Stat 134 lec 40

Warmup 1:00 – 1:10 pm

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

↖ mean ↖ variance

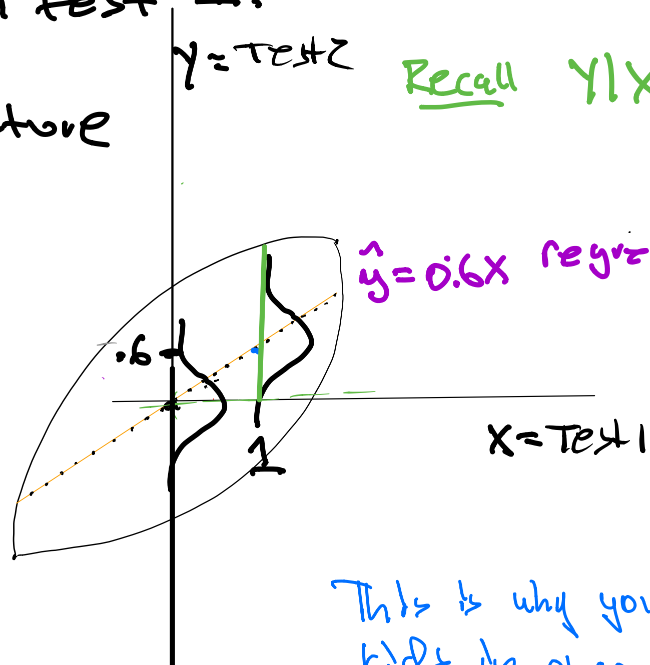
What is greater?

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



This is why you sit next to the successful kids in class. You have a higher chance of doing well on the next test!

Last time

Sec 6.5 Bivariate Normal

Def^n (Standard Bivariate Normal Distribution)

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

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

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

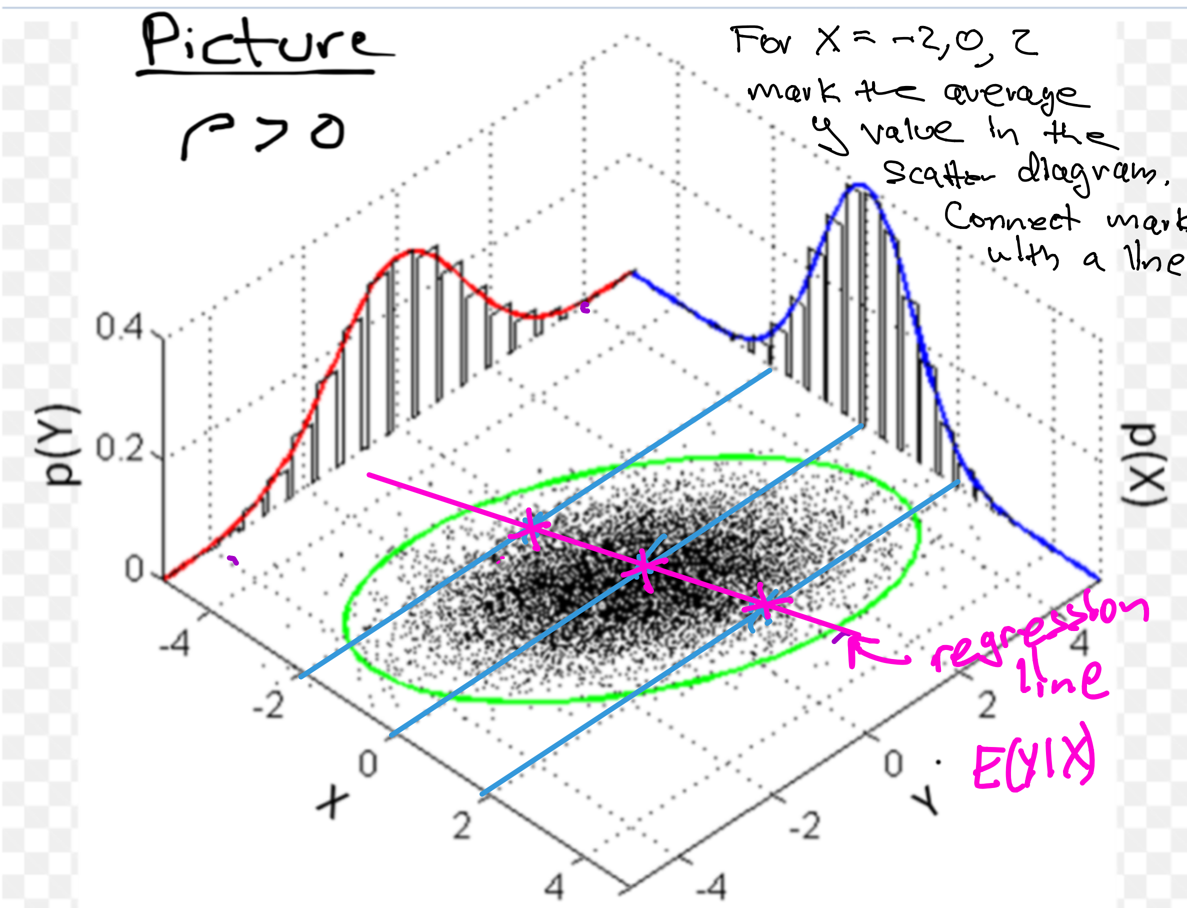
We call the joint distribution (X,Y) the standard bivariate normal with $\mathrm{corr}(X,Y) = \rho$

