Stat 134 lec 38

Warmup 8:00 – 8:10 AM

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

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

① What distribution is Y?

Y is a linear combination of independent normals and hence normal.

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

$$Var(Y) = Var\left(\rho X + \sqrt{1-\rho^2}\, Z\right) = \rho^2 \underbrace{Var(X)}_{=1} + (1-\rho^2)\underbrace{Var(Z)}_{=1} = 1$$

$$\Rightarrow Y \sim N(0,1)$$

② What is $Corr(X, Y)$

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

$$Corr(X,Y) = \frac{\overbrace{Cov(X,Y)}^{=\rho}}{\underbrace{SD(X)}_{=1}\underbrace{SD(Y)}_{=1}} = \boxed{\rho}$$

Final Exam: Monday Dec 16 VLSB 2050 7-10pm

last time

Sec 6.4 Correlation

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{SD(X)\,SD(Y)} = E(X^*Y^*)$$

X in std units

$-1 \le \text{Corr}(X,Y) \le 1$

+ if $ac > 0$

$$\text{Corr}(aX+b, cY+d) = \pm\,\text{Corr}(X,Y) \quad a,b,c,d \in \mathbb{R},\ a,c \neq 0.$$

$N_1 + \cdots + N_k = c$ a sum of exchangeable RVs

$$\text{Corr}(N_1, N_2) = -\frac{1}{k-1}$$

Today

1. sec 6.5 the bivariate normal distribution

① Sec 6.5 <u>Bivariate Normal</u>

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

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

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

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

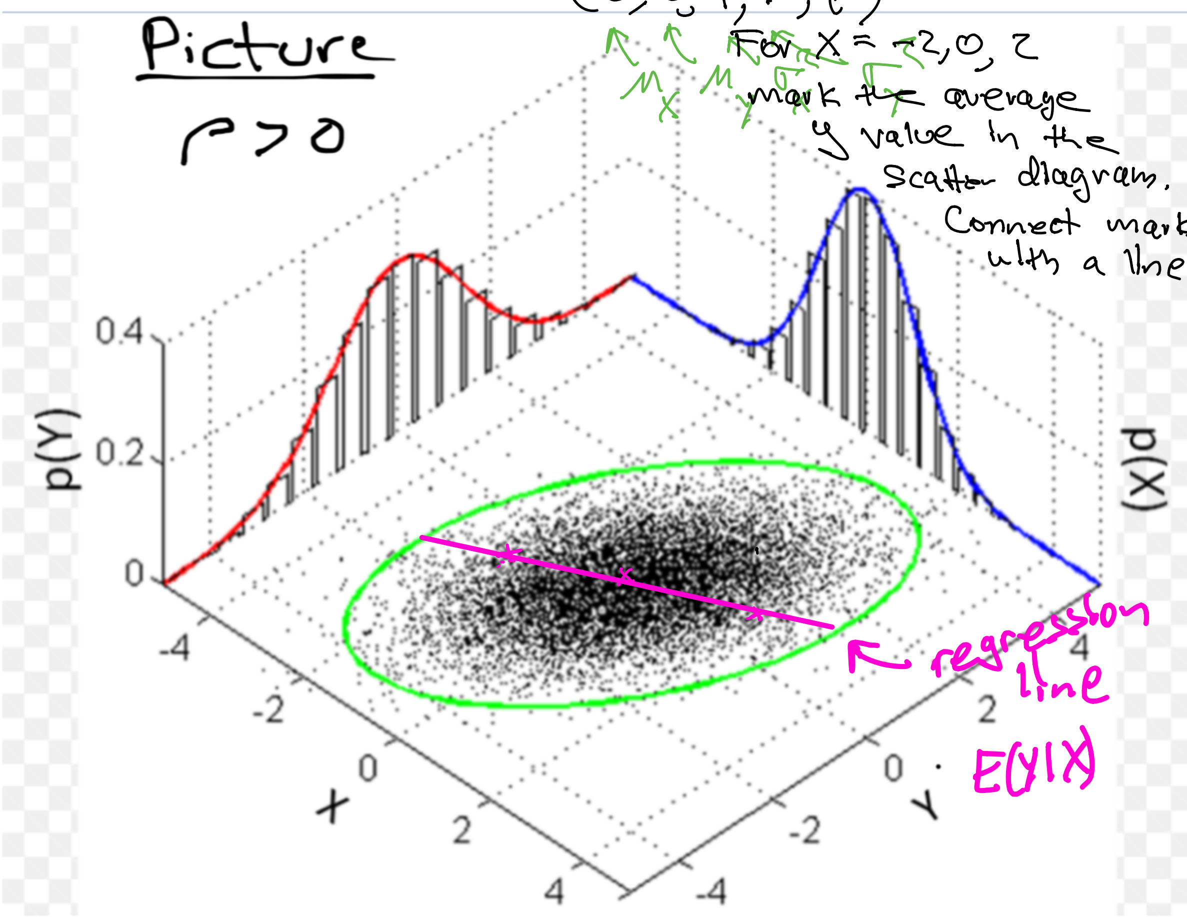
Written $(X,Y) \sim BV(0, 0, 1, 1, \rho)$

