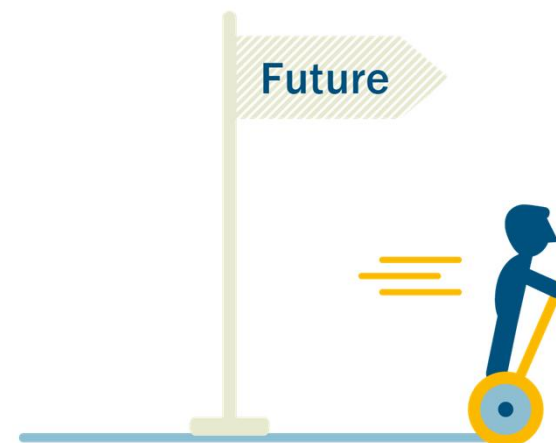


SCOs and statistics

05 November 2020 | Online

Interact



Welcome & housekeeping

- Welcome!
- Please change your name:
first name + programme
Example: Neil (Interreg Outerspace)
- Use the chat for your questions

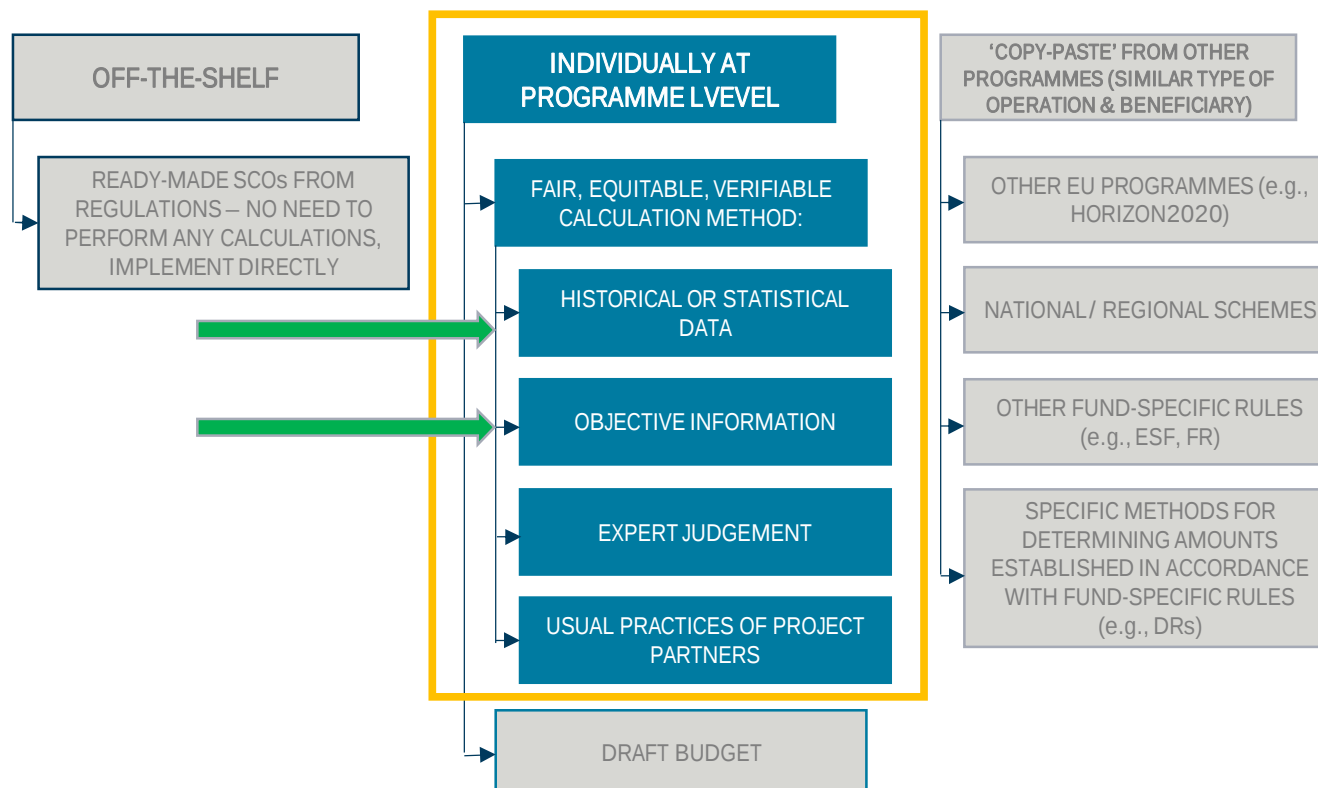
Thanks for spending part of your afternoon with us!

