Stat 134 lec 37

Last time Sec 6.4

Sec 6.4

Covariance

$Cov(X,Y) = E(XY) - E(X)E(Y)$

$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$

$= E(X^* Y^*)$

X^* X in s.u.

Let $N_1, \dots, N_k$ be exchangable RVs.

If $N_1 + \dots + N_k = c$ then

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

X,Y indep $\Rightarrow Corr(X,Y) = 0$

$Corr(X,Y) = 0 \not\Rightarrow X,Y$ indep

ex $X = Unif(-1,1)$, $Y = X^2$

$E(XY) - E(X)(Y) = E(X^3) - E(X)E(X^2) = 0$

$E(X^3) = \frac{1}{2}\int_{-1}^{1} x^3 dx = \frac{x^4}{8}\Big|_{-1}^{1} = 0$, $E(X) = 0$

$\Rightarrow Cov(X,Y) = 0 \Rightarrow Corr(X,Y) = 0$ but X,Y dependent

TODAY

SEC 6.4

1. Review responses to concept test
2. Properties of Correlation

Sec 6.5

3. Bivariate Normal

Stat 134

Monday Nov 26 2018

1. Consider a Poisson(λ) process. Let $T_r \sim$ gamma(r, λ) be the rth arrival time. Cov(T_1, T_3) equals:

   $\text{Var}(T_r) = \frac{r}{\lambda^2}$

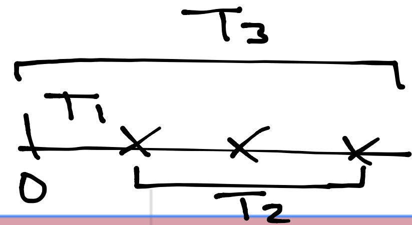
   a λ

   b λ^2

   c $1/\lambda^2$

   d none of the above

Discuss how you did this.



$$T_3 = T_1 + T_2$$

$$\mathrm{Cov}(T_1,T_3) = E(T_1T_3) - \underbrace{E(T_1)}_{\frac{1}{\lambda}}\underbrace{E(T_3)}_{\frac{3}{\lambda}}$$

$$E(T_1T_3) = E(T_1.(T_1+T_2))$$

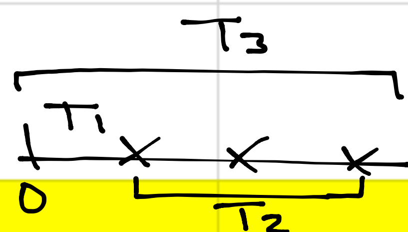
$$= E(T_1^2) + E(T_1)E(T_2)$$

$$= \mathrm{var}(T_1) + E(T_1)^2 + E(T_1)E(T$$

$$= \frac{1}{\lambda^2} + \frac{1}{\lambda^2} + \frac{1}{\lambda}\cdot\frac{2}{\lambda} = \frac{4}{\lambda^2}$$

$$\Rightarrow \mathrm{Cov}(T_1,T_3) = \frac{4}{\lambda^2} - \frac{3}{\lambda^2} = \frac{1}{\lambda^2}$$

57	d	d	
:1	c	c	E(T1+T3) - E(T1)E(T3) = (r+1)/lamda2 - r/lamda2
14	c	c	Cov = var(T1) = 1/lambda^2
21	c	c	
24	d	d	
28	c	c	I actually got -1/lambda^2 but I think I made a mistake somewhere
29	d	c	
33	c	c	It is easier to compute cov T1 T3
34	b	c	
34	d	c	
35	c	c	
38	c	c	
45	c	c	
48	c	c	
50	c	c	
54	c	c	
13	d	d	Var(Tr) = r/(lambda)^2; so D(Tr) = sqrt(Tr)/lambda. Cov(T1,T3) = (1/lambda)(sqrt(3)/lamba) or sqrt(3)/lambda^2
0:	d	c	Write T3 as T1+T2 (time to first arrival + time of next two arrivals). Then, cov(T1,T3) = cov(T1,T1+T2) = Var(T1) + cov(T1,T2). Cov(T1,T2) = 0 (this is clear from the addition of variance formula for covariance). Thus, cov(T1,T3) = var(T1) = 1/lambda^2
54	c	c	



$$T_3 = T_1 + T_2$$

$$\mathrm{Cov}(T_1,T_3)$$

$$= \mathrm{Cov}(T_1, T_1+T_2)$$

$$= \mathrm{Var}(T_1) + 0$$

$$= \frac{1}{\lambda^2}$$

Properties of correlation — Proved end of lecture.

