Lecture 9

Topics:

I) correlation

II) Bivariate Normal

III) prediction with Bivariate Normal.

Recall that the covariance btwn two R.V.s $cov(X,Y)$ is the number which is the missing link for computing the variance of $aX+bY$ when X & Y are dependent.

In some sense $cov(X,Y)$ measures "linear dependence". However, it is somewhat hard to interpret since it changes when the units of measurement are changed.

e.g. Randomly select a from california.

Let X_1 = their ave sodium intake per day (in grams)

X_2 = their ave sodium intake per day (in ounces)

Y = their systolic blood pressure in mm Hg.

The only difference btwn X_1 & X_2 are a change in units:

$$X_1 = 0.035274\, X_2$$

However

$$cov(X_1, Y) \neq \underbrace{cov(X_2, Y)}_{\text{this is actually } 0.035274\, cov(X_1, Y)}$$

This presents a problem in that

$$\begin{pmatrix}\text{cov btwn sodium} \\ \text{\& blood pressure} \\ \text{in USA}\end{pmatrix} \neq \begin{pmatrix}\text{cov btwn} \\ \text{sodium \&} \\ \text{blood pressure} \\ \text{in Europe}\end{pmatrix}$$

So we need a measure of "linear dependence" that doesn't change when changing units (i.e. that is unit free).

Definition: The correlation between two R.V.s (X,Y), denoted ρ or $\rho_{X,Y}$, is defined to be

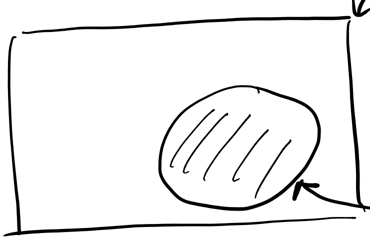
$$\rho = \frac{cov(X,Y)}{sd(X)\, sd(Y)}.$$

Fact 1 ρ is unit free, i.e. ρ stays the same when X or Y (or both) are converted to different units. ← so $\rho_{X_1,Y} = \rho_{X_2,Y}$ in last e.g.

Fact 2

- $-1 \le \rho \le 1$ always
- if $\rho < 0$ then $cov(X,Y) < 0$
- if $\rho > 0$ then $cov(X,Y) > 0$.
- If X & Y are independent then $\rho = 0$ (but not the other way around).

Picture:



all pairs (X,Y) with $\rho=0$

all pairs of R.V.s (X,Y) which are independent.

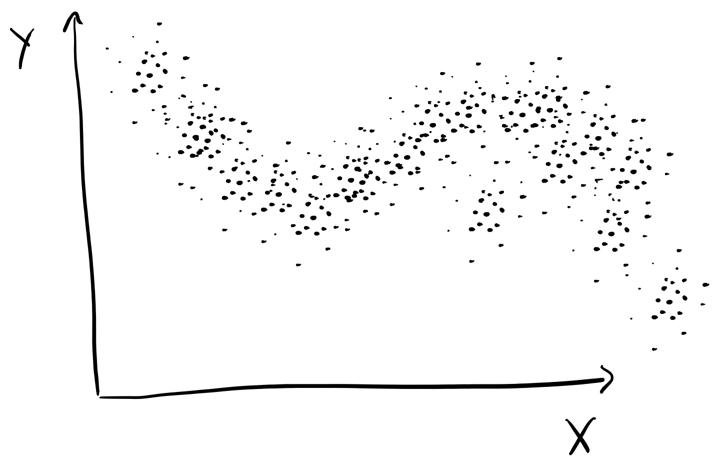
Fact 3

If $\rho = 1$, (X,Y) will always be on a fixed (non-random) line with positive slope

If $\rho = -1$, (X,Y) will always be on a line with negative slope.

