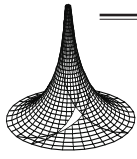


21/2 (2023), 161–183
DOI: 10.5485/TMCS.2023.13291
tmcs@science.unideb.hu
https://ojs.lib.unideb.hu/tmcs

Teaching
Mathematics and
Computer Science

Teaching correlation and regression in three European countries

Balázs Boller

Abstract. In this article, we compare the presence of correlation and regression analysis in secondary education of Ireland, the Netherlands and Luxembourg, through the analysis of final-exam tasks and curricula based on the Anthropological Theory of Didactics (ATD). It points out that the same topic can appear in different ways and extent in curricula, even if the mathematics teaching goals are similar. This article is a kind of introduction to the research that explores the possibilities for the appearance of these concepts in the Hungarian mathematics education. Therefore, in the second part of the article, Hungarian curricular goals are included, and it is shown which methodology of the three studied countries has the greatest curricular basis in Hungary.

Key words and phrases: teaching correlation and regression, comparison of curricula, ATD.

MSC Subject Classification: 97xxx.

Introduction

With all the data generated in the world growing at an astonishing rate, how and what should be taught within statistics in secondary schools have changed significantly in recent decades. It is a real expectation of education stakeholders that students learning statistics should be able to critically analyse data sets and draw conclusions from them. Two important concepts in statistics, correlation and regression, are taught to help to achieve precisely these goals. They can be used to describe useful relationships, however, there are not many countries where they are part of the secondary mathematics curriculum. The teaching

of mathematics has become more practical, and it is common experience that today's students look for usable, practical knowledge. Correlation and regression are just that, as we can determine the degree of relationship (correlation) and make predictions (regression) from a sample. According to the *Encyclopaedia of Mathematics*, data analysis has become a key skill for the future and the present as well (Ben-Zvi, 2020, p. 177). The ability to understand connections is very important in a society where large amounts of data are available and can be processed quickly, easily and with a low chance of error using digital tools. Information thus made available does not always represent reality; this requires a critical attitude (fact-checking, e.g., against graphical manipulations), which can be pointed out already in secondary school. Knowledge of correlation and regression analysis can also help in this respect as a tool for developing covariational thinking (Garfield et al., 2008).

This paper has two objectives. On the one hand, to compare the presence of the topic of correlation and regression in secondary education of three European countries, Ireland, the Netherlands and Luxembourg. On the other hand, to find out which curricular goal system is closest to the Hungarian National Curriculum (HNC). In Hungary, these two concepts are not part of statistics education, and in our broader research we would like to lay the foundations for a possible introduction of the topic. The comparison of the methods and curricula of the three countries provides a good basis for the development of a content-methodological unit that takes into account the traditions and approach of Hungarian mathematics education.

About correlation and regression

Correlation coefficient of Bravais and Pearson describes the existence of a linear relationship between two normally distributed probability variables. It does not take into account whether the variables are causally related, it simply examines whether there is a relationship, and if so, what direction and how strong it is. The degree of linear relationship is indicated by the correlation coefficient, which is a number between -1 and +1.

Regression in its classic form also looks at the relationship between two variables, X and Y. In linear regression analysis, our goal is to find the best fitting regression line for a given sample by means of an empirical function which allows us to estimate for preassigned values (arbitrary values) of the independent variable X the respective dependent variable Y. We can predict the variable Y from X: we do not need to assume that Y is a function of X or X causes Y. For a given

sample, the parameters of the regression line can be calculated by various methods, one of the most common being the method of least squares. For certain samples, higher-order curves can be fitted as well. In these cases, a higher-order regression is performed, such as a second-order fit or exponential regression.

Previous research on the subject

The teaching of correlation and regression has been studied by many. Estepa et al. pointed out the importance of topic structure. They found that it is very easy to develop misconceptions about the representation of tasks in textbooks (Estepa Castro & Sánchez Cobo, 1998). In the words of John Walker, "A picture is worth a thousand words." He has emphasized the importance of simulations in teaching regression analysis (Walker, 2004). He has noted the benefits of simulations in improving understanding of concepts and promoting interactions between students. In another study, Garfield and Ben-Zvi (2008) put the topic in a broader context and explore correlation and regression knowledge in the context of covariational thinking. Their volume of papers aims to promote the renewal of statistics education. They have introduced lessons that build ideas informally, remove calculations, and focus on the concepts and reasoning. Kozak (2008) has described that teaching correlation and regression side by side is problematic because the concepts are easily confused, and he believed that it would be more appropriate to teach them separately. Engel and Sedlmeier (2011) have reviewed common errors and misconceptions about the concepts of correlation and regression, too, and made four recommendations to avoid them. Research summarised in this study suggests that reasoning about correlation and regression is not always easy for laypeople, and even researchers sometimes struggle with a correct understanding. Engel and Sedlmeier have recommended approaches that should be helpful in fostering students' understanding of correlation and regression, such as the use of real data and the integration of modelling experience and technology. Before all this, they encourage teachers and students to dare to make mistakes and to use their experiences for further improvement. Batanero then focused on the development of teachers, aiming to improve their knowledge of correlation and regression by carrying out a statistical project based on real data (Batanero et al., 2014). Their results suggest that prospective teachers also have problems distinguishing between correlation and causation, but that the activities they organised helped to develop prospective teachers' knowledge to teach correlation and regression, and to empower them for their future work in activities such as figuring out what students know. Also, in the context of covariational

thinking, Gea Serrano et al. (2016) have described how the topic is underresearched although it is of great importance. They have also found that some of the prejudices and misconceptions about the subject are resistant to traditional teaching methods and even to experiments based on the use of technology, and the authors therefore recommend better teaching of the subject.