written $(X,Y) \sim BV(\underset{\mu_X}{0}, \underset{\mu_Y}{0}, \underset{\sigma_X^2}{1}, \underset{\sigma_Y^2}{1}, \rho)$

Picture

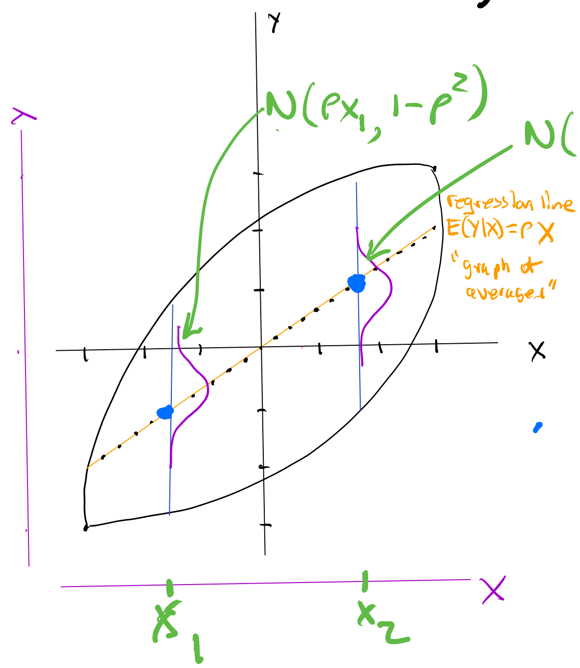
$\rho > 0$

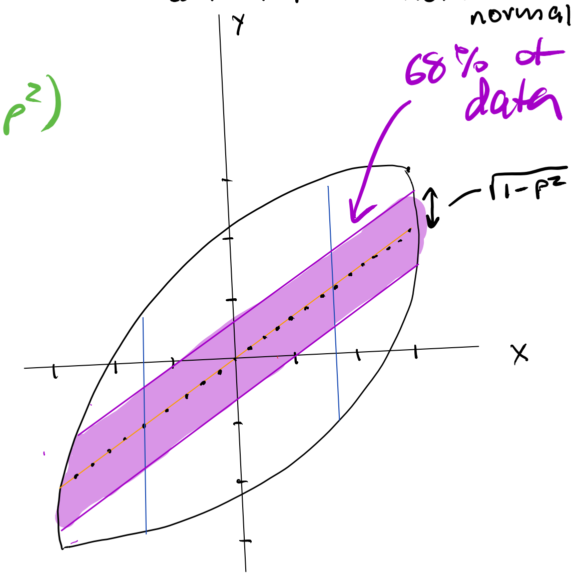
For $X = -2, 0, 2$ mark the average y value in the scatter diagram. Connect marks with a line