- Why are we here?
- What will happen?
 - Part 1: Interact presentation on basics + QA
 - Part 2: Expert presentation on working with statistics + more QA
- Question(!) & Answers(?)
- Follow-up?



Requirements for calculation methodology

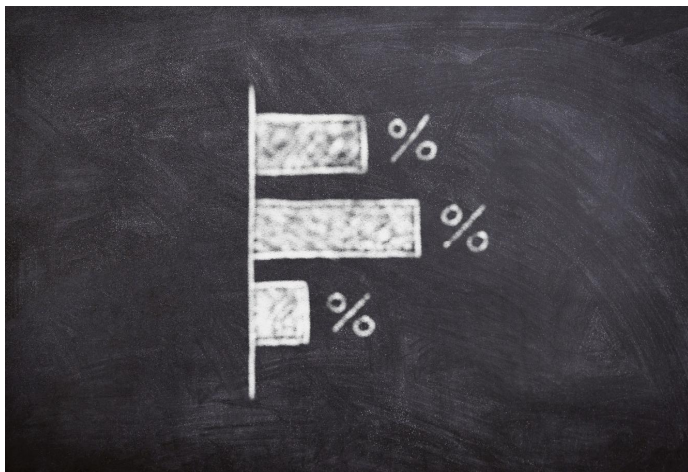
- Ex-ante: established in advanced, at the latest before the signature of the subsidy contract!