Theoretical background and methods

Artigue and Winsløw (2010) have developed the Anthropological Theory of Didactics (ATD) to provide a framework for comparative analysis in mathematics education. The ATD model makes comparisons more informative and precise, as it takes into account a broader perspective than the object of observation. The theory defines different levels for this (Figure 1), where the attitudes and aims of a higher level influence the attitudes and aims of lower levels. Thus, the main idea is that different methodological considerations and objectives cannot be understood without knowledge of the broader context in which they appear. Therefore, institutional analysis appears as a first and necessary step in the comparative work, necessary for making sense of students' behaviour. Another important step is to identify the levels of comparison (Figure 1). Highlighted background indicates the levels at which the current comparison is made.

General levels of ATD theory	Concrete levels of comparison
9. Civilisation	9. "Western culture"
8. Society	8. Situation and maintenance of schools
7. School	7. Autonomy of teachers
6. Pedagogy	6. General pedagogical principles
5. Discipline	5. The purpose of mathematics
4. Domain	4. The role of statistics
3. Sector	3. Descriptive statistics
2. Theme	2. Correlation and regression
1. Subject	1. Typical tasks

Figure 1. General levels of the ATD model and their representation in the context of the present study

The focus of this research is on the analysis of levels 1 and 2, which is the classroom representation of correlation and regression analysis and its type tasks. To understand their development, it is necessary to examine the objectives and trends set at levels 4 and 5, which involves looking at the curriculum environment.

(These levels presumably also implicitly include the approach of the higher levels.) The comparison in this article is made through the analysis of these levels, which is a two-way comparison. On the one hand, there is a horizontal comparison between the three countries at the relevant levels, on the other hand, there is a vertical comparison within a given country through an examination of the coherence of the levels.

The main methods of data collection were document analysis, interviews and mailings. In three European countries, we examined curricula, course structures, final exams, question banks and workbooks. In addition, we contacted didacticians in the three countries to get a better insight into their educational systems and how they present the concepts of correlation and regression in their schools.

Study of the three countries

In this research, we have examined three European countries, Ireland, the Netherlands and Luxembourg. There are of course many countries outside these where correlation and regression analysis are present, the choice of these three countries is partly arbitrary, but the diversity of educational systems nevertheless provides a good basis for the analysis. Ireland and the Netherlands perform markedly better than Hungary in the PISA measures, while Luxembourg does not differ significantly (Oktatási Hivatal, 2019).

The final exams are used to measure the outcome requirements. This is usually a good indicator of what is emphasised in a subject, so we have selected the example tasks for illustration from the most recent final exams available in the three countries studied. Examination of these shows that the different approaches and teaching goals in mathematics in the three countries provide different opportunities for discussing correlation and regression in secondary school.

Ireland

The secondary schools of Ireland are largely state-funded and follow the same state-prescribed curriculum and take the same state public examinations. The

second level school span is predominantly a six-year cycle, taken by ages 12 to 18. The terms "Junior cycle" and "Senior cycle" are commonly used in Ireland[1].

Correlation and regression calculations are introduced in the second half of secondary school (Senior cycle). The Irish statistics curriculum aims to teach students how to identify patterns, make conjectures and draw conclusions (ATD – level 4). Another essential pillar is applicability, since the curriculum states that "learners apply their knowledge and skills to solve problems in familiar and unfamiliar contexts." (Government of Ireland, Department of Education and Skills, 2015)

This is reflected in the fact that, at a basic level, the topic is only presented in a visual way, with the emphasis on its graphic aspects. Specific outcome requirements at the basic level are that the learner recognises the correlation values -1 and $+1$ (based on a point cloud diagram) and knows how to measure the extent of a linear relationship between two variables. This implies knowledge of a specific procedure. It is also expected to correctly pair a correlation coefficient value with the corresponding point cloud diagram. This also serves to strengthen the graphical approach and to develop a deeper understanding of the concept. The differentiation between correlation and causality is also a basic principle in Irish curricula. At the upper level, the most important change is that the correlation coefficient have to be calculated with a calculator. In addition, the best fitting line for a given sample is drawn by eye and the resulting regression line is used for predictions. The topic is also part of the final-exam exercises, the following task (Figure 2) is taken from the 2022 higher level exam. The task requires a complex and in-depth knowledge of correlation and regression analysis. After plotting the last two points (G, H) in a graph, the student has to draw the regression line, which requires a good understanding of the concept and good estimation skills, since all 8 points have to be taken into account. A great advantage of this exercise is that the students can find out the position of the trend line on their own. In addition, the drawing is of great importance, since the equation of the line drawn by eye has to be further calculated. Part iv) measures a deep understanding of the concept by requiring reasoning in the answer. The last part asks the student to give a clear answer, as a check on the previous one, by calculating the correlation coefficient.