Visualizing the joint PMF of a pair of R.V.s (X,Y)

To visualize the dependence between X & Y one can make a scatter plot of all possible pairs (X,Y) with repeats in proportion to probabilities



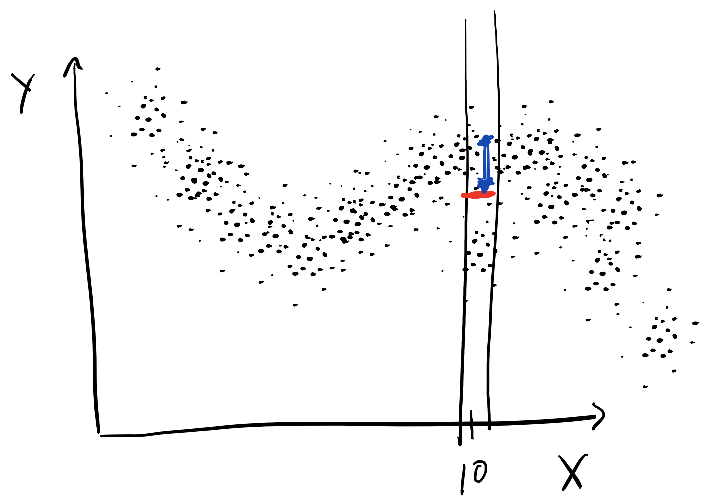
You can imagine that a random sample (X,Y) corresponds to the coordinates of a random dot drawn from this scatter plot.

Equivalently the coordinates of each dot represents the pair of numbers on the tickets in a box model for (X,Y).

With a box model $\boxed{\begin{smallmatrix} X \\ Y \end{smallmatrix}}$ for (X,Y)

Recall that $E(Y|X=x)$ & $sd(Y|X=x)$ are basically the ave & standard deviation of the Y value's on all tickets of the form $\boxed{\begin{smallmatrix} x \\ Y \end{smallmatrix}}$ ← fixed at $X=x$ ← Y is random

$\therefore$ $E(Y|X=x)$ & $sd(Y|X=x)$ can be visualized as follows.

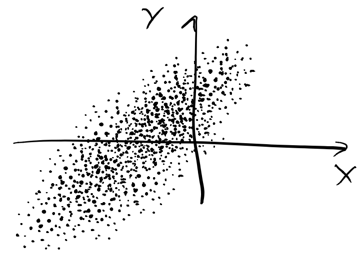


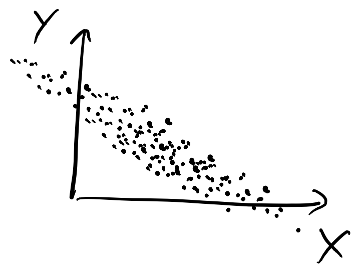
—: $E(Y|X=10) \simeq$ the ave Y value in the column above 10

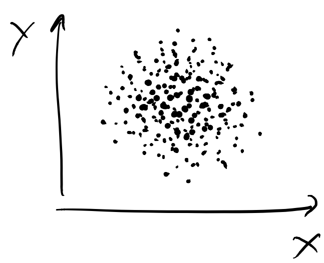
I: $sd(Y|X=10) \simeq$ the s.d. of the Y values in the column above 10.

Bivariate Normal (X,Y)

Bivariate Normal (X,Y) is a special type of (continuous) pair of R.V.s which have scatter plots that look like this:







Def: (X,Y) is Bivariate Normal if

1. X is Normal
2. Y is Normal
3. For any x the conditional PDF of Y given $X=x$ is also Normal

As a consequense of the definition its also true that X given $Y=y$ is also Normal.

Predicting Y given X for Bivariate Normals

It is typically computationally difficult to compute the best prediction & s.d. of Y given $X=x$.

For Bivariate Normals it is extremely easy & all you need to compute these is the following 5 numbers (sometimes called parameters):

$$E(X), \; sd(X)$$

$$E(Y), \; sd(Y)$$

$$\rho \leftarrow \text{The correlation btwn } X, Y.$$

This gives us our next "master formulas".

Suppose (X, Y) are bivariate Normal with correlation ρ.