General recommendation

- Don't overdo it
- Common sense
- Be consistent and coherent
- Know your data

SCOs in Interreg



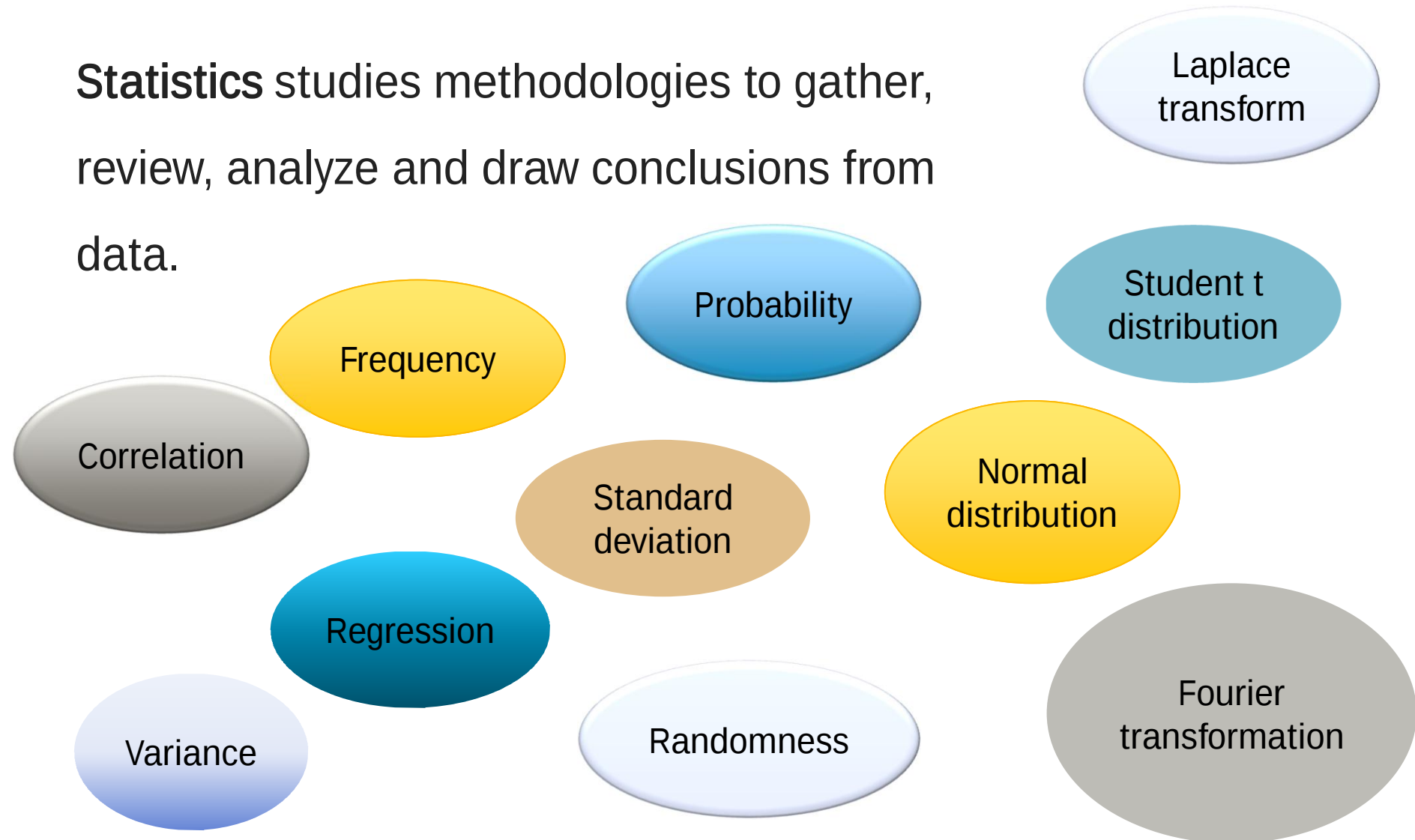
Simplified cost options

- Flat rate financing (Art. 68 CPR*)
- Standard scale of unit costs (Art. 67(b) CPR)
- Lump sum (Art. 67(c) CPR)

**with changes introduced by Omnibus Regulation
(former Articles 48-51)*

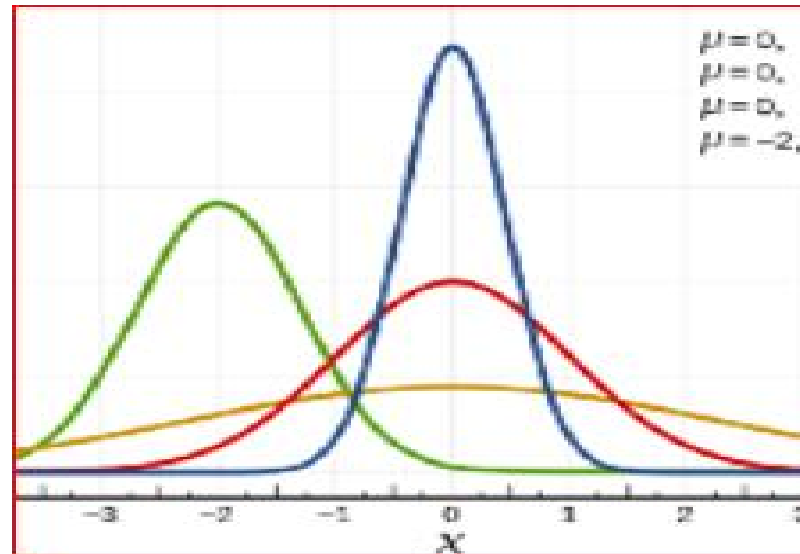
Statistics

Statistics studies methodologies to gather, review, analyze and draw conclusions from data.



Normal distribution

Introducing Mr Johann Carl Friedrich Gauss
(1777 - 1855)



“I am confident that 95% of all data should fall within 2 standard deviations away from the mean”