[1]More information about the Irish education system is available on the following website: `https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=IRL&treshold=5&topic=EO`

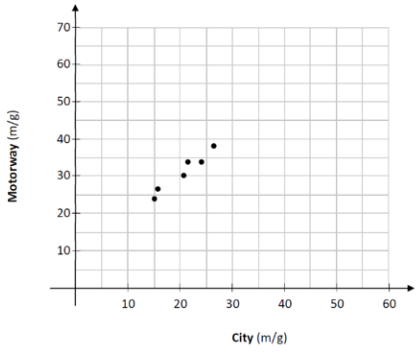
Jena is researching fuel consumption in cars. She finds the following data for the number of miles per gallon (m/g) for eight different cars, labelled A to H, when driving in the city and on the motorway:

Miles per gallon data for city and motorway		
Car	**City** (m/g)	**Motorway** (m/g)
A	22	34
B	27	38
C	24	34
D	16	27
E	15	24
F	21	30
G	**30**	**40**
H	**17**	**30**

The scatterplot below shows this data for cars A to F.

i) *Using the data in the table above, plot and label points to represent cars G and H on the scatterplot below.*

ii) *On the scatterplot, draw the line of best fit for the data, by eye.*



iii) *Two other cars, K and L, have the miles per gallon values given in the following table. Use your line of best fit on the scatterplot to fill in an estimate for each of the two missing values in the table below. Show your work on the scatterplot.*

Car	City (m/g)	Motorway (m/g)
K	20	
L		60

iv) *Based on the data given, would you be more confident in the value you estimated for K or for L? Give a reason for your answer.*

v) *Find the value of r, the correlation coefficient between city and motorway miles per gallon. Use only the values for the 8 cars A to H in the table on the previous page. Give your answer correct to 3 decimal places.*

Figure 2. Task 8 of the 2022 Irish *Higher* Level Mathematics final exam (Leaving Certificate Examination, 2022, p. 20)

Overall, the Irish education system strives for a deep understanding of concepts and applicability, in addition to being conceptually grounded. The concept of correlation is given special emphasis in the curriculum. The theoretical background to the calculation of the regression line is not given, only the drawing of the line by eye is expected from the students.

The Netherlands

There are three types of secondary school in the Netherlands. VMBO schools provide pre-vocational education. HAVO schools are the entry to universities of applied sciences. Graduates from VWO schools usually go to research universities.

Training here lasts one year longer than in HAVO schools, with more in-depth study of the curriculum and a focus on preparation for university-level studies[2].

In 2019, the Netherlands started an education reform involving all stakeholders in education. The new structure is guided and focused by the consultative report "Building tomorrow's primary and secondary education together", which makes several important changes, updates the curriculum, and supports future developments (Curriculum.nu, Leergebied Rekenen & Wiskunde [LRW], 2019). The curriculum emphasises the facilitative role of mathematics with statements such as: all students should be skilled in the functional use of mathematics at their own level, or all students should understand the formal mathematics necessary to achieve the above. Students who can do more formal mathematics should be given the opportunity to develop these skills (level 5). Based on the experts' recommendations, data analysis, statistics and probability will continue to be a priority in the mathematics curriculum (Curriculum.nu, Toelichting Rekenen & Wiskunde [TRW], 2019). As the amount of data and representations increases, it is important that students learn to collect data and draw conclusions (level 4). Students will also learn to estimate the validity of information and representations and how to check their conjectures.

Even though secondary education in the Netherlands is divided into three parts, correlation and regression are to some extent present in all of them. It is a goal for all secondary school students to be able to make a difference between correlation and causation (level 2). For pre-university students, additional aims are set: in both the HAVO and VWO classes, they are expected to use statistical techniques, reason and draw conclusions about reliability and correlation from their results. They are also required to do statistical research with data sets using ICT tools.

Correlation and regression are introduced in the classroom in an application-oriented way, as in the current educational structure, secondary school students must present a detailed chemical, biological or physical measurement in a so-called profile essay. In this essay they have to process and analyse an experiment of their own. The research question is often a comparison of the effectiveness of two different methods or an investigation of the relationship between two quantities. This is where linear regression often comes in (Hemerik, 2003). Here, the teaching guide recommends conducting experiments where, for example, data on

[2]More information about the education system of the Netherlands is available on the following website: `https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=NLD&treshold=5&topic=EO`

the weight and height of the group members are examined, the strength of the relationship between these two quantities is determined, and the fitted line is used to make predictions. Emphasis is also placed on the conditions of the regression calculation, i.e., students are expected to demonstrate the validity of the regression analysis by calculating the correlation coefficient. For this purpose, some workbooks also introduce the concept of covariance (usually in specific classes). For these higher ability groups, the theory of fitting the line is also introduced, most often by using the method of least squares (Math4all – Lineaire regressie, 2021). They do not have to use this method, because both the correlation coefficient and the regression equation are calculated using computer software (e.g., Microsoft Excel), since, as the curriculum says: "The continuous development of ICT tools offers more and more opportunities for practical use" (Curriculum.nu, TRW, 2019). Another benefit of this is that data collection and some statistical tests can be carried out using simulations, making the concept of probability less of a priori knowledge than in the current curriculum. This also illustrates the view in the Netherlands that as a result of ICT, less formal mathematics is needed to do mathematics functionally. Although the topic is presented in-depth in the classroom, in the HAVO and VWO final-exam exercises of the last 5 years, only the already fitted regression line was used for calculations. The reason for this is that a major part of the topic is tested in the so-called school final exams, which are adapted to the curriculum of the class. Thus, in the standardised central final examination, the topic is only presented at a level such as that shown in the third task of the 2022 VWO exam (Figure 3), which states:

> *The points in the figure are approximately on a straight line. This line is also drawn in the figure. The trend line suggests that at some point in the future Sijbrands will win a record 38 games. Calculate from the figure which year this would be.* (The Netherlands "Mathematics A" Final Exam Test – VWO, 2022, p. 4).

The curricula of the Netherlands are therefore dominated by the learning of the regression approach, and this is also the case for the final exam task. Although there is no calculation task on the subject in the final examination of recent years, the tasks look at the conclusions that can be drawn from the regression equation and the understanding of data sets, which is in line with the aims of the curriculum: to understand the concept of regression and apply it in a real-life situation.