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

1) $-1 \le \text{Corr}(x, y) \le 1$

2) correlation is invariant to linear changes of scale except possibly by a sign.

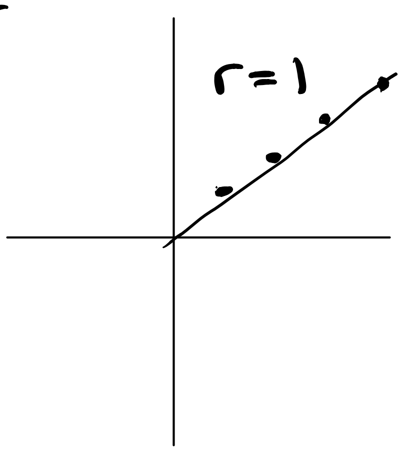
In other words:

$|\text{Corr}(x, y)| = |\text{Corr}(aX+b, cY+d)|$ for constants a, b, c, d with $a \neq 0$, $c \neq 0$.

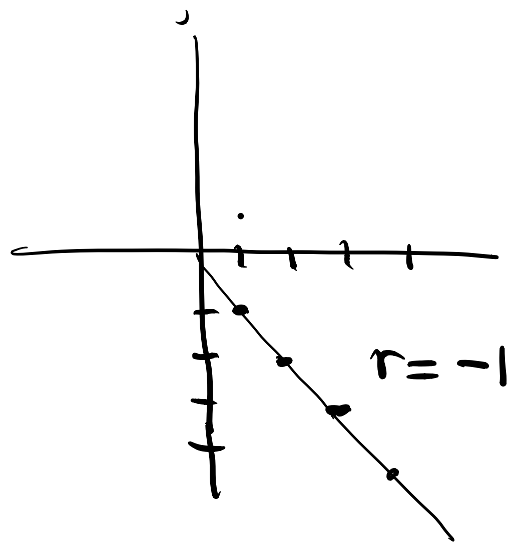
ex Correlation between Boston and NYC temperatures is the same whether temps in $°C$ or $°F = 1.8°C + 32$

