

Applied statistics for business and economics **– a one/two-semester programme for after-school course**

ASSUMPTIONS OF THE COURSE

Main assumption

- WE LEARN FROM EACH OTHER

Detailed assumptions

- WORK COLLABORATIVELY
- IMPLEMENT THEORY IN PRACTICE
- ANALYSE WHAT YOU HAVE DONE

1. Course description

The programme which I present should encourage students to learn from each other and to show the value added from the group competition.

As long as the course starts with the theoretical background supported by practical exercises it is not only for the students very good in mathematics. However, the student work based on the financial statements of the listed companies in Warsaw Stock Exchange should be underlined as the strong advantage of the course.

The presented aspects are not a part of the official programme for entrepreneurship course. This is a proposal for a whole semester after-school programme which contains elements of mathematics, statistics, finance and financial analysis.

The final version of the programme depends on the number of students interested in the participation. The idea of the programme is that there should be at least 6 students in the programme in order to have a chance to create at least 3 groups of 2 persons each.

Depending on the possible implementation of the course it can be modified and applied more from the theoretical perspective or can concentrate more on practical aspects depending to which extent students are familiar with statistics and financial analysis.

The course is divided into several steps. Implementation of each step give the students the appropriate theoretical background and practical experience.

Proposed below steps should be understood as a basic format which can be modified and adjusted depending on the circumstances.

2. Course duration

One or two semesters, depends also how deeply the programme is conducted.

The presented below steps can be applied separately or some parts might be skipped if the students are already familiar with the theoretical aspects.

3. Types of work

At school work – approximately 35 meetings each of 90 minutes – not all meetings has to be conducted with the active participation of the teacher (it is much more about the appropriate introduction to the topics and then to leave the financial analysis for the students)

After school work – preparation of the final poster presentation including the listed companies ranking and the detailed presentation of the selected company – the best one according to the applied taxonomy methods

Note: Students should work collaboratively as much as possible at school, not necessary with the assistance of the teacher. However, they should take the advantage from the team work – this is one of the important assumption of the course.

4. Students task

Organisational – students are working under the time schedule prepared by the teacher

Analytical – students should be ready to make the whole analysis on their own with jus basic supervision of the teacher

5. The final course results

The idea of this programme is to encourage those who are not studying economics to start it as this is a

The strong point of this course is that the results are achieved step by step. This makes the course interesting as the theory is strongly supported by the practical aspects. The final result

Each of the group makes a poster presentation for the listed companies whose company name starts with the selected letter from alphabet.

Brief introduction

The inspiration for those scenarios is based on the International Baccalaurean Programme.

Theoretical background of statistics is a part of the programme for mathematics, however the issues are not that much developed from the practical perspective.

The presented programme concerns the whole semester tasks for the students.

The classes are based on the group tasks - .

The idea of this programme is not only to review the appropriate statistical terms with their explanations but the course is going to have three steps.

6. Programme implementation

Step 1 – theoretical aspects

- statistical terms – descriptive statistics – measures of location and dispersion [theoretical aspects.doc](#)
- taxonomy methods [taksonomia-przykład-polski.doc](#);
- Hellwig method [hellwig method.ppt](#); http://pl.wikipedia.org/wiki/Metoda_Hellwiga
- Financial analysis <http://www.bized.co.uk/compfact/ratios/index.htm> (note: this website uses the format which is different from the Polish financial statements)
-

Step 2 – practical aspects

- applied statistics for the selected aspects in the step 1 – exercises [statistics-exercises.doc](#)
- presentation of the taxonomy methods – the use of existing data base in Excel programme, creation of the synthetic measure [G:\15.05\TAKSON ref.xls](#)
- presentation of Hellwig method – the use of existing data base in excel programme [G:\15.05\taks1.xls](#)
- presentation of the financial indicators – formulas and exercises <http://www.bized.co.uk/compfact/ratios/index.htm> (note: this website uses the

format which is different from the Polish financial statements) [financial analysis.doc](#)

Step 3 – collection of the data

- visit website www.gpwinfostrefa.pl to prepare the list of the companies, select 20 companies (key for selection: random selection, letter of the alphabet selection, sector selection – depends on the group or teacher)
- search for the information of the listed joint-stock companies using the website of the companies
- analyze financial statements and put the financial data in the Excel format [financial statements.xls](#), you may as well prepare your own format if you wish

Step 4 – calculations – financial analysis

- calculate selected financial ratios
- give explanations to the achieved results

Step 5 – implement taxonomy methods + Hellwig method

- implement the theory in order to select the appropriate set of financial ratios
- build synthetic measure

Step 6 – prepare ranking of selected companies

Step 7 – analyze deeply at least one of the selected companies

Step 8 – prepare poster presentation

Note: *programme can be implemented both in English as well as in Polish*

Metody taksonomiczne przeznaczone są do klasyfikacji obiektów opisywanych przez wiele cech.¹ Podstawowymi pojęciami taksonomii są pojęcia obiektu i cechy diagnostycznej.

