Stat 134 Lec 39

Sec 6.4 Correlation

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

last time

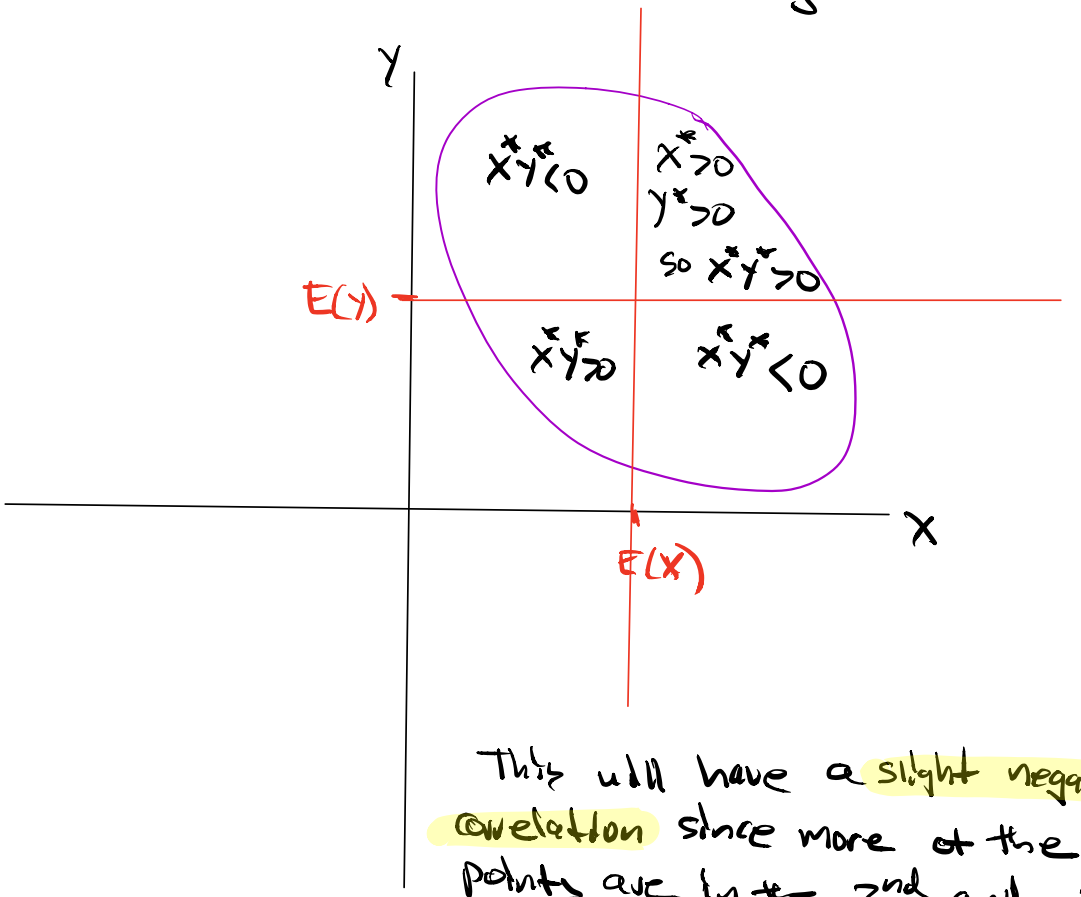
$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = E\left(\left(\frac{X-\mu_X}{SD_X}\right)\left(\frac{Y-\mu_Y}{SD_Y}\right)\right)$$

