

Analysing the effect of the CheckYourSmile (CYS) tool

Dinh Triem Phan, Viet Minh Thong Le. Tutors: Dr. Nadia Yassine-Diab¹, Dr. Sebastien Dejean²

¹Founder of the project CYS, ²Research engineer in scientific calculation at the IMT

Introduction



CheckYourSmile (CYS) is a web platform project for learning specialty vocabulary in foreign languages (e.g., IT English: networks / databases) through a set of "serious" games, founded by Dr. Nadia Yassine-Diab. The

objective is to provide a complement to face-to-face language courses in higher education. Despite the importance of specialty vocabulary in integrating students into their professional careers, only a few hours per week are devoted to this teaching. Note that one of the innovations of CYS is to offer a collaborative system to propose and validate lexical entries: thus, everyone participates in constructing knowledge (cf. crowdsourcing).

Aim

Our objective is to obtain indicators on the plus-value of CYS in a university context and on the combinations of variables needed to obtain the best results and to improve the effects of the tool.

Results of descriptive statistics

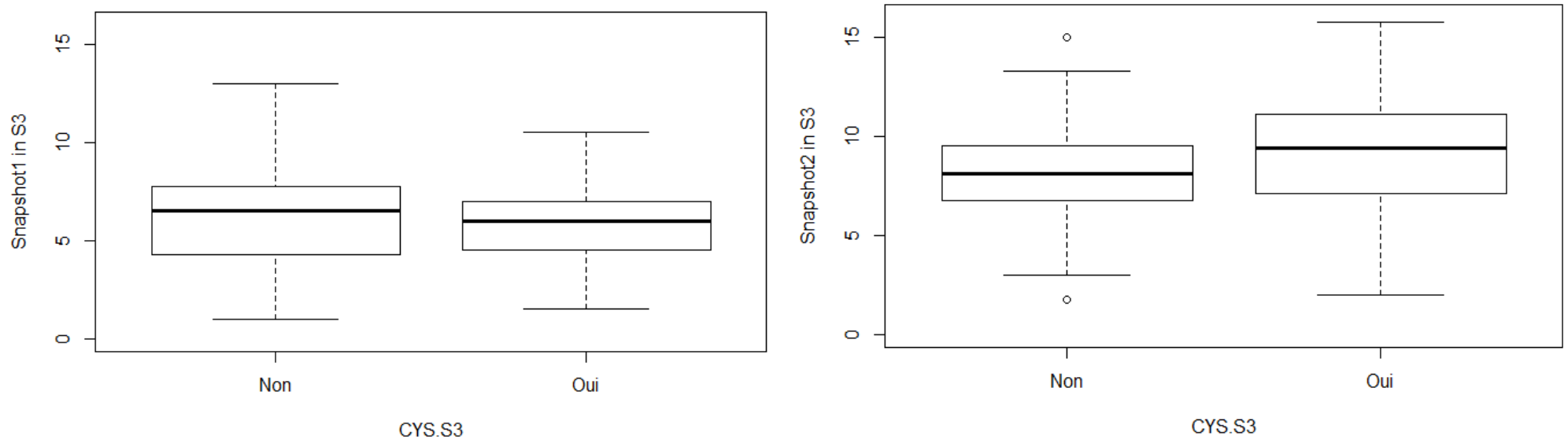


Figure 1: Snapshot1 and Snapshot2 of non-CMI students in semester 3

A non-CMI student has a higher Snapshot 1 score when they used the CYS tool (**Figure 1**). The CYS tool has a positive effect on non-CMI students.