Obiektami są jednostki badania podlegające klasyfikacji. Obiektami badania mogą być np. jednostki administracyjne (kraje, województwa, gminy itp.), jednostki produkcyjne lub usługowe (gałęzie gospodarki, przedsiębiorstwa), wyroby, osoby itp. Specyficznym obiektem badania jest jednostka czasu. Przedmiotem klasyfikacji nie mogą być pojedyncze obiekty, ale ich zbiory $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$.² **Cechami** są właściwości jednostek badanego zbioru rozpatrywane z punktu widzenia zjawiska będącego kryterium klasyfikacji obiektów.³ Każdy obiekt badania ω_i jest opisany przez zestaw cech charakteryzujących ten obiekt (X_j). Badanie taksonomiczne prowadzone jest nie bezpośrednio na obiektach czy cechach, lecz na realizacjach cech diagnostycznych w obiektach.

Prezentowane badanie obejmuje próbę 15 państw członkowskich UE oraz Polskę (łącznie 16 obiektów). Analiza ma charakter statyczny i przeprowadzana została dla roku 2000.

Badanie taksonomiczne przeprowadzane jest zwykle w następujących etapach:⁴

1. wstępna analiza badanego systemu,
2. dobór cech diagnostycznych i skali ich pomiaru,
3. zebranie danych statystycznych,
4. ocena podobieństwa klasyfikowanych jednostek,
5. wybór metody klasyfikacji,
6. klasyfikacja obiektów według wybranej metody,
7. weryfikacja wyników klasyfikacji,
8. interpretacja i wykorzystanie otrzymanych wyników.

Ze względu na stopień skorelowania ich realizacji z poziomem badanego zjawiska, cechy diagnostyczne można podzielić na stymulanty, destymulanty oraz nominanty.⁵ **Stymulanty** to cechy diagnostyczne, których wyższe wartości świadczą o wyższym poziomie "rozwoju" obiektu badania, zatem wartości tych cech są pozytywnie skorelowane z poziomem badanego zjawiska. **Destymulanty** są cechami diagnostycznymi charakteryzującymi się ujemną korelacją z poziomem rozwoju obiektu badania, czyli niższe wartości ich realizacji świadczą o wyższym poziomie rozwoju obiektu badania. **Nominanty** natomiast to cechy diagnostyczne charakteryzujące się tym, że ich wartości powinny zawierać się w pewnym z góry ustalonym przedziale liczbowym lub być równe pewnej z góry ustalonej wartości.

¹ Hellwig Z. [1995]

² Kukuła K. [1994]

³ Nowak E. [1990], s. 13

⁴ Michalski T. [2000], s. 46-50

⁵ Michalski T. [1998]

Dobór cech diagnostycznych jest niezwykle ważnym, a jednocześnie najbardziej kontrowersyjnym etapem badania taksonomicznego. Od właściwego doboru cech zależą bowiem w znacznym stopniu ostateczne wyniki badania – trafność ocen i analiz, dokładność przewidywań, a co za tym idzie również trafność podejmowanych na ich podstawie decyzji. Zbiór cech diagnostycznych powinien zawierać elementy, które jak najpełniej charakteryzowałyby najważniejsze aspekty badanego zjawiska, a więc cechy o największej diagnostyczności. Jednocześnie jest to bardzo subiektywny etap badania, jako że badacz sam decyduje, jaka będzie lista cech diagnostycznych, co zostawia wiele miejsca na dyskusje. Może się wręcz zdarzyć, że badanie tego samego zjawiska przeprowadzone na nieco innym zbiorze cech diagnostycznych da odmienne rezultaty. Stanowi to poważny mankament metod taksonomicznych, dlatego też do otrzymanych przy ich pomocy wyników należy podchodzić z dużą ostrożnością.

W niniejszym badaniu zbiór cech diagnostycznych zawierała następujące wskaźniki:

- 1) udział użytków rolnych w powierzchni kraju (w %),
- 2) liczba gospodarstw rolnych na 100ha UR (w szt.),
- 3) średnia powierzchnia gospodarstw (w ha),
- 4) udział zatrudnionych w rolnictwie w zatrudnionych ogółem (w %),
- 5) wartość produkcji rolniczej na 100ha UR (w mln EURO),
- 6) udział produkcji roślinnej w produkcji rolniczej ogółem (w %),
- 7) udział nakładów w wartości produkcji rolnictwa (w %),
- 8) udział rolnictwa w tworzeniu GDP (w %),
- 9) pomoc państwa na 100ha UR (w mln EURO)
- 10) udział eksportu żywności i produkcji rolnej w eksporcie ogółem (w %),
- 11) udział importu żywności i produkcji rolnej w imporcie ogółem (w %),
- 12) udział wydatków na żywność, napoje i wyroby tytoniowe w wydatkach ogółem gospodarstw domowych (w %),
- 13) roczny trend cen żywności (w %),
- 14) stosunek indeksu cen produkcji sprzedanej rolnictwa do indeksu cen konsumpcyjnych (rok poprzedni=100),
- 15) indeks produktywności majątku trwałego⁶ (rok poprzedni=100)
- 16) indeks realnych dochodów w rolnictwie⁷ (rok poprzedni=100),
- 17) indeks cen produkcji sprzedanej rolnictwa (rok poprzedni=100),
- 18) indeks cen środków produkcji (rok poprzedni=100),
- 19) gęstość zaludnienia (w tys. osób na km²)
- 20) GDP/mieszkańca (w jednostkach PPS),

⁶ Z uwagi na brak analogicznej zmiennej w polskiej statystyce publikowanej, przyjęto indeks stopnia zużycia majątku trwałego

⁷ Z podobnych przyczyn przyjęto indeks dochodów do dyspozycji (w ujęciu nominalnym)

21) stopa bezrobocia w gospodarce (w %).