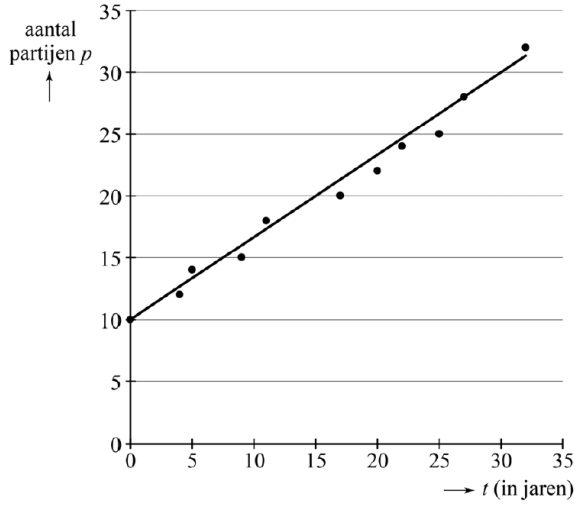


Figure 3. VWO mathematics A 2022 Exercise 3 – The exercise is about the Dutch checkers player Ton Sijbrands. The graph shows how many players Sijbrands defended his world title against in each year from 1982 on (Ton played simultaneous blind checkers, i.e., several matches were played at the same time in parallel, with players keeping track of line-ups by head). (The Netherlands "Mathematics A" Final Exam Test – VWO, 2022)

Luxembourg

In Luxembourg, secondary education is dispensed by more than 50 institutions: public institutions (mainly secondary schools), private institutions applying the ministry's national curriculum, private institutions applying a different curriculum, and European schools. While applying the ministry's official curriculum, each secondary school may, as part of its autonomy, introduce an educational approach, classes or specific measures tailored to its students' needs and expectations. At the end of the third year of general education, students may continue their education either in vocational training or in one of the five streams of general secondary education[3].

[3]More information about the education system of Luxembourg is available on the following website: `https://gpseducation.oecd.org/CountryProfile?plotter=h5&primaryCountry=LUX&treshold=5&topic=EO`

In Luxembourg, content and outcome regulations are not standardised, as different curricula apply to the many different types of classes. However, a common element is that statistics is highlighted as a priority area of knowledge. The aim of the subject is to give students the algebraic, analytical, statistical and probabilistic skills they need to acquire the intellectual tools essential for understanding the economic world and the everyday world (level 4). Correlation and regression are taught in 3 different classes:

1GCG: General education with a focus on accounting, law and economics, correlation and regression in grade 13, in the year of final exam.

1GSO: General education with a social science focus, correlation and regression in grade 13, year of final exam.

2CB: Classical education system with advanced mathematics and advanced language training, correlation and regression in grade 12.

In 1GCG and 1GSO classes, the expected knowledge is basically the same. Students here are considered as users of mathematics, and therefore the practical use of correlation and regression dominates these classes. The graphical introduction to the topic builds heavily on the linear functions that all students learn up to grade 10 at the latest. The formula of the correlation coefficient and the regression line are accepted by the students without a proof, at first, they calculate the formulas using an Excel table, and later the students use a calculator with a graphic screen to work with bivariate statistics, e.g., linear regression.

In the 2CB classes, a deeper understanding of correlation and regression is aimed at. Here, for example, calculus is used to prove that the least squares principle is indeed a correct procedure for fitting a line. Two main additions are the appearance of precision and exponential fitting. At this level, students must use appropriate precision when giving answers or follow the precision required by the task. Also shown is the percentage estimate of the value estimated from the fitted line and the percentage estimate of the error of the true value, which requires a deeper understanding from the students. Another major difference is the need to be able to model exponential relationships in higher mathematics classes. From the dependent variable y in the fitted regression line, $ax + b = y := \ln z$ is transformed into an exponential regression. The resulting equation $z = k \cdot e^{ax}$ (k and a are both real numbers) is used to estimate the value of one variable given the value of the other variable.

The following exercise (Figure 4), taken from the 2021 final exam of the 1GCG general education class, illustrates what is expected of Luxembourg students at the final exams.

Task 10

In this task, all results are rounded to the nearest 0.01. The table below shows the number of visitors to a shopping centre in Luxembourg over the last few years.

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
Rank of the year (x_i)	1	2	3	4	5	6	7	8	9
number of visitors in thousands (y_i)	1840	1890	2170	2430	2650	2980	3550	3990	4370

a) Place the point cloud of coordinates (x_i; y_i) in an orthogonal coordinate system (graphical units: 1 cm per year on the abscissa axis, 2 cm per thousand visitors on the y-axis).
b) Determine the % growth rate of the number of visitors between 2011 and 2019.
c) Check whether the affine correction is valid using the correlation coefficient.
d) Give the equation of the regression line D from y to x.
e) Plot the regression line D.
f) Using the previous fit, estimate the number of visitors to the centre in 2022.
g) Use this fit to estimate the first year in which the number of visitors will exceed 6 000 000.