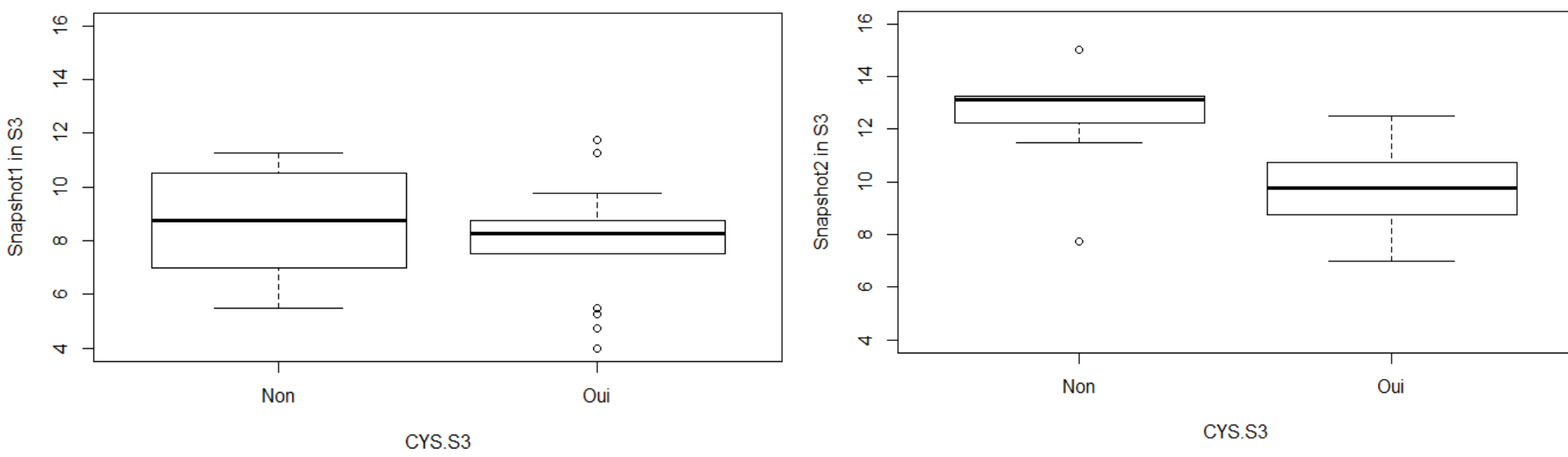


Figure 2: Snapshot1 and Snapshot2 of CMI students in semester 3

A CMI student has a lower Snapshot 1 score when they do not use the CYS tool(**Figure 2**). The CYS tool has a negative effect on CMI students.