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

<u>Picture</u>

$\rho > 0$

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



← regression line

$E(Y|X)$

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

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

Find the "regression line" $\hat{y} = E(Y|X)$

$$\hat{y} = E(Y|X) = E(\rho X + \sqrt{1-\rho^2}\, Z \mid X)$$

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

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

$$= \rho X$$

Find the conditional variance $Var(Y|X)$

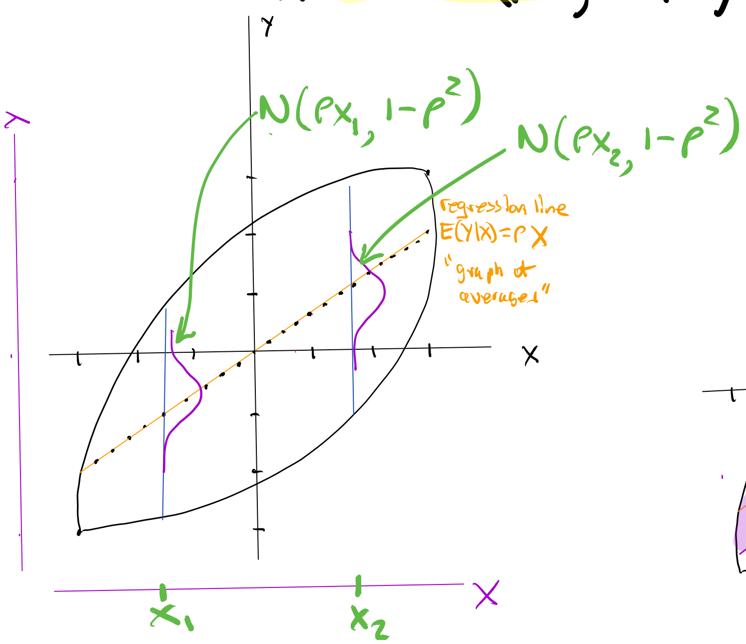
$$Var(Y|X) = Var(\rho X + \sqrt{1-\rho^2}\, Z \mid X)$$

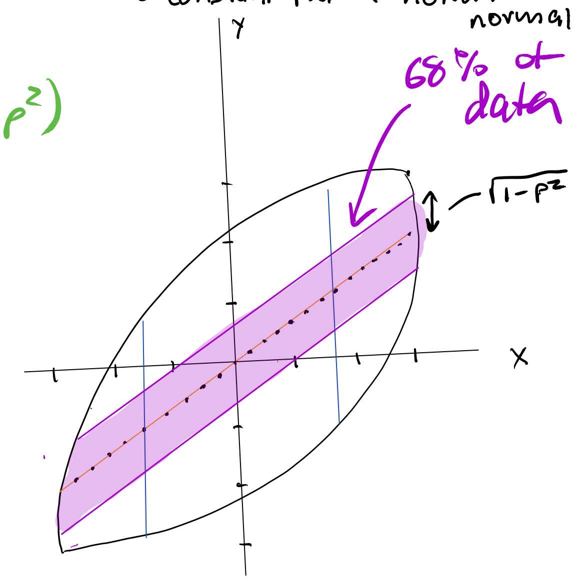
$$= Var(\rho X|X) + Var(\sqrt{1-\rho^2}\, Z \mid X) \quad \text{since } X, Z \text{ indep.}$$

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

$$= 1-\rho^2$$

Hence $Y|X \sim N(\rho X, 1-\rho^2)$ since $Y = \underbrace{\rho X}_{\text{constant given } X,} + \sqrt{1-\rho^2}\, Z$ and a constant plus a normal is normal.