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

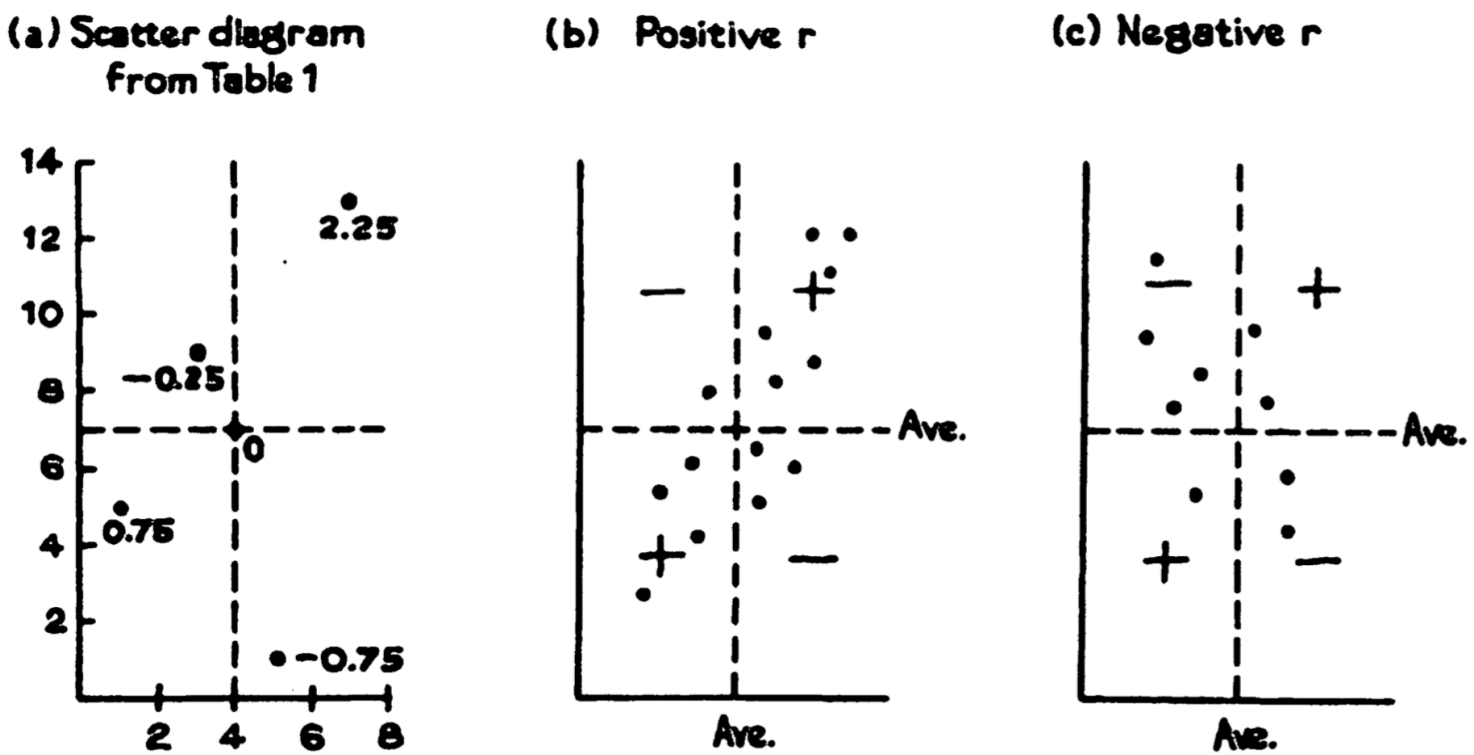
↖ X, Y in S.U.

So suppose you have a scatter diagram



This will have a slight negative correlation since more of the points are in the 2nd and 4th quadrants (of the red axes).

Figure 9. How the correlation coefficient works.



Properties of correlation

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

In other words:

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

Proof

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

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

3) $Corr(X,Y) = Corr(Y,X)$ ✓

4) $-1 \leq Corr(X,Y) \leq 1$

Proof

$E(X^*) = 0 = E(Y^*)$
$SD(X^*) = 1 = SD(Y^*)$
$E(X^{*2}) = 1 = E(Y^{*2})$

} since X^*, Y^* are standard units

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

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

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

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

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

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

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

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

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

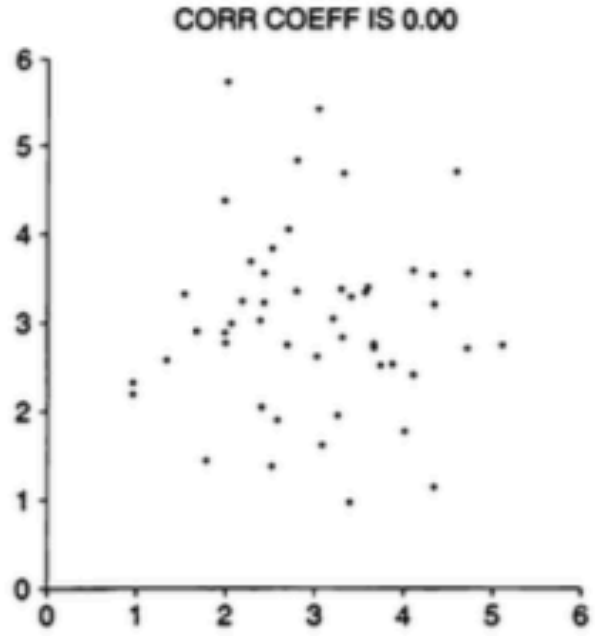
$E(X^*Y^*) \geq -1$

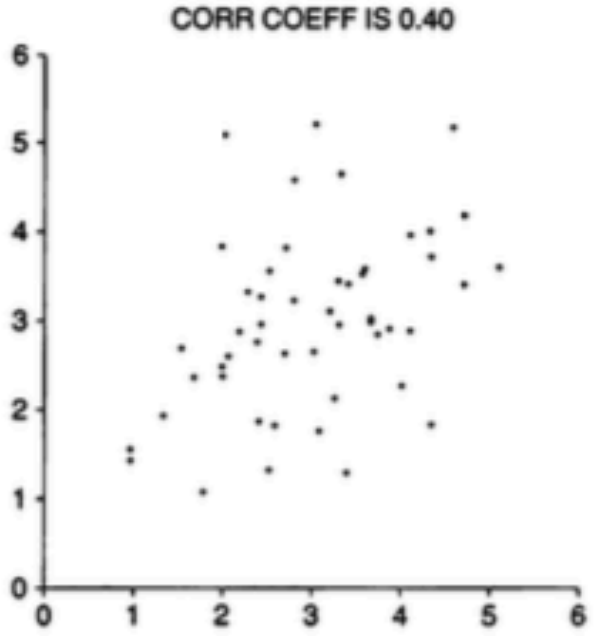
$\Rightarrow \quad -1 \leq E(X^*Y^*) \leq 1$

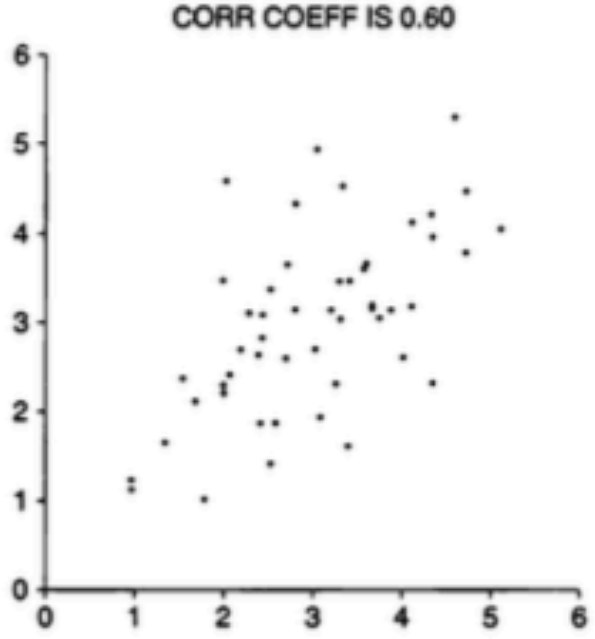
$$\boxed{-1 \leq Corr(X,Y) \leq 1}$$

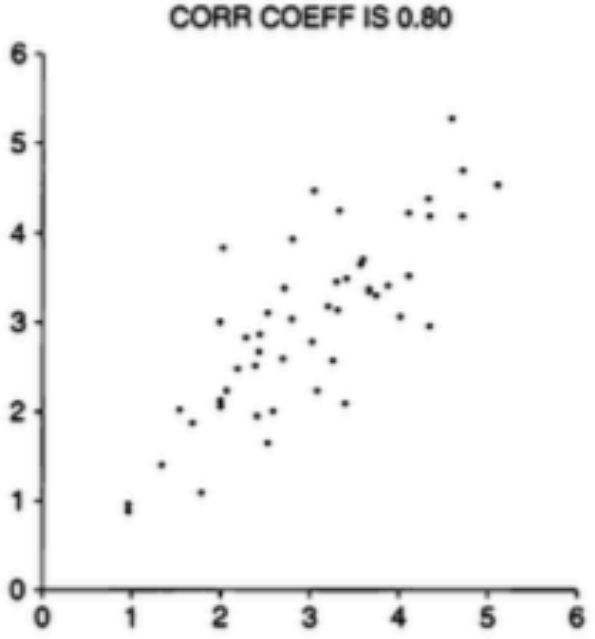
5) Correlation measures <u>linear</u> association, how tightly clustered scatter diagram is around a straight line.

