Stat 134 lec 39

warmup: 10-10:10

let X, Z iid $N(0,1)$,

$Y = \rho X + \sqrt{1-\rho^2}\, Z$ where $-1 \le \rho \le 1$,

greek letter rho

1. What distribution is Y (include parameters)?
2. What is $\mathrm{Corr}(X, Y)$

Today

Sec 6.5 Bivariate Normal

Def^n (Standard Bivariate Normal Distribution)

let X, Z iid $N(0,1)$, $-1 \le \rho \le 1$

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

see warmup

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

see warmup

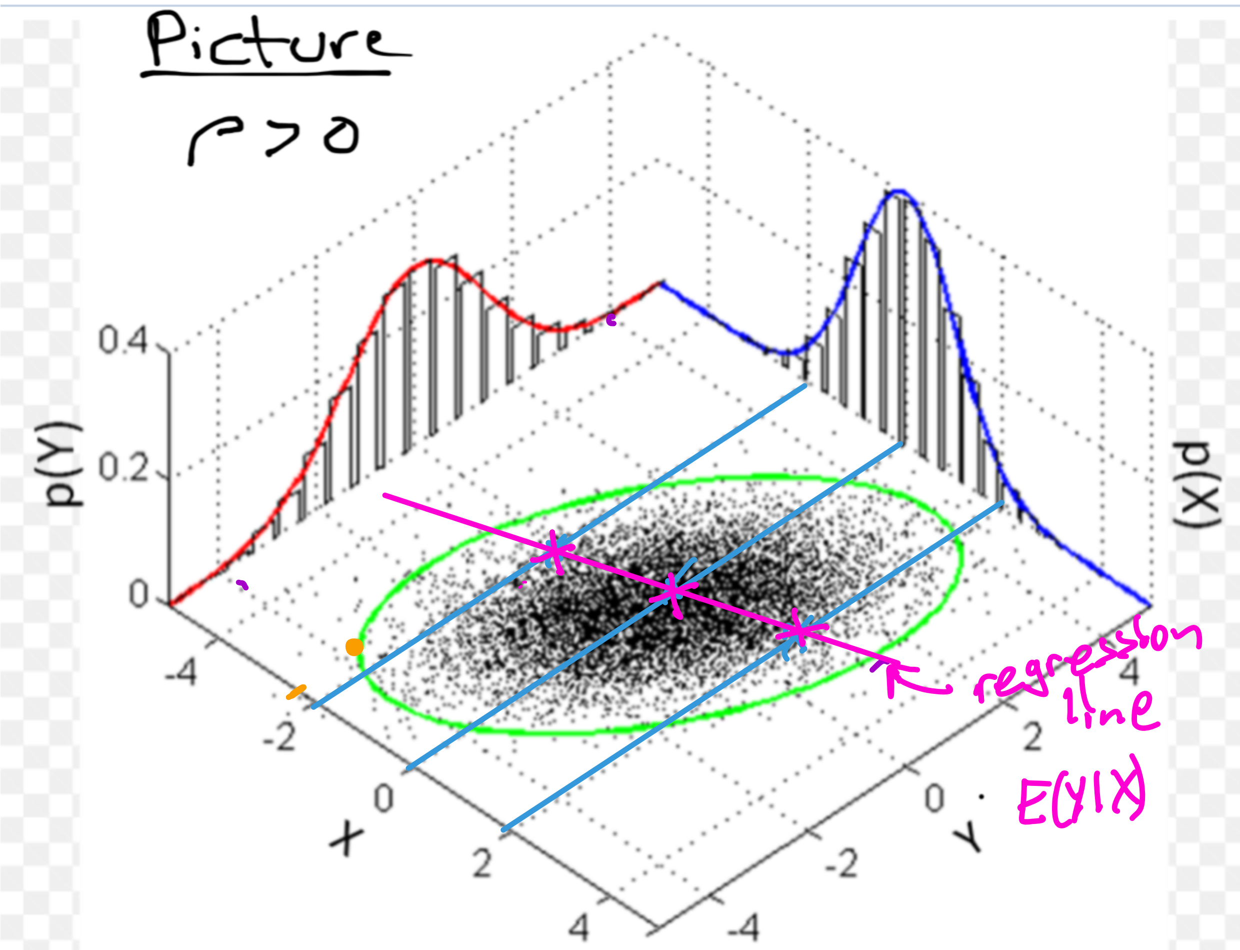
We call the joint distribution (X,Y) the standard bivariate normal with $\text{corr}(X,Y) = \rho$