Większość z wymienionych cech to stymulanty (za destymulanty przyjęto jedynie cechy nr 2, 4, 8, 11, 12, 18, 21).

Cechy diagnostyczne stosowane w badaniu taksonomicznym wyrażone są zwykle w różnych jednostkach, co powoduje, że wyznaczanie miar podobieństwa i odległości obiektów na podstawie bezwzględnych wartości cech ilościowych jest zazwyczaj niemożliwe. W celu doprowadzenia do porównywalności poszczególnych cech diagnostycznych, należy uwolnić je od miana. W tym celu stosowane są różne metody, z których najpopularniejsze to standaryzacja, normalizacja i unitaryzacja.⁸

Standaryzacja cech jest najczęściej stosowanym sposobem uwalniania cech od miana. Przeprowadzana jest według następującego wzoru :

$$z_{ik} = \frac{x_{ik} - \bar{x}_k}{S_k} \quad (1.2)$$

gdzie: z_{ik} – zestandaryzowana wartość cechy X_k w obiekcie O_i ; $i = 1, 2, \dots, N$;

$$k = 1, 2, \dots, K; \quad \bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_{ik} \quad ; \quad S_k = \frac{1}{N} \sqrt{\sum_{i=1}^N (x_{ik} - \bar{x}_k)^2}$$

Normalizacja cech jest przekształceniem ilorazowym, które przeprowadzane jest według ogólnej formuły:

$$z_{ik} = \frac{x_{ik}}{x_{0k}} \quad (k= 1,2, \dots , K, i = 1,2, \dots , N) \quad (1.3)$$

gdzie x_{0k} oznacza tzw. podstawę normalizacji cechy diagnostycznej X_k , przy czym podstawa normalizacji może być ustalana w różny sposób.

Unitaryzacja uwalnia cechy diagnostyczne od miana w ten sposób, że przekształcone cechy mają stały obszar zmienności. Przeprowadza się ją według formuły:

$$z_{ik} = \frac{x_{ik} - \min_i \{x_{ik}\}}{\max_i \{x_{ik}\} - \min_i \{x_{ik}\}} \quad (1.4)$$

W przypadku tego badania, standaryzacja okazała się nie najlepszym rozwiązaniem z uwagi na bardzo wysoki poziom zmienności niektórych z cech, co wiązało się ze znacznym zniekształcaniem wartości średniej (wyniki dla tego podejścia zawiera załącznik 4). Znacznie lepszym sposobem doprowadzania cech do porównywalności okazała się w tym przypadku normalizacja, za podstawę normalizacji przyjmując $\sum_{i=1}^n x_{ik}$.

Ważnym etapem badania taksonomicznego jest także ustalenie obiektu, w stosunku do którego dokonywane będą porównania. Obiekt ten, określany mianem *wzorca*, jest to obiekt

⁸ Nowak E. [1990], s. 38-39, 89

charakteryzujący się optymalnymi z uwagi na badane zjawisko realizacjami cech diagnostycznych,⁹ obrany w następujący sposób:

$$x_{*j} = \max_i \{x_{ij}\} \quad \text{dla cech będących stymulantami} \quad (1.5)$$

$$x_{*j} = \min_i \{x_{ij}\} \quad \text{dla cech będących destymulantami.} \quad (1.6)$$

Podstawową kwestią dla wszelkiego typu badań porównawczych jest problem wypracowania systemu miar i wskaźników umożliwiających porównywanie i klasyfikację rozważanych obiektów badania. Ogół miar stosowanych w badaniach porównawczych można podzielić na miary odległości oraz miary będące wskaźnikami podobieństwa. Do pierwszej grupy zaliczamy miary różnicowania poziomów cech diagnostycznych i miary dystansu czasowego między obiektami, zaś do drugiej między innymi miary określające stopień podobieństwa struktur cech diagnostycznych opisujących rozważane obiekty badania. *Miary odległości* między obiektami charakteryzują się tym, że większa wartość takiej miary dla dwóch badanych obiektów oznacza większe różnicowanie wartości cech diagnostycznych w tych obiektach. *Wskaźniki podobieństwa* są miarami o odmiennej charakterystyce, tzn. większa wartość wskaźnika podobieństwa dla porównywanych obiektów oznacza większe podobieństwo tych obiektów ze względu na uwzględnione przy opisie tych obiektów cechy diagnostyczne.¹⁰

W badaniach taksonomicznych najczęściej korzystamy z miar (wskaźników) podobieństwa, które są różnie definiowanymi miarami umożliwiającymi pomiar stopnia podobieństwa struktur cech diagnostycznych opisujących rozważane obiekty badanej zbiorowości. Jedną z takich miar jest miara μ_{i*} mierząca podobieństwo struktur cech obiektów i -tego oraz $*$ -tego (wzorcowego), definiowana następującym wzorem:¹¹

$$\mu_{i*} = \frac{z_i \circ z_*}{|z_i| |z_*|} \quad (1.7)$$

gdzie: z_i - i -ty wiersz macierzy Z (i -ty obiekt),

z_* - $*$ -ty wiersz macierzy Z ($*$ -ty obiekt),

$z_i \circ z_*$ - iloczyn skalarny wektorów z_i, z_* ,

$|z_i|$ - długość wektora z_i ,

$|z_*|$ - długość wektora z_* ,

$i = 1, 2, \dots, 15, p$, gdzie $i=p$ dla Polski.

