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 $Corr(X, Y)$

Y is a linear combination of indep normals is normal.

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

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

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

$$Corr(X,Y) = \frac{Cov(X,Y)}{\underset{1}{SD(X)}\,\underset{1}{SD(Y)}} = \boxed{\rho}$$

Today

Sec 6.5 <u>Bivariate Normal</u>

<u>Defn</u> (Standard Bivariate Normal Distribution)

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

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

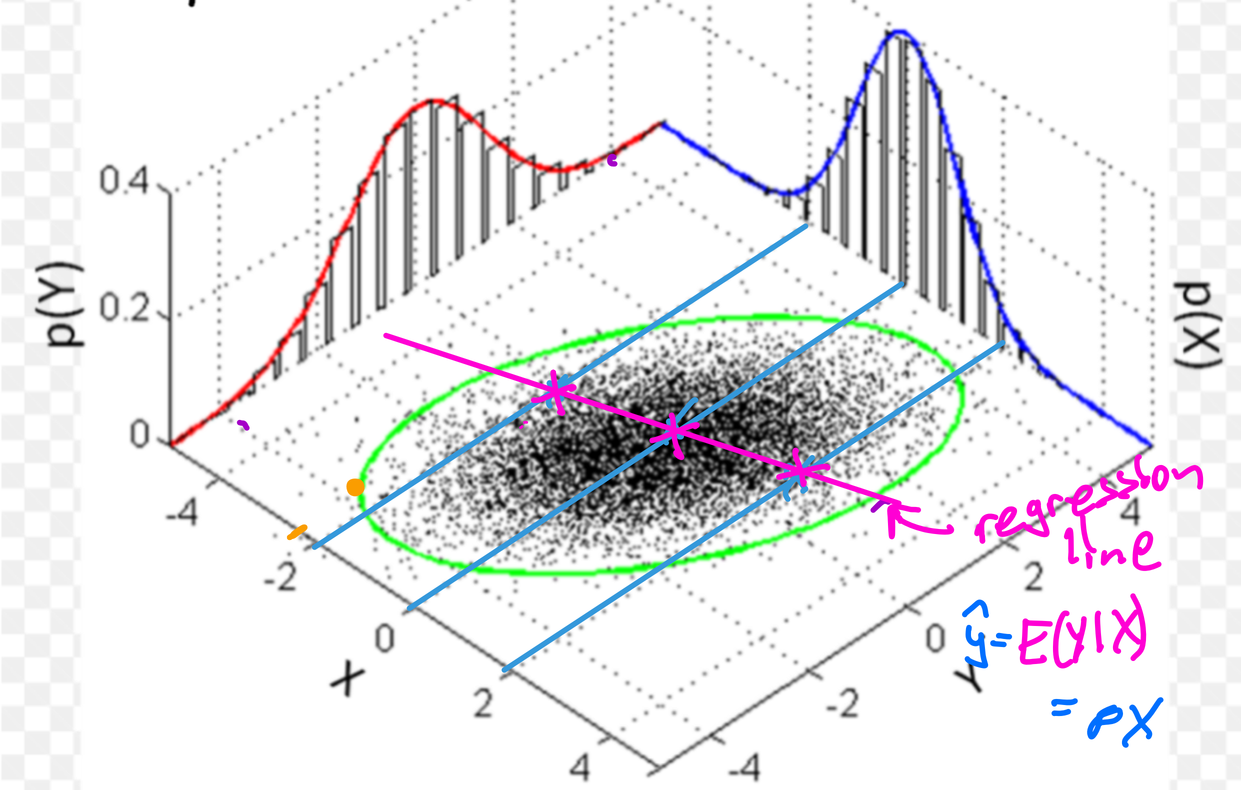
$\mathrm{Corr}(X,Y) = \rho$ see warmup

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

written $(X,Y) \sim BV(\underbrace{0}_{\mu_X}, \underbrace{0}_{\mu_Y}, \underbrace{1}_{\sigma^2_X}, \underbrace{1}_{\sigma^2_Y}, \rho)$

