



eSTEM

université
de BORDEAUX



With the support of the
Erasmus+ Programme
of the European Union

Toolboxes for SuperFastLearning digital contents in STEM

Example presentation IO3

**What is effect size and
how to use effect size
indices**



The topic: What are we talking about?

Effect size is a concept of applied statistics which is more and more used in data processing, and which complements computation of the statistical significance of a result.

Let consider a drug supposed to have anti-cough properties. When you test it on patients, you find a reduction in the cough of the patients who received the drug, as compared to the control patients who did not receive it. Following this observation, biologists will use a statistical test on these data, and if this statistical test indicates a "statistically significant" reduction in cough, they will consider that the drug is efficient. In rigorous words, statistically significance indicates that the decrease in cough is unlikely due to the hazard of patient sampling, and therefore it is likely that it is actually the drug which promoted the decrease.

However, a statistically significant result does not ensure that the new medicine is useful. For example, if you demonstrate that the drug is efficient, but that the drug reduces coughing by 3% only, you will conclude that the comfort of the patient is not really improved. Even, he will probably be unable to perceive the difference. Effect size indices have been developed to quantify the effect, i.e. indicate whether it is strong, weak, neglectable...

Though some scales have been proposed to interpret ES indices, interpretation is never straightforward. We saw that 3% coughing reduction may be considered as uninteresting for a patient. If we find a drug which reduces mortality of 3% for a given lethal sickness it is likely that these 3% will be considered as "very interesting".

The aim of this SBL sequence is to teach how to choose an ES index, how to interpret it, and how to use it to help taking decisions.

Teaching context

Target students are enrolled in the 3rd year University, second semester. All have followed a basic teaching in applied statistics, during which they are trained to choose, use and interpret basic statistical tools (essentially inferential tests). The present UE is optional. About 10% students choose this UE. Since the automatisms have been learned during the 1st semester, the aim of the present UE is i) to train them to get a critic view on statistical practice and ii) to initiate them to other, more specific statistical tools (initiation to multivariate statistics, resampling methods, and tools like ES indices).

Most of these students will apply for masters in bioinformatics, epidemiology or ecology.

This course will be proposed at the end of course sequence, for autonomous completion.

Intended learning outcomes

Intended learning outcomes centered on effect size:

- To be able to choose an appropriate effect size index
- To be able to compute the selected effect size indices with R
- To be able to interpret effect size indices and use them to take decisions

Secondary benefits

- To demonstrate autonomy in basic data statistical processing
- To be able to learn new methods autonomously



The session in practice

- Each student decides whether he wants to perform the sequence alone or by 2 or 3 and choose his co-workers
- We expect the sequence will take about 2hrs.
- The sequence will be implemented on a digital tool and taking the structure of an interactive story with conditional choices. Each page will present a chapter and the students will have to click on hypertext links to go to the next page. Several links will be proposed when students have to choose between several options.
- When documentation is required, hypertext links will point on the appropriate articles, websites, figures...
- A beginning and ending time will be indicated. During that time, a teacher will be available for questions by e-mail and zoom. However, the teacher will answer only to technical questions and will not help about the topic.

Evaluation/remediation

- Evaluation will consist in a quiz proposed after the final step is reached
- Quiz is performed individually
- Each student is informed at the end of the quiz about his note, but not what are the right answer. They can do the quiz as many times as they wants.
- Questions bear on the knowledge which should be mastered thanks to the SBL sequence

Why using an SBL sequence for this lesson?

With the development of informatics, new statistical methods are presently developed. Most often, during their professional life, students will have to learn these new methods by themselves. SBL is a good way to show them that they are able to learn statistics without a teacher.

Effect size is a concept which is intuitive and not difficult to grasp. It provides a good topic for this aim and is easy to understand in 2 hrs.

At the end of a semester, it is a way to show how knowledge acquired during the year may be used later, in professional life.

Effect size was not used 10 years ago. It is a concept that many scientists (biologists, psychologists...) had to learn to master by themselves.

The use of statistical tools in professional life implies that a specific rationale in the course of data processing. The sequential organization of tasks in SBL makes this approach particularly suited to train students to respect this rationale.