Normal distributions

Examples

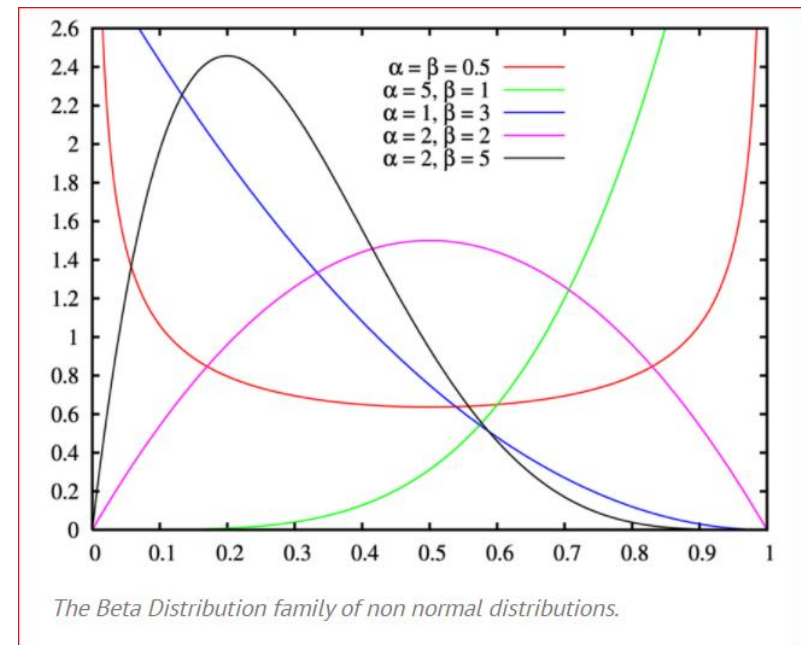
- Rolling a dice
- Tossing a coin
- IQ in a population
- Shoe size
- Birth weight
- Body temperature
- Blood pressure



Not normal distributions

Examples of not normal distributions

- Exponential distribution – bacteria growth
- Poisson distribution – number of accidents
- Weibull distribution – survival time of products
- Gamma distribution – reaction times in psychology



Normal distributions

Basics definitions

- The **mean** (informally, the “average”) is found by adding all of the numbers together and dividing the result by the number of items in the set: $10 + 10 + 20 + 40 + 70 / 5 = 30$.
- The **median** is found by ordering the set from lowest to highest and finding the exact middle. The median is just the middle number: 20.
- The **mode** is the most common number in a set. For example, the mode in this set of numbers is 21:
21, 21, 21, 23, 24, 26, 26, 28, 29, 30, 31, 33

Normal distributions

Features

- Normal distributions are symmetric around their mean
- The mean, median, and mode of a normal distribution are equal
- Normal distributions are denser in the center and less dense in the tails