Results of linear model

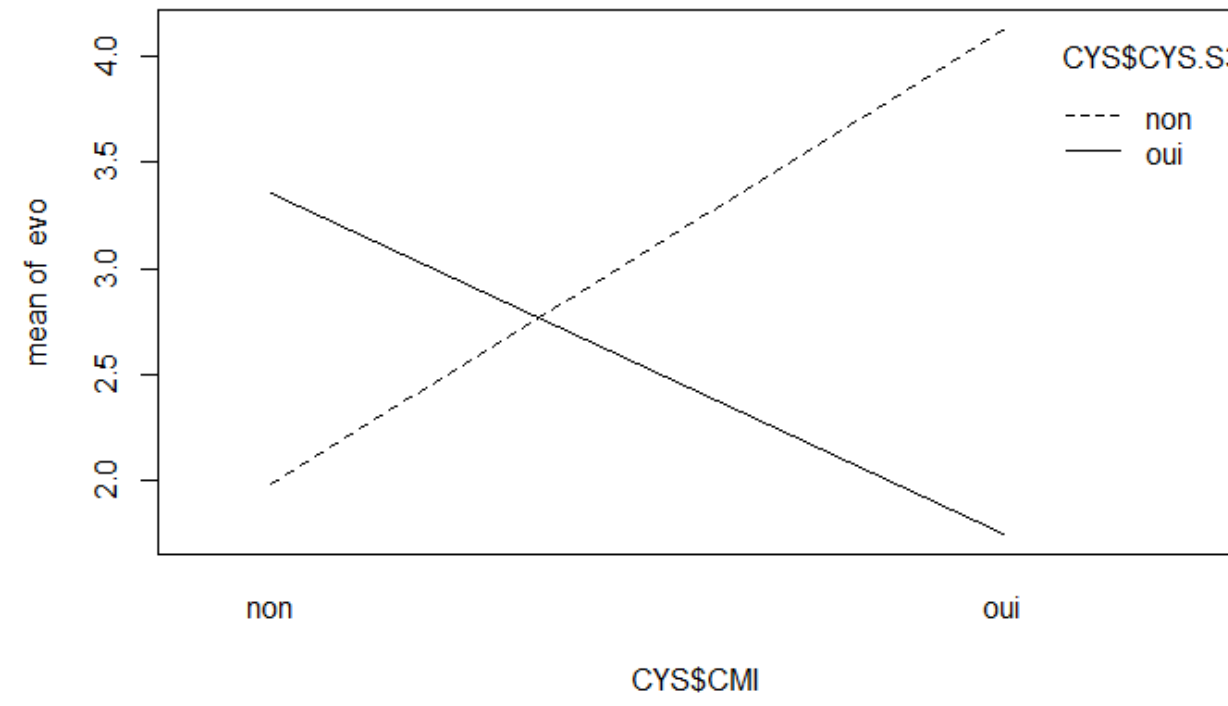


Figure 3: Interaction between CMI and CYS

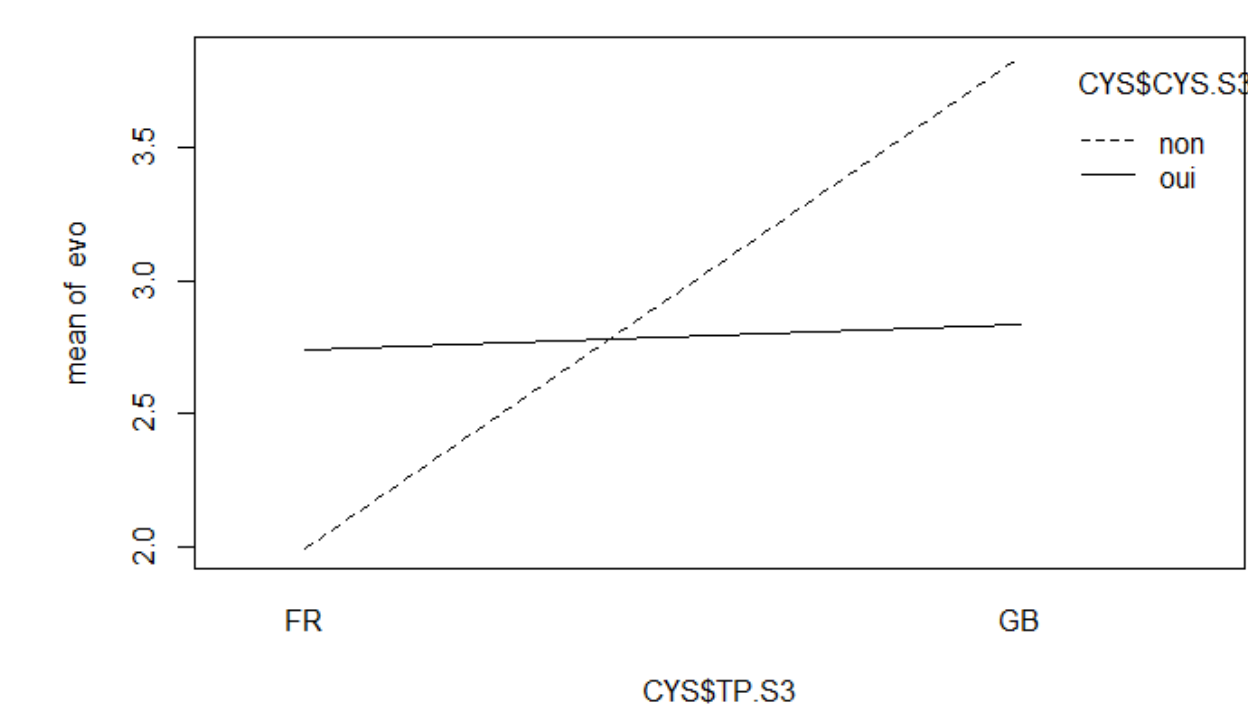


Figure 4: Interaction between CYS and TP

- The interaction between CYS and CMI is strong (**Figure 3**). The CYS tool has a negative effect on CMI students and a positive effect on non-CMI students.
- There is no interaction between TP and CMI because no CMI student uses French for TP.
- There is an interaction between CYS and TP (**Figure 4**). The CYS tool has a positive effect on students using French in TP.

ANCOVA model

$$Snapshot2_{ijkl} = \mu + \alpha Snapshot1_{ijkl} + \beta_i + \gamma_j + \theta_k + \delta_{jk} + \varepsilon_{ijkl}, \forall i = 1, 2, \forall j = 1, 2, \forall k = 1, 2$$

i, j, k : modality indices of TP, CYS and CMI, respectively.

- For CYS and CMI, **1 :Non** and **2 :Oui** (No and Yes, in English).

- For TP, **1: FR** and **2: GB**.