Organigram (decision tree) of the sequence

Screen 1. Presentation of the problem.

You are developing a new fertilizer. You need to know whether it is better than another one which was developed by another company.

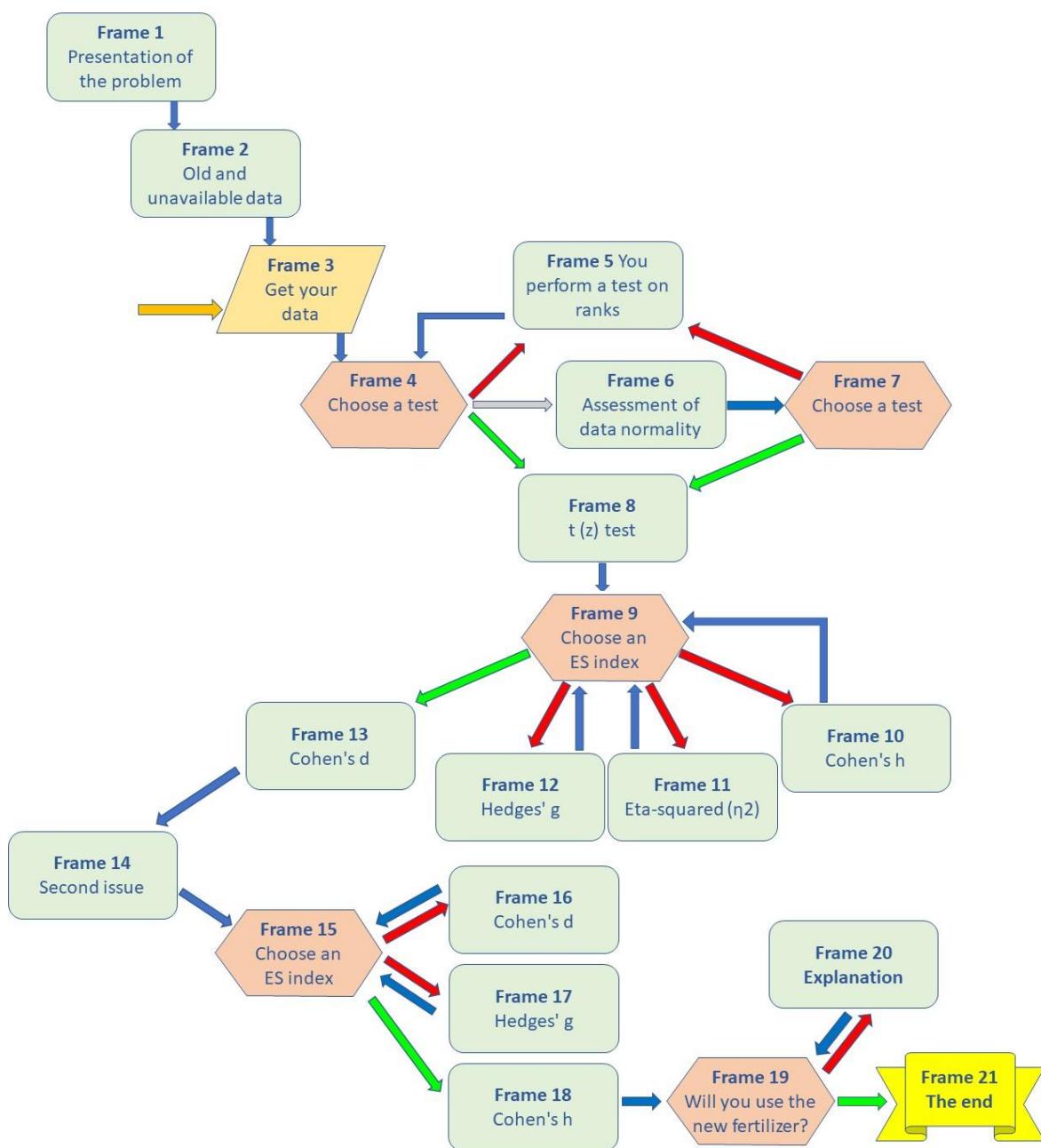
The best way to address this issue would be to try the new fertilizer on a group of plants and to compare the harvest both to a group of plants which did not receive it and to another group of plants receiving the competing one. However, you are not allowed to perform an experiment by using products of the competing company.