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

$\mu_X \quad \mu_Y \quad \sigma_X^2 \quad \sigma_Y^2$

Picture

$\rho > 0$



Let $X, Z \overset{iid}{\sim} N(0,1)$, $-1 \le \rho \le 1$

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

Note

$$E(Z|X) = \int_z z f_{Z|X}(z)\, dz$$

$$f_{Z|X}(z) = \frac{f(z,x)}{f_X(x)} \overset{\text{indep}}{=} \frac{f_Z(z)\cancel{f_X(x)}}{\cancel{f_X(x)}} = f_Z(z)$$

so

$$E(Z|X) \overset{\text{indep}}{=} \int z f_Z(z)\, dz = E(Z).$$

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

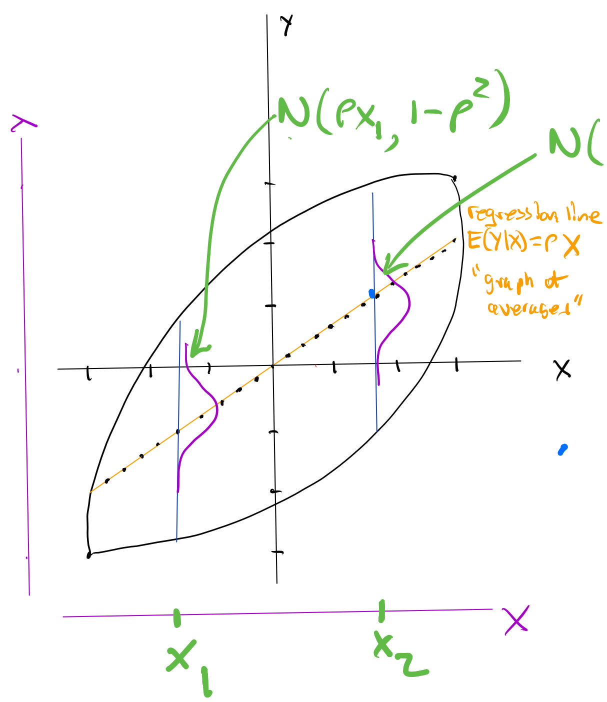
small x is a fixed constant

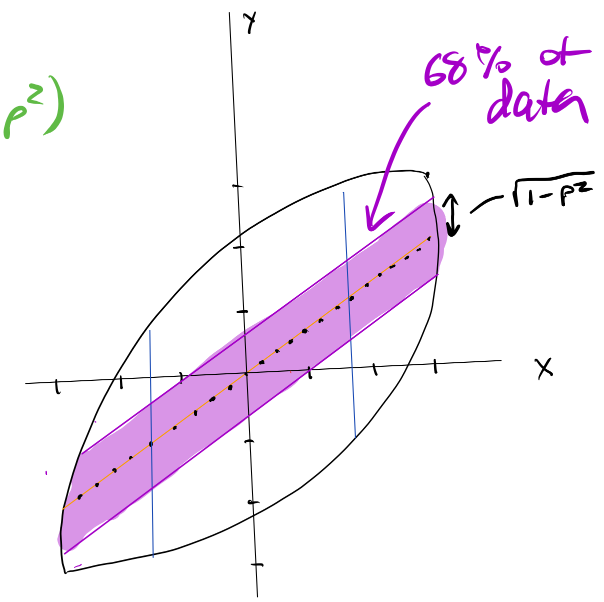
Since $Z \sim N(0,1)$ and a linear combination of normal is normal, $Y \mid X = x$ is normal.