Then

Master Formula 3:

$E(Y|X=x)$ is computed using

$$\left(\frac{E(Y|X=x) - E(Y)}{sd(Y)}\right) = \rho \left(\frac{x - E(X)}{sd(X)}\right)$$

Master Formula 4:

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

e.g. Suppose X & Y are BN with

$E(X)=10$ $\quad$ $sd(X)=1$

$E(Y)=1$ $\quad$ $sd(Y)=5$

$\rho = 0.6$

Suppose you observe $X = 11.5$. Predict Y & find the typical error of this prediction.

$$\frac{x - E(X)}{sd(X)} = \frac{11.5 - E(X)}{sd(X)} = \frac{11.5 - 10}{1} = \underbrace{1.5}_{\text{z-score of your obs.}}$$

$\therefore$ By MF3:

$$\frac{E(Y|X=11.5) - E(Y)}{sd(Y)} = \frac{E(Y|X=11.5) - 1}{5}$$

$$\overset{MF3}{=} \rho\left(\frac{x - E(X)}{sd(X)}\right)$$

$$= 0.6 \cdot (1.5)$$

$$= 0.9$$

$$\therefore \quad \frac{E(Y|X=11.5) - 1}{5} = 0.9$$

$$\therefore \quad E(Y|X=11.5) = 5(0.9) + 1$$

$$= \underbrace{5.5}_{\text{best prediction of } Y \text{ given } X=11.5}$$

Also By MF4

$$var(Y|X=11.5) = (1 - \rho^2)\, var(Y)$$

$$= (1 - 0.6^2) \cdot 5^2$$

$$= 16.0$$

$\therefore$ the typical prediction error using 5.5 to predict Y when X is obs to be 11.5 is

$$sd(Y|X=11.5) = \sqrt{16} = 4$$

Easier way to remember Master formula 3

Recall MF3:

$$\underbrace{\left(\frac{E(Y|X=x) - E(Y)}{sd(Y)}\right)}_{\substack{\text{This is the z-score} \\ \text{for } Y \text{ evaluated at} \\ \text{the prediction} \\ E(Y|X=x).}} = \rho \underbrace{\left(\frac{x - E(X)}{sd(X)}\right)}_{\substack{\text{This is the z-score} \\ \text{for } X \text{ evaluated} \\ \text{at } x.}}$$

Simplified Master formula 3:

$$\left(\text{predicted } \underbrace{\text{z-score}}_{\text{for } Y}\right) = \rho\left(\text{observed } \underbrace{\text{z-score}}_{\text{for } X}\right)$$

Important fact: Since z-scores quantify rarity & correlation is always btwn -1 & 1 the above formula implies that no matter how rare the observed X is among all possible X's, the associated Y value will be predicted to be less rare among all possible Y's.

(this is called the regression effect)

e.g. Suppose (X,Y) are Bivariate Normal with parameters

$$E(X) = 10, \quad sd(X) = 2$$
$$E(Y) = 5, \quad sd(Y) = 100$$

Suppose you observe $X = 16.5$.

Predict the z-score for Y in the following three cases

Case 1: $\rho = 0.01$

Case 2: $\rho = 0.5$

Case 3: $\rho = 0.99$.

Answer:

First lets consider how rare $X = 16.5$ and think about how rare Y will be predicted to be.

$X = 16.5$ corresponds to a z-score of

$$\frac{16.5 - 10}{2} = 3.25$$

So our observed X is fairly rare.

To express it in probabilities

$$P(Z \geq 3.25) = 0.00058$$

so getting a z-score ≥ 3.25 only happen 0.058% of the time (about 1 in 2000).

Now consider case 1:

$$\boxed{\rho = 0.01}$$

By our simplified version of MF3:

(predicted z-score for Y)
$= \rho \times$ (obs z-score for X)
$= 0.01 \times (3.25)$
$= \underbrace{0.0325}$

This z-score for the predicted Y is not rare at all, i.e.

$$P(Z \geq 0.0325) = 0.48$$

i.e. about 1 in 2.

