Stat 134 lec 39

Warmup

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 $\mathrm{Corr}(X,Y)$

Y is a linear combination of independent normals which is normal,

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

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

$$= \rho^2 \underbrace{\mathrm{Var}(X)}_{1} + (1-\rho^2)\underbrace{\mathrm{Var}(Z)}_{1} = 1$$

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

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\underbrace{SD(X)}_{1}\underbrace{SD(Y)}_{1}} = \mathrm{Cov}(X,Y) = \mathrm{Cov}(X, \rho X + \sqrt{1-\rho^2}\, Z)$$

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

$$= \rho$$

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

We will see $(X,Y) \sim BVN$

Announcement

Schedule

- W sec 6.5
- F sec 6.5
- M review + OH
- W off
- F GSI review + OH
- Th Final exam

will post review materials and practice final.

Last time

let
$$X = \frac{U - \mu_U}{\sigma_U} \quad \text{— } U \text{ in std units}$$
$$Y = \frac{V - \mu_x}{\sigma_V} \quad \text{— } V \text{ in std units}$$

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

Today

Sec 6.5 <u>Bivariate Normal</u>

Sec 6.5 Bivariate Normal

1

Def^n (Standard Bivariate Normal Distribution)

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

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

$$\mathrm{corr}(X,Y) = \rho$$

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

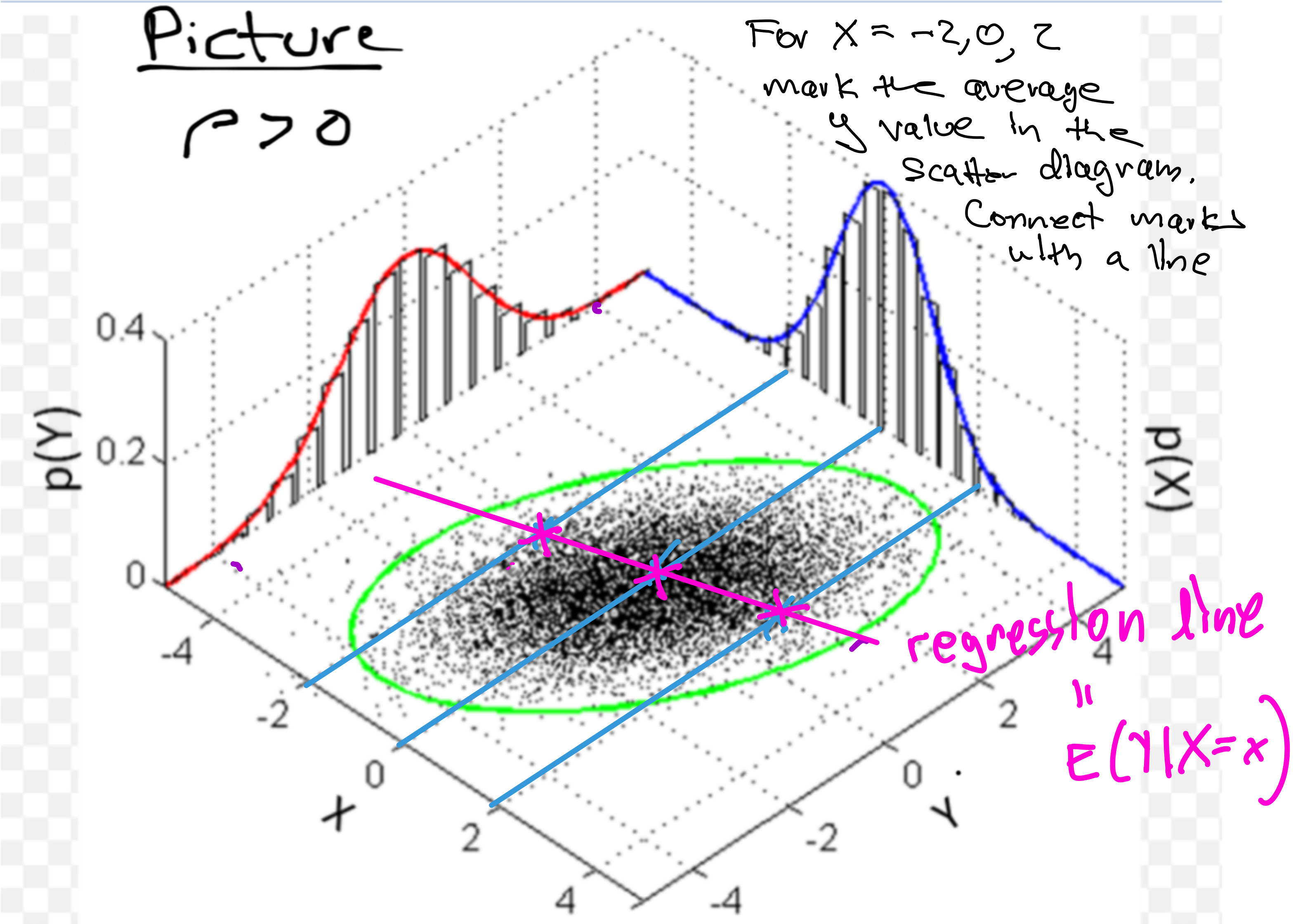
written $(X,Y) \sim \mathrm{BVN}(\underbrace{0}_{\mu_X}, \underbrace{0}_{\mu_Y}, \underbrace{1}_{\sigma_X^2}, \underbrace{1}_{\sigma_Y^2}, \rho)$