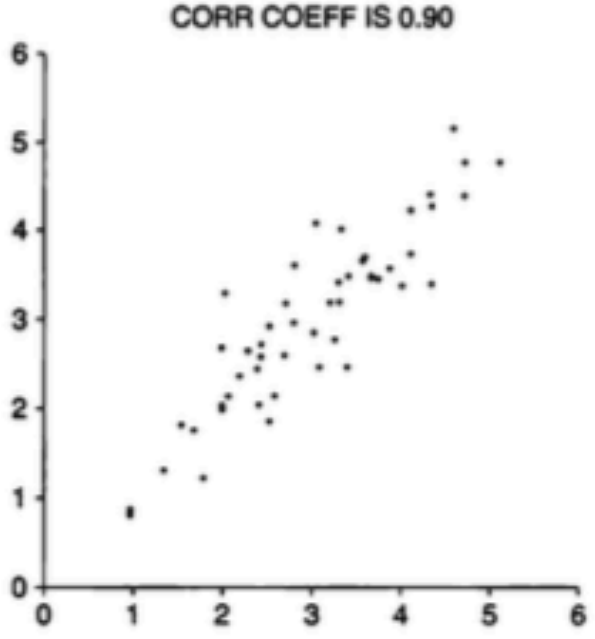
see next page

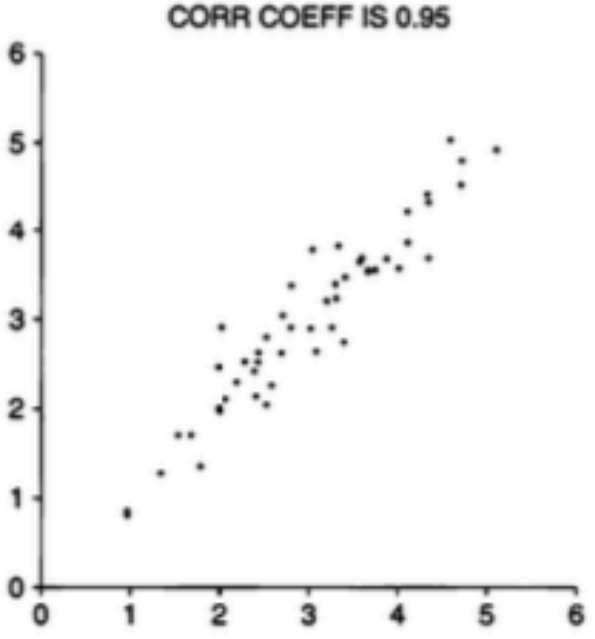

CORR COEFF IS 0.00
0 1 2 3 4 5 6


CORR COEFF IS 0.40
0 1 2 3 4 5 6


CORR COEFF IS 0.60
0 1 2 3 4 5 6


CORR COEFF IS 0.80
0 1 2 3 4 5 6


CORR COEFF IS 0.90
0 1 2 3 4 5 6


CORR COEFF IS 0.95
0 1 2 3 4 5 6

ex Roll a die n times.

N_1 = # of ones

N_6 = # of sixes.