⁹ Michalski T. [2000], s. 93-94

¹⁰ Michalski T. [2000], s. 74-75

¹¹ por. np. Michalski T. [1998], Starzyńska W. *Gospodarka żywnościowa w okresie transformacji. Problemy modelowania i sterowania*, Uniwersytet Łódzki, Fundacja Rozwoju Przedsiębiorczości, Łódź 1995, s.92

Z wzoru (1.7) wynika, że podobieństwo struktur cech diagnostycznych w dwóch badanych obiektach (i -tym oraz $*$ -tym) mierzone jest miarą, która jest cosinusem kąta zawartego między wektorami z_i oraz z_* obrazującymi rozważane obiekty, a tym samym przyjmuje wartości z przedziału $[-1;1]$.

Ograniczanie się w procesie porównywania obiektów tylko do miar podobieństwa, może doprowadzić do wniosków, które przynajmniej w części mogą mijać się z faktyczną sytuacją. Możemy bowiem otrzymać bardzo wysoką wartość miary podobieństwa, a jednocześnie wiemy, że rozważane obiekty różnią się znacznie pod względem poziomu cech diagnostycznych. Oznacza to, że mamy do czynienia z wysokim podobieństwem struktur, przy istotnym zróżnicowaniu poziomu badanego zjawiska w tych obiektach. Aby porównanie rozważanych obiektów badania było bardziej czytelne i kompleksowe, należy obok wartości miar podobieństwa określających podobieństwo struktur cech diagnostycznych w rozważanych obiektach, podawać równoległe wartości miar, które umożliwiają pomiar odległości (dystansu) między dwoma badanymi obiektami.

Miary odległości, jako miary poziomu, konstruowane są w ten sposób, aby charakteryzowały się następującymi własnościami:

- odległość między obiektami jest wielkością nieujemną,
- odległość obiektu od siebie jest równa zero,
- odległość obiektu i -tego od j -tego jest taka sama jak odległość obiektu j -tego od i -tego,
- większa wartość takiej miary odpowiada obiektowi bardziej rozwiniętemu ze względu na rozpatrywaną kategorię, czy badane zjawisko.

Większość propozycji dotyczących konstrukcji odpowiednich miar zróżnicowania, które głównie są miarami dystansu między obiektami, bazuje na podstawowych pojęciach i metodach wielowymiarowej przestrzeni metrycznej. Najbardziej czytelną i najczęściej stosowaną procedurą pomiaru zróżnicowania poziomu cech w rozważanych obiektach jest metoda, w której zróżnicowanie poziomu utożsamiamy z dystansem między tymi obiektami. Dystans między i -tym oraz $*$ -tym obiektem wyznaczamy wykorzystując **metrykę euklidesową** postaci:¹²

$$d_{i*}^{(1)} = \sqrt{\sum_{j=1}^k (z_{ij} - z_{*j})^2} \quad (1.8)$$

lub tzw. **metrykę miejską**:¹³

$$d_{i*}^{(2)} = \sum_{j=1}^k |z_{ij} - z_{*j}|. \quad (1.9)$$

Zarówno metryka euklidesowa jak i metryka miejska należą do szerszej klasy metryk, tzw. **metryk Minkowskiego** postaci:¹⁴

¹² Nowak E. [1990], s. 39

¹³ ibidem, s. 39

¹⁴ ibidem, s. 40

$$d_{i*}^{(3)} = \left[\sum_{j=1}^k |z_{ij} - z_{*j}|^q \right]^{\frac{1}{q}} \quad (1.10)$$

Prezentowane miary odległości charakteryzują się tym, że rosnąca ich wartość wskazuje na zwiększanie dystansu do wzorca i na odwrót – im bliższa 0 wartość miary dystansu, tym badany obiekt jest bardziej podobny do wzorca.

W celu oceny polskiego rolnictwa, wyznaczono miary podobieństwa i dystansu według formuł odpowiednio (5) i (6). Otrzymane mierniki, uszeregowane zostały według ich optymalnych wielkości, co prezentuje tablica 1.10.

Z zaprezentowanego rankingu wynika, iż Polska najbardziej odbiega zarówno strukturą jak i poziomem rozwoju swego rolnictwa od średniej dla UE. W czołówce rankingu znalazły się Belgia, Holandia i Dania, a więc kraje o największej intensywności produkcji. Bardzo wyraźnie odbiega od nich Polska, która w przypadku obu mierników plasuje się na ostatniej pozycji. Oznacza to, że zarówno pod względem struktury rolnictwa, jak i poziomu jego rozwoju Polska znacznie różni się od aktualnych członków UE. Dystans ten jest poważnym problemem, którego skutki nasilą się z pewnością w pierwszych latach członkostwa w UE. Wskazują na to chociażby doświadczenia krajów, które są obecnie członkami Unii (Irlandia czy Hiszpania).¹⁵ Dlatego tak niezwykle ważna wydaje się pomoc ze strony UE jeszcze przed akcesją.

