

# NARZĘDZIA STATYSTYKI MATEMATYCZNEJ WYKORZYSTYWANE W BADANIACH INNOWACYJNOŚCI PRZEDSIĘBIORSTW

Bożena KACZMARSKA, Artur MACIĄG

**Streszczenie:** W pracy przedstawiono narzędzia statystyki matematycznej możliwe do wykorzystania w analizach danych uzyskanych w badaniach własnych lub danych wtórnych, gromadzonych przez powołane do tego celu instytucje. Prezentowane rozważania teoretyczne zastosowano w pogłębionej analizie danych pozyskanych w ramach badań własnych dotyczących innowacyjności przedsiębiorstw.

**Słowa kluczowe:** innowacyjność przedsiębiorstw, ocena innowacyjności, narzędzia statystyki matematycznej.

## 1. Wprowadzenie

Współczesny rozwój państw, regionów i przedsiębiorstw determinują nowe uwarunkowania, odnoszące się przede wszystkim do wyzwań związanych z globalizacją oraz dużą konkurencją we wszystkich obszarach gospodarowania oraz gospodarką opartą na wiedzy. Elementem łączącym te wyzwania jest innowacyjność przedsiębiorstw, a w konsekwencji innowacyjność i konkurencyjność gospodarek regionalnych.

Badanie innowacyjności jest trudnym zadaniem, cechy związane z innowacyjnością są trudno mierzalne, ale mimo braku uniwersalnej metody pomiaru, literatura przedmiotu zawiera różne podejścia i metodyki oceny innowacyjności przedsiębiorstw, regionów, państw [1-6]. Przegląd literatury wskazuje, że w wielu przypadkach badania dotyczą dużej grupy podmiotów, a przeprowadzane analizy ograniczają się do wykorzystania narzędzi statystyki opisowej. W wyniku takich działań, część bardzo istotnych informacji zostaje utracona. Ważne jest zatem przedstawienie możliwości przeprowadzania poszerzonych analiz z wykorzystaniem narzędzi statystyki matematycznej, co jest celem niniejszego artykułu.

## 2. Narzędzia statystyki matematycznej wykorzystane w badaniach

Statystyka matematyczna jest działem matematyki, który z uwagi na swoje liczne zastosowania, rozwija się bardzo dynamicznie. Dotychczas wypracowano wiele technik analizy danych, które znajdują zastosowanie w biznesie, ekonomii, demografii i wielu dziedzinach nauki. Bogactwo dostępnych narzędzi pozwala dobrać odpowiednie techniki do analizowanego problemu. Prezentowana praca ma na celu przedstawienie wybranych technik analizy danych dotyczących innowacyjności przedsiębiorstw. Kluczowym problemem jest jakość posiadanych danych. Aby uzyskać wiarygodne i użyteczne wyniki analiz, jakość danych nie może budzić większych wątpliwości.

Na potrzeby niniejszej pracy przyjęto założenie, że znana jest zmienna określająca, czy rozważane przedsiębiorstwo jest innowacyjne, czy też nie (zmienna decyzyjna, zmienna klasyfikująca). Zmienna ta przyjmuje dwa poziomy wartości (Tak – Nie lub 0 – 1).

Możliwe jest przyjęcie większej liczby poziomów wartości tej zmiennej określających stopień (natężenie) innowacyjności firmy. Mając wartości zmiennej decyzyjnej oraz wartości innych zmiennych charakteryzujących przedsiębiorstwo (lokalizacja, liczba zatrudnionych, rok powstania, przychody, liczba wprowadzanych innowacji, wykorzystywanie ICT, itp.) można przeprowadzić szereg interesujących analiz. Po pierwsze można analizować przedsiębiorstwa innowacyjne pod kątem ich rozmaitych cech (lokalizacja, wielkość, forma prawna, itp.). Po drugie można zbudować klasyfikator pozwalający określić, czy przedsiębiorstwo o pewnych parametrach może być uznane za innowacyjne. Klasyfikator taki na „wejściu” otrzymuje wartości zmiennych charakteryzujących przedsiębiorstwo, a na „wyjściu” podaje decyzję do jakiej klasy przedsiębiorstwo zostało zaliczone. W obu przypadkach wykorzystywane są narzędzia statystyki matematycznej. W trakcie analiz statystycznych wykorzystuje się zwykle zarówno narzędzia statystyki opisowej, wnioskowania statystycznego (często nie ma możliwości poddania badaniom wszystkich działających na rynku firm, natomiast wyodrębniona próba losowa pozwala na ogół wnioskować o całej zbiorowości.), jak również bardziej zaawansowane techniki z zakresu *data mining*. Są to:

- szeregi rozdzielcze i histogramy,
- statystyki zmiennej losowej,
- estymacja oraz testowanie hipotez statystycznych,
- analiza współzależności cech (testy niezależności, korelacja),
- modele ekonometryczne,
- sieci neuronowe oraz reguły i drzewa decyzyjne.