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

a)

$$E(Y|X) = \rho \underbrace{E(X|X)}_{=X} + \sqrt{1-\rho^2}\, \underbrace{E(Z|X)}_{=E(Z)=0}$$

$$= \boxed{\rho X}$$

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

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

b)

$$Var(Y|X) \overset{\text{indep}}{=} \underbrace{Var(\rho X | X)}_{=\rho^2 Var(X|X) = 0} + \underbrace{Var(\sqrt{1-\rho^2}\, Z | X)}_{=1-\rho^2\, Var(Z|X)}$$

$$Var(Z|X) = Var(Z) = 1$$

$$= \boxed{1-\rho^2}$$

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

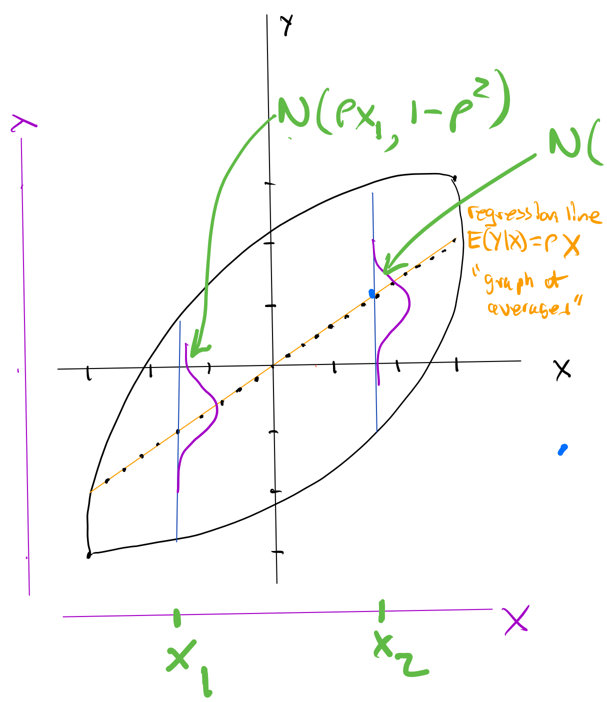
small x is a fixed constant

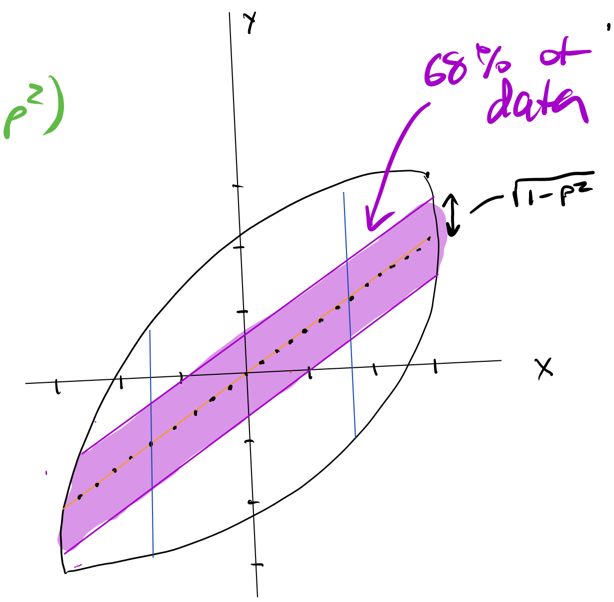
Since $Z \sim N(0,1)$ and a linear combination of normal is normal, $Y \mid X = x$ is normal.

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

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

Picture





ex

$(\text{Test 1}, \text{test 2}) \sim BV(0, 0, 1, 1, 0.6)$ ← standard bivariate normal

↑ mean ↑ variance

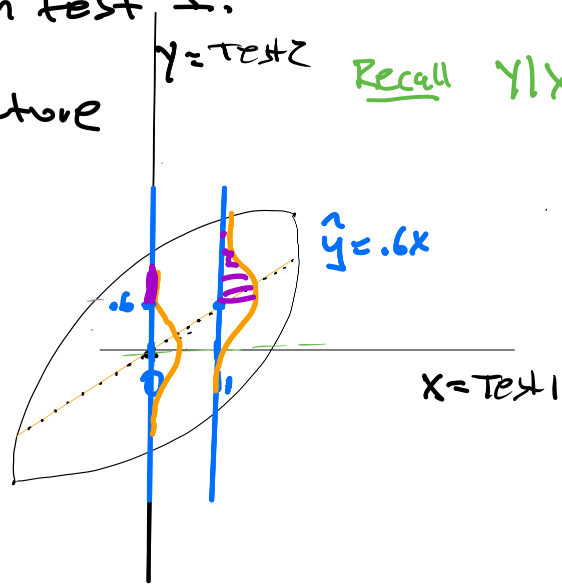
What is greater?