Tablica 1.10. Ranking państw UE i Polski pod względem sytuacji w rolnictwie

Lp.	Miara podobieństwa		Miara dystansu	
	Kraj	Wartość miernika	Kraj	Wartość miernika
1	Belgia	0,8628	Holandia	0,2572
2	Holandia	0,8562	Belgia	0,2828
3	Dania	0,8317	Dania	0,3027
4	Luksemburg	0,8269	Luksemburg	0,3127
5	Niemcy	0,8054	Niemcy	0,3238
6	Irlandia	0,7750	Irlandia	0,3330
7	Francja	0,7626	Francja	0,3359
8	Wielka Brytania	0,7565	Wielka Brytania	0,3374
9	Austria	0,7528	Włochy	0,3540
10	Włochy	0,7009	Austria	0,3566
11	Szwecja	0,7006	Finlandia	0,3649

¹⁵ por. np.: (1) Niemczyk J. [1997]; (2) Liberska B. *Proces integracji Austrii z Unią Europejską*, Gospodarka Narodowa, Numer Specjalny/1998; (3) Kozak-Lisiecka T. *Skutki gospodarcze włączenia Finlandii w strukturę Unii Europejskiej*, Gospodarka Narodowa, Numer Specjalny/1998; (4) Otoliński E. *Procesy i efekty koncentracji produkcji w rolnictwie Danii*, Roczniki Nauk Rolniczych 1999, Seria G, T.88, Z.1; (5) *Portugal Economic Barometer, November 2000*, Economist 28.10-3.11.2000; (6) *Spain. A country of many faces*, Economist 25.11-1.12.2000; (7) Teresa Radzimińska *Nowe oblicze Portugalii*, Biuletyn Analiz UKIE nr 4, sierpień 2000

12	Finlandia	0,6988	Hiszpania	0,3659
13	Hiszpania	0,6669	Szwecja	0,3716
14	Portugalia	0,6375	Portugalia	0,3791
15	Grecja	0,5577	Grecja	0,4320
16	Polska	0,4621	Polska	0,5623

Źródło: własne obliczenia

Podsumowując, mierniki taksonomiczne zastępują opis badanych obiektów za pomocą wielu cech opisem za pomocą jednej agregatywnej wielkości. Dzięki temu oparta na nich klasyfikacja wielocechowych obiektów sprowadza się do podziału zbioru obiektów według jednej zmiennej, co znacznie ułatwia analizę.

Jedną z najpopularniejszych metod klasyfikacji obiektów jest metoda, w której korzysta się z dwóch parametrów taksonomicznego miernika rozwoju: średniej arytmetycznej \bar{z} oraz odchylenia standardowego S_z . Zbiór badanych obiektów dzielony jest na cztery grupy, obejmujące obiekty o wartościach miernika rozwoju z następujących przedziałów:¹⁶

$$\text{grupa 1: } z_i \geq \bar{z} + S_z, \quad (1.11)$$

$$\text{grupa 2: } \bar{z} + S_z > z_i \geq \bar{z}, \quad (1.12)$$

$$\text{grupa 3: } \bar{z} > z_i \geq \bar{z} - S_z, \quad (1.13)$$

$$\text{grupa 4: } z_i < \bar{z} - S_z. \quad (1.14)$$

Dla miar dystansu, które szeregujemy rosnąco – w przeciwieństwie do miar podobieństwa – stosuje się odwróconą kolejność grup, tzn. grupę 1 tworzą obiekty spełniające warunek (1.14), a ostatnią – (1.11).

Metody klasyfikacji oparte na miernikach taksonomicznych posiadają tę cenną zaletę, że pozwalają one na natychmiastową identyfikację poszczególnych grup typologicznych ze względu na poziom w rozwoju badanego zjawiska, osiągnięty przez obiekty wchodzące w skład tych grup. Obiekty należące do danej grupy są bowiem uporządkowane według wartości miernika taksonomicznego. Możliwe jest także natychmiastowe porównanie różnych grup typologicznych pod względem poziomu rozwoju obiektów wchodzących w skład tych grup.

Ponieważ najlepsze wyniki otrzymano przy zastosowaniu metody bezwzorcowej (z normalizacją cech), budując grupy klasyfikacyjne skupimy się na wynikach dla tego właśnie przypadku.

Tablica 1.11. Klasyfikacja krajów UE i Polski pod względem sytuacji w rolnictwie

Klasa	Kraj	
	Struktura rolnictwa	Poziom rozwój rolnictwa
Grupa 1	Belgia, Holandia	Belgia, Holandia
Grupa 2	Dania, Luksemburg, Niemcy, Irlandia, Francja, Wielka Brytania, Austria	Dania, Luksemburg, Niemcy, Irlandia, Francja, Wielka Brytania, Włochy

¹⁶ ibidem, s. 93

Grupa 3	Włochy, Szwecja, Finlandia, Hiszpania, Portugalia	Austria, Szwecja, Finlandia, Hiszpania, Portugalia
Grupa 4	Grecja, Polska	Grecja, Polska

Źródło: własne obliczenia na podstawie tablic 1.9 i 1.10

W najlepszej, pierwszej grupie znalazły się – zarówno pod względem struktury rolnictwa, jak i jego poziomu rozwoju – dwa kraje: Belgia i Holandia. Najslabiej spośród państw Unii wypadła Grecja, która pomimo tego, że już od wielu lat otrzymuje wsparcie z UE, ciągle boryka się z problemami i tak naprawdę niewiele odbiega od Polski, która do tej pory na takie wsparcie nie mogła liczyć.