Stosując narzędzia wnioskowania statystycznego, należy pamiętać, że każdy taki proces obarczony jest pewnym błędem (poziom istotności). Wnioski o całej populacji można uznać za prawdziwe tylko z pewnym prawdopodobieństwem (poziomem ufności).

## 2.1. Statystyka opisowa

Narzędzia statystyki opisowej są powszechnie znane. Najprostszą, a jednocześnie najbardziej obrazową techniką prezentacji rozkładu cechy jest wykres oraz szereg rozdzielczy w postaci tabeli. W przypadku zmiennej losowej o charakterze ilościowym do opisu rozkładu tej zmiennej stosuje się rozmaite liczbowe charakterystyki, do których należą: średnia, odchylenia standardowe, współczynnik zmienności, kwartyle, dominanta, współczynniki asymetrii i spłaszczenia, itp. Dla cech jakościowych wyznaczyć można wskaźnik struktury (frakcję).

## 2.2. Wnioskowanie statystyczne

Do narzędzi wnioskowania statystycznego zalicza się estymację (w szczególności przedziałową) oraz testowanie hipotez statystycznych. Narzędzia te pozwalają formułować sądy o całej zbiorowości na podstawie wyników uzyskanych dla próby losowej.

Testem statystycznym nazywa się regułę postępowania, która każdej możliwej próbie przyporządkowuje decyzje przyjęcia lub odrzucenia hipotezy dotyczącej całej zbiorowości. Hipotezy te mogą być parametryczne (dotyczą parametrów rozkładu, np. średniej lub wariancji) lub nieparametryczne (dotyczą postaci rozkładu). Wyróżnia się trzy grupy testów: zgodności, losowości oraz niezależności. W szczególności test niezależności pozwala rozstrzygnąć istnienie wpływu jednej zmiennej na drugą. Dla cech jakościowych najczęściej stosowany jest test niezależności stochastycznej  $\chi^2$  (chi – kwadrat) (zależność

między zmiennymi ilościowymi stwierdzić można, wykorzystując współczynnik korelacji lub analizę regresji).

### 2.3. Test niezależności stochastycznej $\chi^2$

Niech dane będą dwie, jakościowe zmienne losowe  $X$  i  $Y$ , przy czym zmienna  $X$  ma  $k$  natomiast zmienna  $Y$   $l$  różnych poziomów wartości. Dane dotyczące zmiennych przedstawia się w tablicy dwudzielnej (tabela 1).

Tab. 1. Tablica dwudzielna

$y_j$	$y_1$	$y_2$	...	$y_l$	$\sum_j$
$x_i$					
$x_1$	$n_{11}$	$n_{12}$	...	$n_{1l}$	$n_{1\bullet}$
$x_2$	$n_{21}$	$n_{22}$	...	$n_{2l}$	$n_{2\bullet}$
...	...	...	...	...	...
$x_k$	$n_{k1}$	$n_{k2}$	...	$n_{kl}$	$n_{k\bullet}$
$\sum_i$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet l}$	$n$

Zródło: opracowanie własne

Gdzie:

$x_i, y_l$  – różne poziomami wartości cech odpowiednio  $X$  i  $Y$ ,

$n_{kl}$  – licznosc tych jednostek statystycznych, dla których cecha  $X$  ma wartość  $x_k$  i jednocześnie cecha  $Y$  ma wartość  $y_l$ ,

$n_{k\bullet}, n_{\bullet l}$  – są wartościami brzegowymi odpowiednio zmiennych  $X$  i  $Y$ . Na przykład  $n_{1\bullet}$  oznacza liczbę obserwacji, dla których zmienna  $X$  przyjęła pierwszą wartość, bez względu na to jaka jest wtedy wartość zmiennej  $Y$ .

Hipotezy o niezależności stochastycznej zmiennych mają postać:

$H_0$ : zmienne  $X$  i  $Y$  są niezależne,

$H_1$ : zmienne  $X$  i  $Y$  nie są niezależne.

Statystyką testową jest:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}, \quad (1)$$

gdzie  $\hat{n}_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}$  są tak zwanymi liczebnościami teoretycznymi.

Obszarem krytycznym testu jest przedział  $K = (\chi_{1-\alpha}^2, \infty)$ , gdzie parametr  $\chi_{1-\alpha}^2$  odczytuje się z tablic rozkładu  $\chi^2$  dla  $(k-1)(l-1)$  stopni swobody i dla zadanego poziomu ufności  $\alpha$ . W przypadku stwierdzenia zależności zmiennych można określić siłę tej zależności obliczając wartość współczynnika zbieżności  $V$ -Cramera:

$$V = \sqrt{\frac{\chi^2}{n \cdot [\min(k, l) - 1]}} \in [0, 1]. \quad (2)$$

Bliskie jedności wartości  $V$  wskazują na silną współzależność cech. Im bliższa zera jest wartość  $V$  tym współzależność jest słabsza. Przyjmuje się następującą interpretację współczynnika  $V$ :

- $V < 0,1$  – zależność nikła,
- $0,1 \leq V < 0,3$  – zależność słaba,
- $0,3 \leq V < 0,5$  – zależność przeciętna,
- $0,5 \leq V$  – zależność silna.