Also $E(Y \mid X = x) = \rho x$ and $Var(Y \mid X = x) = 1 - \rho^2$

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

<u>Picture</u>





ex

standard bivariate normal

(Test 1, test 2) ~ BV(0, 0, 1, 1, 0.6)

mean variance

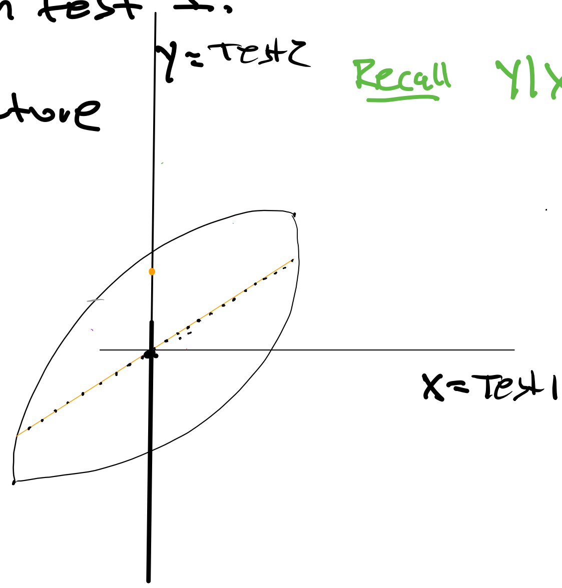
What is greater?

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

Recall $Y|X \sim N(.6x, 1-.6^2)$



$\underline{\text{Def}^n}$ (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 distribution with corr ρ.

Then $\rho = \mathrm{Corr}(X,Y) = \mathrm{Corr}(U,V)$.

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

<u>regression line of bivariate normal distribution</u>

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

$$X = \frac{U-\mu_U}{\sigma_U} \qquad Y = \frac{V-\mu_V}{\sigma_V}$$

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

$\hat{y} = \rho X$ is regression line in s.v.

$$\hat{y} = \frac{\hat{V} - \mu_V}{\sigma_V}, \qquad X = \frac{U - \mu_U}{\sigma_U}$$

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

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

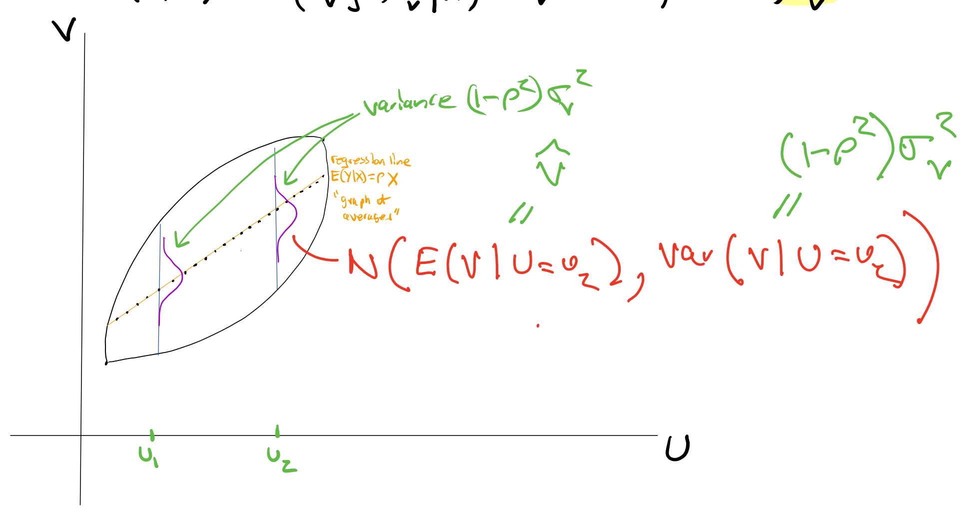
regression line.