Now consider case 2:

$$\boxed{\rho = 0.5}$$

(predicted z-score for Y)
$= \rho \times$ (obs z-score for X)
$= 0.5 \times (3.25)$
$= \underbrace{1.625}$

still not that rare for Y.

$$P(Z \geq 1.625) = 0.052$$

i.e. about 1 in 20

Now consider case 3:

$\boxed{\rho = 0.99}$

(predicted Z-score for Y)

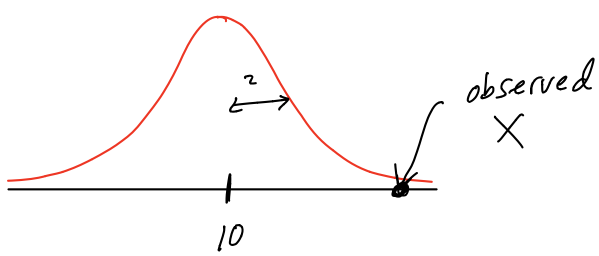
$= \rho \times$ (obs Z-score for X)

$= 0.99 \times (3.25)$

$= \underbrace{3.2175}$

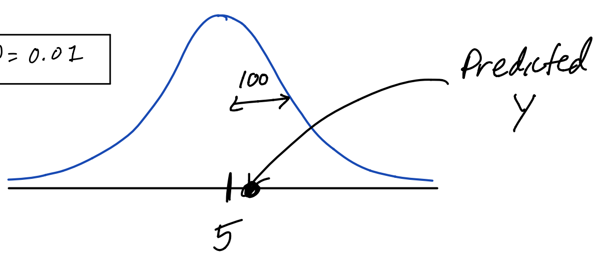
almost exactly as rare as X

The picture

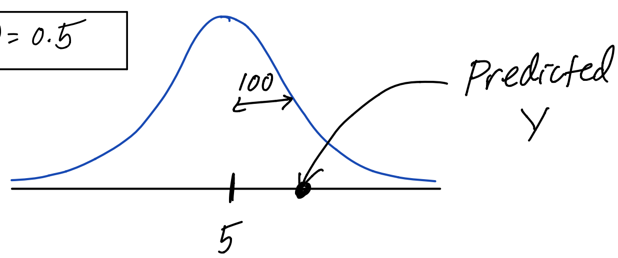


Here are the predictions in 3 cases:

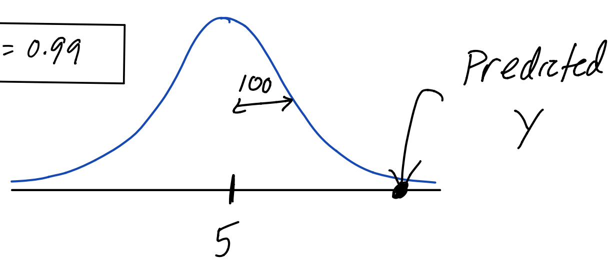
$\boxed{\rho = 0.01}$



$\boxed{\rho = 0.5}$



$\boxed{\rho = 0.99}$



This illustrates the Regression effect

Example:

It seems that whenever my son has a high fever in the middle of the night, when I bring him to the doctor the next day his fever isn't as bad.

...Regression effect!

Example:

why does it always seem that after winning an oscar an actors follow up movie isn't as good...

Regression effect!

Example:

If I look at the top scorers on midterm 1, why are many of them No longer top scorers on midterm 2?

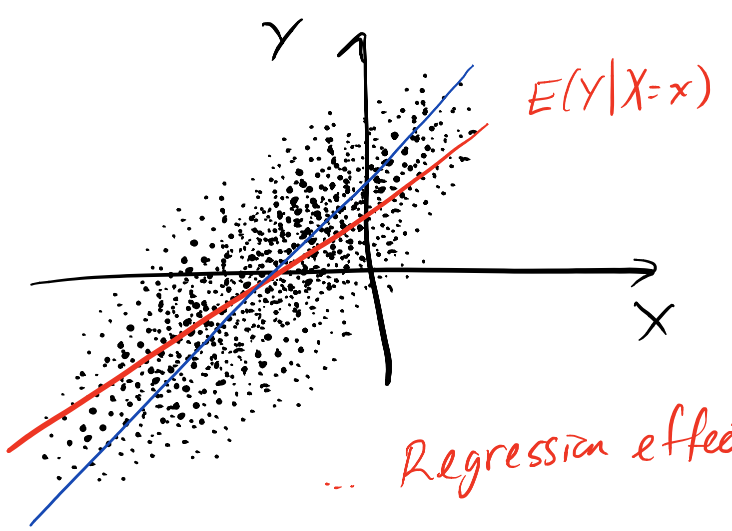
Regression effect!