Wygodnym narzędziem przy przeprowadzaniu rozmaitych testów statystycznych jest tak zwana  $p$  – wartość ( $p$  – value). Interpretacja tej liczby jest następująca: jeśli jest ona mniejsza aniżeli założony poziom istotności  $\alpha$ , to hipoteza zerowa winna by odrzucona i są podstawy, aby przyjąć hipotezę alternatywną. W przeciwnym wypadku hipoteza zerowa odrzucona być nie może.

#### 2.4. Drzewa i reguły decyzyjne, sieci neuronowe

Do bardziej zaawansowanych metod analizy danych należą drzewa i reguły decyzyjne oraz sieci neuronowe. Złożoność obliczeniowa tych techniki wymaga do ich realizacji odpowiednich programów komputerowych (w niniejszej pracy wykorzystano system SAS [7]). W pierwszej kolejności należy określić zmienną celu (target), która stanowi podstawę do budowy drzewa decyzyjnego oraz uczenia sieci neuronowej (uczenie z nauczycielem). Pozostałe zmienne nazywane są wejściowymi i pełnią niejako rolę zmiennych objaśniających. Następnie zbiór danych dzielony jest losowo na trzy części: treningową, walidacyjną i testową. Zbiór treningowy wykorzystywany jest do wstępnego dopasowania modelu, który następnie jest weryfikowany na zbiorze walidacyjnym. Na zbiorze testowym sprawdza się poprawność uzyskanych wyników. Uzyskane drzewo decyzyjne stanowi podstawę do wygenerowania reguł decyzyjnych. Nauczona sieć neuronowa może pełnić rolę klasyfikatora.

### 3. Studium przypadku

W celu ilustracji opisanych powyżej narzędzi przeprowadzono pogłębione analizy danych, które są wynikami dwóch badań zrealizowanych w 2008 i 2012 roku, w których brali udział autorzy artykułu. Celem badań była ocena innowacyjności i konkurencyjności przedsiębiorstw z województwa świętokrzyskiego. Obie bazy danych zawierały zmienną decyzyjną zaliczającą przedsiębiorstwo do kategorii innowacyjnych lub nieinnowacyjnych. W dotychczasowych analizach i publikacjach wyników badań ograniczono się do opisów z wykorzystaniem tylko narzędzi statystyki opisowej [3, 8, 9].

#### 3.1. Przykład I

##### Charakterystyka badanej grupy przedsiębiorstw [8]

W celu promowania działalności innowacyjnej i zbadania warunków dla rozwoju inicjatyw innowacyjnych w regionie świętokrzyskim, zrealizowano projekt pn.: „Rozwój

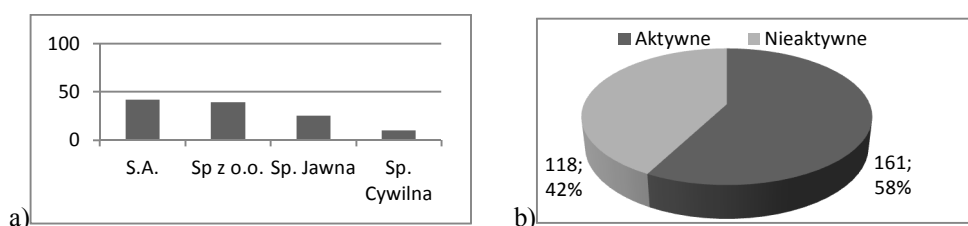
Regionalne Strategii Innowacji Województwa Świętokrzyskiego poprzez promocję i monitoring – RSI PROMONIT” - współfinansowany przez Unię Europejską. W ramach projektu w 2008 roku przeprowadzono badania w zakresie warunków konkurencyjności przedsiębiorstw. Badaniom poddanych zostało 500 przedsiębiorstw województwa świętokrzyskiego aktywnych w zakresie innowacyjności i pozyskiwania nowych technologii. Próba została dobrana celowo. Do badań wytypowano przedsiębiorstwa różnej wielkości (duże, średnie, małe i mikro), których obszar działalności zawiera się w sekcjach PKD: C, D, E, F, G, I, K, O. Liczba wytypowanych przedsiębiorstw była proporcjonalna do ich liczby w poszczególnych powiatach województwa świętokrzyskiego, ponadto baza przedsiębiorstw została uzupełniona o przedsiębiorstwa, które aplikowały o środki pochodzące z Unii Europejskiej oraz przedsiębiorstwa, które znalazły się w rankingach firm wyróżniających się w działalności gospodarczej. Wypełniony kwestionariusz ankietowy otrzymano od 279 przedsiębiorstw.