$E(V|U)$

furthermore,

$(1-\rho^2)$

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



ex Test 1 is $\mu_U = 60$
$\sigma_U = 20$
Test 2 is $\mu_V = 60$
$\sigma_V = 20$

$\rho = .6$

a) Find the regression line $\hat{V}$

$$\overbrace{\frac{\hat{V} - \mu_V}{\sigma_V}}^{\hat{y}} = \rho \overbrace{\frac{U - \mu_U}{\sigma_U}}^{x}$$

$$\hat{V} = \underbrace{\left(\frac{\sigma_V}{\sigma_U}\rho\right)}_{m} U + \underbrace{\mu_V - \frac{\sigma_V}{\sigma_U}\rho\,\mu_U}_{b}$$ regression line.

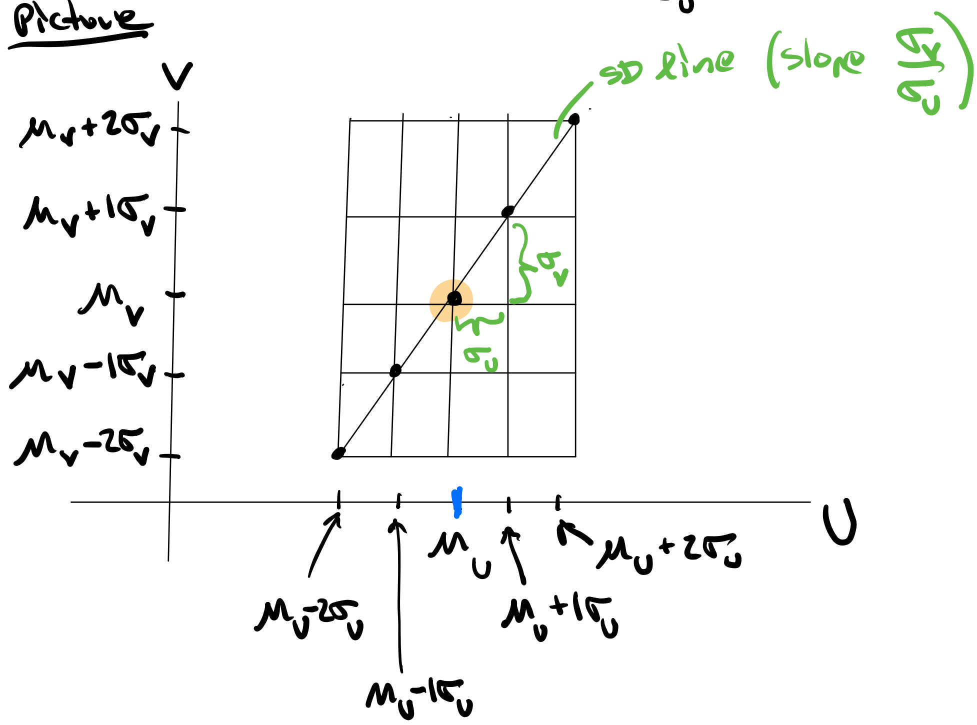
b) If you get a 70 on Test 1 what score do you predict to get on ~~test~~ 2?

We discuss "regression effect" below:

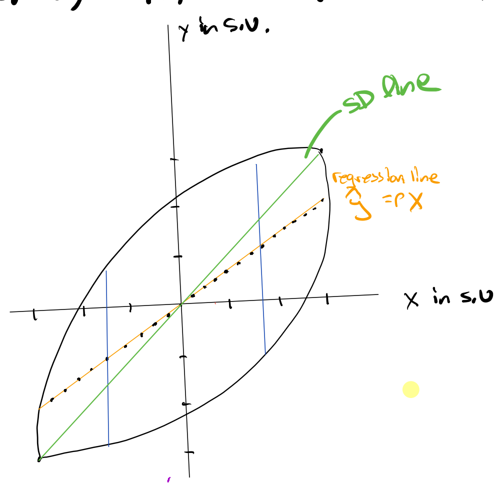
Regression line vs. SD line and regression effect