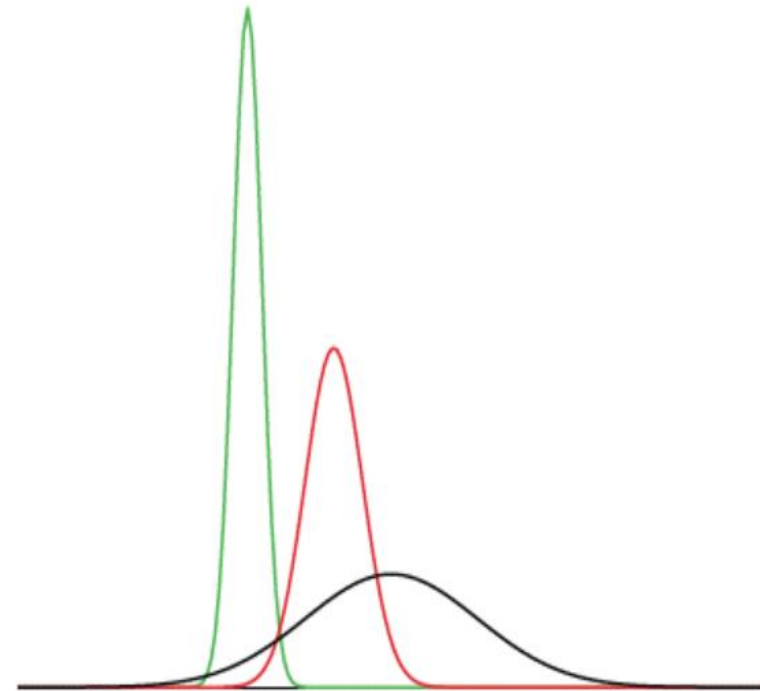
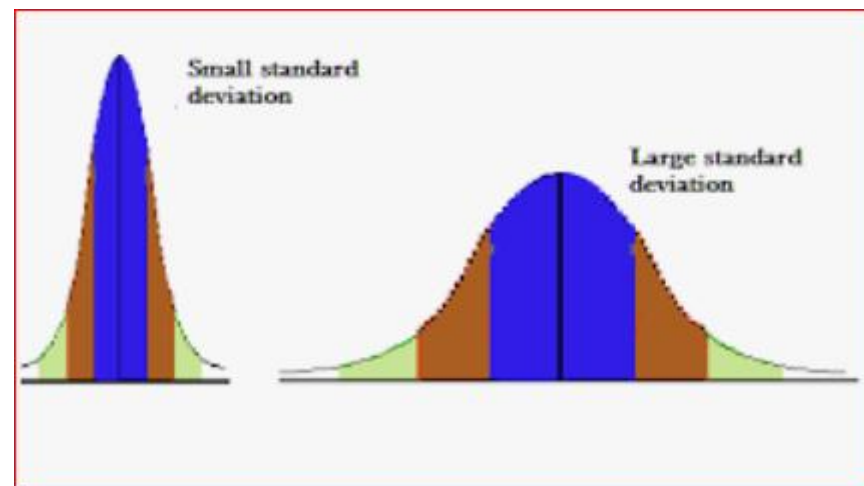
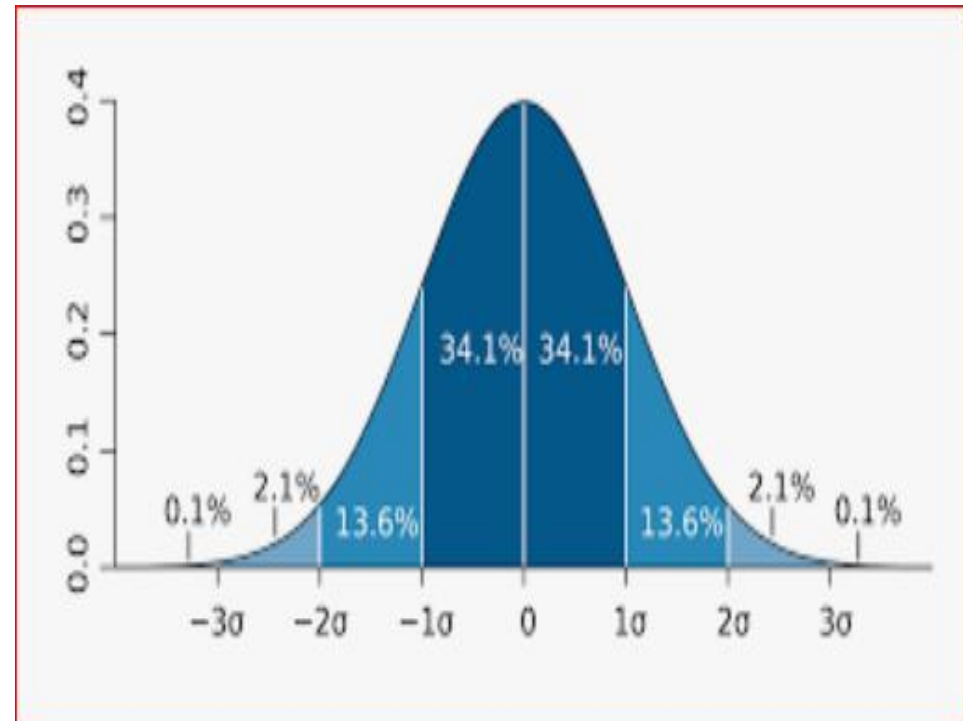


Figure 1. Normal distributions differing in mean and standard deviation.

Normal distributions

Features

- Normal distributions are defined by two parameters, the mean and the standard deviation.
- 68% of the area of a normal distribution is within one standard deviation of the mean.
- Approximately 95% of the area of a normal distribution is within two standard deviations of the mean.



Some other useful statistical terms

- A common aim of statistical analysis is to produce information about some chosen **population**
- A **sample** is collected, and statistics are calculated from the samples and extrapolated to the population
- **Margin of error** is the range of values below and above your sample statistic, how many % your calculation will differ from real costs



Why to use normal distributions in our calculations?