#### Metodyka badań

W prowadzonych badaniach zastosowano metodę ankietową, uzupełnioną techniką obserwacji w oparciu o wyznaczone wskaźniki. Kwestionariusz ankietowy zawierał pytania zamknięte i otwarte, które dotyczyły cech jakościowych i ilościowych dotyczących innowacyjności przedsiębiorstwa. W celu weryfikacji poprawności kwestionariusza ankietowego przeprowadzono badania pilotażowe. Wskaźniki wynikowe dotyczyły: liczby i rodzaju wprowadzonych w przedsiębiorstwach innowacji zwiększających ich konkurencyjność, finansowania działalności innowacyjnej oraz współpracy poziomej i pionowej podmiotów w zakresie działalności innowacyjnej. Ankieta obejmowała innowacje technologiczne (produktowe, procesowe) i nie technologiczne (organizacyjne, marketingowe).

#### Analiza

Baza danych zawierała rozmaite dane dotyczące 279 przedsiębiorstw. Były to między innymi: forma prawna, forma organizacyjna, rok powstania, zaangażowanie kapitału zagranicznego, szkolenia, posiadanie strony www, oferowanie pracownikom konta poczty elektronicznej (e-mail). Przykład graficznej prezentacji danych przedstawia rysunek 1.



Rys. 1. Przykładowe histogramy:

- a) przedsiębiorstwa innowacyjne w ramach szczegółowych form prawnych,
- b) przedsiębiorstwa posiadające aktywną stronę www.

Źródło: opracowanie własne.

Uzyskanie histogramu jest jedynie krokiem początkowym do dalszych analiz. Histogramy takie dostarczają informacji i powinny być odpowiednio zinterpretowane i odczytane. Widoczne jest, że we wszystkich rodzajach form prawnych, jest poniżej 50% firm uznanych za innowacyjne. Najlepiej pod tym względem wyglądają spółki akcyjne, a najgorzej spółki cywilne. Prawie 60% analizowanych firm posiadało aktywną stronę internetową.

Rozważmy dwie zmienne:  $X$  – forma prawna przedsiębiorstwa,  $Y$  – innowacyjność przedsiębiorstwa. Stawiamy hipotezy o niezależności stochastycznej:

$H_0$ : zmienne  $X$  i  $Y$  są niezależne,

$H_1$ : zmienne  $X$  i  $Y$  nie są niezależne.

Tablica dwudzielna jest następująca (Tabela 2):

Tabela 2. Tabela dwudzielna innowacyjności przedsiębiorstw w zależności od roku powstania

$Y \backslash X$	Innowacyjne	Nieinnowacyjne	Suma
Osoba fizyczna	23   31,72	109   100,28	132
Spółka cywilna	3   7,45	28   23,55	31
Spółka jawna	3   2,88	9   9,12	12
Spółka z o.o.	23   14,18	36   44,82	59
Spółka akcyjna	10   5,77	14   18,23	24
Suma	62	196	258

Zródło: opracowanie własne.

Zmienna  $X$  przyjmuje  $k=5$  natomiast zmienna  $Y$  przyjmują dwa  $l=2$  poziomy. Na przykład liczba 23 oznacza, że tyle odnotowano osób fizycznych prowadzących działalność gospodarczą, które uznano za innowacyjne. Ostatnia kolumna zawiera rozkład brzegowy zmiennej  $X$ . Przykładowo obserwujemy 132 osoby fizyczne prowadzące działalność gospodarczą, bez względu na to, czy są one innowacyjne, czy też nie. Analogicznie w rozważanej próbie występuje 62 podmioty innowacyjne bez względu na formę prawną. Tabela zawiera również liczebności teoretyczne, które interpretujemy jako takie, które powinny wystąpić przy całkowitym braku zależności zmiennych. Przykładowo

$$\hat{n}_{11} = \frac{n_{1\cdot} \cdot n_{\cdot 1}}{n} = \frac{132 \cdot 62}{258} = 31,72, \quad \hat{n}_{12} = \frac{n_{1\cdot} \cdot n_{\cdot 2}}{n} = \frac{132 \cdot 196}{258} = 100,28.$$

Pozostałe liczebności liczone są analogicznie. Obszar krytyczny:  $K = (\chi^2_{1-\alpha}, \infty) = (9,49, \infty)$ .

