

Quantitative Methods for Political Economy

Week 1: Review of probability I

Wouter Leenders

wouter.leenders@kcl.ac.uk

Random variable

A **random process** is a process with an uncertain outcome, where we describe the process's possible outcomes using probabilities.

The **probability** of an outcome is the proportion of the time that the outcome occurs in the long run.

A **random variable** is a numerical summary of a random outcome.

How many hours of sleep did you get last night?

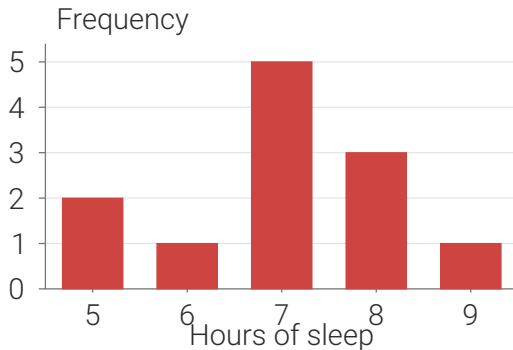
Data Frame

Student	Hours
A	8
B	5
C	8
D	8
E	7
F	7
G	6
H	7
I	9
J	7
K	5
L	7

The sleep data

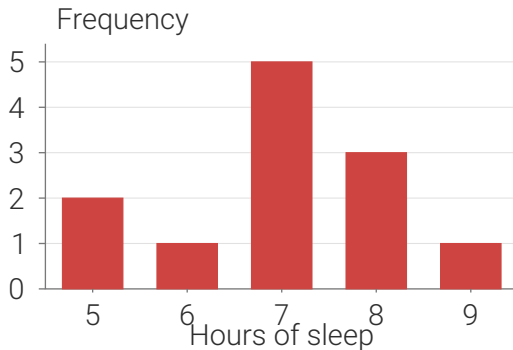
- Unit of analysis: student
- Random variable: number of hours of sleep
- Observed values: sample of twelve students

Sleep: observed frequencies



What is the probability that a student slept 6 hours?

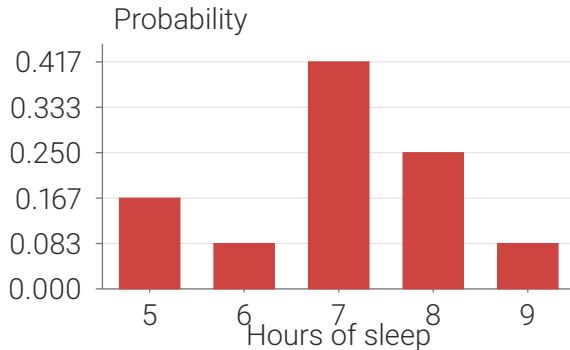
Sleep: observed frequencies



What is the probability that a student slept 6 hours?

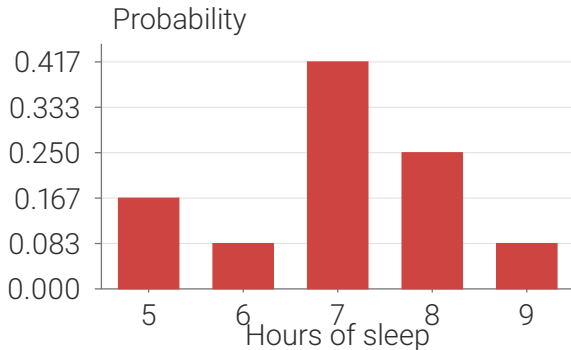
$$\Pr(X = 6) = 1/12 \approx 0.083.$$

Sleep: from frequency to probability



What is the probability that a student slept at most 6 hours?

Sleep: from frequency to probability



What is the probability that a student slept at most 6 hours?

$$\Pr(X \leq 6) = (2 + 1)/12 = 0.25.$$

Probability distribution

The **probability distribution** of a discrete random variable is the list of all possible values of the variable and the associated probabilities

The **cumulative probability distribution** is the probability that the random variable is less than or equal to a particular value

Hours of sleep	$\Pr(\text{hours of sleep} = x)$	$\Pr(\text{hours of sleep} \leq x)$
5	0.167	0.167
6	0.083	0.250
7	0.417	0.667
8	0.250	0.917
9	0.083	1.000

Expected value

The **expected value** of a random variable X , $E(X)$, is the long-run average value of the random variable over many repeated trials or occurrences:

$$\begin{aligned} E(X) &= x_1p_1 + x_2p_2 + \cdots + x_kp_k = \sum_{i=1}^k x_ip_i = \mu_X \\ &= 5\frac{2}{12} + 6\frac{1}{12} + 7\frac{5}{12} + 8\frac{3}{12} + 9\frac{1}{12} \\ &= \frac{84}{12} = 7.0 \text{ hours.} \end{aligned}$$

Variance

The **variance** of a random variable measures the spread of a probability distribution and equals the expected squared deviation of X from its mean.