- Relatively simple calculations of mean (average)
- Can be done in excel, no need for sophisticated software
- We are not creating scientific papers
- We are not having huge populations or samples
- Simplification!



What can go wrong?

Insufficient data

- Too small sample, not enough data
- We may get very scattered data e.g. 43 - 55 - 78

Multimodal distribution

- Maybe your data covers 2 or more types
- Costs covering catering and costs not covering catering
- Analyse separately

Measurements mistakes (accountancy, reporting)

- Make sure your data is clean

Existing boundaries

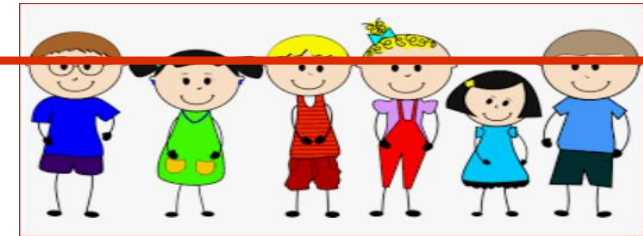
- Events cannot cost $< 0\text{€}$

What can go wrong?

Outliers

What do we do with them

- If this is error in measurement, data collection and you know it and can prove it, delete it
- There are many, very sophisticated numerical methods on how to diminish their influence
- Data within 2 standard deviations are OK, and some (ca 5% of total should be 3 SD away from the mean and it is still OK)
- Use common sense



Interact data

What we wanted to check

- Possibility to set up unit costs and lump sums for organisation of meetings/ events in Interreg

Which data we used

- The maximum number of events by Interact were organized in Belgium

Why Interact data?

- We needed to start from somewhere
- We (used to) organise plenty of events and meetings all over Europe



Interact data

Initial dataset

- 68 events organized by Interact 2017-2019

Which data we eliminated

- Events organised by IO Valencia (specific conditions of hosting institution)
- Events where the number of participants was not known / clear

What is our sample size for calculations

- 56 datasets



Our assumptions

Population

- All events organised within 3 years horizon (in line with EC's Guidance for SCOs)
- Only events organised in MS
- All events organised by a programme (TA) and all it's projects

Size of population

- 2000 events / 1 year / 1 programme
- 6000 events / 3 years / 1 programme
- X 100 programmes

Interreg population size 600 000

Our assumptions and calculations

Population size for Belgium

- All MS population 600 000
- Divide by 27 MS (simplicity)
- Belgium population size 22 222
- What sample size would be ideal?

Sample size: **378**

This means 378 or more measurements/surveys are needed to have a confidence level of 95% that the real value is within $\pm 5\%$ of the measured/surveyed value.

Confidence Level: ?	95%	▼
Margin of Error: ?	5%	
Population Proportion: ?	50%	Use 50% if not sure
Population Size: ?	22222	Leave blank if unlimited population size.

Calculate ▶ **Clear**

Our calculations

The smaller the sample the bigger the margin of error

Margin of error: **13.08%**

This means, in this case, there is a 95% chance that the real value is within $\pm 13.08\%$ of the measured/surveyed value.

Confidence Level:  95% 

Sample Size:  56

Population Proportion:  50%

Population Size:  22222 Leave blank if unlimited population size.

Calculate



Clear

Acceptable margin of error

It depends....