Statystyka testowa ma postać:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = \frac{(23 - 31,72)^2}{31,72} + \frac{(109 - 100,28)^2}{100,28} + \dots + \frac{(10 - 5,77)^2}{5,77} + \frac{(14 - 18,23)^2}{18,23} = 17,97 \in K$$

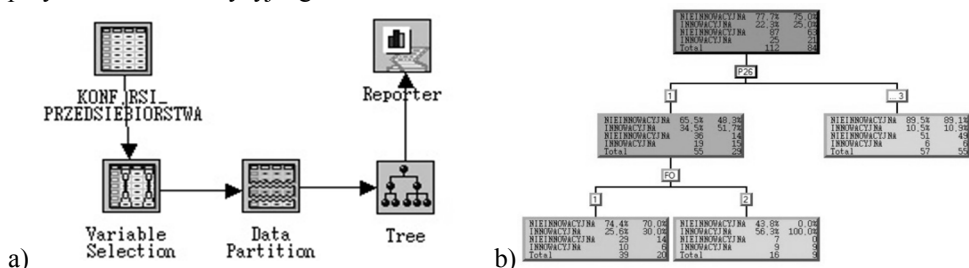
Odrzucamy hipotezę zerową uznając za prawdziwą hipotezę alternatywną, zatem można twierdzić, że zmienne są zależne. Badamy siłę zależności licząc współczynnik zbieżności

$$V = \sqrt{\frac{\chi^2}{n \cdot [\min(k, l) - 1]}} = \sqrt{\frac{17,97}{258 \cdot 1}} = 0,26.$$

Forma prawna ma wpływ na innowacyjność, jednak zależność jest dość słaba. Dane zawarte w tabeli 2 wskazują, że mniej innowacyjne są osoby fizyczne prowadzące działalność gospodarczą oraz spółki cywilne. Do bardziej innowacyjnych zaliczyć należy

spółki kapitałowe – spółki z ograniczoną odpowiedzialnością i spółki akcyjne.

W celu ilustracji wykorzystania drzew decyzyjnych bazę danych poddano analizie z wykorzystaniem modułu Enterprise Miner w programie SAS [7]. W tym celu wystarczy utworzyć odpowiedni diagram procesu i go wykonać. W pierwszej kolejności należy wybrać zbiór danych do analizy. Następnie wybiera się i określa rolę zmiennych oraz dzieli zbiór danych na część treningową, walidacyjną i testową. W szczególności powinna być wskazana jedna zmienna celu (target). Podobnie jak wcześniej, jest to zmienna określająca, czy przedsiębiorstwo jest innowacyjne. Po wykonaniu procesu otrzymuje się drzewo decyzyjne oraz wynikający z niego zbiór reguł. Wyniki zawarte są w wygenerowanym raporcie. Rysunek 2a przedstawia diagram wykonywanego procesu, natomiast rysunek 2b przykład drzewa decyzyjnego.



Rys. 2. Diagram procesu oraz przykład drzewa decyzyjnego  
Źródło: opracowanie własne w module Enterprise Miner programu SAS.

Uzyskane drzewa decyzyjne pozwalają sformułować reguły decyzyjne, które mają postać zdań typu „Jeżeli ... to...”. Przykładowe reguły zamieszczono niżej.

### Reguła 1

JEŻELI **Formą prawną** jest osoba fizyczna prowadząca działalność gospodarczą lub spółka cywilna lub spółka jawna

TO

N : 67  
NIEINNOWACYJNA: 88,1%  
INNOWACYJNA: 11,9%

### Interpretacja

W następniku implikacji (po słowie „TO”) widoczne są interesujące informacje. Liczba N (tutaj 67) określa ile firm spełnia warunek następujący po “JEŻELI” Spośród tych firm 88.1% jest nieinnowacyjnych, natomiast 11,9% jest innowacyjnych. Oznacza to, że forma prawna określona poprzednikiem implikuje w większości przypadków nieinnowacyjność. Reguła ta potwierdza wniosek uzyskany dzięki testowi  $\chi^2$ .

### Reguła 2

JEŻELI Firma nie posiada strony www ORAZ nie oferuje pracownikom konta poczty elektronicznej ORAZ nie jest zaangażowany kapitał zagraniczny

TO

N: 31  
NIEINNOWACYJNA: 100,0%  
INNOWACYJNA: 0,0%

### **Interpretacja**

Firmy, które nie posiadają strony www i jednocześnie nie oferują pracownikom konta poczty elektronicznej i jednocześnie nie mają zaangażowanego kapitału zagranicznego nie są innowacyjne.

### **Reguła 3**

JEŻELI Pracownicy poświęcają mało czasu na szkolenia ORAZ firma nie oferuje pracownikom konta poczty elektronicznej

TO

N: 20

NIEINNOWACYJNE: 75,0%

INNOWACYJNE: 25,0%

### **Interpretacja**

Większość firm, które nie dba o szkolenia pracowników i jednocześnie nie oferuje pracownikom konta poczty elektronicznej jest nieinnowacyjna.

### **Reguła 4**

JEŻELI Firma posiada udział kapitału zagranicznego

TO

N: 17

NIEINNOWACYJNE: 47,1%

INNOWACYJNE: 52,9%

### **Interpretacja**

Nieco ponad połowa przedsiębiorstw z udziałem kapitału zagranicznego jest innowacyjna.