Example:

If you plot the prediction line

$$E(Y|X=x) = sd(Y)\,\rho\left(\frac{x - E(X)}{sd(X)}\right) + E(Y)$$

The slope looks too shallow.



... Regression effect

Prediction error

Fact: when X,Y are Bivariate Normal the typical prediction error for Y is always smaller when incorporating information on X.

i.e. $$sd(Y|X=x) \overset{MF4}{=} \underbrace{\sqrt{(1-\rho^2)}}_{\le 1}\, sd(Y)$$

$$\xrightarrow{\text{only equal when } \rho=0} \; \le sd(Y)$$

Fact: when X,Y are Bivariate Normal the prediction error gets smaller as $\rho \to 1$ or $\rho \to -1$.

$$sd(Y|X=x) = \underbrace{\sqrt{(1-\rho^2)}}_{\to 0 \text{ as } \rho \to 1 \text{ or } -1.}\, sd(Y)$$

e.g. Midterm 1 scores: X

$E(X) = 10.62$, $sd(X) = 1.46$

Midterm 2 scores: Y

$E(Y) = 8.54$, $sd(Y) = 2.08$

Correlation btwn X & Y

$\rho = 0.155$.

Find $E(Y|X=11)$ & $sd(Y|X=11)$

For $E(Y|X=11)$ use MF3

$$(\text{predicted z-score}) = 0.155(\text{obs z-score})$$
$$= 0.155\Big(\underbrace{\tfrac{11-10.62}{1.46}}_{0.26}\Big)$$
$$= 0.04$$

$$\therefore E(Y|X=11) = sd(Y)(\text{predicted z-score}) + E(Y)$$
$$= 2.08(0.04) + 8.54$$
$$= \underbrace{8.623}_{\text{best prediction for } Y \text{ given } X=11}$$

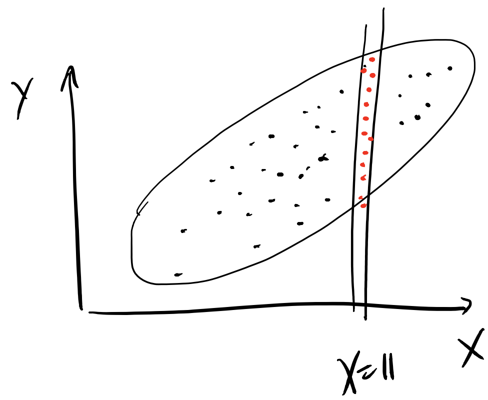
slightly above average.

The typical prediction error is

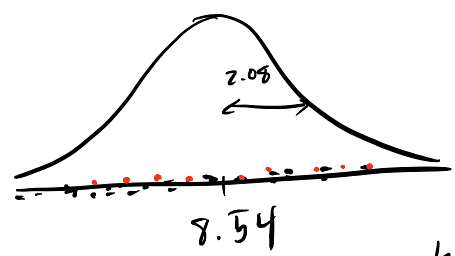
$$sd(Y|X=11) = \sqrt{(1-\rho^2)}\, sd(Y)$$
$$= \sqrt{1-0.155^2}\; 2.08$$
$$\approx 2.05$$

smaller than $sd(Y)$.

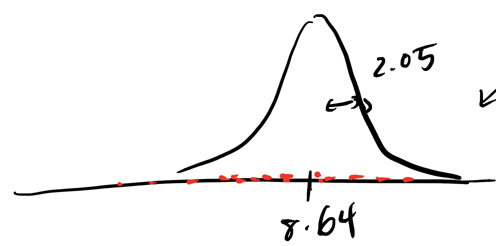
Here is the picture



All the Y's



The Y's for the subgroup that got X=11



given X=11 this is the subgroup you use to do prediction.