Student	Hours	Deviation from the mean
A	8	1
B	5	-2
C	8	1
D	8	1
E	7	0
F	7	0
G	6	-1
H	7	0
I	9	2
J	7	0
K	5	-2
L	7	0
Total	84	

Variance

The **variance** of a random variable measures the spread of a probability distribution and equals the expected squared deviation of X from its mean.

Student	Hours	Deviation from the mean
A	8	1
B	5	-2
C	8	1
D	8	1
E	7	0
F	7	0
G	6	-1
H	7	0
I	9	2
J	7	0
K	5	-2
L	7	0
Total	84	0

Variance

The **variance** of a random variable measures the spread of a probability distribution and equals the expected squared deviation of X from its mean.

Student	Hours	Deviation from the mean	Squared deviation
A	8	1	1
B	5	-2	4
C	8	1	1
D	8	1	1
E	7	0	0
F	7	0	0
G	6	-1	1
H	7	0	0
I	9	2	4
J	7	0	0
K	5	-2	4
L	7	0	0
Total	84	0	16

Variance

The **variance** measures spread: it is the expected squared deviation from the mean.

$$\begin{aligned}\text{Var}(X) &= E[(X - \mu_X)^2] = \sigma_X^2 \\&= \sum_{i=1}^k (x_i - \mu_X)^2 p_i \\&= (5 - 7)^2 \frac{2}{12} + (6 - 7)^2 \frac{1}{12} \\&\quad + (7 - 7)^2 \frac{5}{12} + (8 - 7)^2 \frac{3}{12} + (9 - 7)^2 \frac{1}{12} \\&= \frac{16}{12} \approx 1.33 \text{ hours}^2.\end{aligned}$$

Standard deviation

The **standard deviation**, σ_X , is the square root of the variance:

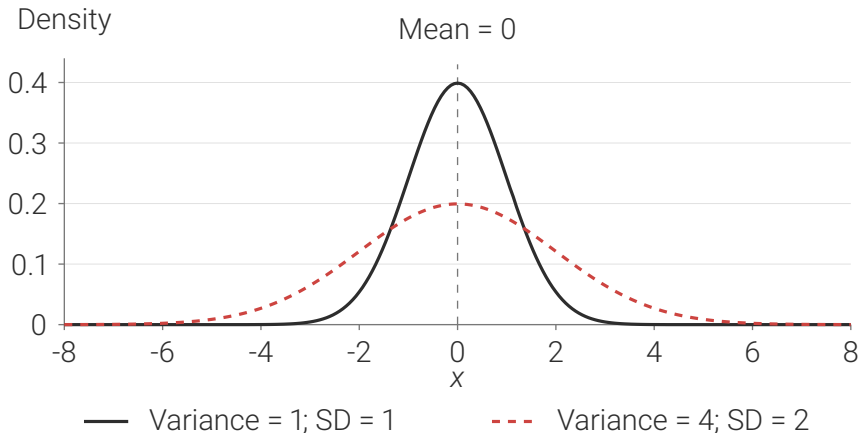
$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{4/3} \approx 1.155 \text{ hours.}$$

The advantage of the standard deviation is that it is measured in the same units as the random variable (hours rather than hours²).

Standard deviation: comparing distributions



Standard deviation: comparing distributions



Squares, cubes and fourth powers

z	z^2	z^3	z^4
3	9	27	81
-3	9	-27	81
5	25	125	625

Suppose z measures an observation's deviation from the mean

- Squares: negative and positive deviations do not offset each other
- Cubes: they preserve a number's sign, so negative and positive deviations offset each other, but they also capture the degree of asymmetry.
- Fourth powers: similar to squares, but extreme deviations are given greater weight.