ε_{ijkl} is error in the estimation of the individual with the index *ijkl*. From where:

$$\begin{aligned} \mu &= 4.62 \\ \alpha &= 0.57 \\ \beta_1 &= \gamma_1 = \theta_1 = \delta_{11} = \delta_{12} = \delta_{21} = 0 \\ \beta_2 &= 1.90 \\ \gamma_2 &= 0.44 \\ \delta_{22} &= -3.03 \end{aligned}$$

Three qualitative factors have an impact on the Snapshot2 result. Snaphot1 has a strong effect on the result of Snapshot2. There is an interaction term between the variable CYS and the variable CMI and the latter has a significant negative effect on Snapshot2.

Results of non-linear model

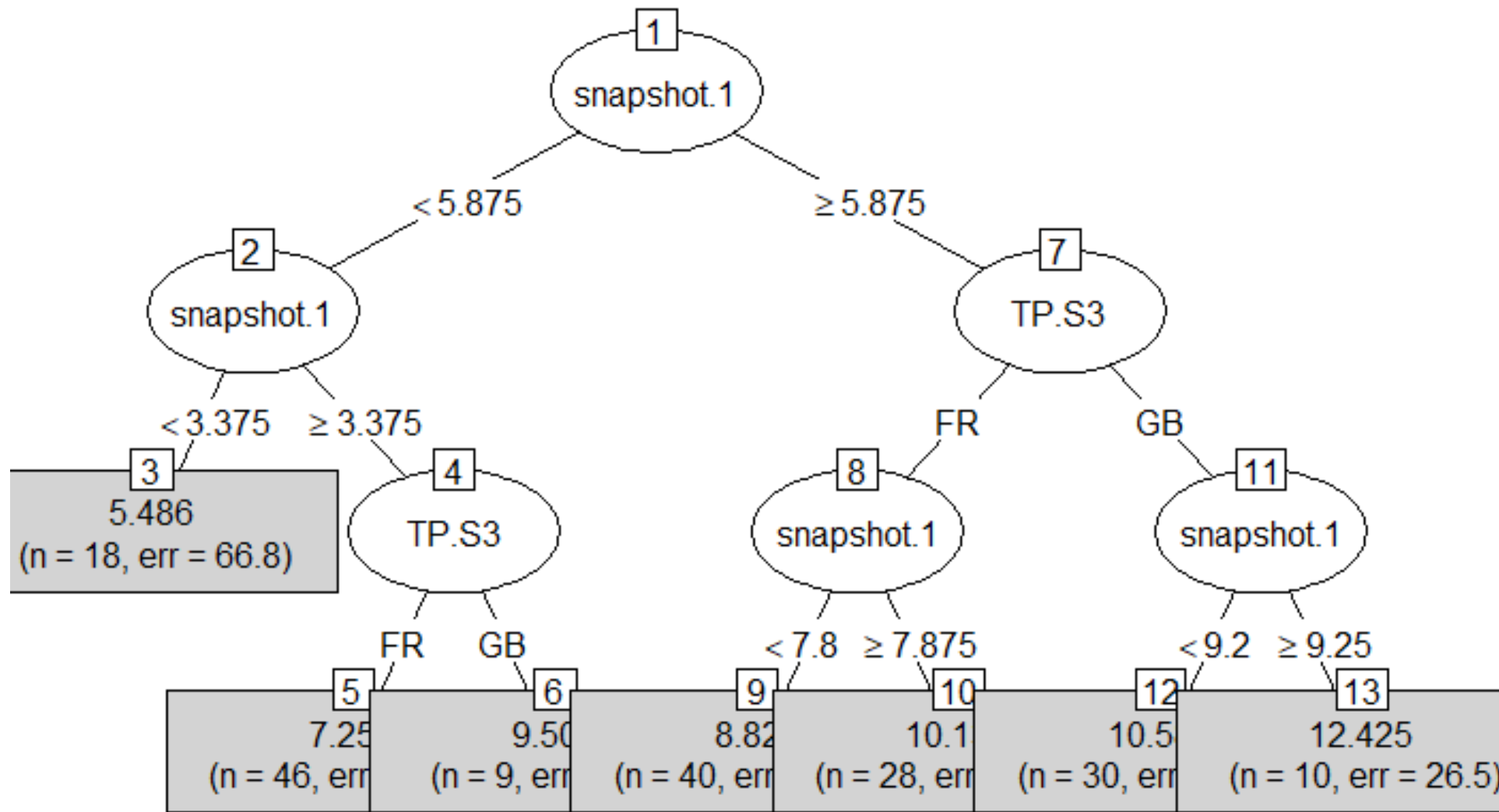


Figure 5: Regression tree of students in semester 3-first case

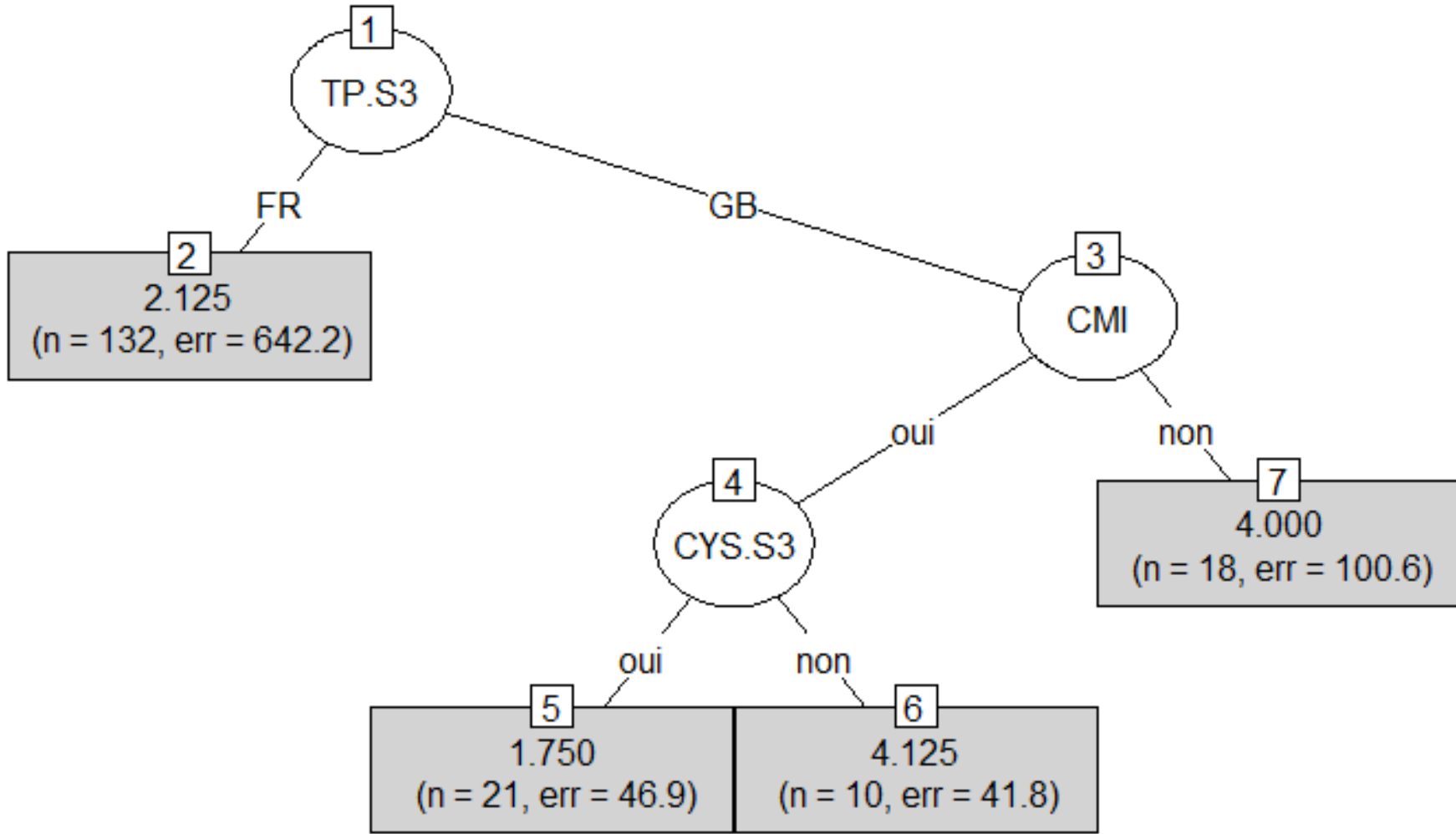


Figure 6: Regression tree of students in semester 3-second case

We did not see the impact of CYS in the first case (**Figure 5**).