The only thing that you can do is to compare the harvest of a group of plants which did not receive your new fertilizer with the harvest of the group receiving your new fertilizer, and then, to compare the results to the results presented in a press release published by the competing company.

Screen 2. Old and unavailable data

By reading the press release of the other company, you understand that their fertilizer increases significantly the harvested mass as compared to plants which were not treated. They indicate that the Cohen's d reached 0.79 and that the number of plants which died before harvest was about 20.1% with the fertilizer, a proportion which was about the same as the percentage without the fertilizer (20.3% ; no statistical difference between the two samples).





Screen 3. Get your own data

You design an experiment where 1000 feet of the plant are planted and treated with your new fertilizer, and 1000 feet are planted without fertilizer. Other experimental conditions are counterbalanced between the two groups. After a given time lapse, you harvest and measure the weight of each plant.

Download the datafile "Fertilizer.txt"

The first column of this file indicates the group ("Control" or "Fertilizer"), the second indicates whether the plant survived until the harvest ("Died" or "Alive"), and the third one indicates the weight of the harvested plants.

Screen 4. Choose a test.

Which statistical procedure will you use to know whether the fertilizer increases the mean weight of the plants?

- Test on ranks (Wilcoxon-Mann-Whitney) ↗Fr5 (↗ symbols indicate hypertext links)
- Assessment of normality (Shapiro test and/or draw a frequency histogram) ↗Fr6
- Student t test ↗Fr8

Screen 5. You perform a Wilcoxon-Mann-Whitney test.

The Wilcoxon-Mann-Whitney test is significant but doesn't give you information about the differences in means ↗Fr4

Screen 6. You perform a Shapiro test and/or draw a frequency histogram.

Screen 7. The test indicates a significant deviation from normality. Will you do

- a Wilcoxon-Mann-Whitney test ↗Fr5
- or a Student t-test? ↗Fr8

Screen 8. Student t-test

Congratulations : You know that the samples are sufficiently large (each sample size >>30). Therefore, whatever the distribution of the data, the Student t-test will be valid.

↗Fr9

Screen 9. You have now to decide which effect size index you will use.

- Cohen's d ↗Fr13
- Cohen's h ↗Fr10
- Hedges' g ↗Fr12
- Eta-squared ↗Fr11

Screen 10. Wrong answer:

Cohen's h can be used only to compare proportions and not means. ↗Fr9



Screen 11. Wrong answer:

Eta-squared may be used if your comparison of means was designed as a one-way ANOVA. Rather, you used a Student t-test. ↗Fr9

Screen 12. Not bad but wrong answer:

