CORRELATION or DEPENDENCE

is any statistical relationship, whether CAUSAL or not, b/w 2 random variables or bivariate data.

correlation may indicate any type of association, but in statistics, it usually refers to the degree to which a pair of variables are linearly related.

Example Dependent phenomena include the correlation

1. b/w the height of parents and their offspring,

2. correlation b/w the price of product and the quantity the consumers are willing to purchase, as it is depicted in DEMAND CURVE (self study)

CORRELATIONS are useful, since they can indicate a predictive relationship that can be exploited in practice.

E.g. An electrical utility may produce less power on a mild day based on the correlation b/w electricity demand and weather.

Here, a causal relationship, because extreme weather causes people to use more electricity for cooling.

However, in general, the presence of correlation is not sufficient to infer the presence of CAUSAL relationship.

CORRELATION does not imply CAUSATION.

(Think!!)

In logic, the technical use of word "implies" means is a "sufficient condition" for.

$$\boxed{p \longrightarrow q}$$

p implies q

if p then q

if p is true, then q follows

Cause can refer to necessary sufficient or contributing causes.

CAUSAL ANALYSIS (self study)

Example: The faster that windmills are observed to rotate, the more wind is observed.
Therefore wind is caused by the rotation of windmills.

Here, the correlation (simultaneity) b/w windmills activity and wind velocity does not imply that wind is caused by windmills.

It is rather the other way around, as suggested by the fact that wind does not need windmills to exist, while windmills need wind to rotate.

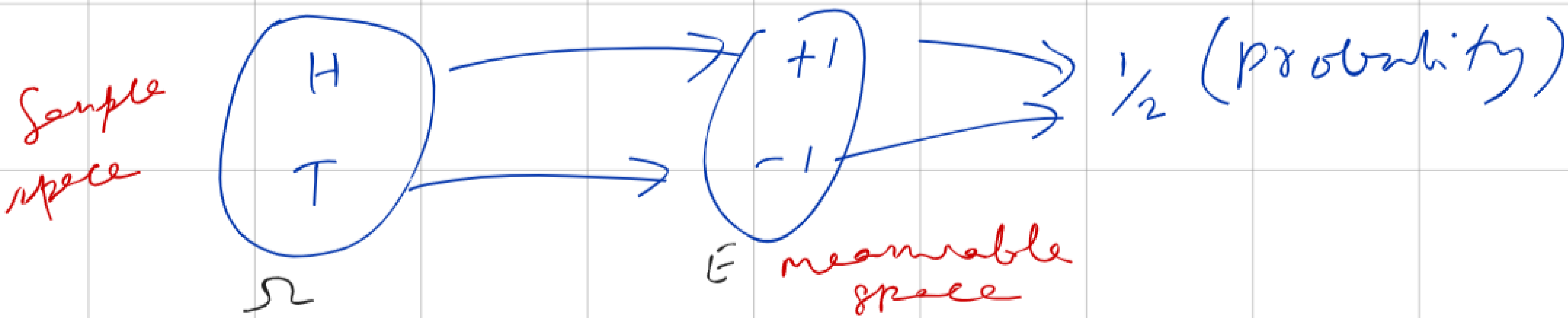
Wind can be observed in places, where there are no windmills.
(wind existed before the windmill invention)

RANDOM variable / Random quantity / ALEATORY variable / Stochastic variable

is a mathematical formalization of a quantity which depends on random events.

The term 'Random variable' in its mathematical definition refers to neither randomness nor variability, but instead of a MATH FUNCTION in which

- The domain is a set of possible outcomes in a sample space (e.g. Head / Tail from coin flipping) $\{H, T\}$
- The range is a measurable space (e.g. $\{-1, 1\}$, if H maps to -1, T maps to 1)