Find $\text{Corr}(N_1, N_6)$

method 1 (indicators, bilinearity of Cov)

$N_1 = \sum_{j=1}^{n} I_{1,j}$ ← get 1 on j^{th} roll

$N_6 = \sum_{j=1}^{n} I_{6,j}$ ← get 6 on j^{th} roll.

$$\text{Cov}(N_1, N_6) = \text{Cov}\left(\sum_{j=1}^{n} I_{1,j}, \sum_{j=1}^{n} I_{6,j}\right)$$

$$= n \cdot \text{Cov}(I_{1,1}, I_{6,1}) + n(n-1)\text{Cov}(I_{1,1}, I_{62})$$

etc

method 2 (var of $N_1 + N_6$)

$N_1 \sim \text{Bin}(n, 1/6)$ $\quad \text{Var}(N_1) = n\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)$

$N_6 \sim \text{Bin}(n, 1/6)$

$N_1 + N_6 \sim \text{Bin}\left(n, \frac{2}{6}\right)$ $\quad \text{var}(N_1 + N_6) = n\left(\frac{2}{6}\right)\left(\frac{4}{6}\right)$

$$\text{Var}(N_1 + N_6) = \text{Var}(N_1) + \text{Var}(N_6) + 2\text{Cov}(N_1, N_6)$$

$$\text{Cov}(N_1, N_6) = n\left(\frac{2}{6}\right)\left(\frac{4}{6}\right) - 2n\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = -\frac{n}{36}$$

$$\text{Corr}(N_1, N_6) = \frac{-\frac{n}{36}}{\left(\sqrt{n\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)}\right)^2} = \frac{-\frac{n}{36}}{n\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)} = -\frac{1}{5}$$

More generally; (sum of exchangeable RVs)

See p238

Suppose the sum of k exchangeable (i.e identically distributed) RVs is a constant

$$N_1 + N_2 + \cdots + N_k = c$$

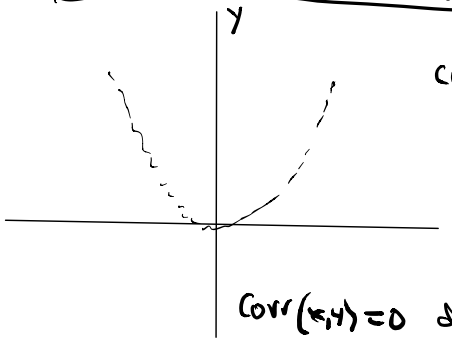
then $\boxed{\text{Corr}(N_1, N_2) = -\frac{1}{k-1}}$

ex You roll a die n times.
Let N_i = the number of rolls where you get i

$$N_1 + N_2 + N_3 + N_4 + N_5 + N_6 = n$$

then $\text{Corr}(N_1, N_2) = \frac{-1}{5}$ ✓

Conditions for X and Y to be uncorrelated



$\text{corr}(x, y) = 0$
$y = x^2 \Rightarrow x, y$ are dependent.

$\text{Corr}(x,y) = 0$ doesn't imply independence. of x, y.

$\left.\begin{array}{l} \text{corr}(x,y) = 0 \\ \text{Cov}(x,y) = 0 \\ E(xy) = E(x)E(y) \end{array}\right\}$ equivalent.

x, y indep $\Rightarrow E(xy) = E(x)E(y) \Leftrightarrow \text{Corr}(x,y) = 0.$

Stat 134

Friday April 23 2018

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

review if $T_r = \text{gamma}(r, \lambda)$

$E(T_r) = \frac{r}{\lambda}$

$SD(T_r) = \sqrt{r}/\lambda$.

$T_3 = T_1 + w$

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

$= \text{Cov}(T_1, T_1) + \text{Cov}(T_1, w)$

$\text{Var}(T_1) = \frac{1}{\lambda^2}$, $= 0$

$\text{Corr}(T_1, T_3) = \frac{\text{Cov}(T_1, T_3)}{SD(T_1)SD(T_3)} = \frac{\frac{1}{\lambda^2}}{\frac{1}{\lambda} \cdot \frac{\sqrt{3}}{\lambda}} = \frac{1}{\sqrt{3}}$

