



SO5041 Unit 7: Distributions

Brendan Halpin, Sociology

Autumn 2023/4

SO5041 Unit 7

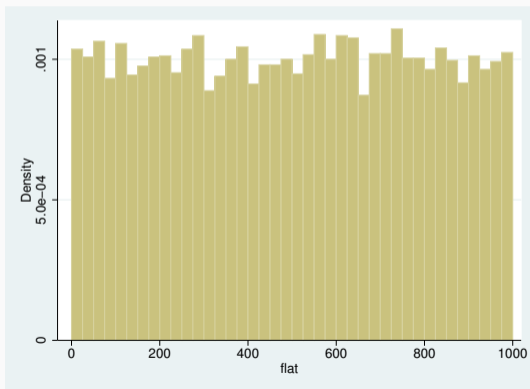
Characteristics of distributions

Distributions

- We have seen how to display and summarise the distribution of variables:
 - Categorical: frequency distribution, percentage distribution, bar and pie charts
 - Quantitative (interval/ratio): mean, median, IQR, standard deviation, histogram, box-plot

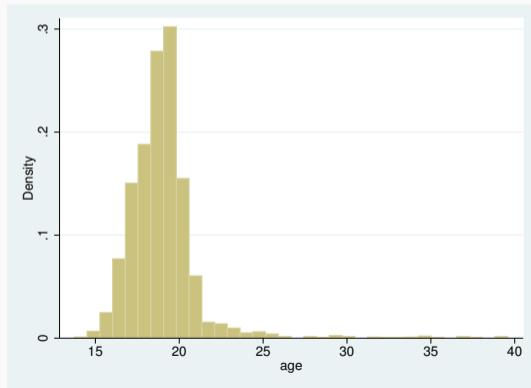
Distributions have shapes

- The shape of the histogram tells us about the distribution of the variable
- If a variable is “uniformly” distributed we see a flat distribution between the extremes:



Heaped distributions

- More often we see “heaped distributions” where more of the observations cluster around the centre, like this example (age) from the 1999 school-leavers’ survey:



Many patterns

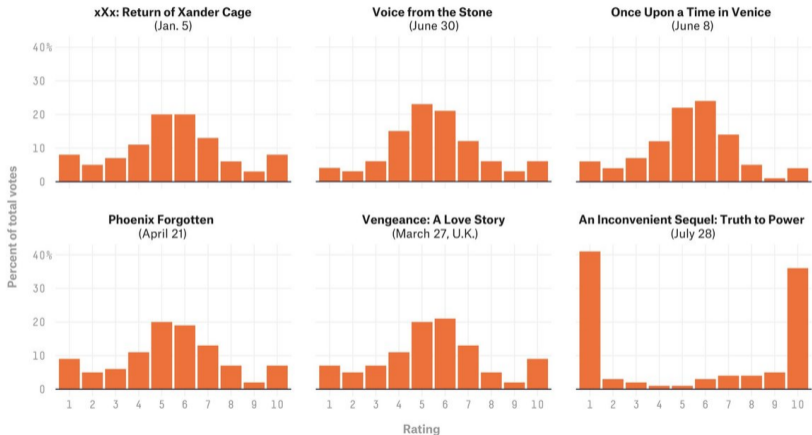
There are many patterns we might see:

- Uniform
- Extremes
- Bimodal
- Uni-modal

Polarisation

One of these films is not like the others

The six movies from 2017 with an IMDb average user rating of 5.2 on Aug. 16.