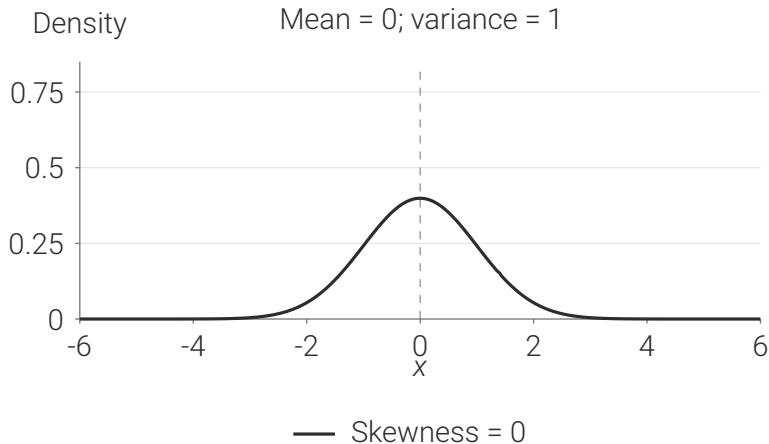
Skewness

Skewness measures the degree of a distribution's asymmetry:

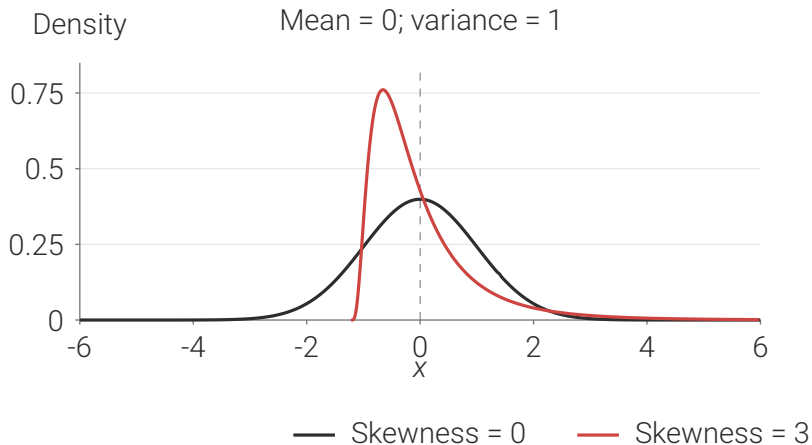
$$\begin{aligned}\text{Skewness} &= \frac{E[(X - \mu_X)^3]}{\sigma_X^3} \\&= \frac{(-2)^3(2/12) + (-1)^3(1/12) + 0^3(5/12)}{(\sqrt{4/3})^3} \\&\quad + \frac{1^3(3/12) + 2^3(1/12)}{(\sqrt{4/3})^3} \\&= \frac{-1/2}{(\sqrt{4/3})^3} \approx -0.325.\end{aligned}$$

The sleep distribution has negative skewness.

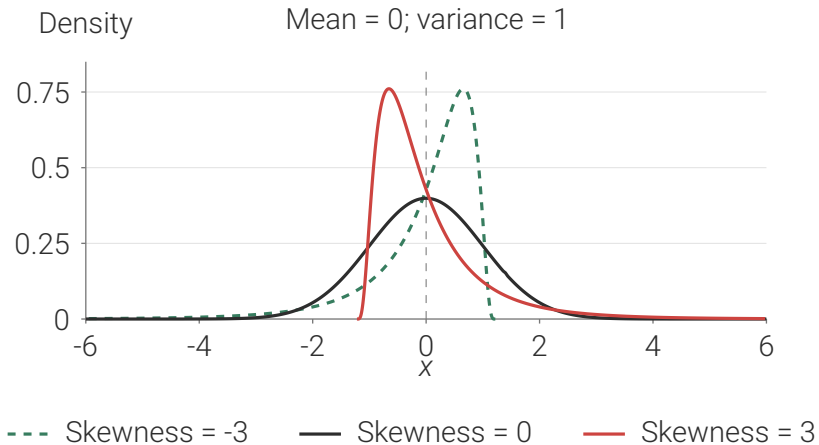
Skewness: comparing distributions



Skewness: comparing distributions



Skewness: comparing distributions



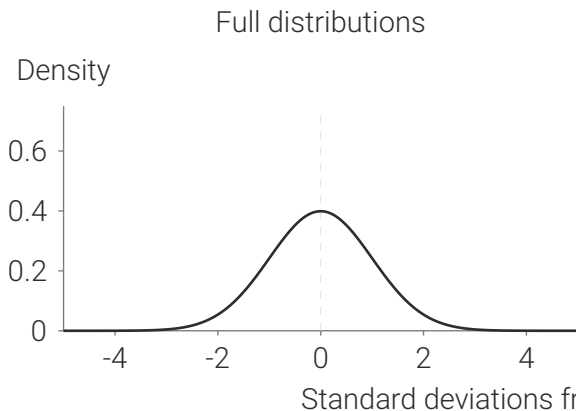
Kurtosis

The **kurtosis** of a distribution is a measure of how much mass is in its tails. Relative to the variance, it is more sensitive to extreme values.