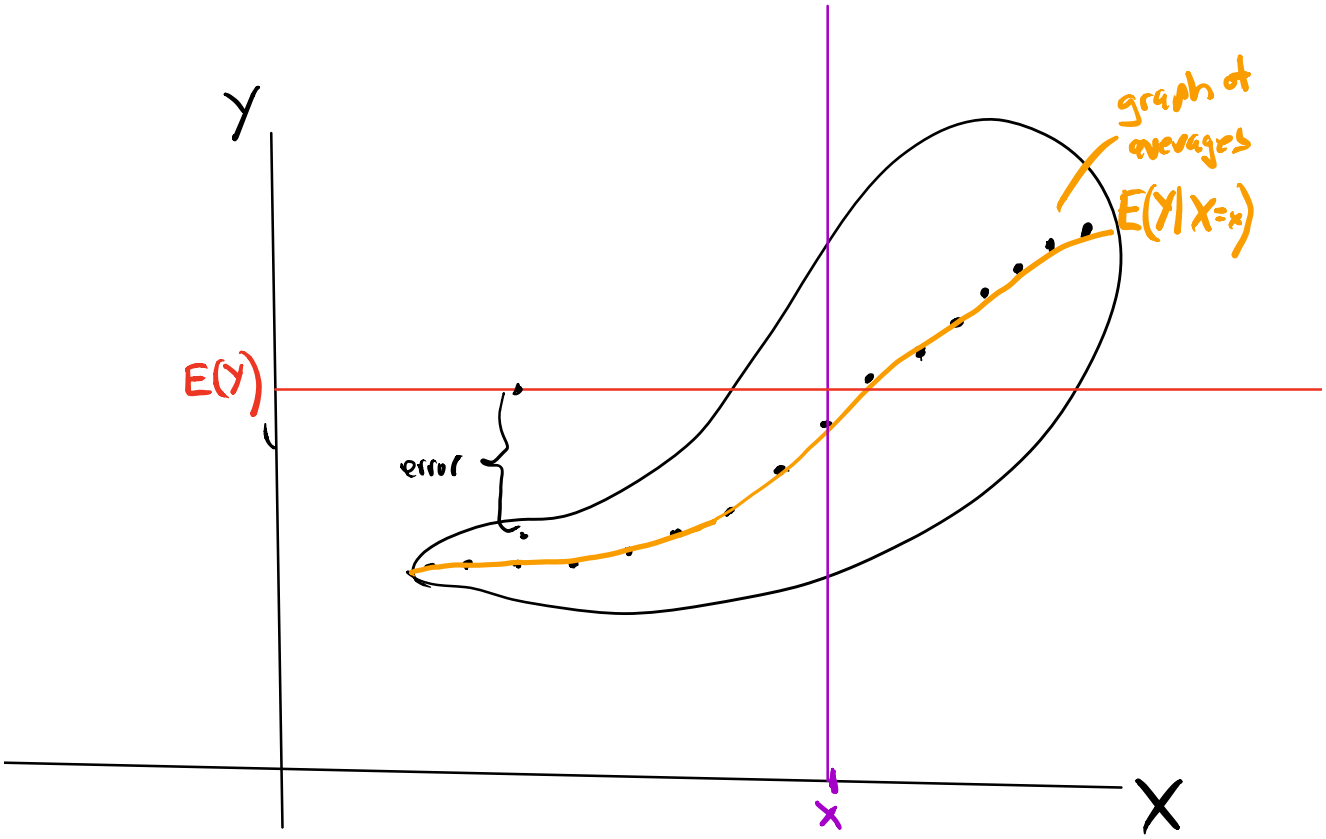
a $\sqrt{\frac{3}{1}}$

b $\sqrt{\frac{1}{3}}$

c $\sqrt{\frac{2}{3}}$

d none of the above

Predicting Y from X



We wish to predict Y from X.

What is the best predictor? — smallest mean square error (mse)

Predictor #1 E(Y) straight line.

for every x same prediction.

$error = Y - E(Y)$

$E(error) = E(Y) - E(Y) = 0$

$E(error^2) = Var(error) = Var(Y - E(Y))$

$= \boxed{Var(Y)}$

Predictor #2

graph of averages, $E(Y|X)$ — see picture above.

error $= Y - E(Y|X)$

$$\text{mse}_{\text{Predictor 2}} = E(\text{error}^2) = E\left((Y - E(Y|X))^2\right).$$

We will show that not only is $E(Y|X)$ a better predictor of Y than $E(Y)$ but it is a better predictor of Y than any function $h(X)$.

That is cool!!

Stay tuned...