Materiały do opracowania tej części zostały udostępnione przez dr Justynę Wiktorowicz i pochodzą z opracowań pracy doktorskiej.

Theoretical aspects

FINANCIAL ANALYSIS

Financial ratios are based on summary data presented in financial statements. This summary is based on the accounting method and accounting standards used by the organization.

Examples of the possible for use ratios are named below in the table.

- Gross Profit Margin
- Net Profit Margin
- ROCE
- Gearing Ratio
- Asset Turnover
- Stock Turnover
- Debtor Days
- Current Ratio

Exercise 1 - Ratio Analysis

The aim of this Activity is to look at some financial data from some businesses and to use ratio analysis to analyse their performance and to assess the value of using ratios in drawing conclusions about business performance.

One of the important aspects of ratio analysis is to know what ratio to use and what the resulting figure tells you. Some ratios can be used on their own; others need to be compared with previous years or other companies to make any sense. The most important thing to remember is that a ratio is not used to demonstrate your mathematical prowess (or lack of it as the case may be) but to provide you with some information. All the ratios used are no more difficult than dividing one number by another or doing a simple bit of multiplication. But make sure that you know what the number means!

Step 1:

Go to the www.gpwinfostrefa.pl and choose one company on the list

Step 2:

The database has a list of businesses to choose from and the financial accounts for those businesses. Use these accounts to calculate the following ratios for two businesses for at least a two-year period.

- Gross Profit Margin
- Net Profit Margin
- ROCE
- Gearing Ratio
- Asset Turnover
- Stock Turnover
- Debtor Days
- Current Ratio
- Acid Test

Ratio	Company1, Year 1	Company 2, Year 2
Gross Profit Margin		
Net Profit Margin		
ROCE		
Gearing Ratio		
Asset Turnover		
Stock Turnover		
Debtor Days		
Current Ratio		
Acid Test		

Step 3:

Once you have your completed table, comment on the performance of the business over the two years highlighting the possible reasons for the changes that you identify.

Step 4:

Draw some comparisons between the performance of the businesses you have chosen and offer some explanations for the differences that you might identify.

<p>Skonsolidowany rachunek zysków i strat</p> <ol style="list-style-type: none"> 1. Przychody ze sprzedaży usług 2. Przychody ze sprzedaży towarów 3. Przychody ze sprzedaży 4. Amortyzacja 5. Zużycie materiałów i energii 6. Usługi obce 7. Podatki i opłaty 8. Wynagrodzenia i inne świadczenia 9. Pozostałe koszty rodzajowe 10. Wartość sprzedanych towarów 11. Zysk brutto ze sprzedaży 12. Pozostałe przychody operacyjne 13. Pozostałe koszty operacyjne 14. Zysk z działalności operacyjnej 15. Przychody finansowe 16. Koszty finansowe 17. Zysk brutto 18. Podatek dochodowy 19. Zysk netto za okres obrotowy 	<p>Consolidated income statement</p> <ol style="list-style-type: none"> 1. Sales of services 2. Sales of goods 3. Revenue 4. Depreciation 5. Materials and energy 6. External services 7. Taxes and charges 8. Payroll and social security 9. Other costs 10. Cost of sold goods for resale and materials 11. Gross profit on sales 12. Other operating income 13. Other operating expenses 14. Operating profit 15. Finance income 16. Finance costs 17. Profit before tax 18. Income tax expense 19. Net profit for the year
<p>Skonsolidowany bilans</p> <ol style="list-style-type: none"> 1. AKTYWA 2. Aktywa trwale 3. Rzeczowe aktywa trwale 4. Wartości niematerialne 5. Wartość firmy 6. Pozostałe aktywa finansowe (długoterminowe) 7. Aktywa z tytułu podatku odroczonego 8. Aktywa obrotowe 9. Zapasy 10. Należności z tytułu dostaw i usług oraz pozostałe należności 11. Należności z tytułu podatku dochodowego 12. Inwestycje krótkoterminowe 13. Krótkoterminowe rozliczenia międzyokresowe 14. Środki pieniężne i ich ekwiwalenty 15. SUMA AKTYWÓW 16. PASYWA 17. Kapitał własny ogółem 18. Kapitał własny (przypadający akcjonariuszom jednostki dominującej) 	<p>Consolidated balance sheet</p> <ol style="list-style-type: none"> 1. ASSETS 2. Non-current assets 3. Property, plant and equipment 4. Intangible assets 5. Goodwill 6. Other financial assets 7. Deferred tax asset 8. Current assets 9. Inventories 10. Trade and other receivables 11. Income tax receivables 12. Short-term investments 13. Prepayments 14. Cash and cash equivalents 15. TOTAL ASSETS 16. EQUITY AND LIABILITIES 17. Total equity 18. Equity attributable to equity holders of the parent