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



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

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

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

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

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

$\hat{y} = \dfrac{\hat{V} - \mu_V}{\sigma_V}, \qquad x = \dfrac{U - x_\mu}{\sigma_U}$

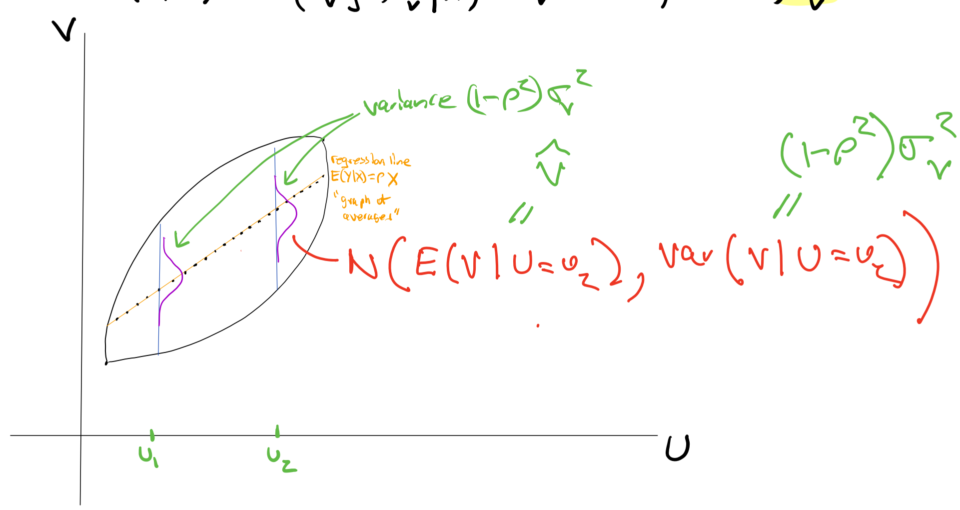
$$\iff \hat{V} - \mu_V = \frac{\sigma_V}{\sigma_U}\rho\,(U - \mu_u)$$

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

regression line.

furthermore,

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



ex Test 1 is $\mu_U = 60$, $\sigma_U = 20$

Test 2 is $\mu_V = 60$, $\sigma_V = 20$ $\Big\}$ $\rho = .6$

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.

$$\hat{V} = \frac{20}{20}(.6)U + 60 - \frac{20}{20}(.6)(60)$$

$$= \boxed{.6U + 24}$$

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

$$\hat{V} = E(V \mid U=70) = .6(70) + 24 = \boxed{66}$$

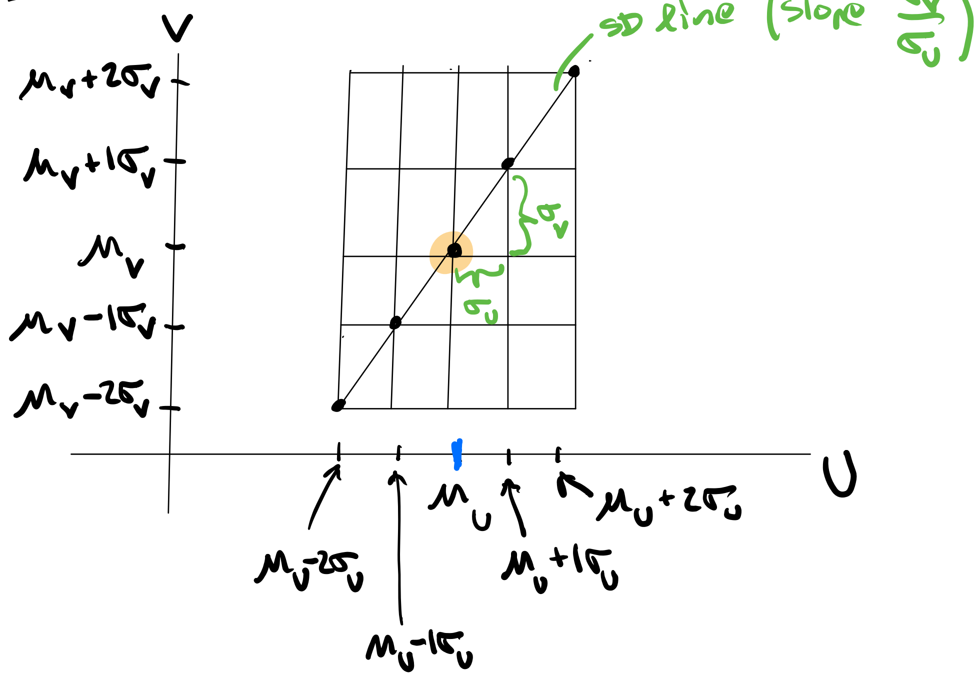
We can see the "regression effect" here. 70 is 1SD above the mean and 66 is 6/20 SD above the mean. On Test 2 your predicted score is smaller (regressed towards the mean of 60).