ex

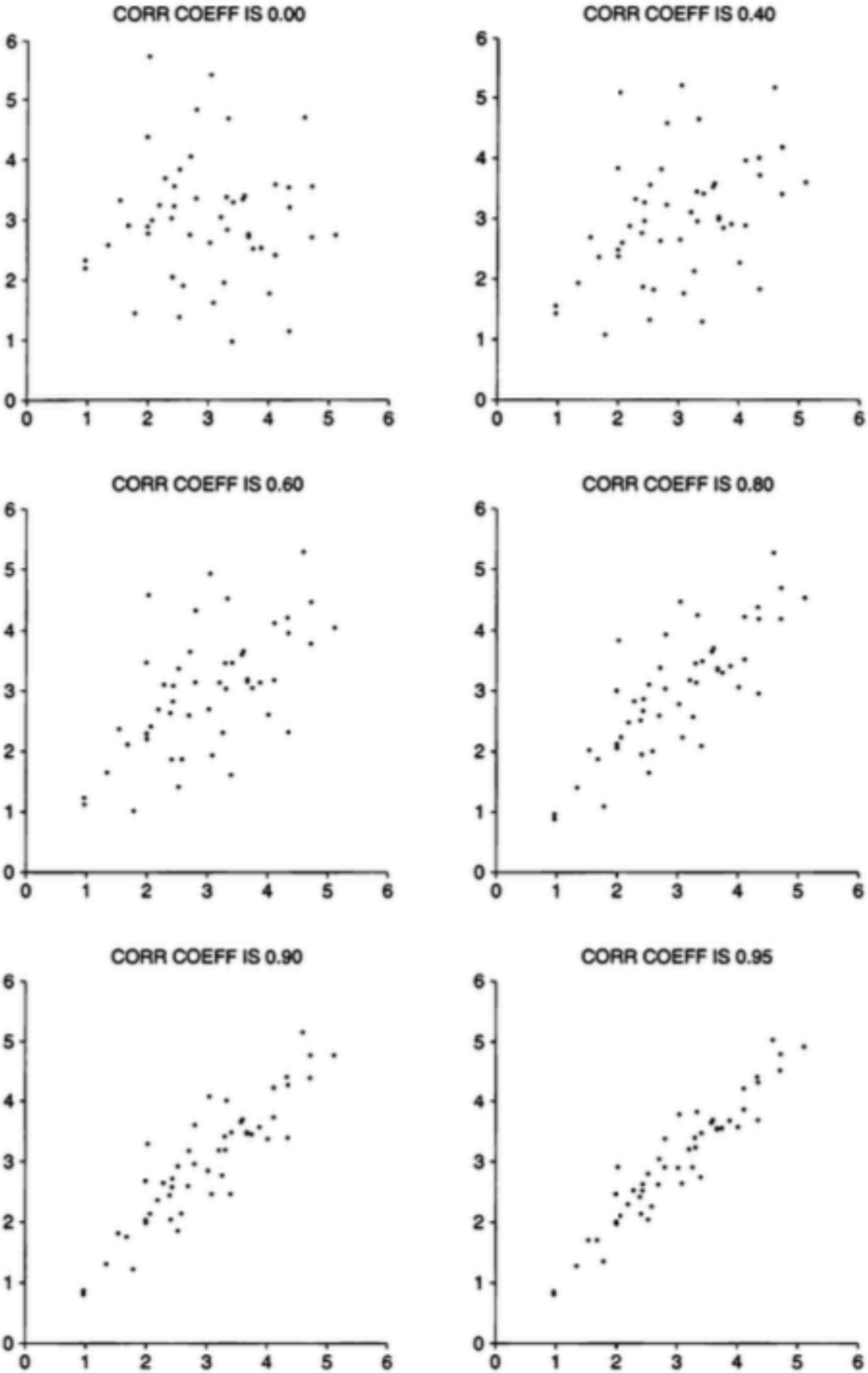
X	X
1	1
2	2
3	3
4	4



X	-X
1	-1
2	-2
3	-3
4	-4



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


CORR COEFF IS 0.00
CORR COEFF IS 0.40
CORR COEFF IS 0.60
CORR COEFF IS 0.80
CORR COEFF IS 0.90
CORR COEFF IS 0.95

Stat 134

Wednesday Nov 28 2018

1. An urn contains 90 marbles, of which there are 20 green, 25 black, and 45 red marbles. Alice randomly picks 10 marbles without replacement, and Bob randomly picks another 10 marbles without replacement. Let X_1 be the number of green marbles that Alice has and X_2 the number of green marbles that Bob has.
   To find $Corr(X_1, X_2)$ is

   $$X_1 + X_2 + \cdots + X_9 = 20$$

   a true identity that is useful? Explain.

   a yes

   b no

   c not enough info to decide

sum of exchangable X_i is a constant.

X_i = # green of i^{th} person

2. An urn contains 90 marbles, of which there are 20 green, 25 black, and 45 red marbles. Alice randomly picks 10 marbles without replacement, and Bob randomly picks another 10 marbles without replacement. Let X_1 be the number of green marbles that Alice has and X_2 the number of green marbles that Bob has.
Find $Corr(X_1, X_2)$.

a $-1/8$

b $-1/9$

c $-1/10$

d none of the above

$$X_1 + \cdots + X_9 = 20$$

$$Corr(X_1, X_2) = -\frac{1}{9-1} = -\frac{1}{8}$$

If asked to find $Cov(X_1, X_2)$ we need to find $SD(X_1)\, SD(X_2)$

$$Cov(X_1,X_2) = Corr(X,Y)\, SD(X_1)\, SD(X_2)$$

$N=90$, $G=20$, $B=70$

$X_1, X_2 \sim \text{Hypergeom}(n=10, N=90, G=20)$

$$SD(X_1)^2 = n\frac{G}{N}\frac{B}{N}\left(\frac{N-n}{N-1}\right)$$

$$= 10\left(\frac{20}{90}\right)\left(\frac{70}{90}\right)\left(\frac{90-10}{90-1}\right)$$