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

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

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

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

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

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

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

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

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

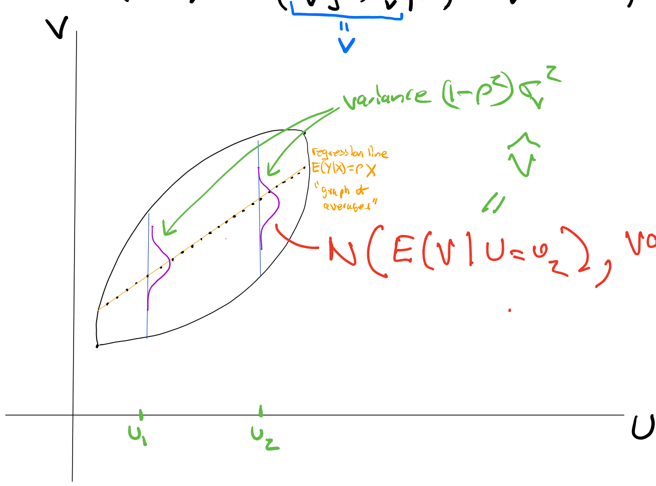
$$\Longleftrightarrow \quad \hat{V} - \mu_V = \frac{\sigma_V}{\sigma_U}\rho(U-\mu_U)$$

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

regression line.

furthermore,

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



$$N\Big(\underbrace{E(V|U=u_2)}_{=\hat{V}},\ \underbrace{\mathrm{var}(V|U=u_2)}_{=(1-\rho^2)\sigma_V^2}\Big)$$

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 $\hat{V}$

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

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

$$= \frac{20}{20}(.6)U + 60 - \frac{20}{20}(.6)(60) = \boxed{.6U + 24}$$

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

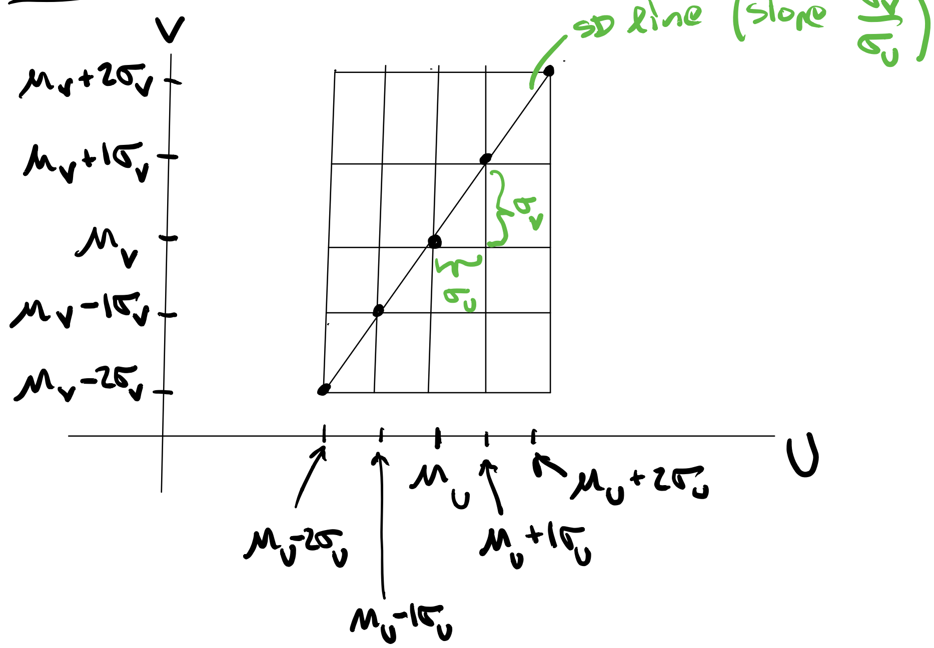
$$E(V \mid U = 70) = .6(70) + 24 = \boxed{66}$$

Notice you did relatively worse. Your test 1 score was $\frac{10}{20}$ sd above average but your test 2 score was only $\frac{6}{20}$ sd above average.
This is the "regression effect".