Among students having TPs in English, a student using the CYS tool progresses less than another not using CYS. Among students having TPs in English and using the CYS tool, a CMI student progresses less than another non-CMI (**Figure 6**).

We applied the same method on the data in semester 4. We divided the students into two groups: medical and non-medical .

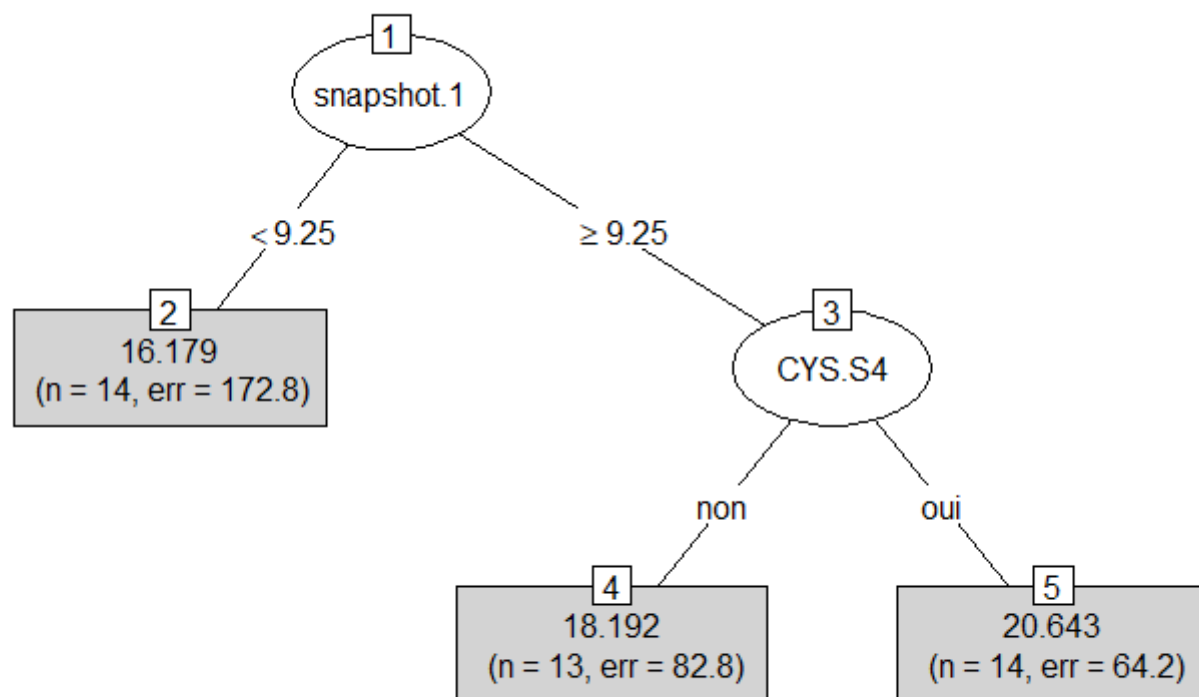


Figure 7: Regression tree of medical students in semester 4

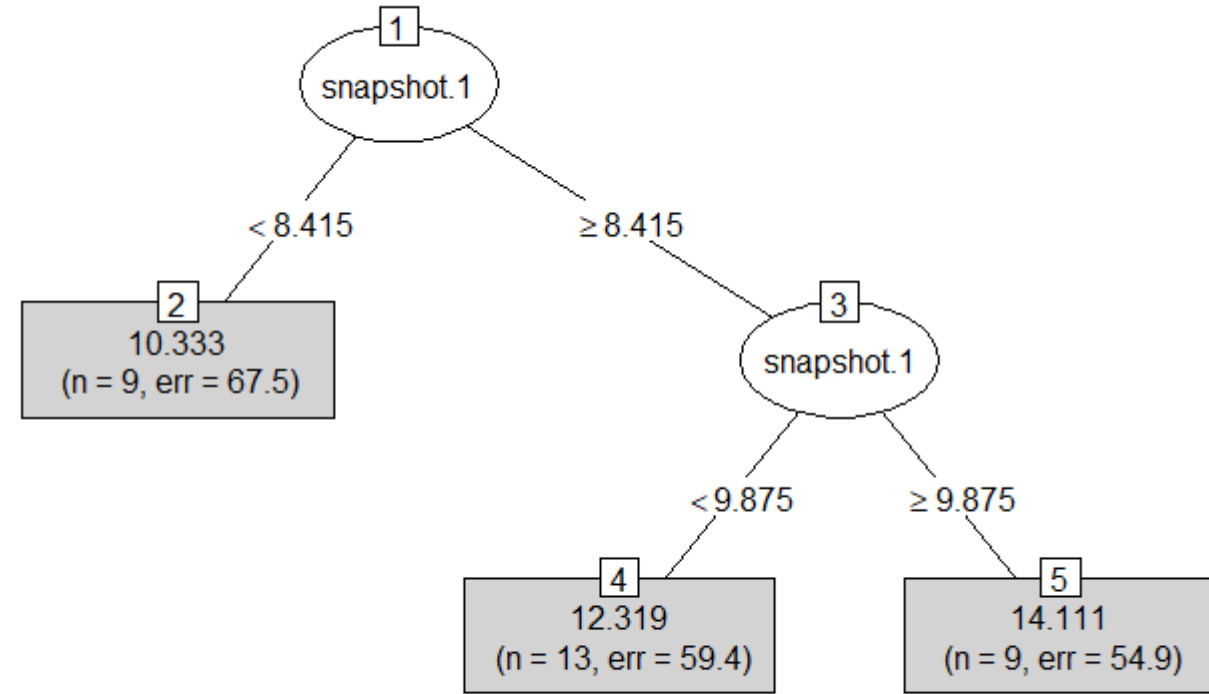


Figure 8: Regression tree of non-medical students in semester 4

In semester 4, modelling the score of snapshot2 from snapshot1 and CYS does not ensure a good result (**Figure 7 and 8**). We cannot conclude on the effect of CYS on student progression.

Discussion

In semester 3, both linear (ANCOVA) and non-linear (decision tree) models showed that CYS only had a positive impact on non-CMI students. The result agrees with those found in the descriptive statistics.

However, the non-linear model showed that the TP variable impacted on the result of students using CYS while the linear model did not.

In summary, modelling the score of Snapshot2 from Snapshot1 and other variables (TP, CYS, CMI) in semester 3 are not enough to claim the effectiveness of CYS. We have to add more variables, which resulted in very small R² values.

References

- [1] R. J. Lewis, An introduction to classification and regression tree (cart) analysis, in Annual meeting of the society for academic emergency medicine in San Francisco, California, vol. 14, 2000.
- [2] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to linear regression analysis, vol. 821, John Wiley & Sons, 2012.

Acknowledgements