Figure 4. Luxembourg final exam, mathematics, GCG class (The Luxembourg Government, Ministry of Education, Children and Youth, 2021, p. 4)

A common feature of the 1GCG and 1GSO classes is that there is a general logical arc to solve the problems. To solve a final exam problem, the curriculum expects students to do the following:

- *Represent a bivariate data set related to a task (statistical situation) as a point cloud.* From the graphical representation the students can estimate the correlation coefficient at the beginning of the task and see the trends on the point cloud. This is taken to a more conscious level in part b) of the task, where students have to calculate the percentage increase.
- *The correlation coefficient as a tool to quantify the degree of linear relationship between two variables.* The curriculum describes the need to stick to the limits of use of the correlation coefficient, although we could not find a counterexample. The calculation of the coefficient is a skill expected of students.
- *Determination and representation of the centre of sample.*
- *Using a calculator, find the reduced equation of the regression line and use it to estimate the value of y given x and vice versa.* The calculation requires students to know the method of least squares but does not require its use at a skill level, the parameters being calculated here by skipping the formulae

and using a calculator. It is a kind of counting algorithm. Part (d) clarifies the role of the dependent and independent variables, and the plot of the line is also a good self-checking tool. Further counting with the line makes the task more valuable for the learner, as it leads to a conclusion supported by calculations.

- *Solving statistical problems that can be reduced to linear fitting, problems with other indicated corrections.* There is the possibility to choose a chapter in general education in which students can learn about second-order regression.

The presence of correlation and regression in the Luxembourg curriculum is diverse and depends largely on the specialisation of the class. The curriculum takes into account the needs of the students. In the higher mathematics class, it expects a deeper, theoretical knowledge, while in general, mathematics education is more of a problem-solving algorithm with practical implications, as the final exam task showed. Correlation and regression are present in almost equal weight, although the general task arc shows a much stronger focus on regression, which is what the tasks are based on, with correlation appearing in a subordinate role in most cases.

Differences and similarities in curriculum aims

The comparison has 3 parts. In the first part, we will look at the curricular aims and curricular differences between the three countries. The second part compares the results according to some key aspects, and in the third part, we will make some comments from the Hungarian perspective.

Differences in curriculum aims (level 4-5) and subject knowledge (level 1-2)

In all three countries, adaptability is a key concept. In Ireland, mathematics is a tool for developing skills that are essential for everyday life and work, in the Netherlands, one of the aims of teaching statistics is to help people manage taxes, pay scales, insurance and other social issues. In Luxembourg, the aim of mathematics is to support the professional and everyday life of the class and to provide the tools needed to understand the economic world. This common purpose is likely to mean that a third, higher level is also similar, such as the pedagogical principles (level 6). In addition, there are other similarities between pairs, which make the comparison more detailed. In both Ireland and the Netherlands, for example,

the importance of recognising connections is reflected, with the former focusing on connections outside and within mathematics, and the latter more on the analysis of causality. The Netherlands and Luxembourg are linked by a functional mathematical approach. In these two countries, the emphasis is more on the use of correlation and regression, as the Luxembourg curriculum says, "the student is the user of mathematics". A specific feature of the Irish curriculum is strategic competence, while in the Netherlands the curriculum writers have specifically set the maintenance and promotion of motivation as a goal. The complex set of objectives of the different countries together shape the emerging subject knowledge in this area. A summary of these is shown in Figure 5:

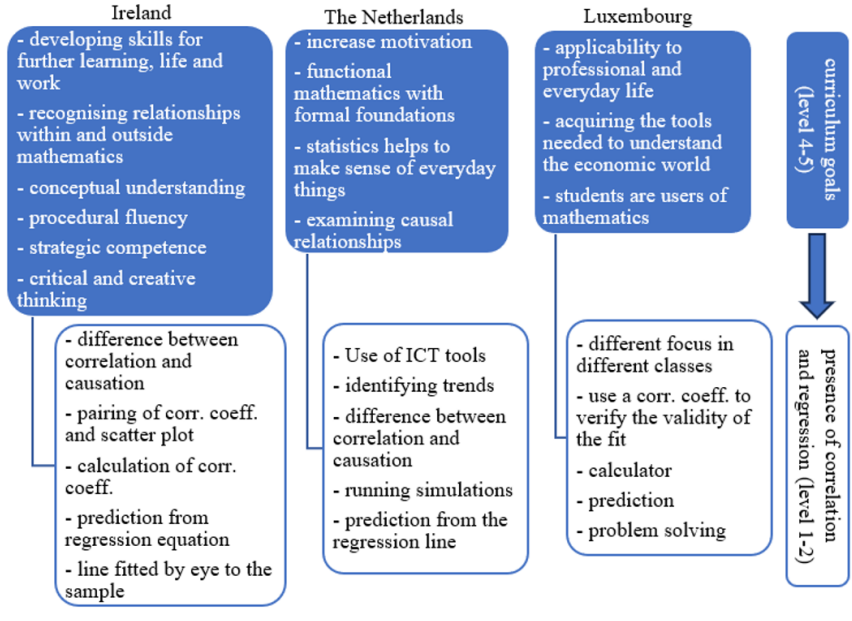


Figure 5. Relationship between levels of ATD theory for the three countries studied. The role of mathematics and related general objectives and the role of statistics correspond to ATD levels 4 and 5, and the learning goals and knowledge items on correlation and regression that these require belong to levels 1 and 2.

Curricular differences therefore also affect the knowledge content (level 2), and the next paragraph will compare these. The differences between basic and higher levels in the different countries can be identified based on the final exam

tasks and curricula analysed. It is important to note that, due to the diversity of education systems in the countries, some countries have different basic levels (e.g., Luxembourg), where the topics are presented in different ways. To avoid this, we have interpreted as a basic level all classes that are not mathematics specialisations. Thus, the following picture emerges (Figure 6).

Basic level	Correlation (concept)	Calculation of corr. coeff.	Regression	
			linear	non-linear
Ireland	yes	-	-	-
The Netherlands (VMBO, HAVO)	yes	-	yes	-
Luxembourg (1GCG, 1GSO)	yes	yes	yes	exp.