Regression line vs. SD line and regression effect

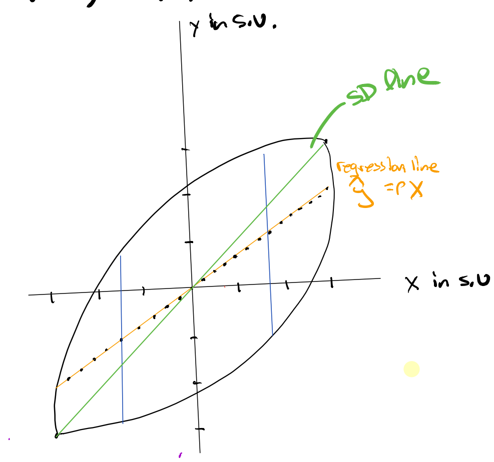
Def^n The SD line is $V - \mu_V = \frac{\sigma_V}{\sigma_U}(U - \mu_U)$.

Picture



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

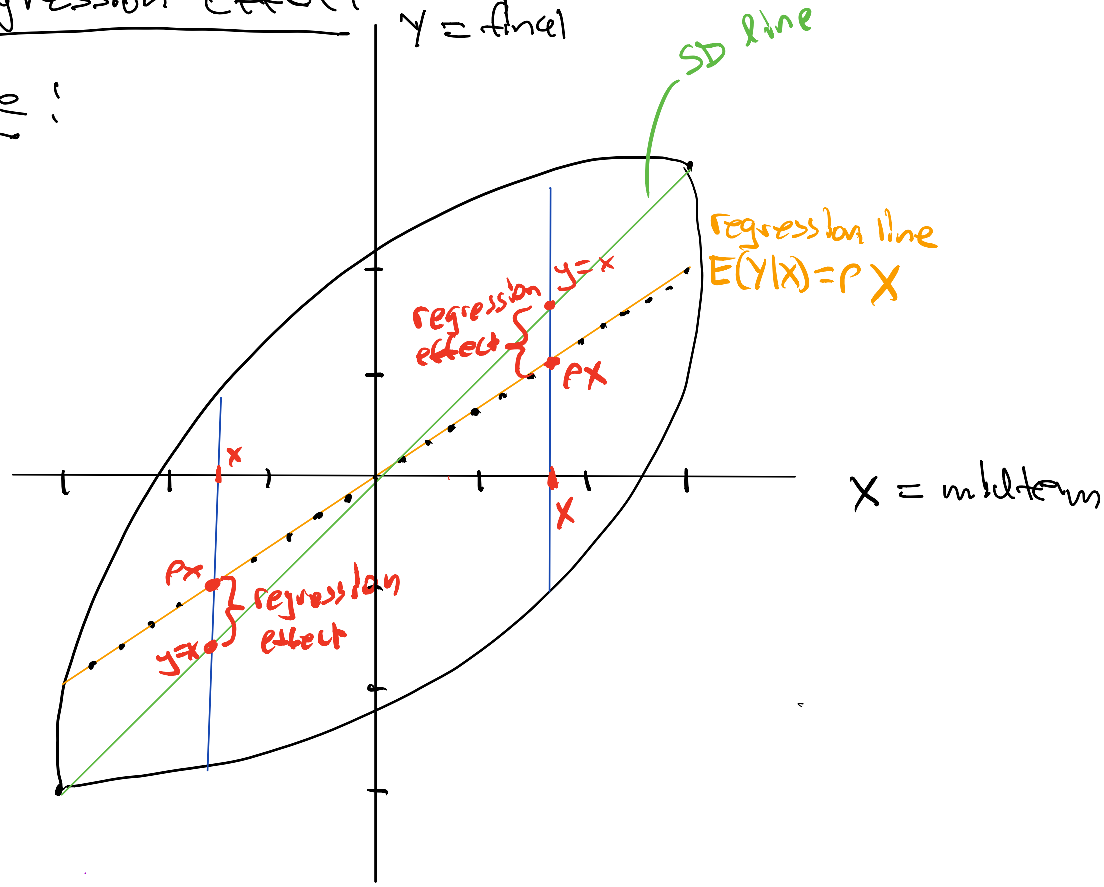
$y = V^*$, $x = U^*$



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).