We discuss "regression effect" below:

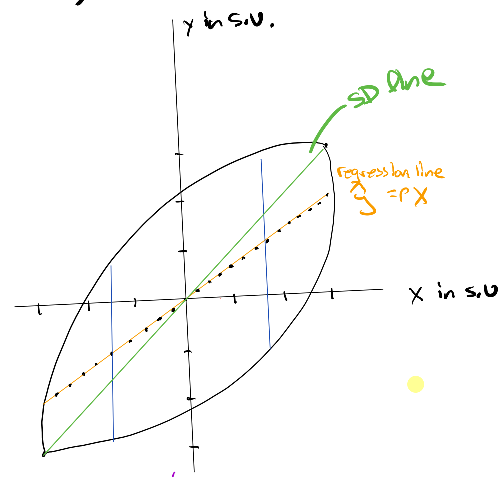
Regression line vs. SD line and regression effect

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

Picture



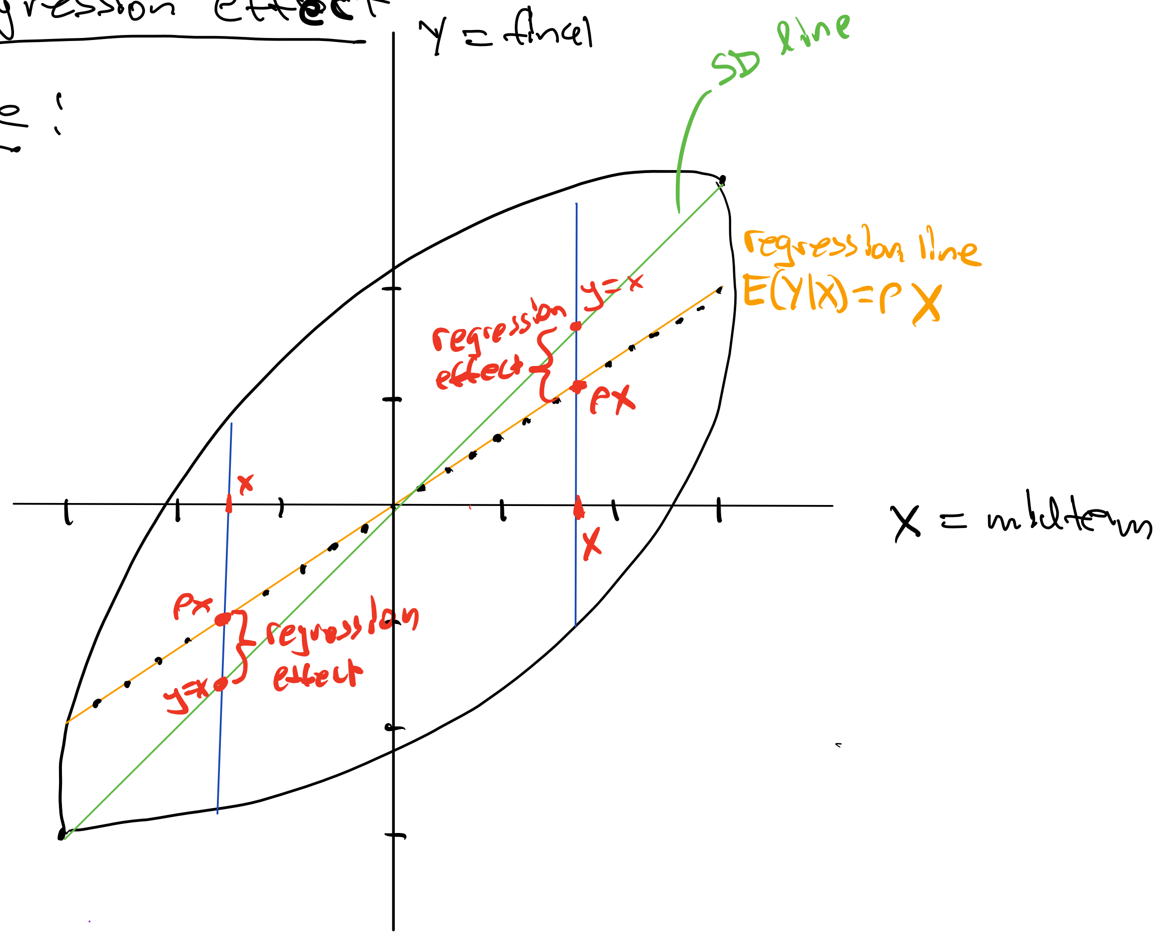
For U, V in s.u. the SD line is $y = 1.x$, with $y = V^*$ and $x = U^*$



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).

ex

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

The area in green below represents the percentage of students getting over 75 on Physics who got 75 in Math. This is $< 50\%$