## **3.2. Przykład II**

### **Charakterystyka badanej grupy przedsiębiorstw [3, 9]**

Na potrzeby badań empirycznych, prowadzonych w 2012 roku, z systemu Krajowego Rejestru Podmiotów Gospodarki Narodowej REGON wybrano 929 przedsiębiorstw z województwa świętokrzyskiego (baza przedsiębiorstw została przygotowana w ramach projektu badawczego pn., „Polskie Młode Zaawansowane Technologicznie Firmy”). Są to przedsiębiorstwa powstałe i zarejestrowane w latach 2001-2010 i deklarujące na etapie uruchamiania działalności gospodarczej wiodący przedmiot działalności w branżach określanych jako branże wysokiej techniki.

Sposób wyboru przedsiębiorstw do badań wynikał z celu badań, jakim jest ocena poziomu innowacyjności przedsiębiorstw regionu świętokrzyskiego, albowiem poziom innowacyjności gospodarki jest bezpośrednio powiązany z liczbą przedsiębiorstw, które prowadzą działalność innowacyjną. Dlatego pominięto podmioty przypisane do innych branż (niewysokiej techniki), jako te, u których występuje znacznie mniejsze prawdopodobieństwo występowania innowacyjnych technologii.

Wykorzystywane w badaniach informacje dotyczące przedsiębiorstw pozyskiwane były z następujących źródeł:

- Rejestr REGON.
- Baza Centralna Ewidencja i Informacja o Działalności Gospodarczej.
- Baza Podmiotów Gospodarczych.
- Krajowy Rejestr Sądowy on-line.
- Strony internetowe badanych przedsiębiorstw.



- Strona internetowa Kieleckiego Parku Technologicznego.
- Inne źródła internetowe związane pośrednio z badanymi przedsiębiorstwami.

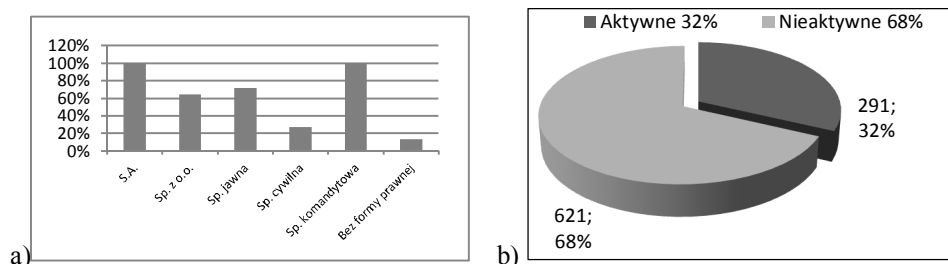
### Metodyka badań

W prowadzonych badaniach zastosowano clipping, jako technikę zbierania danych, uzupełniony techniką obserwacji. Przyjęta metodyka prowadzonych badań obejmowała analizę treści stron (witryn) internetowych podmiotów gospodarczych województwa świętokrzyskiego. Analiza treści ukierunkowana była na zagadnienia innowacyjności. Informacja uzyskana ze stron internetowych badanych podmiotów była uzupełniana też z Internetu, ale w sposób pośredni poprzez informacje zamieszczane w tematycznych portalach, listach dyskusyjnych, itp.

Analizując informacje internetowe, dokonywano dwustanowej klasyfikacji, wskazując przedsiębiorstwa stosujące innowacje. W ocenie poziomu innowacyjności i nowości wykorzystywanych w działalności przedsiębiorstw stosowano standardy OECD (*Organization for Economic Co-operation and Development*)i metodykę przedstawioną w Podręczniku Oslo [10] w stosunku do parametrów, takich jak: produkowane wyroby, stosowane technologie, prowadzone badania oraz współpracopresiębiorstw z ośrodkami naukowo- badawczymi.

### Analiza

W bazie danych uzyskanej w przeprowadzonych badaniach, oprócz zmiennej decyzyjnej, znalazły się takie zmienne, jak: numer Regon, NIP, adres, powiat, gmina, rok powstania, forma prawna, forma własności, PKD, grupa zatrudnienia, email, strona www, przychody, zatrudnienie oraz aktywa. Wzięto pod uwagę 929 przedsiębiorstw. Przykład graficznej prezentacji danych przedstawia rysunek 3.



Rys. 3. Przykładowe histogramy:

- przedsiębiorstwa innowacyjne w ramach szczegółowych form prawnych,
- przedsiębiorstwa posiadające aktywną stronę www.

Źródło: opracowanie własne

Analizując rysunek 3, widoczne jest, że wszystkie spółki akcyjne i komandytowe zaliczone zostały do grupy przedsiębiorstw innowacyjnych. Najniższy odsetek takich firm występuje w przypadku spółek cywilnych. Jedynie 291 przedsiębiorstw posiada aktywną stronę www. Stanowi to 32% przebadanych firm.

Rozważmy dwie zmienne:  $X$  – rok powstania przedsiębiorstwa,  $Y$  – innowacyjność przedsiębiorstwa. Stawiamy hipotezy o niezależności stochastycznej:

- $H_0$ : zmienne  $X$  i  $Y$  są niezależne,  
 $H_1$ : zmienne  $X$  i  $Y$  nie są niezależne.