$$(X,Y) \sim \mathrm{BVN}(0,0,1,1,\rho)$$

The scatter diagram of BVN always, with $\rho \neq 0$ looks like an upwards or downwards facing football:

Picture

$\rho > 0$



When (X,Y) is standard BVN, the best fitting line through your data is approximately $E(Y|X=x) = \rho x$.

We call $E(Y|X=x) = \rho x$ the regression line

Note The regression line is actually the best fitting line through your data but with many points this is approx $E(Y|X=x)$

ex

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

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

a) Show that $Y \mid X=x$ is normaly distributed.

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

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

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

$= \rho x + \sqrt{1-\rho^2}\, Z \leftarrow$ this is normal.

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

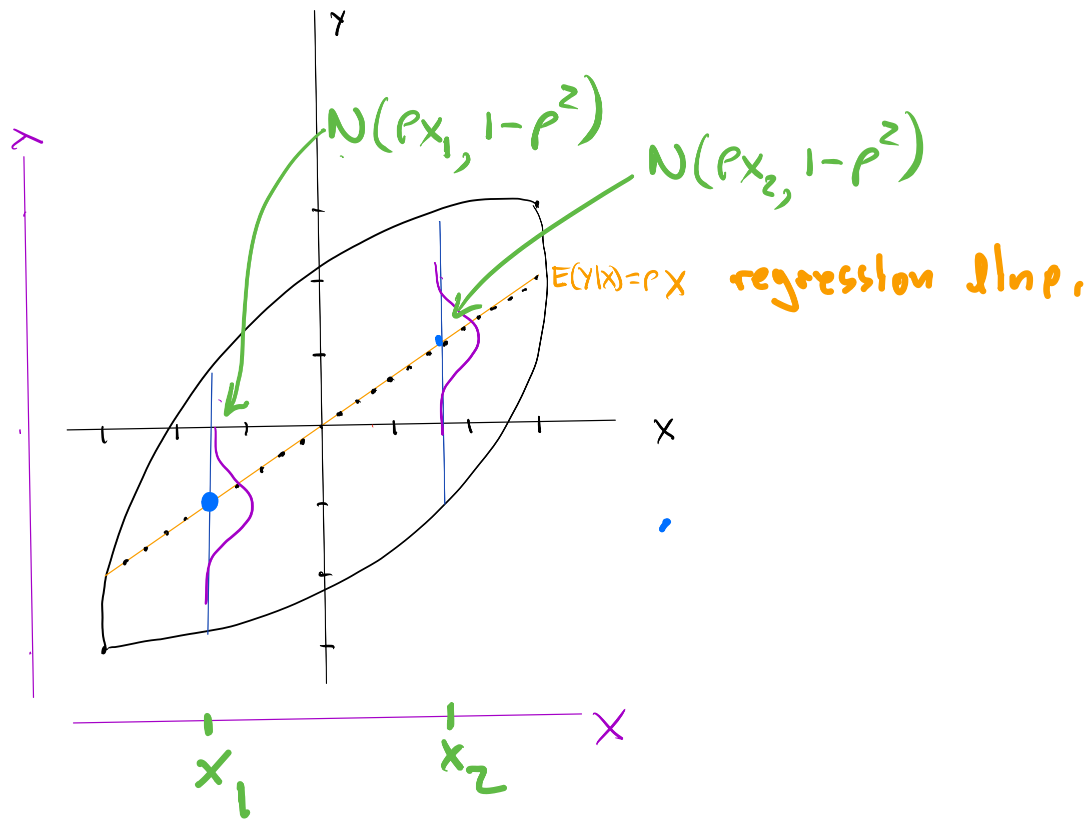
c) $Var(Y \mid X=x) = Var(\rho x + \sqrt{1-\rho^2}\, Z) = 1-\rho^2$