$$Cov(X_1,X_2) = Corr(X,Y)\, SD(X_1)\, SD(X_2)$$

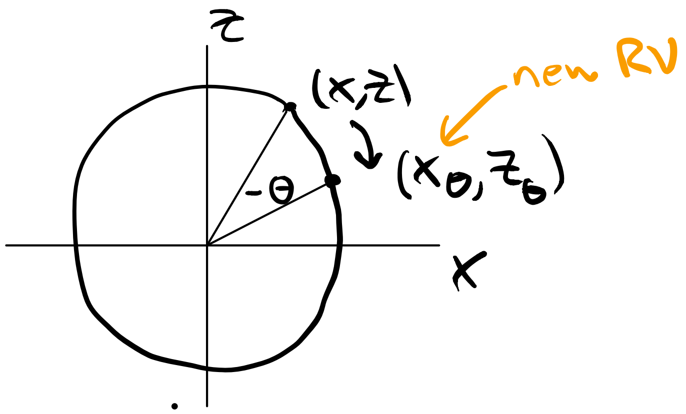
$$= -\frac{1}{8}\cdot 10\left(\frac{20}{90}\right)\left(\frac{70}{90}\right)\left(\frac{90-10}{90-1}\right)$$

$$= \boxed{\frac{-1400}{7209}}$$

Sec 6.5 Bivariate normal

Here is a recipe how to make two correlated std normals!

let X, Z iid standard normals



We saw in lecture 30

$\cancel{X_\theta} = \underbrace{\cos\theta}_{\rho}\, X + \underbrace{\sin\theta}_{= \sqrt{1-\rho^2}}\, Z \sim N(0,1)$, $\quad -1 \le \rho \le 1$

call this Y

r

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

Def^n

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

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

We call the joint distribution (X,Y) the standard bivariate normal with corr ρ.