<p>19. Kapitał podstawowy</p> <p>20. Nadwyżka ze sprzedaży akcji powyżej ich wartości nominalnej</p> <p>21. Inne elementy kapitału własnego</p> <p>22. Zyski zatrzymane</p> <p>23. Różnice kursowe z przeliczenia jednostek zagranicznych</p> <p>24. Kapitały akcjonariuszy mniejszościowych</p> <p>25. Zobowiązania</p> <p>26. Zobowiązania długoterminowe</p> <p>27. Oprocentowane kredyty bankowe i pożyczki</p> <p>28. Inne zobowiązania długoterminowe</p> <p>29. Rezerwy</p> <p>30. Rezerwa z tytułu odroczonego podatku dochodowego</p> <p>31. Zobowiązania krótkoterminowe</p> <p>32. Bieżąca część oprocentowanych kredytów bankowych i pożyczek</p> <p>33. Zobowiązania z tytułu emisji obligacji</p> <p>34. Rezerwy</p> <p>35. Zobowiązania z tytułu dostaw i usług oraz pozostałe zobowiązania</p> <p>36. Zobowiązania z tyt. podatku dochodowego</p> <p>37. Bierne rozliczenia międzyokresowe kosztów</p> <p>38. Przychody przyszłych okresów</p> <p>39. SUMA PASYWÓW</p>	<p>19. Issued capital</p> <p>20. Share Premium</p> <p>21. Other capital elements</p> <p>22. Retained earnings</p> <p>23. Foreign currency translation</p> <p>24. Minority interest</p> <p>25. Total liabilities</p> <p>26. Non-current liabilities</p> <p>27. Interest-bearing loans and borrowings</p> <p>28. Other liabilities</p> <p>29. Provisions</p> <p>30. Deferred income tax liabilities</p> <p>31. Current liabilities</p> <p>32. Interest-bearing loans and borrowings</p> <p>33. Liabilities due to bonds issue</p> <p>34. Provisions</p> <p>35. Trade and other payables</p> <p>36. Income tax payable</p> <p>37. Accruals</p> <p>38. Deferred income</p> <p>39. TOTAL EQUITY AND LIABILITIES</p>

Choosing independent variables

- The main idea of variables selection is to reduce the number of independent variables.
- The goal is to identify the independent variables that are decently correlated with dependent variable and possibly not correlated among themselves.
- Otherwise the independent variables might be of linear relationship which may seriously damage the model.

Choosing independent variables

Two popular methods of choosing independent variables are:

- Hellwig's method
- Graphs analysis method.

Hellwig's method

Three steps:

1. *Number of combinations:* $2^m - 1$
2. *Individual capacity of every independent variable in the combination:*

$$h_{kj} = \frac{r_{0j}^2}{\sum_{i \in I_k} |r_{ij}|}$$

3. *Integral capacity of information for every combination:*

$$H_k = \sum h_{kj}$$

Hellwig's method

1. *Number of combinations*

In Hellwig's method the number of combinations is provided by the formula $2^m - 1$ where m is the number of independent variables.

Hellwig's method

2. Individual capacity of each independent variable in the combination is given by the formula:

$$h_{kj} = \frac{r_{0j}^2}{\sum_{i \in I_k} |r_{ij}|}$$

where:

h_{kj} – individual capacity of information for j -th variable in k -th combination

Hellwig's method

2. Individual capacity of each independent variable in the combination is given by the formula:

$$h_{kj} = \frac{r_{0j}^2}{\sum_{i \in I_k} |r_{ij}|}$$

where:

r_{0j} – correlation coefficient between j -th variable (independent) and dependent variable

Hellwig's method

2. Individual capacity of each independent variable in the combination is given by the formula:

$$h_{kj} = \frac{r_{0j}^2}{\sum_{i \in I_k} |r_{ij}|}$$

where:

r_{ij} – correlation coefficient between i -th and j -th variable
(both independent)

Hellwig's method

2. Individual capacity of each independent variable in the combination is given by the formula:

$$h_{kj} = \frac{r_{0j}^2}{\sum_{i \in I_k} |r_{ij}|}$$

where:

I_k – the set of numbers of variables in k -th combination

Hellwig's method

3. Integral capacity of information for every combination

The next step is to calculate H_k – integral capacity of information for each combination as the sum of individual capacities of information within each combination:

$$H_k = \sum h_{kj}$$

Hellwig's method

- *Q: HOW TO CHOOSE INDEPENDENT VARIABLES?*
- **A: LOOK AT INTEGRAL CAPACITIES OF INFORMATION. THE GREATEST H_k MEANS THAT VARIABLES FROM THIS COMBINATION SHOULD BE INCLUDED IN THE MODEL.**

Example

- Let's choose independent variables, using Hellwig's method.

Y	X1	X2	X3
12	1	5	8
14	1	6	7
17	3	6	6
20	2	8	5
25	4	6	6
30	4	9	5
36	6	9	5

Example

- First we need to have vector and matrix of correlation coefficients.
 - Correlation coefficients between every independent variable X_1 , X_2 and X_3 and dependent variable Y are provided in vector R_0 .

	0,9467	$r_{yx1} (r_{01})$
R_0	0,8374	$r_{yx2} (r_{02})$
	-0,7895	$r_{yx3} (r_{03})$

Example

- First we need to have vector and matrix of correlation coefficients.
- ⑩ □ Correlation matrix R includes correlation coefficients between independent variables.

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

	$r_{x1x1} (r_{11})$	$r_{x1x2} (r_{12})$	$r_{x1x3} (r_{13})$
R	$r_{x2x1} (r_{21})$	$r_{x2x2} (r_{22})$	$r_{x2x3} (r_{23})$
	$r_{x3x1} (r_{31})$	$r_{x3x2} (r_{32})$	$r_{x3x3} (r_{33})$

Example

1. *Number of combinations*

We have 3 independent variables X_1 , X_2 and X_3 . Thus we may have $2^m - 1 = 2^3 - 1 = 8 - 1 = 7$ combinations of independent variables.