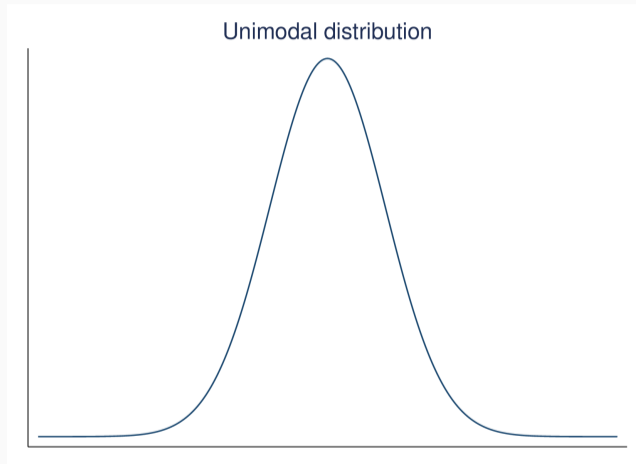
FiveThirtyEight

SOURCE: IMDb

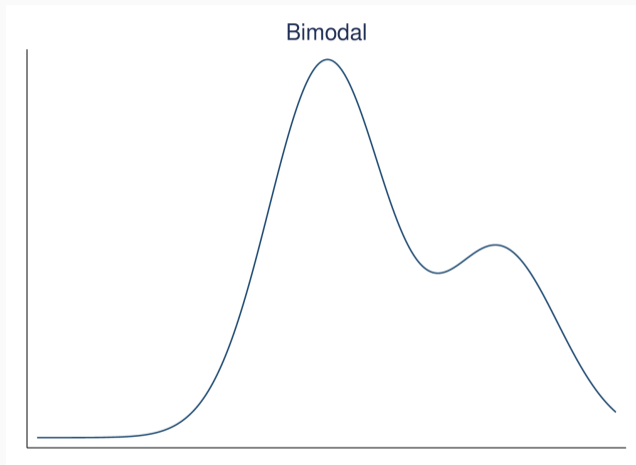
Uni-modal differences

- Symmetric (with different levels of **kurtosis**)
 - platykurtic – flatter
 - mesokurtic – average
 - leptokurtic – very concentrated around centre
- Asymmetric
 - Positively skewed (to right)
 - Negatively skewed (to left)