Tablica dwudzielna jest następująca (Tabela 3):

Tab. 3. Tabela dwudzielna innowacyjności przedsiębiorstw w zależności od roku powstania

$X \backslash Y$	Innowacyjne	Nieinnowacyjne	Suma
2001–2005	63  51,72	252  263,28	315
2006–2010	81  92,28	481  469,72	562
Suma	144	733	877

Źródło: opracowanie własne

Obie zmienne  $X$  i  $Y$  przyjmują dwa poziomy wartości ( $k=l=2$ ). Obszar krytyczny:  $K = (3,84, \infty)$ , statystyka testowa ma postać:  $\chi^2 = 4,59 \in K$ . Odrzucamy hipotezę zerową uznając za prawdziwą hipotezę alternatywną, zatem można twierdzić, że zmienne są zależne. Badamy siłę zależności licząc współczynnik zbieżności  $V=0,07$ . Rok powstania firmy ma wpływ na innowacyjność, jednak zależność jest nikła. Dane zawarte w tabeli 3 wskazują, że bardziej innowacyjne są firmy starsze.

Rozważmy dwie zmienne:  $X$  – forma prawna przedsiębiorstwa,  $Y$  – innowacyjność przedsiębiorstwa. Stawiamy hipotezy o niezależności stochastycznej:

$H_0$ : zmienne  $X$  i  $Y$  są niezależne,

$H_1$ : zmienne  $X$  i  $Y$  nie są niezależne.

Tablica dwudzielna jest następująca:

Tab. 4. Tabela dwudzielna innowacyjności przedsiębiorstw w zależności od formy prawnej

$X \backslash Y$	Innowacyjne	Nieinnowacyjne	Suma
Spółki kapitałowe	60  17,76	30  72,24	90
Spółki osobowe	20  11,84	40  48,16	60
Osoby fizyczne prowadzące działalność gospodarczą	100  150,39	662  611,61	762
Suma	180	732	912

Źródło: opracowanie własne.

Obliczenia prowadzimy analogicznie, jak w przykładzie poprzednim, pamiętając, że teraz zmienna  $X$  przyjmuje trzy poziomy wartości. Obszar krytyczny:  $K = (5,991, \infty)$ , statystyka testowa:  $\chi^2 = 264,26 \in K$ , współczynnik zbieżności  $V=0,54$ . Odrzucamy hipotezę zerową, uznając za prawdziwą hipotezę alternatywną. Zatem można twierdzić, że forma prawna firmy ma wpływ na innowacyjność, przy czym zależność jest silna. Analizując dane zamieszczone w tabeli 4 widoczne jest, że za najbardziej innowacyjne należy uznać spółki kapitałowe, natomiast najmniej innowacyjne są osoby fizyczne prowadzące działalność gospodarczą.

Zbiór danych zawierający poza zmienną decyzyjną zmienne: rok powstania, podstawowa forma prawna oraz szczegółowa forma prawna poddano analizie za pomocą drzew decyzyjnych (dokładniejszy opis – patrz Badanie II) z wykorzystaniem modułu Enterprise Miner w programie SAS [7]. Uzyskano następujące reguły:

#### **Reguła 1**

JEŻELI **Podstawową formą prawną** jest Osoba fizyczna prowadzące działalność gospodarczą lub Jednostka organizacyjna niemająca osobowości prawnej

TO

N: 323  
TAK: 14,2%  
NIE: 85,8%

#### **Interpretacja**

Osoby fizyczne prowadzące działalność gospodarczą oraz jednostki organizacyjne niemające osobowości prawnej są w przeważającej większości nieinnowacyjne.

#### **Reguła 2**

JEŻELI **Podstawową formą prawną** jest Osoba prawna

TO

N: 32  
TAK: 59,4%  
NIE: 40,6%

#### **Interpretacja**

Osoby prawne są w większości przypadków innowacyjne.

#### **Reguła 3**

JEŻELI **Szczegółową formą prawną** jest spółka z ograniczoną odpowiedzialnością

TO

N: 31,0  
TAK: 58,1%  
NIE: 41,9%

#### **Interpretacja**

Spółki z ograniczoną odpowiedzialnością są w większości przypadków innowacyjne.

Uzyskane reguły potwierdzają wnioski płynące z testu niezależności: innowacyjności od formy prawnej.

W trakcie analiz nie uzyskano reguł dotyczących zależności innowacyjności od roku powstania firmy. Należy przypomnieć, że test  $\chi^2$  wskazał na bardzo słabą współzależność tych zmiennych. Zatem działalność innowacyjna może być prowadzona w różnych lokalizacjach w województwie świętokrzyskim. Miasto Kielce z uczelniami wyższymi i parkiem technologicznym nie determinuje innowacyjności przedsiębiorstw.