constant when X is fixed

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





Today ① sec 6.5 Bivariate Normal

Next week

① M 1–2pm Properties of Bivariate Normal + review

② W 1–2pm Review

③ F 1–2pm Review

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

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

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

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

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

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 $\hat{V}$

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

V intercept

slope

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

regression line

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

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

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

Summary

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

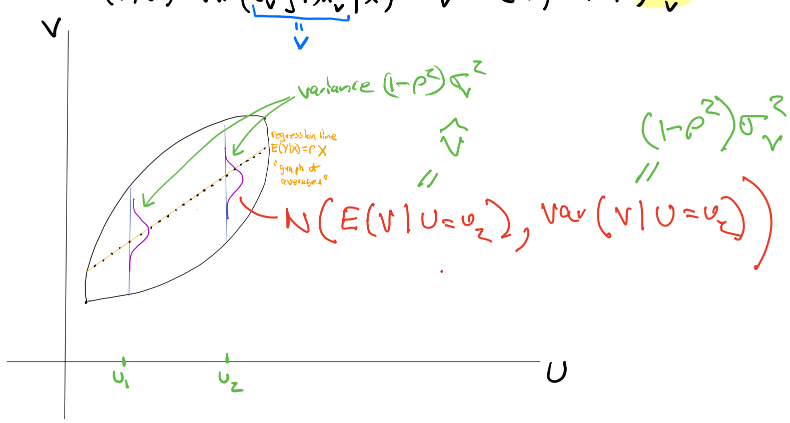
$\hat{y} = \frac{\hat{V} - \mu_V}{\sigma_V}$, $\quad x = \frac{U - \mu_U}{\sigma_U}$

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

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

furthermore,

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



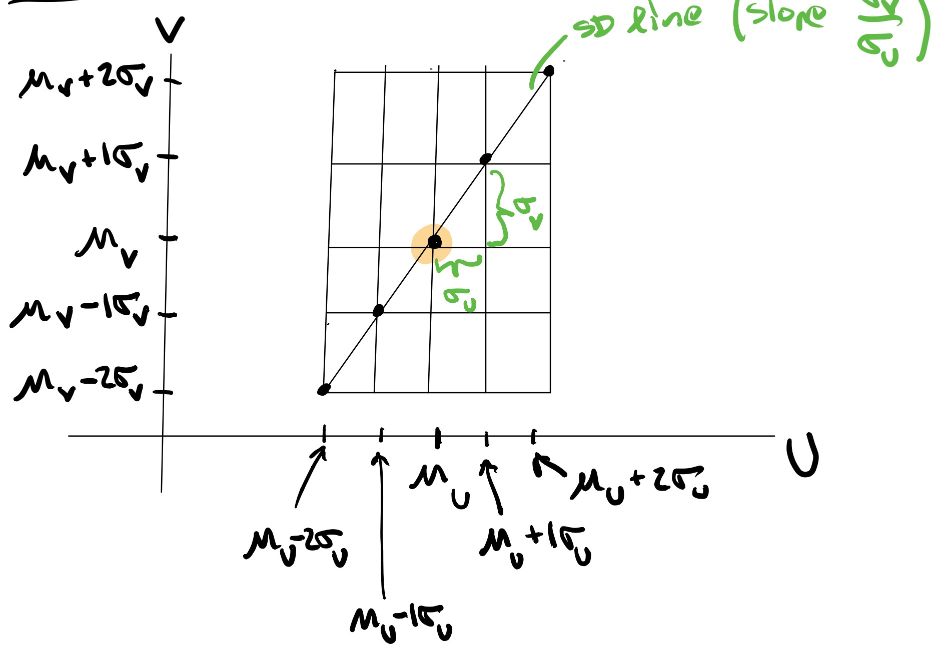
Notice you did relatively worse. Your test 1 score was $\frac{10}{20}$ sd above average but your test 2 score was only $\frac{6}{20}$ sd above average.

This is the "regression effect".

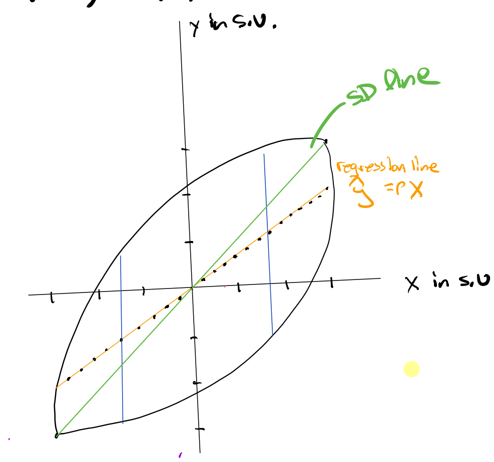
Regression line vs. SD line and regression effect