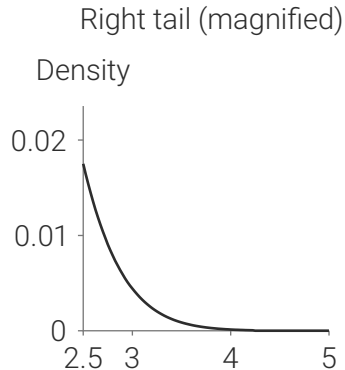
$$\begin{aligned}\text{Kurtosis} &= \frac{E[(X - \mu_X)^4]}{\sigma_X^4} = \frac{\sum_{i=1}^k (x_i - \mu_X)^4 p_i}{\sigma_X^4} \\ &= \frac{(-2)^4(2/12) + (-1)^4(1/12) + 0^4(5/12)}{(\sqrt{4/3})^4} \\ &\quad + \frac{1^4(3/12) + 2^4(1/12)}{(\sqrt{4/3})^4} \\ &= \frac{13/3}{(\sqrt{4/3})^4} = \frac{39}{16} \approx 2.438.\end{aligned}$$

A normal distribution has kurtosis 3.

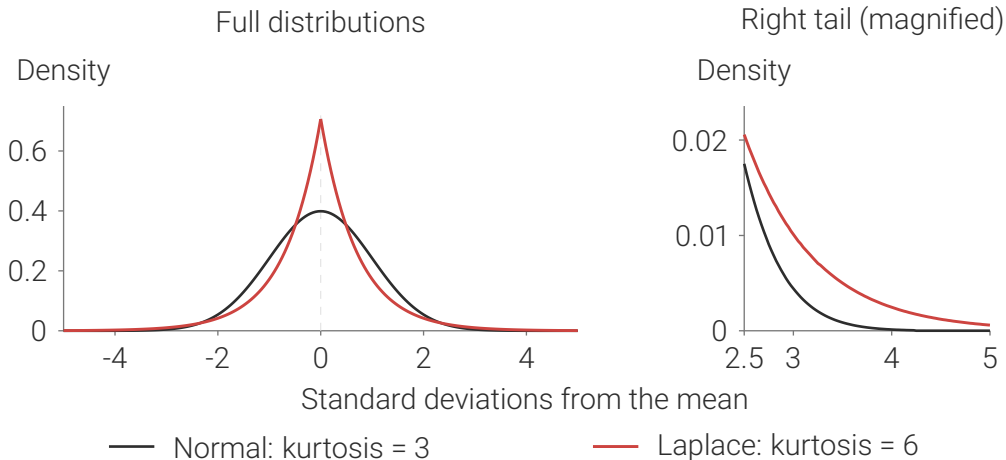
Kurtosis: normal and Laplace distributions



— Normal: kurtosis = 3



Kurtosis: normal and Laplace distributions



Two random variables

- Data from the stevedata R package
- American National Election Study, 1984
- *E*: bachelor's degree or higher (**1**); below bachelor's (**0**)
- *V*: voted (**1**); did not vote (**0**)

Education and voting: observed counts

	Did not vote	Voted	Total
Below bachelor's	600	1,054	1,654
Bachelor's or higher	46	296	342
Total	646	1,350	1,996

Joint distribution

The **joint probability distribution** of two discrete random variables, say E and V , is the probability that the random variables simultaneously take on certain values, say e and v : $\Pr(E = e, V = v)$.

	$V = 0$	$V = 1$	Total
Below bachelor's ($E = 0$)	0.301	0.528	0.829
Bachelor's or higher ($E = 1$)	0.023	0.148	0.171
Total	0.324	0.676	1.000

30.1% of all people had less than a bachelor's degree and did not vote.

Marginal distributions

The **marginal probability distribution** of a random variable V is simply the probability distribution, $\Pr(V = v)$. The term is only used to distinguish it from the joint and conditional distributions.

	$V = 0$	$V = 1$	Total
Below bachelor's ($E = 0$)	0.301	0.528	0.829
Bachelor's or higher ($E = 1$)	0.023	0.148	0.171
Total	0.324	0.676	1.000

32.4% of all people did not vote.