### **4. Wnioski**

W pracy przedstawiono podejście do oceny innowacyjności w ujęciu ilościowym, z wykorzystaniem narzędzi statystyki matematycznej. Szerokie spektrum metod statystycznych znajduje zastosowanie również w przypadku badania innowacyjności przedsiębiorstw. W prezentowanym artykule pokazano wybrane narzędzia, które mogą być w takim przypadku stosowane. Należy pamiętać, że wykorzystanie rozmaitych metod statystycznych jest możliwe, gdy dysponuje się odpowiedniej jakości zbiorem danych. Jest to kluczowy problem, na który należy zwrócić uwagę. Prezentowane analizy przeprowadzono przy założeniu znanych wartości zmiennej decyzyjnej. Osobny problem stanowi uzyskanie wartości takiej zmiennej. Próba odpowiedzi na pytanie, które czynniki decydują o innowacyjności firmy oraz kiedy firma może być uznana za innowacyjną jest zadaniem niebanalnym. Narzędzia statystyki matematycznej bez wątpienia mogą być wykorzystane również w tym przypadku, którego opis przekracza ramy niniejszej pracy.

Skuteczne zarządzanie innowacjami wymaga stałej analizy, działań kontrolnych i monitoringu/ewaluacji. Temu służą powtarzane badania, których celem jest ocena innowacyjności przedsiębiorstw, regionów, państw. Badania prowadzone są przez instytucje z założenia do tego powołane (np. GUS, Eurostat) oraz przez osoby/zespoły realizujące własne badania naukowe. Część badań dotyczy dużej populacji przedsiębiorstw, a ocena dotyczy ogólnego poziomu innowacyjności regionu, państwa. Odbiorcami tego typu badań są władze regionu lub państwa. Jednak poprzestanie w procesie analiz tylko na narzędziach statystyki opisowej nie pozwala na wydobycie z gromadzonych danych pełnej informacji oraz słabych sygnałów o możliwości pojawienia się nowych szans lub zagrożeń.

### Literatura

1. Gierulski W., Kaczmarek B.: *Methodology for Evaluating Organization Development State. An Application of the DEA Method*, LAP Lambert Academic Publishing, Germany 2012, ISBN 978-3-659-22975-6.
2. Gierulski W., Kaczmarek B.: *Evaluating the level of technology development – a mathematical model*, w: *Innovations in Management and Production Engineering*, ed. R. Knosala, Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, Opole 2013, ISBN 978-83-930399-9-9, s. 29-40.
3. Kaczmarek B.: *Ocena poziomu innowacyjności przedsiębiorstw na podstawie zasobów internetowych*, w: *Innowacje w zarządzaniu i inżynierii produkcji*, pod red. R. Knosali, Oficyna Wydawnicza Polskiego Towarzystwa Zarządzania Produkcją, Opole 2013, ISBN 978-83-930399-5-1, s. 112-123.
4. Kaczmarek B.: *Macierz stanów jako model działalności ośrodków innowacji i przedsiębiorczości*, w: P. Łebkowski (red. nauk.) *Aspekty inżynierii produkcji*, Wydawnictwo AGH, Kraków 2010, s. 115-128.
5. Kaczmarek B., Gierulski W.: *Analiza porównawcza poziomu innowacyjności państw Unii Europejskiej*, w: *Zbiór prac naukowych – nauki ekonomiczne*, Uniwersytet Chmielnicki, Ukraina 2012, s. 28-38.
6. Knosala R., Boratyńska-Sala A., Jurczyk-Bunkowska M., Moczala A.: *Zarządzanie innowacjami*, PWE Warszawa 2014.
7. Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA.
8. Nowak P., Miłek D., Kaczmarek B., (red. nauk.): *Analiza wyników monitorowania rozwoju Regionalnej Strategii Innowacji Województwa Świętokrzyskiego*, Kielce, 2008, s. 17, 113 - 186.
9. Kaczmarek B., Sulerz A.: *Ocena innowacyjności przedsiębiorstw informatycznych*, w: *Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne*, Wyd. WSH Kielce 2013, s. 155-165.
10. *Podręcznik Oslo. Zasady gromadzenia i interpretacji danych dotyczących innowacji*, Wyd. trzecie, ([www.nauka.gov.pl](http://www.nauka.gov.pl)).

Dr inż. Bożena Kaczmarek  
Katedra Inżynierii Produkcji, Wydział Zarządzania i Modelowania Komputerowego  
Dr hab. Artur Maciąg, prof. PŚk  
Katedra Matematyki, Wydział Zarządzania i Modelowania Komputerowego  
Politechnika Świętokrzyska, 25-314 Kielce, Aleja Tysiaclecia PP 7  
e-mail: [bozena.kaczmarek@tu.kielce.pl](mailto:bozena.kaczmarek@tu.kielce.pl)  
[matam@tu.kielce.pl](mailto:matam@tu.kielce.pl)