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

Picture



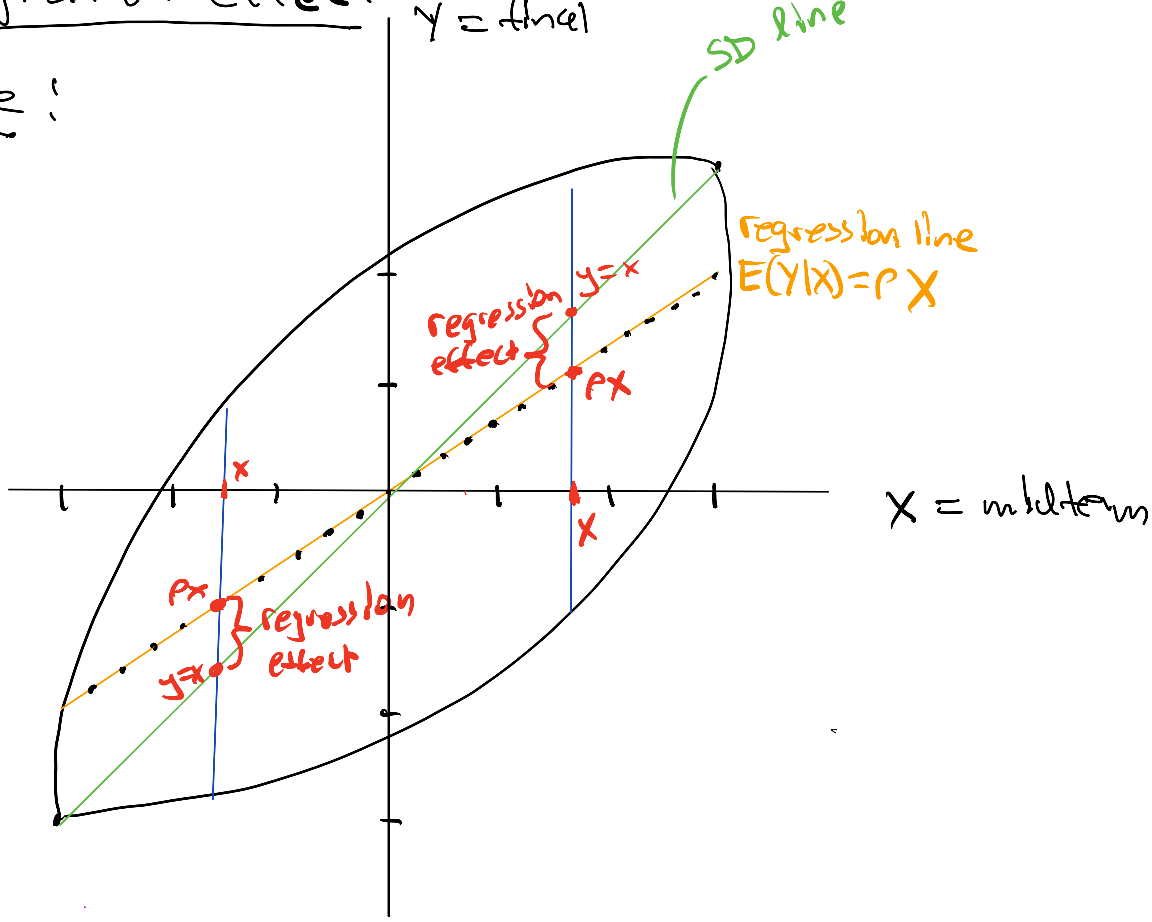
For U, V in s.u. the SD line is $y = 1.x$ ($y = V^*$, $x = U^*$)



Steeper than regression line which has slope ρ.

regression effect

Picture:



Regression effect,

Corr (test 1, 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).

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