Example:

Suppose (X,Y) are Bivarite Normal s.t.

$$X \sim N(2, 5^2)$$
$$Y \sim N(-5, 1)$$
$$\rho = -\tfrac{1}{4}$$

Question 1:

Predict Y when $X = 7$ ← above average

$$(\text{Prediction z-score}) = (-\tfrac{1}{4})(\text{obs z-score})$$
$$= -\tfrac{1}{4}\left(\frac{7-2}{5}\right)$$
$$= -\tfrac{1}{4}$$

Predicted $Y = (-\tfrac{1}{4})(1) - 5 = -5.25$

↑ below average

Question 2:

Predict X when $Y = -5.25$

$$(\text{Prediction z-score}) = (-\tfrac{1}{4})(\text{obs z-score})$$
$$= (-\tfrac{1}{4})(-\tfrac{1}{4})$$
$$= \tfrac{1}{16}$$ ← so predicted X will be above ave.

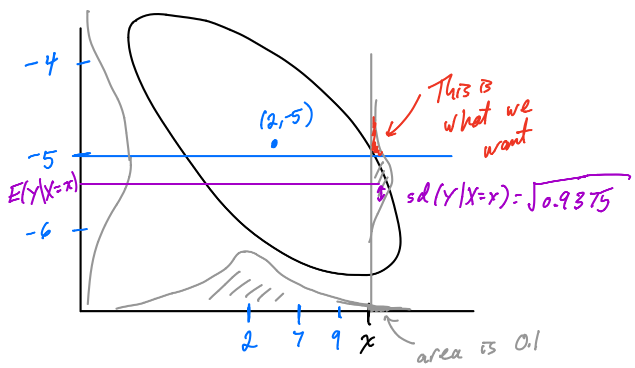
Predicted $X = \tfrac{1}{16} \cdot 5 + 2$
$= 2.3125$

Question 3:

Suppose the X value was observed to be at the 99^{th} percentile. Find the probability that the corresponding Y value is greater than -5.

Find $P(Y \geq -5 \mid X = x)$

↑ x is the 99^{th} percentile for the X values.



Observed z-score is at the 99^{th} percentile
This corresponds to $z = 2.326$

∴ predicted z-score for Y (for the whole range of Y values) is

$$-\tfrac{1}{4}(2.326) = -0.5815$$

∴ predicted Y value is

$$E(Y|X=x) = -0.5815 \underbrace{(1)}_{sd(Y)} + \underbrace{(-5)}_{E(Y)}$$

$$= -5.5815$$

Also $\text{var}(Y|X=x) = (1-\rho^2)\,\text{var}(Y)$

$$= 1 - \left(\tfrac{-1}{4}\right)^2$$

$$= 0.9375$$

$\therefore$ Y given $X=x$ behaves like a draw from $N(\underbrace{-5.5815}_{E(Y|X=x)}, \underbrace{0.9375}_{\text{var}(Y|X=x)})$

$\therefore$ $P(Y \geq -5 \mid X=x)$

$$= P\Big(Z \geq \underbrace{\frac{-5-(-5.5815)}{\sqrt{0.9375}}}\Big)$$

$$= P(Z \geq 0.6)$$

$$= 0.2743$$

Example:

Let M_2 = midterm 2 scores

M_3 = midterm 3 scores

suppose (M_1, M_2) are Bivariate Normal with

$$M_2 \sim N(8.5, 2^2)$$

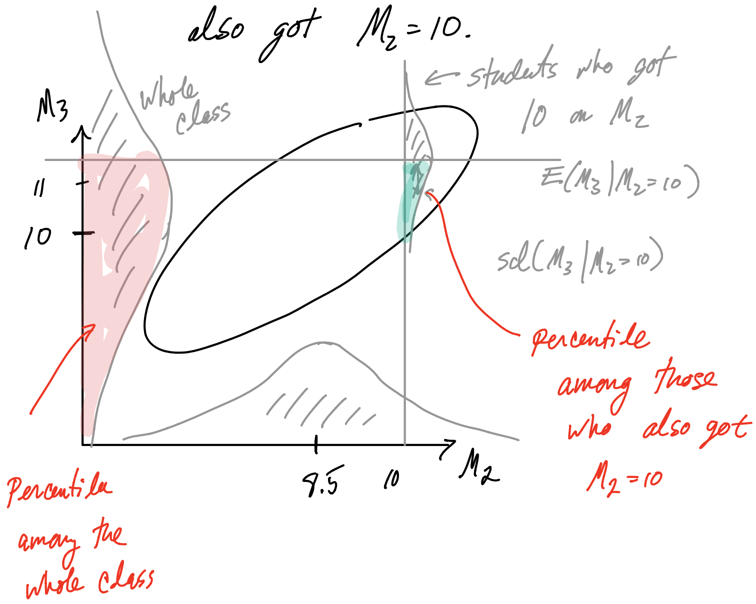
$$M_3 \sim N(10, 1^2)$$

$$\rho = 0.4$$

If you got $M_2 = 10$ Predict the percentile for M_3 among:

(a) the whole class

(b) the subgroup of students who also got $M_2 = 10$.



Answer for (b) is 50%

Answer for (a)

$$100 \cdot P(M_3 \leq E(M_3|M_2=10)) = 100\, P\Big(Z \leq \overbrace{\frac{E(M_3|M_2=10)-E(M_3)}{sd(M_3)}}^{\text{Predicted z-score for } M_3}\Big)$$

$$= 100\, P(Z \leq \rho\,(\text{observed z-score for } M_2))$$

$$= 100 \cdot P\Big(Z \leq 0.4\,\underbrace{\Big(\frac{10-8.5}{2}\Big)}_{0.75}\Big)$$

$$= 100 \cdot P(Z \leq 0.3)$$

$$= 61.79^{th} \text{ percentile.}$$

An equivalent way to specify a Bivariate Gaussian pair

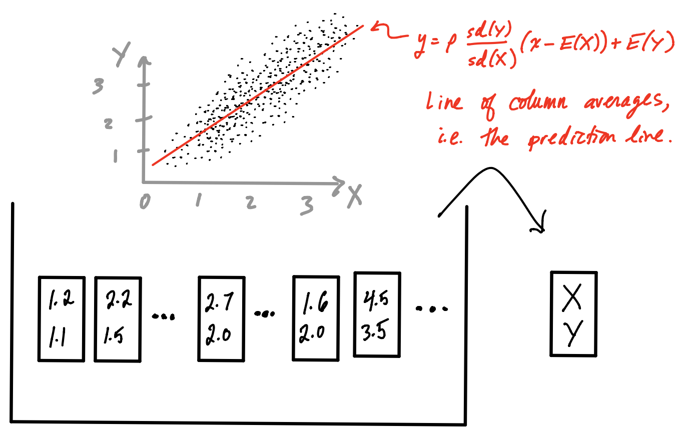
We have seen that to determine the joint distribution of a Bivariate Gaussian pair of r.v.s (X,Y) you just need to specify 5 parameters

$E(X)$ $sd(X)$

$E(Y)$ $sd(Y)$

ρ

The 5 numbers give the following picture



$E(X) = 2.2$, $sd(X) = 1.1$

$E(Y) = 3.0$, $sd(Y) = 1.9$

$\rho = 0.6$

In words: the parameters tell you.

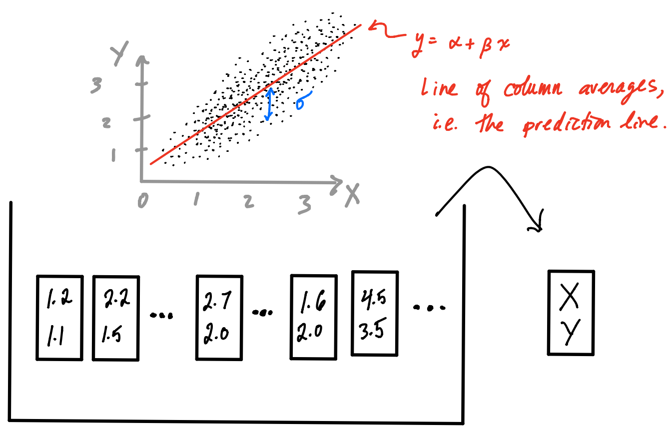
- $X \sim N(E(X), var(X))$
- $Y \sim N(E(Y), var(Y))$
- (X,Y) is Bivariate Gaussian with
- $\rho =$ correlation btwn X & Y

A completely equivalent way to specify the distribution of a Bivariate Gaussian pair (X,Y) with 5 different parameters:

α, β, σ^2, $E(X)$, $sd(X)$.

