Probability review III: Measuring dependence

Topics we'll cover

1. When are two random variables **independent**?
2. Qualitatively assessing dependence
3. Quantifying dependence: **covariance** and **correlation**

Independent random variables

Random variables X, Y are **independent** if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Pick a card out of a standard deck.
X = suit and Y = number.

Independent random variables

Random variables X, Y are **independent** if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

Flip a fair coin 10 times.
$X = \#$ heads and Y = last toss.

Independent random variables

Random variables X, Y are **independent** if
$\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$.

$X, Y \in \{-1, 0, 1\}$, with these probabilities:

			Y	
		-1	0	1
	-1	0.4	0.16	0.24
X	0	0.05	0.02	0.03
	1	0.05	0.02	0.03

Dependence

Example: Pick a person at random, and take

$$H = \text{height}$$
$$W = \text{weight}$$

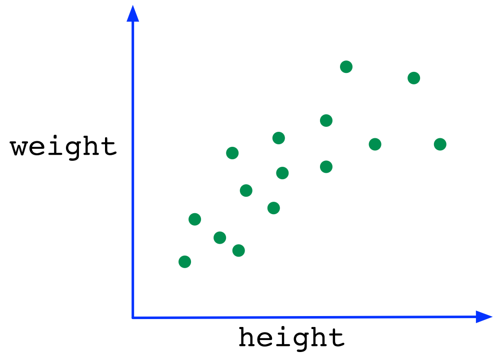
Independence would mean

$$\Pr(H = h, W = w) \;=\; \Pr(H = h)\,\Pr(W = w).$$

Not accurate: height and weight will be **positively correlated**.

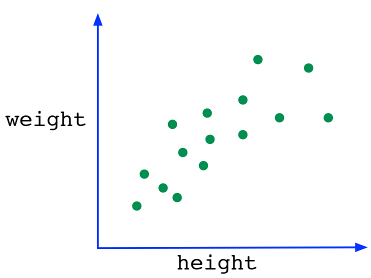
Positive correlation

H, W are **positively correlated**



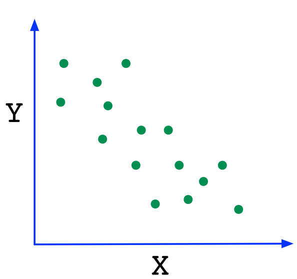
This also implies $\mathbb{E}[HW] > \mathbb{E}[H]\,\mathbb{E}[W]$.

Types of correlation

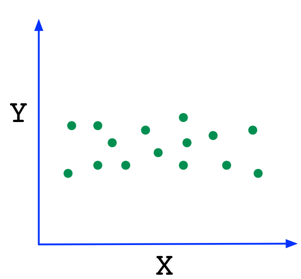


H, W **positively correlated**
This also implies

$$\mathbb{E}[HW] > \mathbb{E}[H]\,\mathbb{E}[W]$$

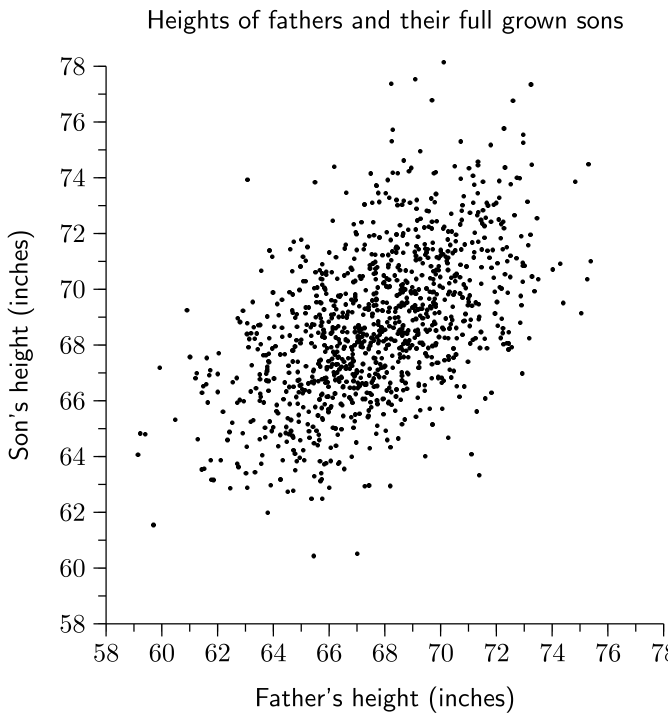


X, Y **negatively correlated**
$\mathbb{E}[XY] < \mathbb{E}[X]\,\mathbb{E}[Y]$



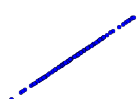
X, Y **uncorrelated**
$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

Pearson (1903): fathers and sons

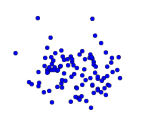


Correlation coefficient: pictures

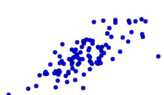
$r = 1$



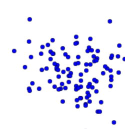
$r = 0$



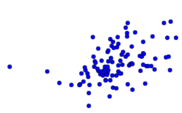
$r = 0.75$



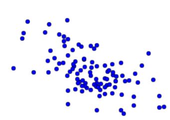
$r = -0.25$



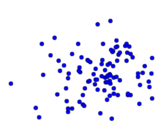
$r = 0.5$



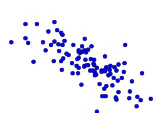
$r = -0.5$



$r = 0.25$



$r = -0.75$



Covariance and correlation

- Covariance

$$\begin{aligned}\mathrm{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

  Maximized when $X = Y$, in which case it is $\mathrm{var}(X)$.
  In general, it is at most $\mathrm{std}(X)\mathrm{std}(Y)$.

- Correlation

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\mathrm{std}(X)\mathrm{std}(Y)}$$

  This is always in the range $[-1, 1]$.

If X, Y independent then $\mathrm{cov}(X, Y) = 0$.
But the converse need not be true.

Covariance and correlation: example

Find $\mathrm{cov}(X, Y)$ and $\mathrm{corr}(X, Y)$

x	y	$\Pr(x, y)$
-1	-3	$1/6$
-1	3	$1/3$
1	-3	$1/3$
1	3	$1/6$