A Random Variable X is a measurable function

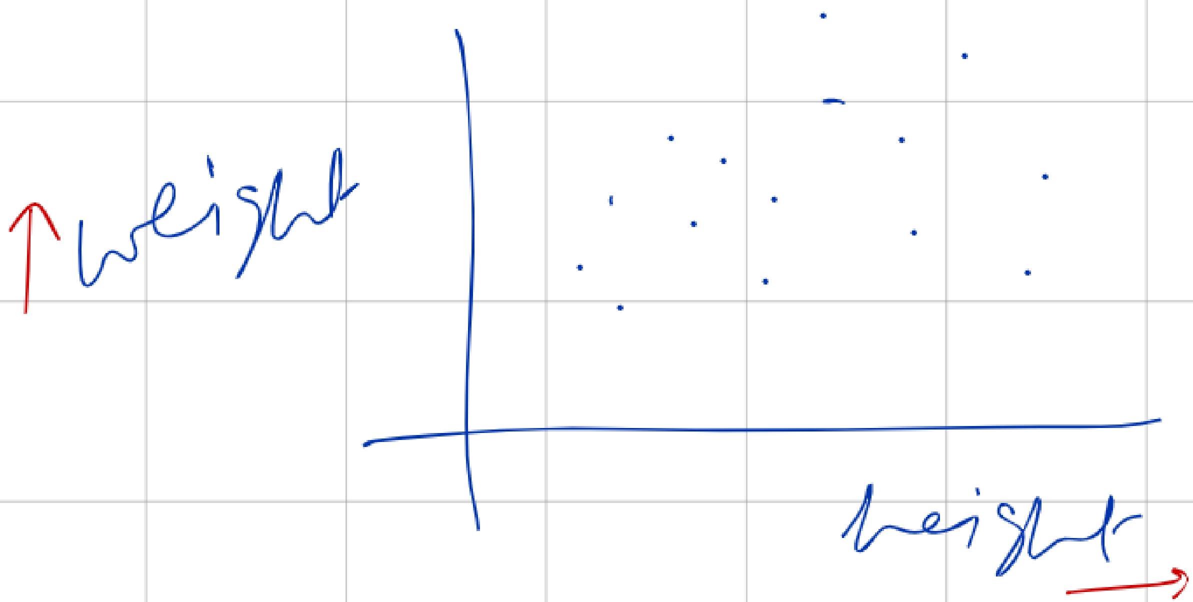
$$X : \Omega \longrightarrow E$$

from a sample space Ω (set of possible outcomes of an event)

to a measurable space E

Bivariate data is data on each of 2 variables, where each value of one of the variables is paired with a value of the other variable.

Scatter plot



Multivarite data (Think!!)

Dependent and independent variable
(weight) (height)

Think!!

PEARSON correlation coefficient : (PCC) ρ_{XY}

PCC measures linear correlation b/w 2 sets of data.

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \, \sigma_Y}$$

normalized measurement of covariance

s.t. $-1 \leq \rho_{XY} \leq 1$

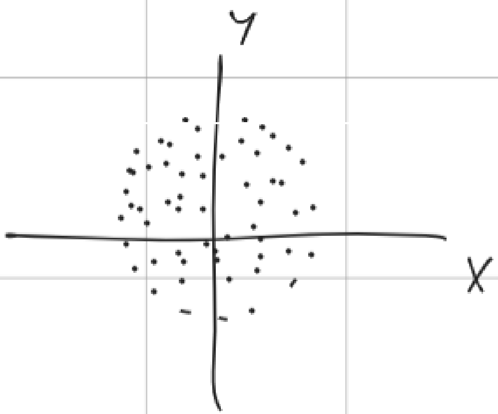
Correlation and independence

ρ_{XY} is 1 for perfect direct ($\uparrow$ i.e.) linear rel^n/ correlation

ρ_{XY} is -1 for perfect inverse ($\downarrow$ i.e.) linear rel^n/ anti-correlation

X, Y independent $\longrightarrow$ $\rho_{XY} = 0$ (X, Y uncorrelated)

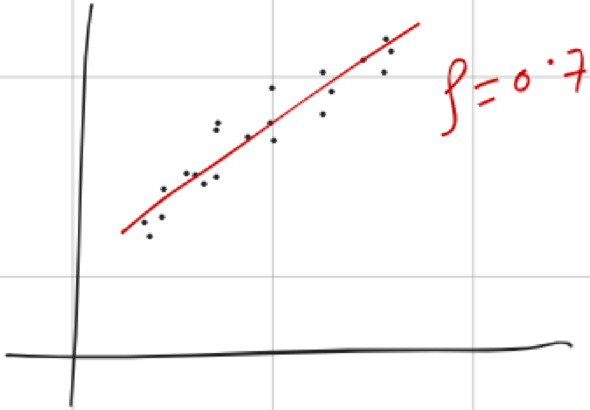
$\rho_{XY} = 0$ (X, Y uncorrelated) $\not\longrightarrow$ X, Y independent