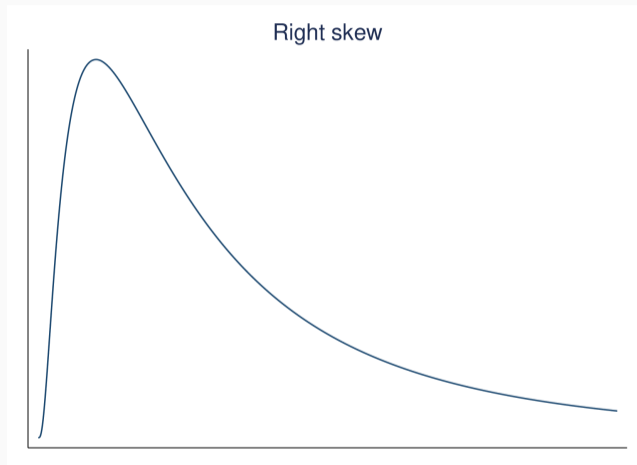
Symmetric unimodal



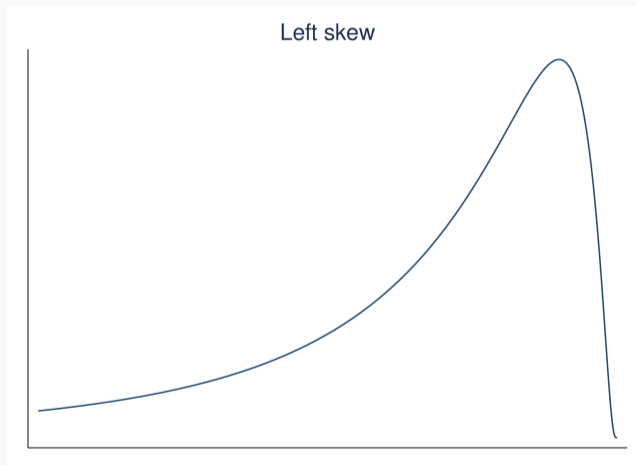
Asymmetric bimodal



Asymmetric: right skew



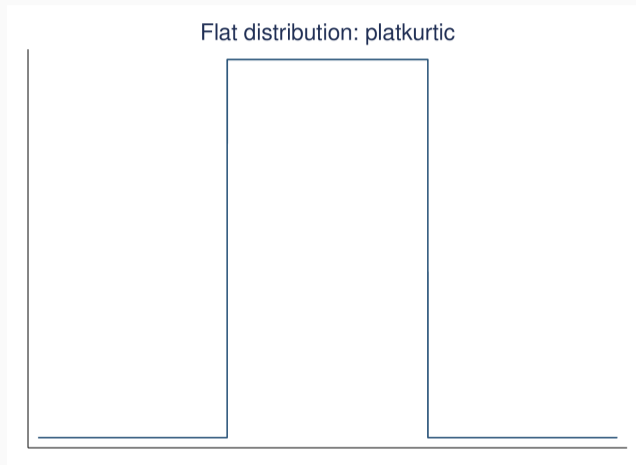
Asymmetric: left skew



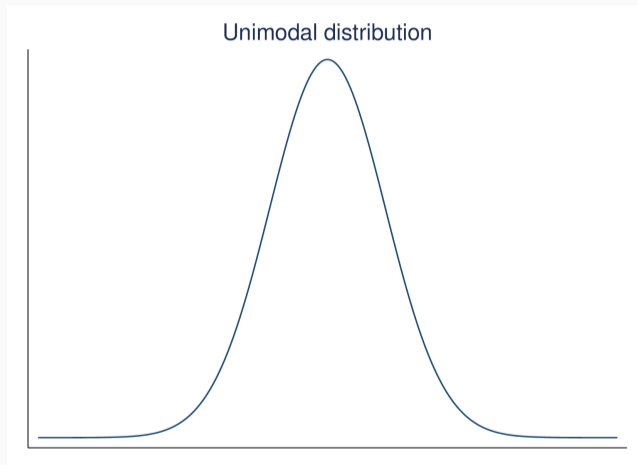
Different symmetric shapes: kurtosis

- Different distributions with the same mean and standard deviation can have different shapes
- Kurtosis: balance between peak, shoulders and tails