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

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

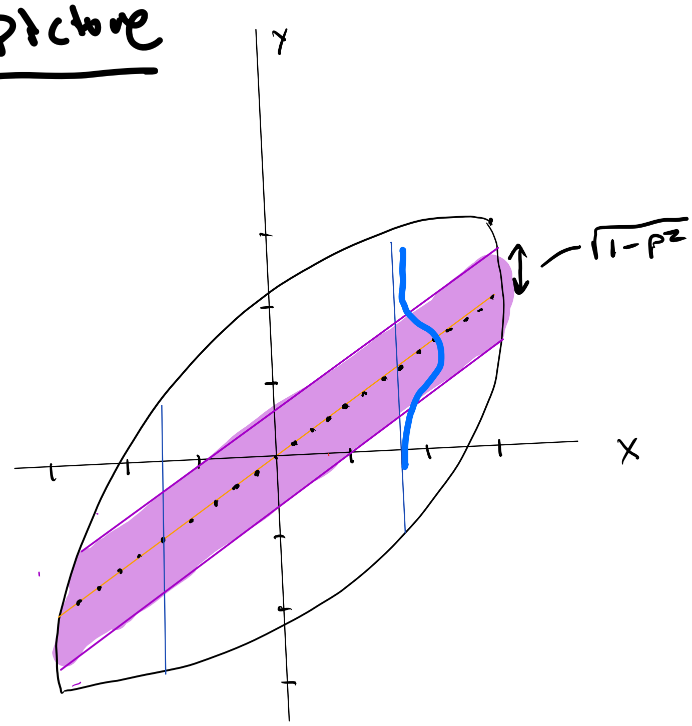
We call the line $E(Y|X=x) = \rho x$ the **regression line**

Picture



What is the chance your data is more than 1 SD outside of the regression line? — $1/3$

Picture



For every x, the chance you are within 1SD of the mean is $2/3$

Hence $1/3 = 1 - 2/3$

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

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

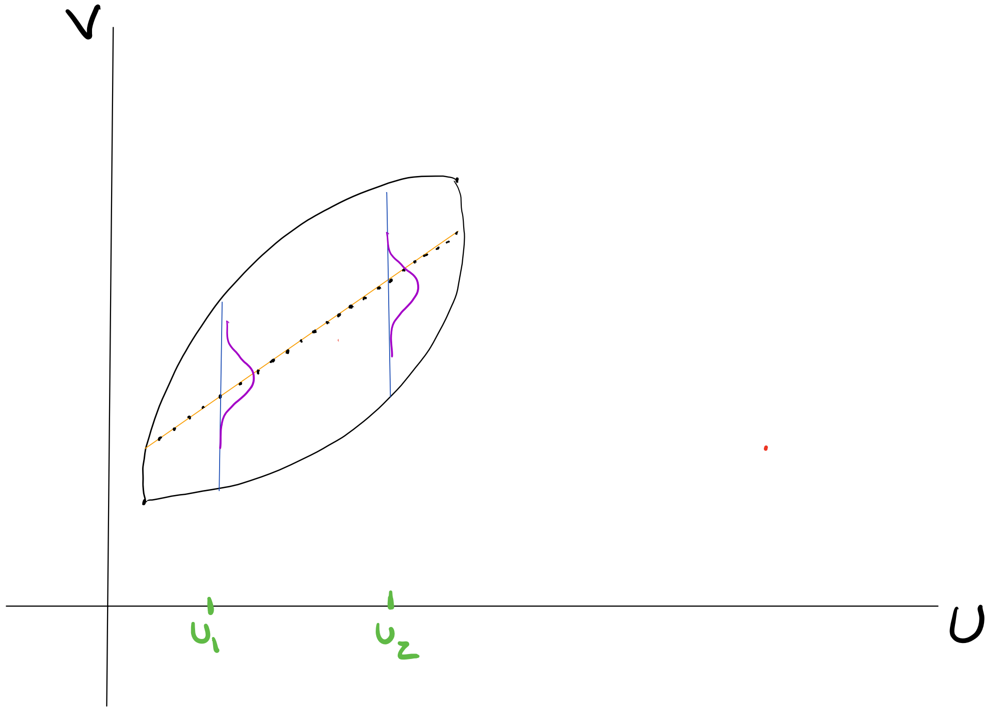
have std. bivariate normal distribution with corr ρ.

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

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

See the appendix for the equation of the regression line $E(U|V=v)$.

