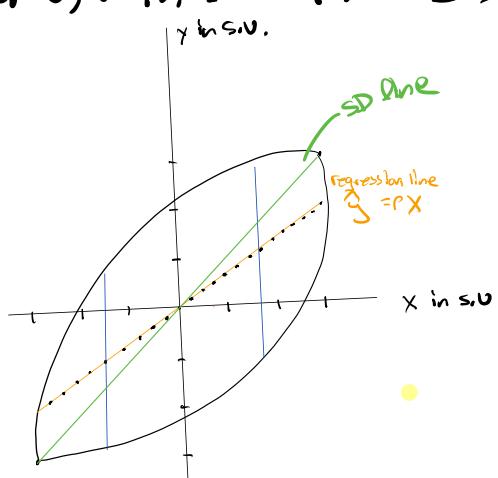
Regression line vs. SD line and regression effect

Def'n the SD line is $V - \mu_V = \frac{\sigma_V}{\sigma_U} (U - \mu_U)$.

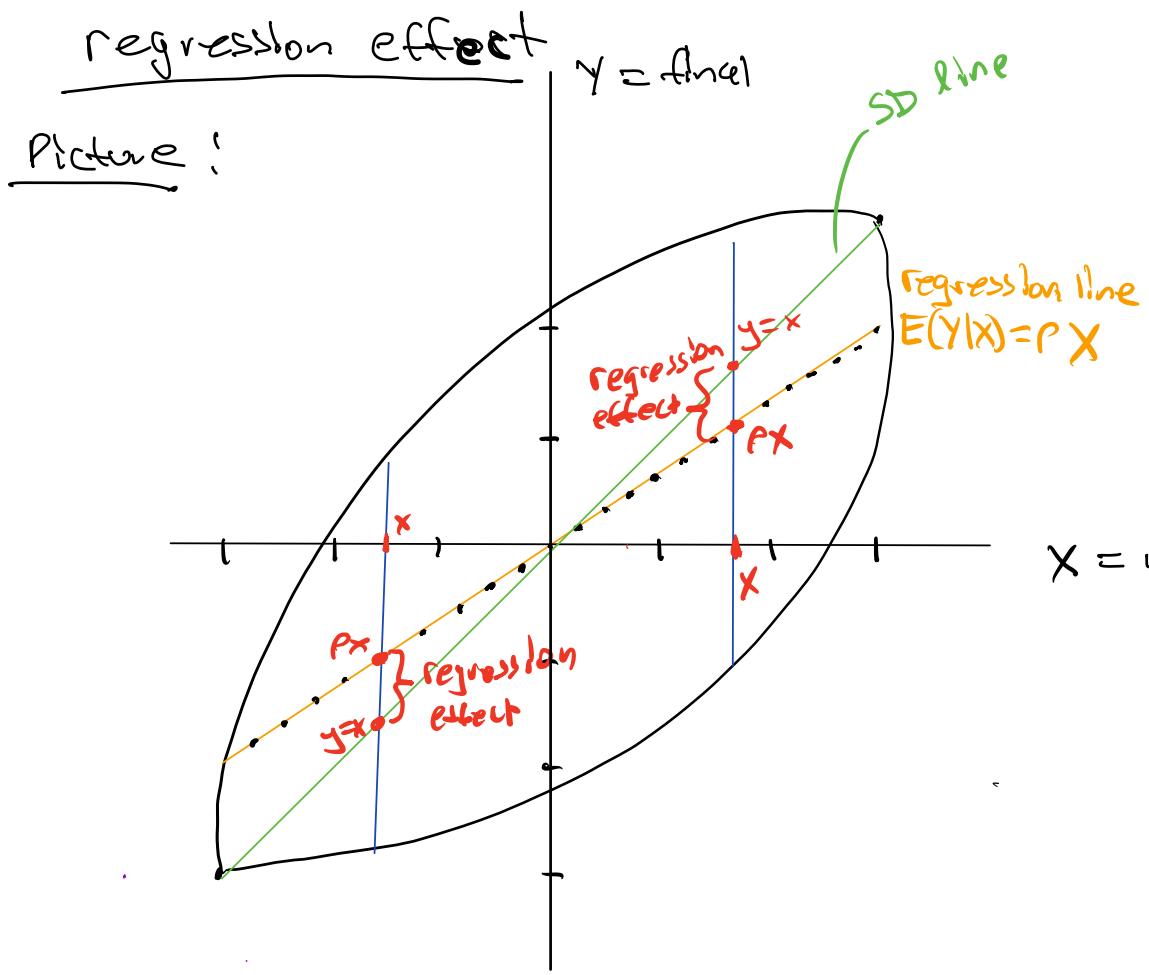
Picture



For U, V in s.u. the SD line is $y = 1 \cdot x$



$y = 1 \cdot x$
 \downarrow
 V^* U^*
 \uparrow
 Steeper than
 regression line
 which has
 slope ρ .



Regression effect,
 $\text{Corr}(\text{test 1}, \text{test 2}) = .6$
 If 1 SD above mean
 on test 1 then on average
 you will be less than 1 SD
 above average on test 2.
 (Regression line is less steep
 than SD line),

ex

Exercises

1. A test score in Math and Physics is bivariate normal, $\rho > 0$. The average is 60 on both tests and the SDs are the same. Of students scoring 75 on the Math test:
- a about half scored over 75 on Physics
 - b more than half scored over 75 on Physics
 - c less than half scored over 75 on Physics