Higher level	Correlation (concept)	Calculation of corr. coeff.	Regression	
			linear	non-linear
Ireland	yes	yes	yes	-
The Netherlands (VWO)	yes	yes	yes	-
Luxembourg (2CB)	yes	yes	yes	spec.

Figure 6. Difference of knowledge content in correlation and regression at basic (intermediate) and higher levels in the three studied countries

As the tables above show, the amount of correlation and regression knowledge in general secondary education varies. In the general curriculum classes, knowledge of the concept of correlation can be considered as a minimum level, and it is present everywhere. Ireland and Luxembourg are at the two extremes in this aspect: in Ireland, no more than this is required at basic level, whereas in Luxembourg, nonlinear fitting may already be introduced at basic level (e.g., in the 1GCG economics specialisation). The calculation of the correlation coefficient is not an explicit requirement at basic level. Linear fitting is central in Luxembourg, while in the Netherlands it is more often referred to as a trend line and is included in the final exam exercises.

In the classes with higher mathematics, we see a more consistent picture. In all three countries, the ability to calculate the correlation coefficient is required, linear regression is part of the curriculum, and a common point is that there is a purpose to the fitting, i.e., students draw conclusions from the regression line to out-of-sample values. This can be explained by the common focus on application at higher levels (levels 4-5). The deepest subject knowledge that emerges is nonlinear regression. In Luxembourg, the fitting of the second-order curve can be introduced in the 2CB classes.

Comparison by a few selected aspects

In the second part of the comparison, we compare the presence of correlation and regression across countries according to four criteria. We have chosen these aspects arbitrarily, but they are all relevant to our research.

(1) *The age range of students*
The subject is complex and requires a lot of prior knowledge to understand, so it is basically relevant for the second half of secondary school. This is true for all three countries, with the depth of the subject coming closer to the final examination. This is particularly true in Luxembourg, where the earliest that students are exposed to the concepts is in the year before the final exam. In the Netherlands, on the other hand, preparation begins in the first half of secondary education, with the separation of correlation and causation.

(2) *The role of calculators and digital processing programs*
Complex formulas, such as the correlation coefficient, require a calculator, which is used in all three countries. In Luxembourg, the coefficient is first calculated using a table method, and later students can use the built-in functions of their calculator, as in the other two countries. To find the equation of the regression line, in the Netherlands and Luxembourg, students learn the method of least squares using different programs such as Excel. In the Netherlands, ICT tools are used for simulations and experiments with large samples, as in the case of Walker, who also considers simulations to be important in teaching this subject (Walker, 2004).

(3) *The role of regression line fitted by eye*
A free-fitting line is of great importance, especially for estimation. Drawing such a line requires analysing the sample data and drawing conclusions from it. In this way, it is possible to ensure that the student understands the concept of regression, the location of some sample data, and to draw attention to the bias of outliers. In Ireland, drawing a regression line by eye is only required at higher level, but the workbook exercises suggest that in the Netherlands this type of exercise is also found in general education (but not in the central final examination). In Luxembourg, we could not find any such exercises.

(4) *The role of distinguishing causality and correlation*
One of the key ideas in teaching correlation may be to understand that a correlation coefficient close to 1 does not necessarily mean a causation. This

allows learning goals to be achieved, such as: students being able to estimate the correctness of information and representations, and to draw correct conclusions from small sets of data (Curriculum.nu, LRW, 2019).

The four aspects are summarised in Figure 7:

	Ireland	The Netherlands	Luxembourg
Age of students (years)	15-18	14-15, then 16-18	17-18
Calculator and ICT tools	calculator (corr. coeff.)	ICT-tools, simulations, Excel	formulas, then calculator built-in function
Estimated regression line by eye	only at higher level	also at basic level	is not present
Separation of causality and correlation	also at basic level	from the start of secondary education	not pronounced

Figure 7. The four selected aspects in the examined countries

From the table, we can see that in Ireland there is a fundamental separation of correlation and causation, and much of the curriculum is more expected at the higher level, so we can say that there is more of a focus on deepening basic knowledge in general education. In addition, in the Netherlands there is a greater emphasis on preparation and activity-based learning, which fits in with the motivational objective. The curriculum-based approach, where students are users of mathematics, can also be observed here. The wide range of ICT tools and analysis programmes allows students to actively learn and understand key concepts through their own statistical experiments. The diversified teaching structure of Luxembourg tries to provide students with what they need most in each specialisation. At the end of mathematics education, correlation and regression calculations are introduced, building on the knowledge of several topics. The focus here is not on the separation of correlation and causality, but on the different types of regression (linear, second-order, exponential), correlation being the most important to verify them.

Foreign and Hungarian curriculum goals

At the end of the comparison, we should also mention some Hungarian features, since all the above research is the first step towards how to integrate this topic into Hungarian curriculum. We will show that many of the objectives of the Hungarian National Curriculum are in line with the curricular objectives of the studied countries.

In both Luxembourg and Ireland, the development of critical thinking in mathematics lessons is essential. In the former, this is facilitated by skills such as: reasoning, interpreting data from a proposed model, or examining the relevance of a reasoning and argument. The Hungarian curriculum uses similar language, although it does not go into details: "One of the aims of teaching mathematics is to develop the student's reflective thinking, a system of skills that enables the analysis, synthesis and evaluation of data."

The curriculum of the Netherlands is committed to teaching statistics through simulations and experiments, while the Irish curriculum also uses experimentation as a primary tool for knowledge learning and understanding. We find the HNC states that "the active participation of students in learning activities is a key factor and therefore [...] teachers should always give preference to activity-based forms of learning organisation." Or elsewhere, where it mentions the introduction of new concepts: "New concepts and knowledge should continue to be introduced through demonstration, experience, building on learning activities and relating them to reality!"