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

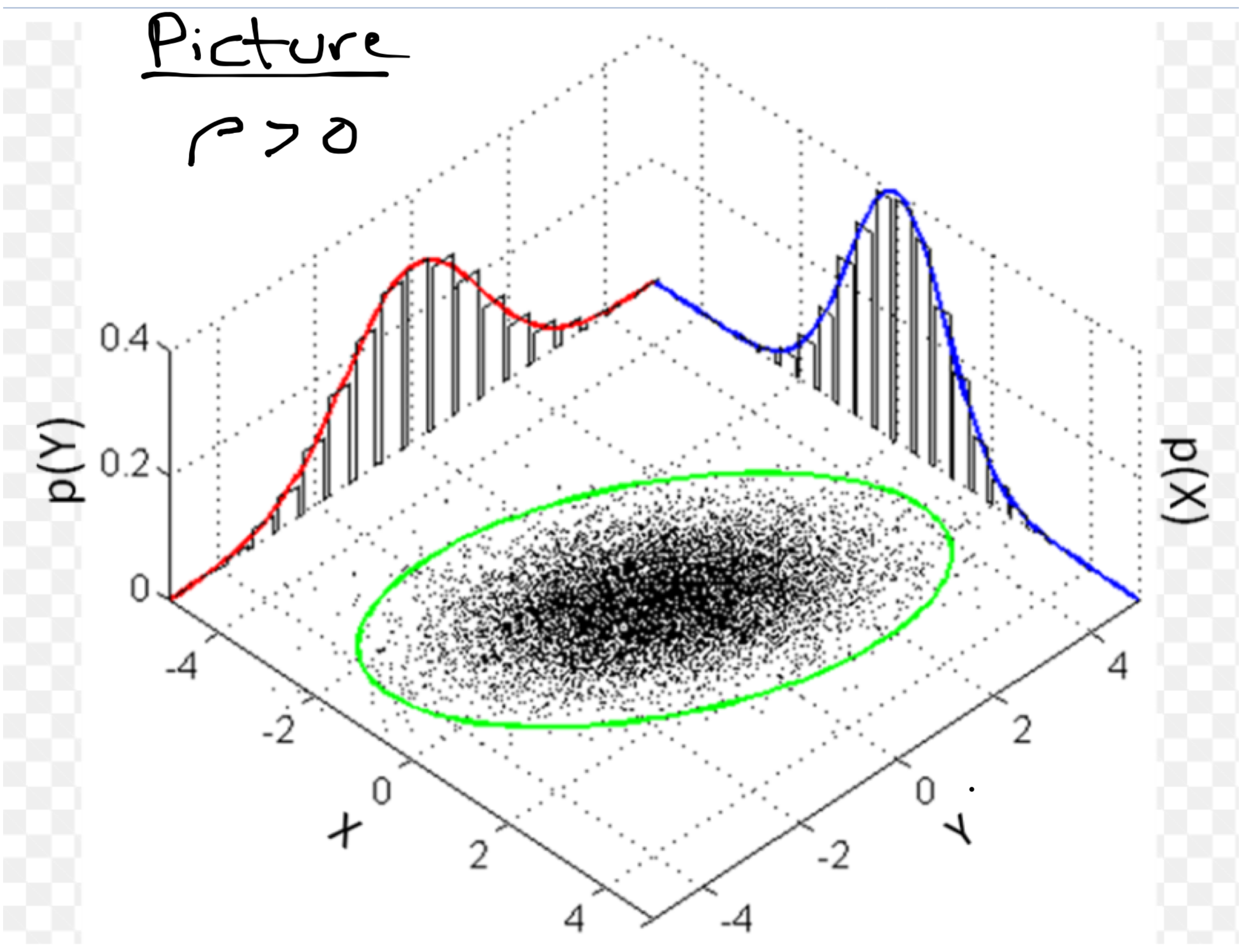
$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{Cov}\left(X, \rho X + \sqrt{1-\rho^2}\, Z\right) \\
&= \mathrm{Cov}(X, \rho X) + \mathrm{Cov}\left(X, \sqrt{1-\rho^2}\, Z\right) \\
&= \rho \underbrace{\mathrm{Var}(X)}_{=1} + \sqrt{1-\rho^2}\, \underbrace{\mathrm{Cov}(X,Z)}_{=0} \\
&= \boxed{\rho}
\end{aligned}$$