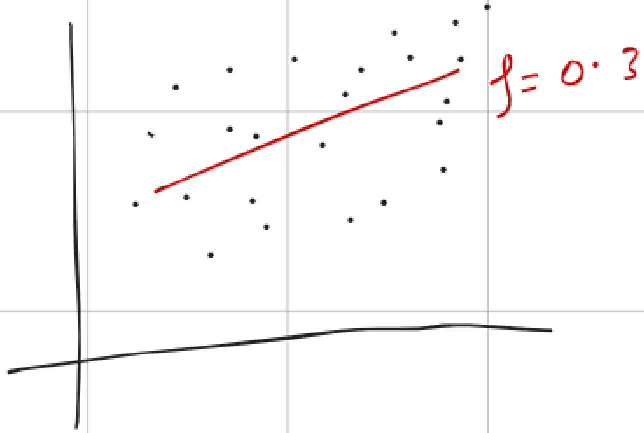


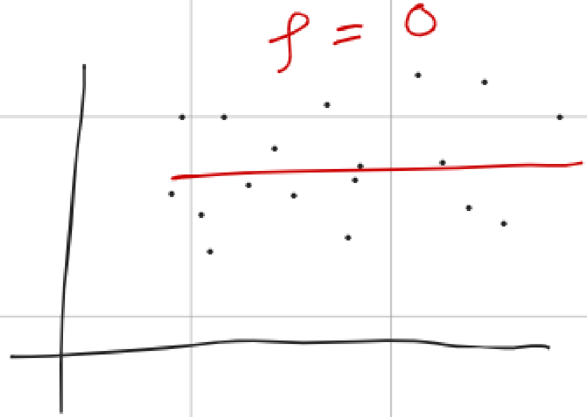
$Y = X^2$

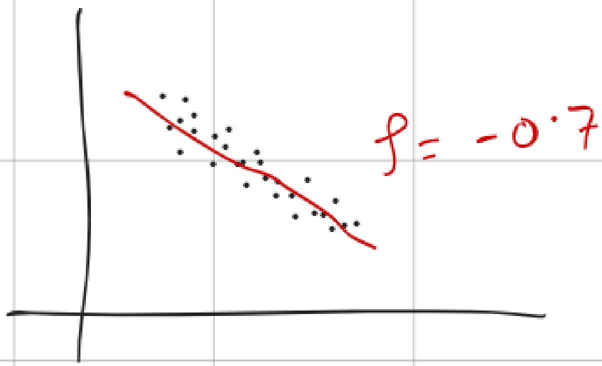
Y is completely determined by X,
X, Y are perfectly dependent,
but $\rho_{XY} = 0$
they are uncorrelated.

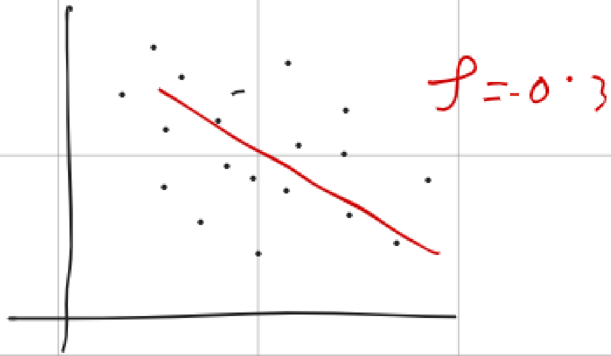
Special case:
[when X, Y are jointly normal, uncorrelatedness is equivalent to Independence]