In Luxembourg, ICT tools are seen as a way of diversifying learning methods and adapting teaching to the needs and speed of all students. According to the HNC: "An essential element of the 21st century learning environment is the diversity of teaching methods supported by digital technology for school learning." And while Luxembourg students are able to create graphical representations (e.g., diagrams and graphs) from collected data, on paper and digitally, Hungarian students "use dynamic geometric, graphical and table management softwares for experimentation, conjecturing, graphical equation solving and verification." (Hungarian National Curriculum [HNC], 2020)

One of the basic principles of correlation and regression analysis is the efficient data management. In both the Netherlands and Ireland, students are expected to be able to go beyond the collection of data to the selection of the right mathematical methods for processing and the drawing of the correct conclusions. The relevant learning outcomes in the HNC are "collect, organise, represent and interpret data."

Correlation and regression combine a wide range of knowledge. In Ireland, one of the goals of teaching mathematics is for students to see connections within mathematics, just as in HNC, where the output requirement for students is "to recognise the connections between different areas of mathematics."

In statistics in the Netherlands, students learn to estimate the validity of information and representations and how to check their assumptions. This not

only provides important knowledge in the classroom, but also encourages good behaviour in everyday life. The Irish curriculum explicitly says that students should acquire knowledge and skills that will directly benefit them in other areas of their everyday lives. Whereas in Luxembourg, statistical knowledge is seen as an essential tool for understanding the economic and everyday world. This is our aim when we think "Mathematics helps us to understand and deal with everyday problems." (HNC, 2020)

In these three countries, correlation and regression calculations are used as a tool to achieve the goals. Could this topic be included in the Hungarian curriculum? The question is not so simple, of course, but it is easy to imagine another parallel in the curricula. According to the Hungarian NAT, students "understand logical, quantitative, functional, spatial and statistical relationships in their environment". From there, it is only a step to the goal formulated in the Netherlands, where statistics is given a special place in the mathematics curriculum: "The student knows the concept of correlation", or "The student distinguishes between correlation and causation." (Curriculum.nu, LRW, 2019, p. 52)

The examples above show that the curricular objectives of the three studied countries correspond to the Hungarian curricular objectives in many areas, i.e., the curricular foundation of the subject is quite good. The most similarities are with Ireland's system of objectives, so their approach may be the best fit for a possible Hungarian concept.

Conclusion

Learning statistics should provide students with the tools and ideas they need to respond effectively to the information in the world around them. One of the tools can be the two concepts introduced in secondary school: correlation and regression. This idea is new in Hungary, but it is already part of the curriculum in many countries. However, the learning and teaching of statistics may differ significantly in these countries due to cultural, pedagogical and curricular differences, as well as the availability of qualified teachers, resources and technology.

In this paper, we first examined how correlation and regression are present in the curricula and final exams of Ireland, the Netherlands and Luxembourg, based on ATD theory. The focus of the research was on levels 4-5 and 1-2. The comparison revealed similarities in the learning objectives set at curriculum levels 4-5, which were also observed in the common elements of subject knowledge. However, there were also significant differences in the thematic focus at levels 1-2, which

may be explained by differences in mathematical approach. While in Ireland, a deep understanding of the basics is sought, for example, by emphasising the separation of correlation and causation, in the Netherlands the focus is on the application of mathematics (functional mathematics), but with thorough, early preparation and the use of ICT tools, simulations and experiments by students. In Luxembourg, the focus is also on application, with a particular emphasis on the use of regression curves up to a higher-order within the subject. These differences were also reflected in the final examination tasks, which suggests their coherence with the curricular goals.

In our second research question, we looked at which of the three teaching practices in the three countries would be most accessible in Hungary. In answering this research question, the most similar approach to a Hungarian approach to the subject might be that of Ireland, both because of its way of understanding basic concepts, which might be ideal when introducing a new subject, and because of the high degree of similarity between the Irish and Hungarian curricular objectives. The applicability of the topic should not be dismissed either; of course, purpose and meaning must be shown when teaching a new concept. Therefore, it may be necessary to use certain elements of the Netherlands and Luxembourg methodologies when developing a concept that also considers Hungarian specificities.

Introducing a new topic is not easy, there are many other aspects to consider. Two major difficulties can be the emergence of experimentation with digital tools and the knowledge and attitudes of teachers. Although new knowledge must be linked to old knowledge in the learning process, the novelty of these concepts must not exclude the use of innovative tools and methods. The role of the computer has changed today, especially when working with big data, and it can be a very useful tool for learning a subject. On the other hand, researches have shown that in many cases the knowledge of teachers in the field of statistics and pedagogy is weak/lacking, and many teachers do not feel prepared to teach statistics (Ben-Zvi, 2020). These questions remain to be answered at a later stage of the research but can be resolved in line with the triple approach of content – pedagogy – technology and the professional development of teachers (Engel & Sedlmeier, 2011). The introduction of new concepts should therefore not be rejected, especially as the HNC also calls for the occasional innovative rethinking of the curriculum, when it states that "It is necessary to introduce and define new concepts when well-chosen problems are being discussed." And finding the actual problem leading to correlation and regression is not difficult.

Overall, it can be said that the topic of correlation and regression would be a useful topic that would fit well with the aims of the Hungarian National Curriculum. The examples of the three countries presented here support the validity of this idea and their examples can be inspiring and reinforcing. It may also be worthwhile to make ourselves aware that statistics should be an essential part of education, since the multiplicity of data has become a completely ordinary phenomenon in science, society, everyday life and almost every profession.