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

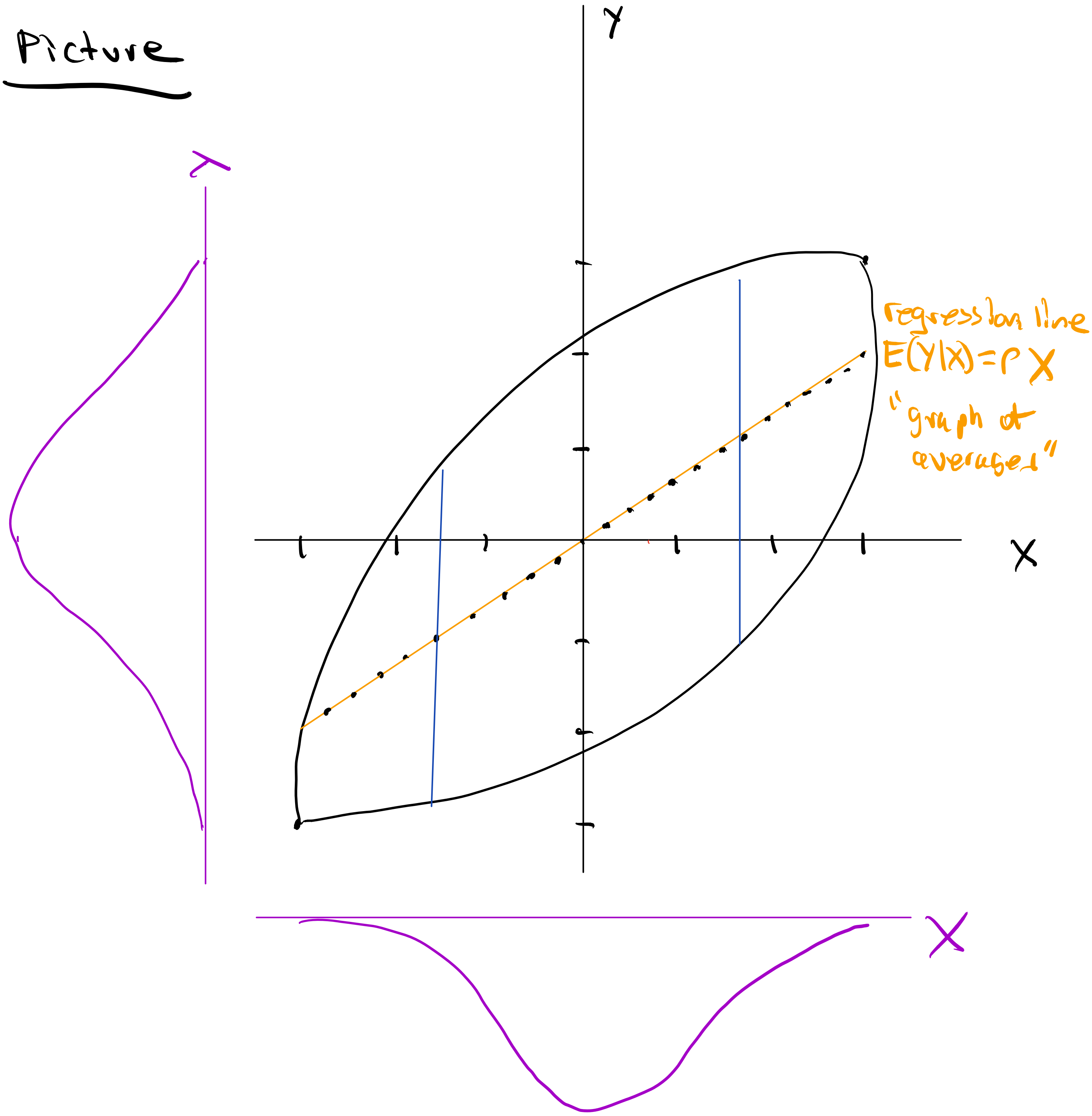
correlation coeff.

Picture

$\rho > 0$



Picture



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

Find $E(Y|X)$

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

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

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

$$= \rho X$$

since X, Z indep.

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

Find $Var(Y|X)$:

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

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

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

$$\Rightarrow \boxed{SD(Y|X) = \sqrt{1-\rho^2}}$$

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

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

is normal

we have

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

Appendix

Properties of correlation

2) correlation is invariant to linear changes of scale except possibly by a sign.

Proof

$$|\mathrm{Corr}(aX+b, cY+d)| = \left|\frac{\mathrm{Cov}(aX+b, cY+d)}{SD(aX+b)SD(cY+d)}\right|$$

$$= \left|\frac{ac\,\mathrm{Cov}(X,Y)}{|a||c|\,SD(X)SD(Y)}\right| = \mathrm{Corr}(X,Y).$$

1) $-1 \leq \mathrm{Corr}(X,Y) \leq 1$

Proof

$$\left.\begin{aligned} E(X^*) &= 0 = E(Y^*) \\ SD(X^*) &= 1 = SD(Y^*) \\ E(X^{*2}) &= 1 = E(Y^{*2}) \end{aligned}\right\} \text{ since } X^*, Y^* \text{ are standard units}$$

$(X^*+Y^*)^2 \geq 0$

so $E((X^*+Y^*)^2) \geq 0$

$E(X^{*2} + Y^{*2} + 2X^*Y^*) \geq 0$

$1 + 1 + 2E(X^*Y^*) \geq 0$

$E(X^*Y^*) \geq -1$

$\Rightarrow \quad -1 \leq E(X^*Y^*)$

$\boxed{-1 \leq \mathrm{Corr}(X,Y)}$

$(X^*-Y^*)^2 \geq 0$

so $E((X^*-Y^*)^2) \geq 0$

$E(X^{*2} + Y^{*2} - 2X^*Y^*) \geq 0$

$1 + 1 - 2E(X^*Y^*) \geq 0$

$E(X^*Y^*) \leq 1$

$\Rightarrow \quad \boxed{\mathrm{Corr}(X,Y) \leq 1}$

$\square$