Sampling and hypothesis tests
Part 4: Data distributions (2/2)
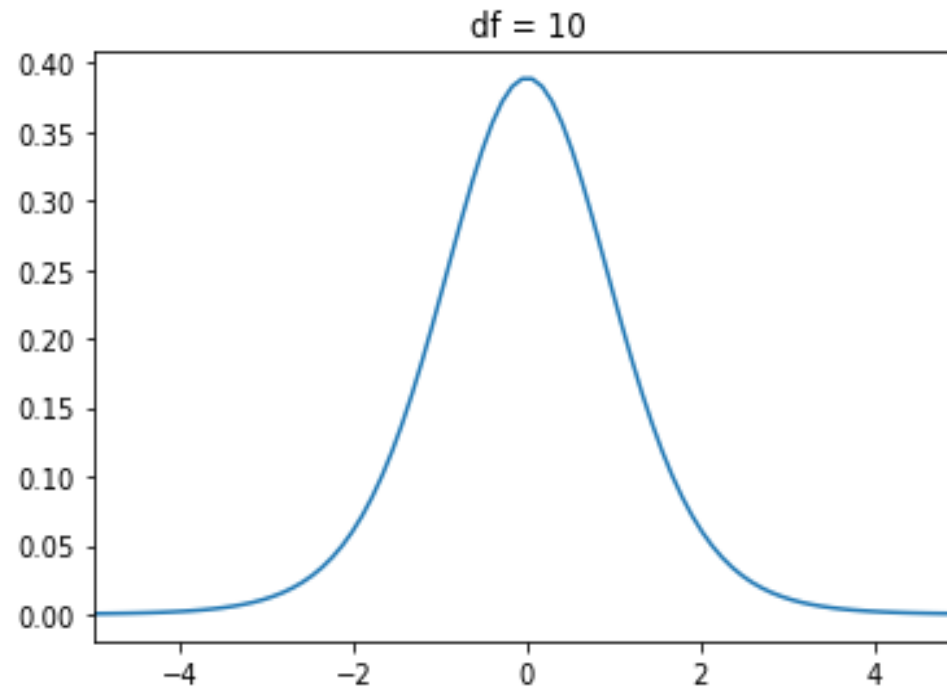
By: Noureddin Sadawi, PhD

University of London

Student's t-distribution

- 'The t-distribution is a normally shaped distribution, except that it is a bit thicker and longer on the tails.

- It is used extensively in depicting distributions of sample statistics.

- Distributions of sample means are typically shaped like a t-distribution, and there is a family of t-distributions that differ depending on how large the sample is.

- The larger the sample, the more normally shaped the t-distribution becomes.'

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Example

The t-distribution is a normally shaped distribution, except that it is a bit thicker and longer on the tails.

Key terms and ideas for student's t-distribution

- '**n:** Sample size.
- **Degrees of freedom:** A parameter that allows the t-distribution to adjust to different sample sizes, statistics, and numbers of groups.
- The t-distribution is actually a family of distributions resembling the normal distribution but with thicker tails.
- The t-distribution is widely used as a reference basis for the distribution of sample means, differences between two sample means, regression parameters, and more.'

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Binomial distribution

This distribution to directly related to experiments where the outcome is binary (e.g. Yes/No, Success/Fail).

It is based on the idea of having a set of trials, where each trial can have one of two possible outcomes with definite probabilities.

'For example, flipping a coin 10 times is a binomial experiment with 10 trials, each trial having two possible outcomes (heads or tails).'

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Key terms binomial distribution

- '**Trial:** An event with a discrete outcome (e.g., a coin flip).
- **Success:** The outcome of interest for a trial.
  - Synonym: "1" (as opposed to "0")
- **Binomial:** Having two outcomes.
  - Synonyms: yes/no, 0/1, binary
- **Binomial trial:** A trial with two outcomes.
  - Synonym: Bernoulli trial
- **Binomial distribution:** Distribution of number of successes in x trials.
  - Synonym: Bernoulli distribution.'

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).

Other distributions

- '**The chi-square distribution:** is typically concerned with counts of subjects or items falling into categories.

- **The F-distribution:** is used with experiments and linear models involving measured data.

- **Poisson distribution:** the frequency distribution of the number of events in sampled units of time or space.

- **Exponential distribution:** the frequency distribution of the time or distance from one event to the next event.

- **Weibull distribution:** a generalized version of the exponential distribution in which the event rate is allowed to shift over time.'

(Bruce and Bruce *Practical statistics for data scientists*, second edition, 2020).