

Maciej Oziembłowski

Uniwersytet Przyrodniczy we Wrocławiu
e-mail: maciej.ozieblowski@upwr.edu.pl
ORCID: 0000-0001-7369-6051

Tomasz Lesiów

Uniwersytet Ekonomiczny we Wrocławiu
e-mail: tomasz.lesiow@ue.wroc.pl
ORCID: 0000-0002-1284-5874

Cornelia Šabanagić

e-mail: cornelia.sabanagic@gmail.com

Metodologia testu chi-kwadrat na przykładzie badań ankietowych dotyczących europejskich serów regionalnych

The Methodology of the Chi-Square Test on the Example of Surveys on Regional Cheeses

DOI: 10.15611/nit.2022.38.09
JEL Classification: C19 i Q19

Streszczenie: Celem badań było praktyczne wskazanie metodologii testu niezależności chi-kwadrat z wykorzystaniem pakietu programu *Statistica* na przykładzie wyników ankiety dotyczącej europejskich serów regionalnych. Otrzymano trzy grupy danych: 10 zmiennych jakościowych (kategoryzujących), 4 zmienne jako wielokrotne odpowiedzi, 46 zmiennych jako wielokrotne dychotomie – łącznie 60 zmiennych. Każdej odpowiedzi ankietowej przyporządkowano 4 wyróżniki opisujące respondenta, tj. płeć, wiek, wykształcenie oraz jego miejsce zamieszkania. Przeprowadzono analizę testu chi-kwadrat Pearsona oraz chi-kwadrat największej wiarygodności (NW), łącznie 240 podwójnych testów, oraz obliczono nieparametryczne korelacje rang Spearmana pomiędzy wybranymi zmiennymi. Na podstawie uzyskanych wyników udowodniono istnienie 25 statystycznie istotnych zależności pomiędzy analizowanymi zmiennymi a jednym z czterech wyróżników metryczki respondentów. Analiza korelacji rang Spearmana pomiędzy wybranymi zmiennymi wykazała co najwyżej umiarkowaną korelację pomiędzy kilkoma parami zmiennych. Uzyskane wyniki potwierdziły przydatność testu niezależności chi-kwadrat oraz analizy korelacji rang Spearmana w badaniach ankietowych.

Słowa kluczowe: test chi-kwadrat Spearmana, test chi-kwadrat największej wiarygodności, korelacje rang Spearmana, europejskie sery regionalne, badania ankietowe.

Summary: The aim of the research was to present the methodology of the chi-square test of independence with the use of the Statistica package on the example of the results of a survey on European regional cheeses. Three groups of data were obtained: 10 qualitative (categorizing)

variables, 4 variables as multiple responses, 46 variables as multiple dichotomies, 60 variables in total. Each questionnaire answer was assigned 4 differentiators describing the respondent, i.e. gender, age, education, and his place of residence. The Pearson chi-square test and the maximum likelihood chi-square (NW) test for a total of 240 duplicate tests were analyzed and non-parametric Spearman rank correlations between the selected variables were calculated. On the basis of the obtained results, the existence of 25 statistically significant relationships between the analyzed variables and one of the four characteristics of the respondents' record was proved. The analysis of Spearman's rank correlation between the selected variables showed at most a moderate correlation between several pairs of variables. The obtained results confirmed the usefulness of the chi-square test of independence and the analysis of Spearman's rank correlation in the survey.

Keywords: Spearman's chi-square test, maximum likelihood chi-square test, Spearman's rank correlations, European regional cheeses, surveys.

1. Wstęp

W badaniach ankietowych wykorzystuje się różne analizy statystyczne. Stosunkowo często można spotkać informację, że w badaniach przeprowadzono test niezależności chi-kwadrat, zwany również testem chi-kwadrat Pearsona. Metodę tę wykorzystuje się w celu zbadania zależności pomiędzy dwiema zmiennymi nominalnymi. Ideą tej analizy jest porównanie wartości obserwowanych (uzyskanych w ankiecie) z wartościami oczekiwanymi (najbardziej prawdopodobnymi, gdyby nie było zależności pomiędzy zmiennymi). W przypadku uzyskania dużych różnic pomiędzy wartościami obserwowanymi a oczekiwanymi można wnioskować, że zachodzi statystycznie istotna relacja pomiędzy analizowanymi zmiennymi.