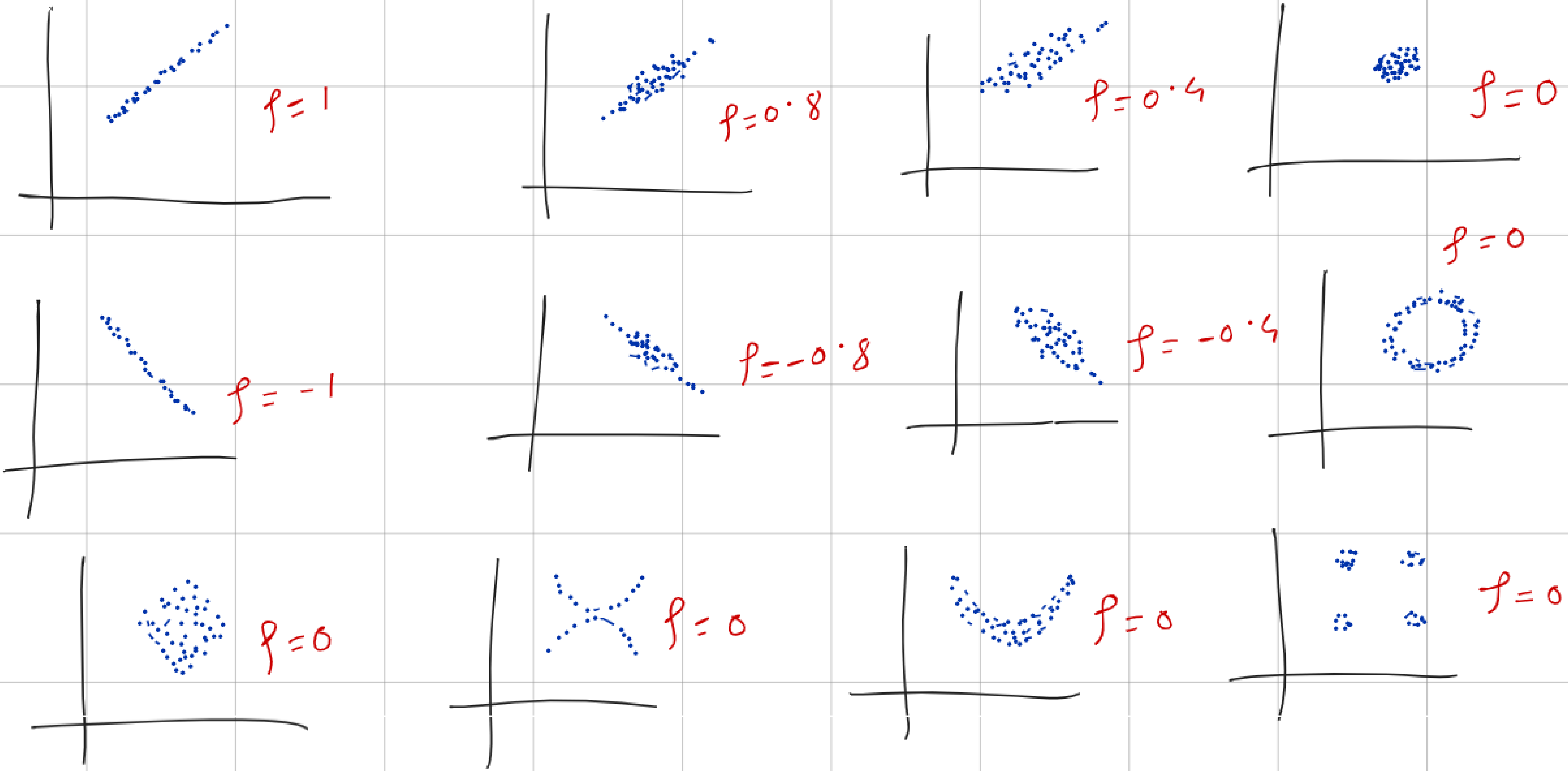

ρ=0·7


ρ=0·3


ρ = 0


ρ= -0·7


ρ=-0·3


$\rho = 1$
$\rho = 0 \cdot 8$
$\rho = 0 \cdot 4$
$\rho = 0$
$\rho = -1$
$\rho = -0 \cdot 8$
$\rho = -0 \cdot 4$
$\rho = 0$
$\rho = 0$
$\rho = 0$
$\rho = 0$
$\rho = 0$

SPEARMAN'S RANK Correlation coefficient

(Self Study)