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

Picture



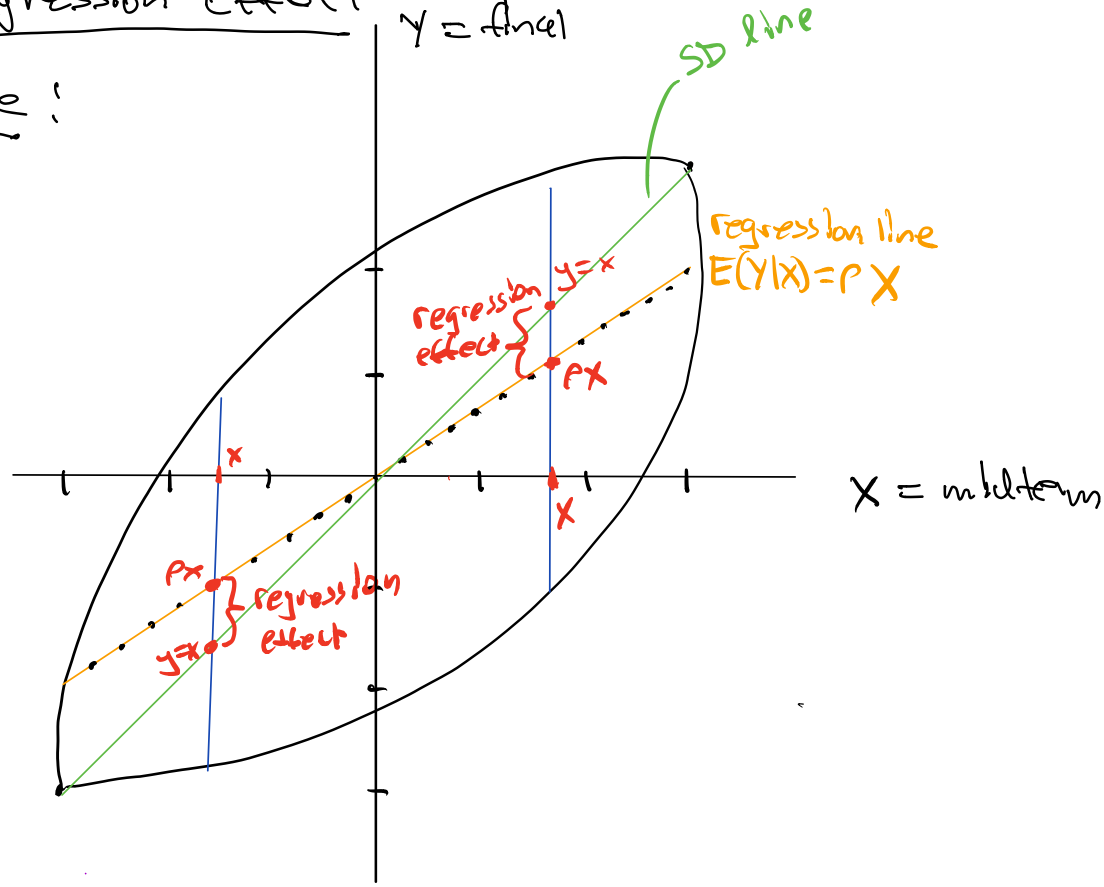
For U, V in s.u. the SD line is $y = 1 \cdot x$, where $y = V^*$ and $x = U^*$.



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).