Picture

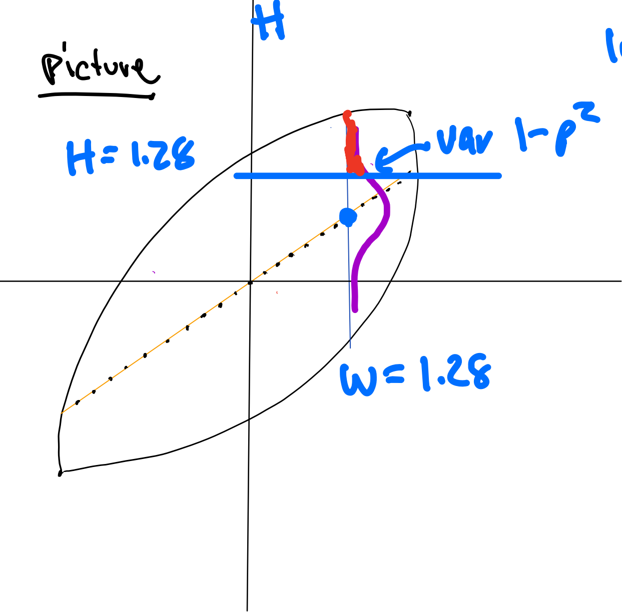


ex

3. Heights and weights of a large group of people follow a bivariate normal distribution, with correlation 0.75. Of the people in the 90th percentile of weights, about what percentage are above the 90th percentile of heights?

Fact $\Phi^{-1}(.9) = 1.28$

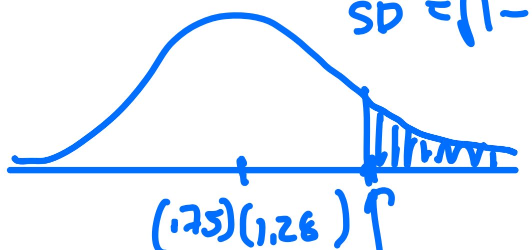
Picture



let W = weight in S.U.
H = height in S.U

$\Phi^{-1}(.9) = 1.28$

$SD = \sqrt{1-\rho^2} = \sqrt{1-(.75)^2} = .66$



H | W = 1.28

$$\boxed{1 - \Phi\left(\frac{1.28 - (.75)(1.28)}{.66}\right)}$$

1. Here is a summary of Pre-SAT and SAT scores of a large group of students.

$X =$	PSAT scores:	average: 1200	SD: 100
$Y =$	SAT scores:	average: 1300	SD: 90
		correlation: 0.6	

Assume the data are approximately bivariate normal in distribution.

Of the students who scored 1000 on the PSAT, about what percentage scored above average on the SAT?

Find $P(Y > 1300 \mid X = 1000)$

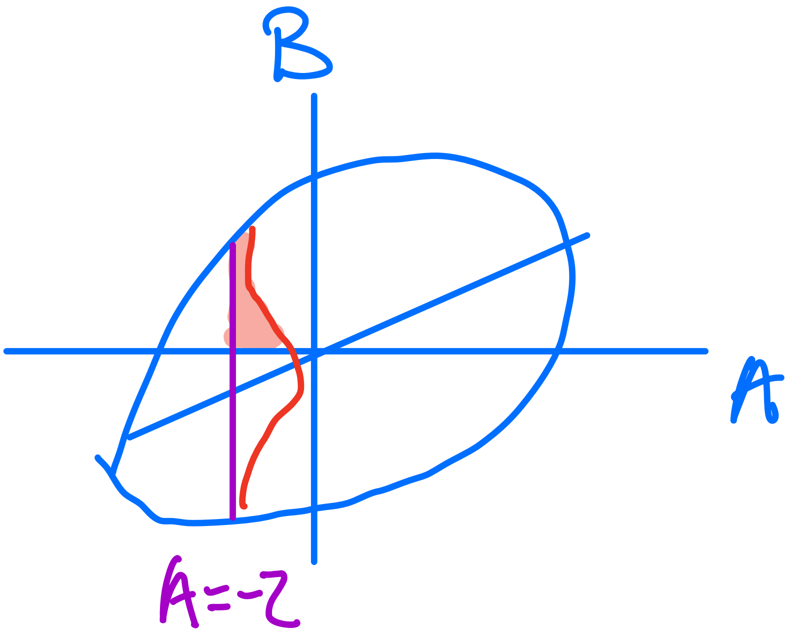
let B = Y in s.u

A = X in s.u.

Find $P\left(B > \frac{1300-1300}{90} \;\middle|\; A = \frac{1000-1200}{100}\right)$

$= P(B > 0 \mid A = -2)$

$SD = \sqrt{1-(.6)^2} = .8$



$(.6)(-2) = -1.2$ 0

$1 - \Phi\left(\frac{0-(-1.2)}{.8}\right)$

Appendix

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

The regression line of (U,V) is

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

Pf/

let $(X,Y) \sim BV(0,0,1,1,\rho)$ where $X = \dfrac{U-\mu_U}{\sigma_U}$, $Y = \dfrac{V-\mu_V}{\sigma_V}$

$E(Y|X) = \rho X$ is regression line in s.v.

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

so $$\frac{E(V|U)-\mu_V}{\sigma_V} = \rho\,\frac{U-\mu_U}{\sigma_U}$$

$$\Leftrightarrow \quad E(V|U) - \mu_V = \frac{\sigma_V}{\sigma_U}\rho(U-\mu_u)$$

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

$\square$