Acknowledgments

The author would like to express his special thanks to his supervisors Csaba Csapodi and Ödön Vancsó for their advice, and to the three didactics (Jos Tolboom – Netherlands, Claude Hinger – Luxembourg, Rachel Linney – Ireland).

References

Artigue, M., & Winsløw, C. (2010). International comparative studies on mathematics education: A viewpoint from the anthropological theory of didactics. *Recherche En Didactique des Mathématiques*, *30*(1), 47–82.

Batanero, C., Gea Serrano, M. M., Díaz, C., & Cañadas, G. R. (2014). Building high school pre-service teachers' knowledge to teach correlation and regression. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, 2014)*. International Statistical Institute. https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_1A3_BATANERO.pdf?1405041555

Ben-Zvi, D. (2020). Data handling and statistics teaching and learning. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (2nd ed.) (pp. 177–181). Springer.

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2009). The transitivity misconception of Pearson's correlation coefficient. *Statistics Education Research Journal*, *8*(2), 33–55.

Curriculum.nu (October, 2019). Leergebied Rekenen & Wiskunde. https://curriculum.nu/download/rw/Voorstellen-ontwikkelteam-Rekenen-en-Wiskunde.pdf

Curriculum.nu (October, 2019). Toelichting Rekenen & Wiskunde. https://www.eerstekamer.nl/nonav/overig/20191010/toelichting_rekenen_wiskunde/document

Engel, J., & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. In C. Batanero, G. Burrill, C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education.* New ICMI Study Series, vol. 14. (pp. 247–258). Springer.

Estepa Castro, A., & Sánchez Cobo, F. T. (1998). Correlation and regression in secondary school text books. In : Pereira-Mendoza, L. Seu-Kea, T. Wee Ke,& W. Wong (Eds.), Proceedings of the Fifth Conference on Teaching Statistics (ICOTS 5 1998) (pp. 672 677). International Statistical Institute. https://iase-web.org/documents/papers/icots5/Topic6d.pdf?1402524957

Garfield, J., & Ben-Zvi, D. (Eds.) (2008). *Developing students' statistical reasoning.* Springer. doi:10.1007/s11858-009-0176-6

Garfield, J. B., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., & Zieffler, A. (2008). Learning to reason about covariation. In J. B. Garfield & D. Ben-Zvi (Eds.). *Developing students' statistical reasoning* (pp. 289–308). Springer.

Gea Serrano, M. M., Batanero, C., López Martin, M. M., & Arteaga, P. (2016). Research on the perception and learning of correlation and regression. *BEIO, Boletín de Estadística e Investigación Operativa*, *32*(3), 234–256. https://www.researchgate.net/publication/312419434_Research_on_the_perception_and_learning_of_correlation_and_regression

Government of Ireland, Department of Education and Science. (2004). A brief description of the Irish education system. https://assets.gov.ie/24755/dd437da6d2084a49b0ddb316523aa5d2.pdf

Government of Ireland, Department of Education and Skills. (2015). Mathematics Syllabus – Foundation, Ordinary and Higher Level (2015). https://curriculumonline.ie/getmedia/f6f2e822-2b0c-461e-bcd4-dfcde6decc0c/%20SCSEC25_Maths_syllabus_examination-2015_English.pdf

Hemerik, L. (2003). Lineaire regressie: het toetsen van samenhang tussen twee variabelen. (Lesbrief; No. 7). Wageningen University, VWO-campus. http://www.vwo-campus.net/lesbrief/7

Hungarian National Curriculum (2020). *Magyar Közlöny*, *17*. https://magyarkozlony.hu/dokumentumok/3288b6548a740b9c8daf918a399a0bed1985db0f/letoltes

Kozak, M. (2008). Correlation and regression: Similar or different concepts? *Statistics in Transition*, *9*(1), 159–162.

Leaving Certificate Examination 2022 – Mathematics, Higher Level (2022). https://www.examinations.ie/archive/exampapers/2022/LC003ALP200EV.pdf

Math4all – Lineaire regressie (2021). https://www.math4all.nl/overzichten/vwo-a/20

Oktatási Hivatal. (2019). PISA2018 Summary Report. https://www.oktatas.hu/pub_bin/dload/kozoktatas/nemzetkozi_meresek/pisa/PISA2018_v6.pdf

The Luxembourg Government, Ministry of Education, Children and Youth. (2020). The Luxembourg Education System (2020). https://men.public.lu/dam-assets/catalogue-publications/divers/informations-generales/the-luxembourg-education-system-en.pdf

The Luxembourg Government, Ministry of Education, Children and Youth. (2021). Examen de fin Détudes secondaires générales. https://portal.education.lu/Portals/0/Documents/examens/ESG/2021/GCG/MATHE_GCG_\'ecrit_juin.pdf?ver=uzETgjH7bQhw4XGh3vyXow%3d%3d

The Netherlands "Mathematics A" Final Exam Test – VWO (2022). https://www.examenblad.nl/examendocument/2022/cse-1/wiskunde-a-vwo/opgaven/2022/vwo/f=/VW-1024-a-22-1-o.pdf

Walker, J. H. (2004). Teaching regression with simulation. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), *Proceedings of the 2004 Winter Simulation Conference, USA* (pp. 2096–2102). IEEE.

BALÁZS BOLLER
EÖTVÖS LORÁND UNIVERSITY
BUDAPEST, HUNGARY

E-mail: bollerbalazs@gmail.com

(Received September, 2023)