"intercept" "slope" "error variance"

The 5 numbers give the following picture



$E(X) = 2.2$, $sd(X) = 1.1$

$$\beta = \rho\frac{sd(Y)}{sd(X)} = 0.6\,\frac{1.9}{1.1} = 1.036$$

$$\alpha = E(Y) - \beta E(X) = 3.0 - 1.036(2.2) = 0.72$$

$$\sigma^2 = (1-\rho^2)\,var(Y) = (1-0.6^2)(1.9)^2 = 2.31$$

In words: the parameters tell you:

- $X \sim N(E(X), var(X))$
- $Y = \alpha + \beta X + Z$
- $Z \sim N(0, \sigma^2)$
- X and Z are in

Benefits of the old parameterization

(i) Exposes the insightful relation between correlation, prediction, z-scores and the regression effect:

$$\begin{pmatrix}\text{predicted}\\ \text{z-score}\\ \text{for } Y\end{pmatrix} = \rho \begin{pmatrix}\text{observed}\\ \text{z-score}\\ \text{for } X\end{pmatrix}$$

(ii) Allows z-score-to-z-score prediction or percentile-to-percentile prediction with just ρ.

i.e. If X and Y are Bivariate Normal with $\rho = -0.6$ and X is observed at the 90^{th} percentile (i.e. $Z = 1.282$)

... then we predict Y to be at the 22^{nd} percentile (i.e. $Z = -0.7692 = \rho \cdot (1.282)$)

(iii) Makes clear the relation between correlation and the reduction of uncertainty (for Y) when observing X.

$$\underbrace{var(Y)}_{\text{uncertainty for } Y \text{ without } X} = \underbrace{(1-\rho^2)var(Y)}_{\text{left over uncertainty for } Y \text{ after observing } X\text{: i.e. } var(Y|X=x)} + \underbrace{\rho^2 var(Y)}_{\text{amont of variability in } Y \text{ that is "explained by } X\text{"}}$$

$\Rightarrow$

Benifits of the new parameterization

(i) More visual/Geometric specification

(ii) The prediction equation is much easier:

$$E(Y|X=x) = \alpha + \beta x \quad \text{easy}$$

$$= E(Y) + \rho\frac{sd(Y)}{sd(X)}(x - E(X))$$

not so much

(iii) Allows prediction of Y given X using only α, β (without knowing $\sigma^2, E(X), sd(X)$).

e.g. If X & Y are Bivariate Normal s.t.

$$Y = 1.1 + 10.27X + Z$$

indep $N(0, \sigma^2)$

... Then given $X = 22$ we predict Y to be $1.1 + 10.27(22)$.

(iv) The model $Y = \alpha + \beta X + Z$ extends to the case that (X, Y) are not bivariate Normal.

Going from old parameterization to new

Suppose (X,Y) are Bivariate Gaussian with parameters $E(X), E(Y), sd(X), sd(Y), \rho$.

How can one find α, β, σ^2 in the new parameterization?

The new parameterization stipulates

indep

$$Y = \alpha + \beta X + Z$$

$N(\cdot,\cdot)$ $\qquad N(0,\sigma^2)$

Old parameterization

$$E(Y|X=x) = \underbrace{sd(Y)\left(\rho\frac{x-E(X)}{sd(X)}\right) + E(Y)}$$

this slope is $\rho\frac{sd(Y)}{sd(X)}$

intercep is $-\beta E(X) + E(Y)$

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

New parameterization

the "column ave" of the Y's above $X=x$ is

$$E(Y|X=x) = \alpha + \beta x + 0$$

since $E(Z)=0$

$$var(Y|X=x) = \sigma^2$$

since the only thing random in the Y's above $X=x$ is Z with $var(Z)=\sigma^2$

Matching these formulas gives:

$$\beta = \rho\frac{sd(Y)}{sd(X)}$$

$$\sigma^2 = (1-\rho^2)\,var(Y)$$

$$\alpha = E(Y) - \beta E(X)$$