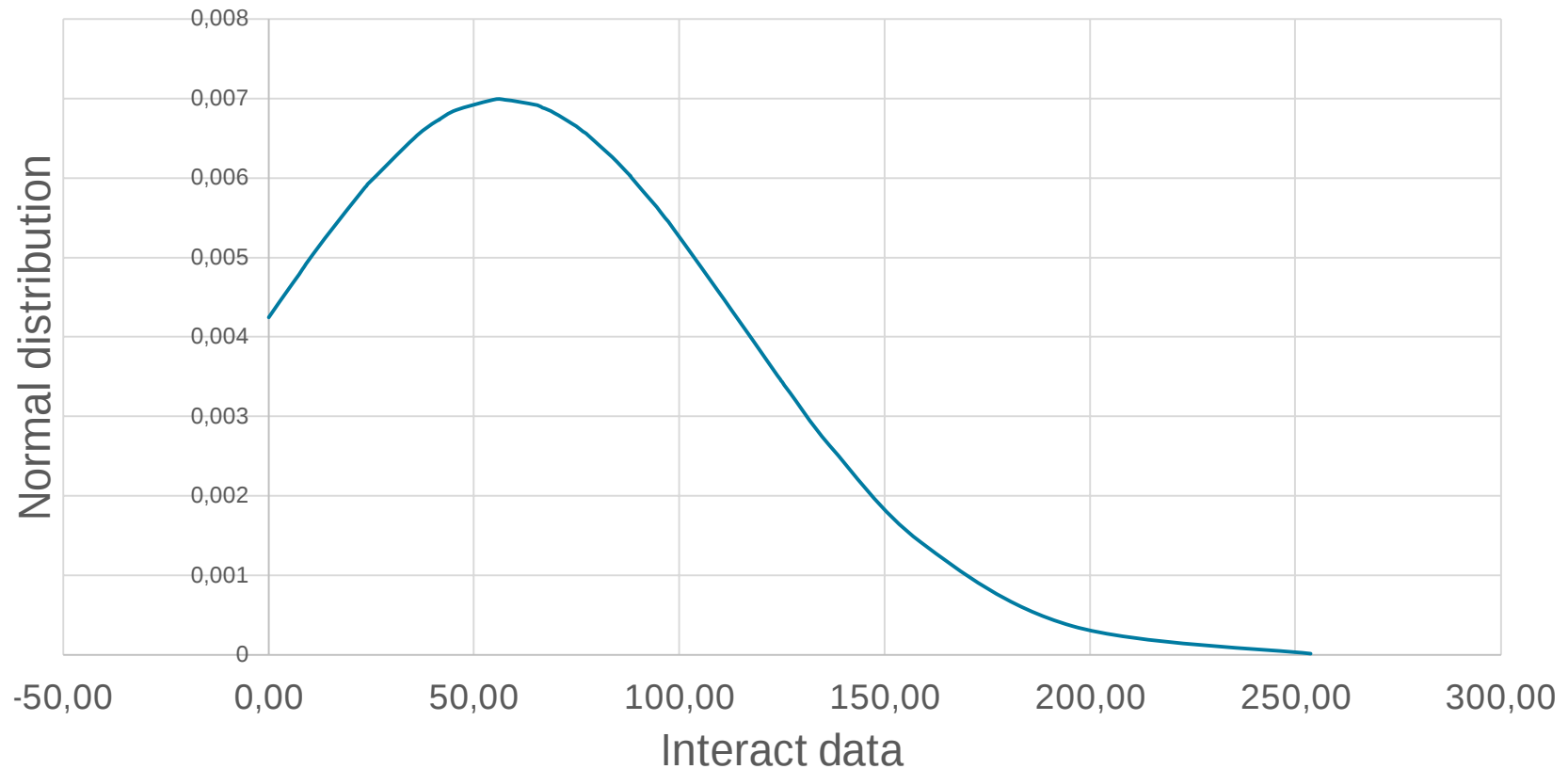
When the confidence level is 95%

- 4% - 8 % in surveys (who will win election)
- Up to 10% in lab science
- Up to 5% in scientific papers

In our case

- The population is bigger than sample
- The bigger sample the smaller error
- ... increase a sample to >300 and the margin of error will be ca 5%