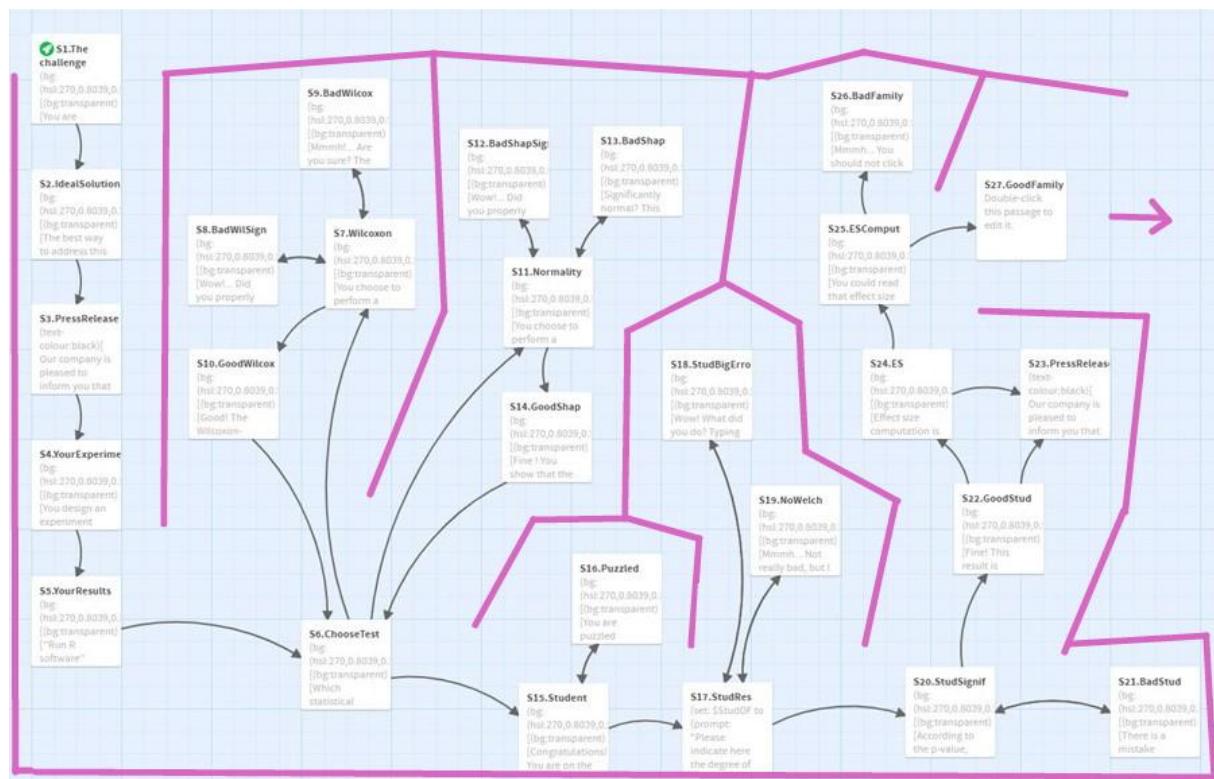
Hedges' g is presently one of the best indices to compare two means. However, in order to compare your results with those of the study of the other company, you need to use the same index. ↗Fr9

Screen 13. Good !

Cohen's d is suited for this kind of comparison. It is not the best index but you need to compare your results to the results of the other company. They used Cohen's d, so, you must use the same index.

Your result is Cohen's d = 0.78 [to be continued...]

Please note that this is a first version of the example presentation. About $\frac{1}{3}$ of the interactive story remains to be written. The actual story has been implemented on the software Twine, in which the story is richer and the sequencing more detailed than in the example presentation (see below).



Future implementations:

- a budget mechanic will be added for each question, as if the teacher would charge for his consulting sessions
- points (representing the trainee's worth towards their boss), that will go up or down depending on the right or wrong choices that had been made.



Play the story with Twine

To play the story on your browser:

1. Download the Twine folder, unzip it, and then run the html file (keep the images files in the same folder). You may have to enable Javascript to be able to run the story in your browser by clicking "enable" on the pop up.
2. The interactive links bring the story further and you will have to make the right choices by clicking on the correct link(s)
3. At one moment in the story, there will be a dialog box. The correct answer is 1322.8 (1566 is a small error). The other answers are wrong.

Final evaluation/remediation

Final evaluation is done via a quiz. This quiz will be presented for the first time "as this".

Items and proposed responses are presented in a random order.

After completion, students receive a first feed-back indicating which items are validated. For the other items, the references of the text where the response can be found is given. All these texts have already been downloaded by students during the SBL sequence.

Then, students can respond a second time to the quiz.

Examples of questions for the quiz (in green, the correct answers)

- 1) The effect size indice[s] that can be used to compare 2 means is [are]:
- Cohen's d
 - Cohen's h
 - Hedges' g
 - R²

Evaluated ILO : choose an appropriate effect size index

- 2) The effect size indice[s] that can be used to compare 2 proportions is [are]:
- Cohen's d
 - Cohen's h
 - Hedges' g
 - R²

Evaluated ILO : choose an appropriate effect size index

- 3) What are the true propositions?
- An effect size index allows comparing the amplitude of the effects between 2 or more experiments
 - Together with statistical tests, an effect size index is a way to compute statistical significance
 - A large effect size index indicates a statistically significant result



- A large effect size index associated with a statistically significant result indicates a strong influence of the factor on the dependent variable
- A high statistical significance indicates a strong influence of the factor on the dependent variable

Evaluated ILO : Interpret effect size indices