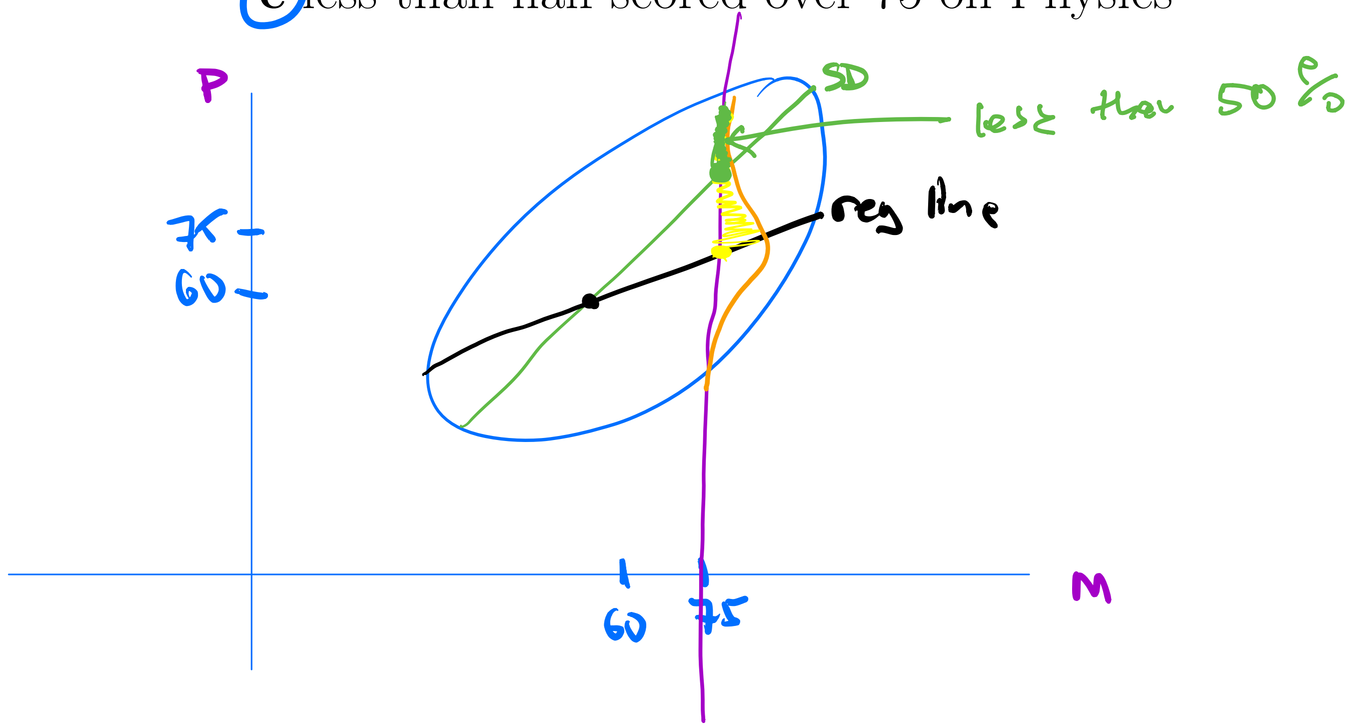
ex

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



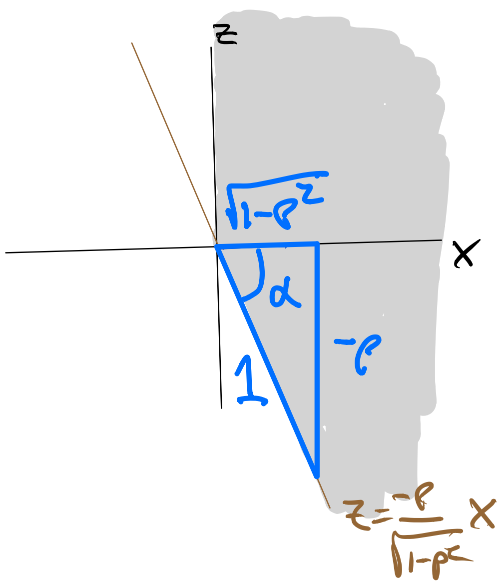
eg: X, Y std bivariate normal, $\rho > 0$

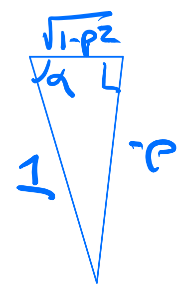
Find $P(X>0, Y>0)$

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

$$= P\left(X>0,\ Z > \frac{-\rho}{\sqrt{1-\rho^2}} X\right)$$

→ note X and Z are uncorrelated so joint (X,Z) is a symmetric bell over x,z plane.





$$\tan\alpha = \frac{-\rho}{\sqrt{1-\rho^2}}$$

$z = \frac{-\rho}{\sqrt{1-\rho^2}} x$ $\quad \alpha = \tan^{-1}\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)$

$$P\left(X>0,\ Z > \frac{-\rho}{\sqrt{1-\rho^2}} X\right) = \frac{90 + |\alpha|}{360} = \boxed{\frac{90 + \left|\tan^{-1}\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)\right|}{360}}$$