Our data normal distribution



- Flat distribution (high standard deviation)
- Relatively high margin of error

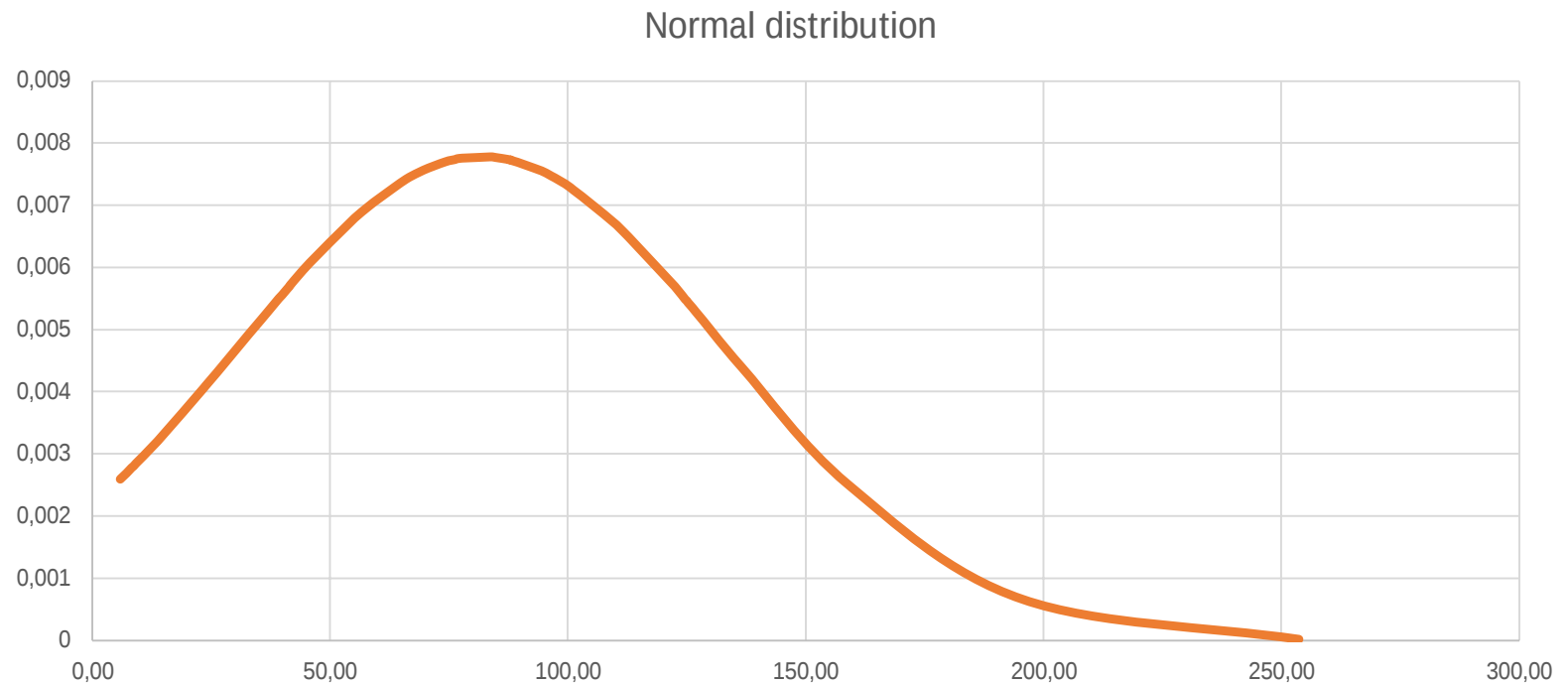
Our calculations results

- Calculated mean value: 56,99€
- Calculated standard deviation: 57,06
- When we compare with ERASMUS+

Country / €	Interact	Jean Monnet (ERASMUS+)	% (Interact/Erasmus)
Belgium	56,99	88,00	64,76%

Eliminating 0€ events

- 39 data points
- Calculated mean value: 81,84€
- Calculated standard deviation: 51,25
- Calculated margin of error ca 16% (smaller sample)



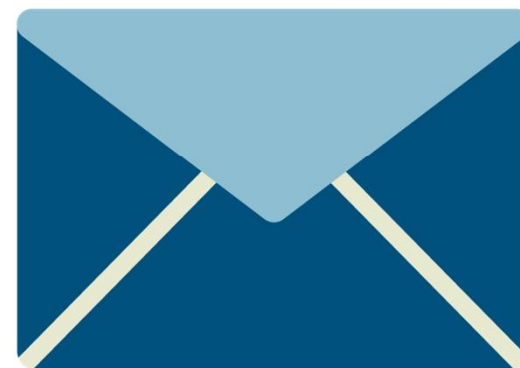
Useful links

- [One of the sample size on line calculators](#)
- [Another sample size calculator](#)
- [Margin of error vs standard deviation](#)
- [Margin of error calculator](#)
- [Statistics for dummies](#)



Any SCO-related questions ...

sco@interact-eu.net



Cooperation works

All materials will be available on:

www.interact-eu.net