We would like to express our sincere gratitude to our supervisors Dr. Nadia Yassine-Diab & Dr.Sebastien Dejean for providing their invaluable guidance, comments and suggestions throughout the course of the project. We would specially thank Dr. Sebastien Dejean for constantly motivating us to work harder and Dr. Nadia Yassine-Diab for getting us the data samples.

Materials & Methods

Materials

We study a database acquired in three academic years (2016-7, 2017-8 and 2018-9). It contains the evaluation results of students (Snapshot1 and Snapshot2) of different courses and the language teaching they received (English or French, engineering courses (CMI), use of CYS) (**Table 1**)

Semestre	Filière	snapshot.1	snapshot.2	CYS.S3	TP.S3	CMI
S3 2017-18: 38	EEA:181	Min. : 1.000	Min. : 1.75	non:120	FR:132	non:150
S3 2018-19:143		1st Qu.: 4.500	1st Qu.: 7.00	oui: 61	GB: 49	oui: 31
		Median : 6.750	Median : 8.75			
		Mean : 6.442	Mean : 8.82			
		3rd Qu.: 8.000	3rd Qu.:10.50			
		Max. :13.000	Max. :15.75			

Table 1: Summary of data in semester 3

Methods

Descriptive statistics determined the most influential factors among the variables. Charts such as boxplot illustrated which variables are more important than others. "interaction.plot" showed how variables interacted mutually.

Linear model [2] reported how multiple variables affected the progression of students simultaneously, including their mutual interactions. Model ANCOVA showed the effect of three qualitative variables and Snapshot1 on Snapshot2.

Non-linear model (decision tree [1]) with the tree graph demonstrated how multiple variables affected the progressions of students. According to cross-validated predictions, we kept the model which possessed the minimal complexity parameters. We studied in two cases the effect of:

- 3 qualitative variables on the progression of students as a term of difference between 2 Snapshots.
- 3 qualitative variables and Snapshot1 on Snapshot2 .