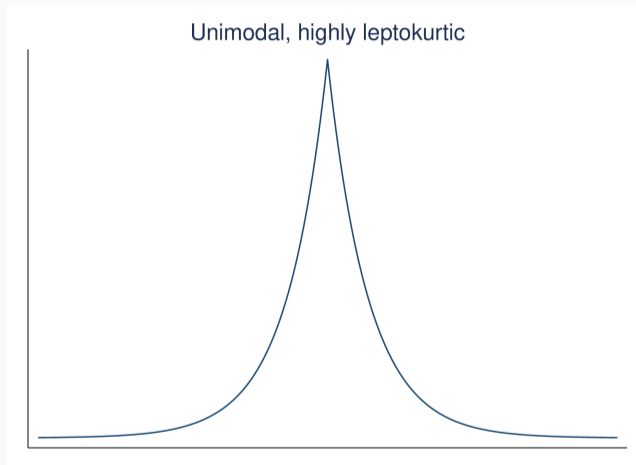
Flat: low kurtosis



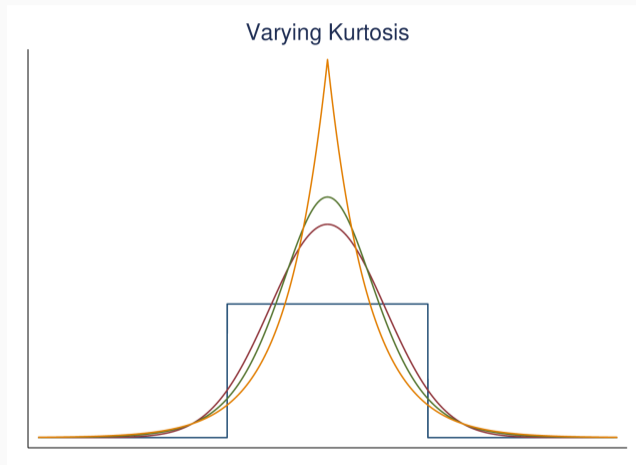
Normal: mid-kurtosis



Very peaky: high-kurtosis



Varying kurtosis

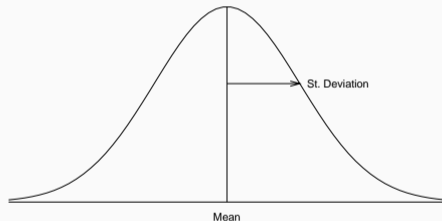


SO5041 Unit 7

The Normal distribution

Mathematically defined distributions

- There are some patterns that are defined “theoretically” or mathematically
- Among these the most important is the **Normal Distribution**:



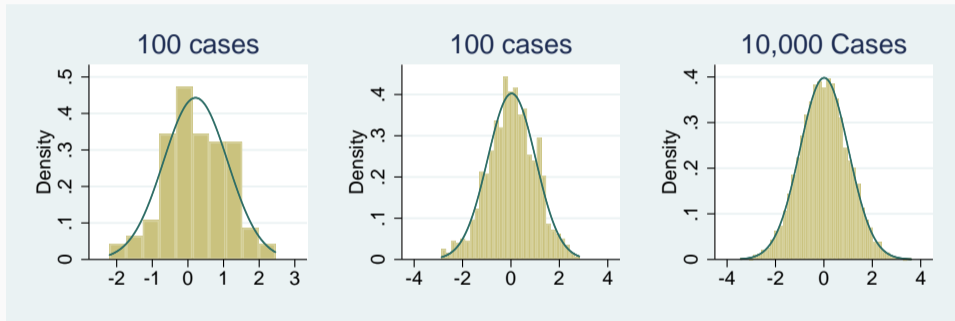
- This is the famous “bell-shaped curve”

The Normal Distribution

- The normal distribution is
 - symmetric (no skew)
 - mesokurtic (between flatter and pointier)
- The mean, mode and median are the same
- The farther you go from the mean, the lower the proportion of cases, in each direction symmetrically

Histograms display distributions

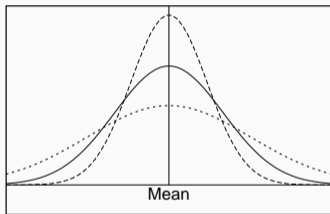
- We can see what a histogram of a variable drawn from a normally distributed population looks like:



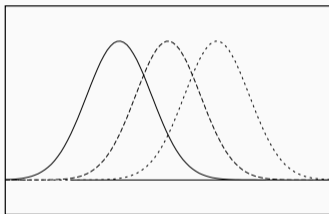
- As the sample gets bigger, the histogram approximates the theoretical distribution better

Normal: mean and standard deviation

- What makes the normal distribution useful is that its form is well understood:
 - It is completely characterised by its mean and its standard deviation



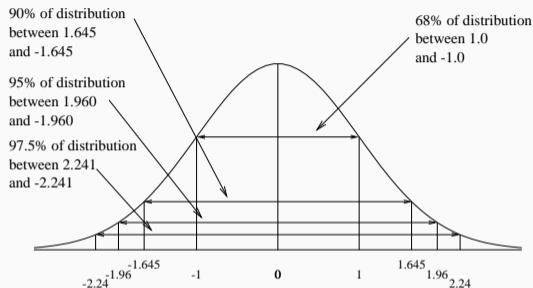
Same mean, different SD



Same SD, different mean

Online app: <http://teaching.sociology.ul.ie:3838/apps/normsd>

Normal distribution: well-understood



- About 68% of the cases in a normal distribution are within 1 SD on either side of its mean
- 95% are within ± 1.96 standard deviations of the mean
- 97.5% are within ± 2.24 standard deviations of the mean