Conditional distribution

The distribution of a random variable V conditional on another random variable E taking on a specific value is the **conditional distribution** of V given E :

$$\Pr(V = v \mid E = e) = \frac{\Pr(V = v, E = e)}{\Pr(E = e)}$$

	$V = 0$	$V = 1$	Total
Below bachelor's ($E = 0$)	0.363	0.637	1.000
Bachelor's or higher ($E = 1$)	0.135	0.865	1.000

36.3% of all people with less than a bachelor's degree did not vote.

Conditional probabilities

	Did not vote	Voted	Total
Below bachelor's	600	1,054	1,654
Bachelor's or higher	46	296	342
Total	646	1,350	1,996

$$\Pr(V = 1 \mid E = 1) = \frac{296}{342} \approx 86.5\%,$$

$$\Pr(E = 1 \mid V = 1) = \frac{296}{1350} \approx 21.9\%,$$

$$\Pr(E = 1, V = 1) = \frac{296}{1996} \approx 14.8\%.$$

Independence

Two random variables X and Y are **independently distributed** if knowing the value of one of the variables provides no information about the other:

$$\Pr(Y = y \mid X = x) = \Pr(Y = y)$$

Equivalently, the joint distribution is the product of the marginal distributions

$$\Pr(X = x, Y = y) = \Pr(X = x) \Pr(Y = y)$$

Independent and dependent random variables

Consider a fair die that is rolled twice. Let X be the result of the first and Y the result of the second roll. Let $S = X + Y$ be the sum of the results.

- X and Y are **independent**:

$$\Pr(Y = 6 \mid X = 6) = \Pr(Y = 6) = \frac{1}{6}.$$

- X and S are **dependent**:

$$\Pr(S = 2) = \frac{1}{36} \neq \Pr(S = 2 \mid X = 6) = 0.$$

In the education-voting data:

$$\Pr(V = 1 \mid E = 1) = 0.865 \neq \Pr(V = 1 \mid E = 0) = 0.637.$$

Education and voting are not independent.

Covariance and correlation

The **covariance** of two random variables measures the extent to which they move together:

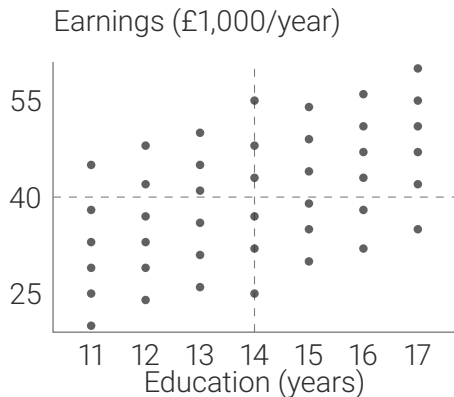
$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= \sum_{i=1}^k \sum_{j=1}^l (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i, Y = y_j).\end{aligned}$$

If Y is above average when X is above average, X and Y have a positive covariance.

The **correlation** between X and Y is the covariance between X and Y divided by their standard deviations

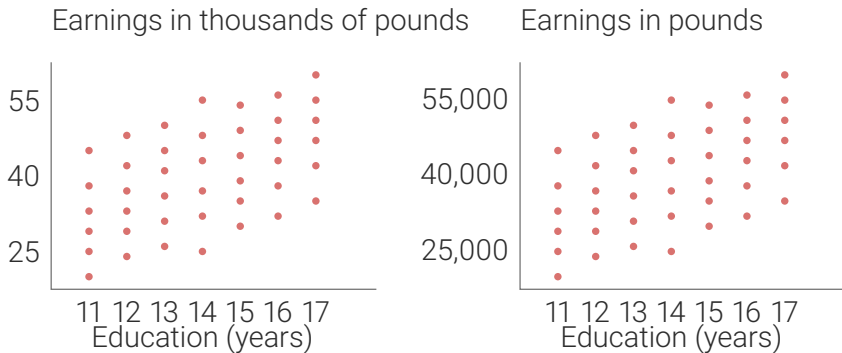
$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Covariance: deviations from the means



When education is above average ($\mu_X = 14$), earnings also tend to be above average ($\mu_Y = 40$).

Correlation: invariant to unit of measurement



$$\text{Cov}(X, Y) \approx 10.24$$

$$\text{Cov}(X, 1000Y) \approx 10,238.10$$

$$\text{corr}(X, 1000Y) = \frac{1000 \text{Cov}(X, Y)}{\sigma_X(1000\sigma_Y)} = \text{corr}(X, Y) \approx 0.519.$$