$\{X_1\}$ $\{X_2\}$ $\{X_3\}$ $\{X_1, X_2\}$ $\{X_1, X_3\}$ $\{X_2, X_3\}$
 $\{X_1, X_2, X_3\}$

Example

2. Individual capacity of independent variable in the combination 1

Combination 1

$$\{X1\} \quad h_{11} = \frac{0,9467^2}{1} = 0,8962 = H_1$$

	0,9467
R ₀	0,8374
	-0,7895

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

2. Individual capacity of independent variable in the combination 2

Combination 2

$$\{X_2\} \quad h_{22} = \frac{0,8374^2}{1} = 0,7013 = H_2$$

	0,9467
R ₀	0,8374
	-0,7895

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

2. Individual capacity of independent variable in the combination 3

Combination 3

$$\{X3\} \quad h_{33} = \frac{(-0,7895)^2}{1} = 0,6234 = H_3$$

	0,9467
R ₀	0,8374
	-0,7895

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

2. Individual capacity of every independent variable in the combination 4

Combination 4

$$\{X_1, X_2\} \quad h_{41} = \frac{0,9467^2}{1+0,671} = 0,5364; \quad h_{42} = \frac{0,8374^2}{1+0,671} = 0,4197$$

	0,9467	
R ₀	0,8374	
	-0,7895	

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

2. Individual capacity of independent variables in the combination 5

Combination 5

$$\{X1, X3\} \quad h_{s1} = \frac{0,9467^2}{1 + |-0,712|} = 0,5236; \quad h_{s3} = \frac{(-0,7895)^2}{1 + |-0,712|} = 0,3642$$

	0,9467
R ₀	0,8374
	-0,7895

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

2. Individual capacity of every independent variables in the combination 6

Combination 6

{X2, X3}
$$h_{62} = \frac{0,8374^2}{1 + |-0,884|} = 0,3723; \quad h_{63} = \frac{(-0,7895)^2}{1 + |-0,884|} = 0,3309$$

	0,9467
R ₀	0,8374
	-0,7895

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

Combination 7

{X1, X2, X3}

$$h_{71} = \frac{0,9467^2}{1 + 0,671 + |-0,712|} = 0,3762; \quad h_{72} = \frac{0,8374^2}{0,671 + 1 + |-0,884|} = 0,2745$$

$$h_{73} = \frac{(-0,7895)^2}{|-0,712| + |-0,884| + 1} = 0,2402$$

	0,9467
R ₀	0,8374
	-0,7895

	1	0,671	-0,712
R	0,671	1	-0,884
	-0,712	-0,884	1

Example

3. Integral capacity of information for each combination

Combination number	Independent variables for combination	Individual capacity of information h		Integral capacity of information H	
1	$\{X_1\}$	h_{11}	0,8962	H_1	0,89621
2	$\{X_2\}$	h_{22}	0,7013	H_2	0,7013
3	$\{X_3\}$	h_{33}	0,6234	H_3	0,62338
4	$\{X_1, X_2\}$	h_{41}	0,5364	H_4	0,95612
		h_{42}	0,4197		
5	$\{X_1, X_3\}$	h_{51}	0,5236	H_5	0,88786
		h_{53}	0,3642		
6	$\{X_2, X_3\}$	h_{62}	0,3723	H_6	0,70316
		h_{63}	0,3309		
7	$\{X_1, X_2, X_3\}$	h_{71}	0,3762	H_7	0,89089
		h_{72}	0,2745		
		h_{73}	0,2402		

The greatest integral capacity is for combination C4. Independent variables - X_1, X_2 - will be included in model.

Exercise 1

A sample of 9 petrol stations in Paris had the following prices for a litre of petrol:

Petrol station	Price per litre in Euro
1	1,14
2	1,19
3	1,25
4	1,21
5	1,17
6	1,19
7	1,22
8	1,24
9	1,19

Questions:

- What is the mean price per liter among 9 stations?
- What is the mode for the above data?
- What is the median of the above data?
- Compute the 25th percentile
- Compute the 75th percentile
- Compute the lower and the upper hinges
- Compute the variance for the above data
- Compute the standard deviation for the above data

THEORETICAL ASPECTS

STATISTICS

Calculation depending on the data set can be computed both for the sample as well as for the population (remember that there are differences in the formulas depending on the equation for the sample or for the population)

- mean
- median
- mode
- quartiles
- hinges
- variance
- standard deviation
- coefficient of variation
- Z-Score

Sample

Population

Mean

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\mu = \frac{\sum X_i}{N}$$

Interquartile range

$$IQR = Q_3 - Q_1$$

same as for sample

Variance

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$$

Standard Deviation

$$S = \sqrt{S^2}$$

$$\sigma = \sqrt{\sigma^2}$$

Coefficient of Variation (C.V.)

$$C.V. = \left(\frac{S}{\bar{X}} \right) (100)$$

$$C.V. = \left(\frac{\sigma}{\mu} \right) (100)$$

Z-Score

$$Z = \frac{X_i - \bar{X}}{S}$$

$$Z = \frac{X_i - \mu}{\sigma}$$