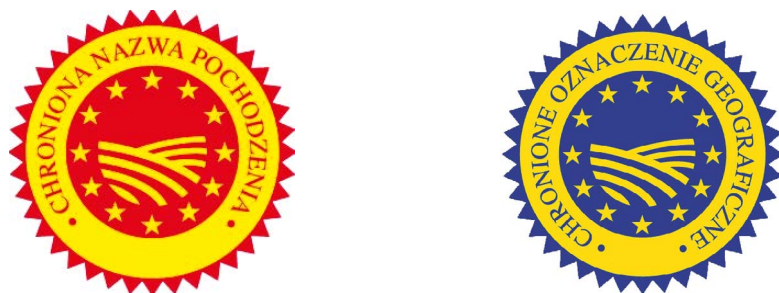
Osobną kwestią w badaniach ankietowych jest liczebność próby. Generalnie można stwierdzić, że im więcej reprezentatywnych respondentów bierze udział w ankiecie, tym uzyskuje się statystycznie pewniejsze wyniki. Więcej o tym aspekcie badań ankietowych znaleźć można w pracy Słowińskiej (2019). Zakładając reprezentatywność i odpowiednią liczbę ankietowanych (Gadzała i Lesiów, 2019), uzyskane w niniejszej pracy wyniki dotyczące europejskich serów regionalnych poddano testowi niezależności chi-kwadrat z wykorzystaniem programu *Statistica 13.3*.

Statystyka chi-kwadrat Pearsona jest podstawą najbardziej rozpowszechnionego testu istotności dla zmiennych jakościowych (skategoryzowanych). Miara ta oparta jest na możliwości obliczenia liczebności oczekiwanych w tabeli dwudzielczej, to znaczy liczebności, jakich oczekiwano by, gdyby nie istniała zależność między zmiennymi (Stat-Soft, 2022).

W pakiecie statystycznym *Statistica*, oprócz możliwości obliczenia klasycznych parametrów testu chi-kwadrat, istnieje również możliwość obliczenia chi-kwadrat największej wiarygodności (NW), gdzie testuje się tę samą hipotezę co statystyka chi-kwadrat Pearsona, jednak sposób jej obliczania oparty jest na teorii największej wiarygodności. W praktyce statystyka chi-kwadrat NW jest pod względem warto-

ści zbliżona do statystyki chi-kwadrat Pearsona (Bishop, Fienberg i Holland, 1975; Fienberg, 1977). Z tego też względu możliwe jest – z wykorzystaniem programu *Statistica* – obliczenie m.in. tych dwóch rodzajów statystyk. W praktyce interpretacja i wielkości obu statystyk chi-kwadrat są zasadniczo identyczne. Oba testy służą do oceny, czy oczekiwane licznosci komórek, zgodnie z odpowiednim modelem, są istotnie różne od licznosci obserwowanych. Jeśli tak, dany model dla tabeli jest odrzucany (Stat-Soft, 2022) i przyjmowana jest hipoteza alternatywna, tzn. potwierdzająca istnienie zależności pomiędzy badanymi zmiennymi.

Metodologia testu chi-kwadrat z wykorzystaniem programu *Statistica* została przedstawiona w niniejszej pracy na przykładzie danych ankietowych dotyczących m.in. znajomości, kupowania czy przygotowywania potraw z wykorzystaniem wybranych europejskich serów regionalnych. Badania ankietowe przeprowadzono w 2016 roku w województwie dolnośląskim na grupie 101 osób, w większości nie starszych niż 35 lat. Ankieta składała się z 15 pytań zamkniętych z możliwością wyboru jednej lub kilku odpowiedzi. Autorzy ankiety dali możliwość odniesienia się bezpośrednio do wskazanych trzynastu europejskich serów regionalnych mających unijny certyfikat (Oziębłowski i Salejda, 2021) Chroniona Nazwa Pochodzenia (10 serów) lub Chronione Oznaczenie Geograficzne (3 sery). Znaki graficzne produktów posiadających unijny certyfikat żywności regionalnej przedstawiono na rys. 1.



Rys. 1. Znaki graficzne produktów mających unijny certyfikat żywności regionalnej: Chroniona Nazwa Pochodzenia oraz Chronione Oznaczenie Geograficzne

Fig. 1. Symbols of products authorised under the EU Regional Food Certificate: Protected Designation of Origin and Protected Geographical Indication

Źródło/Source: serwis polskasmakuje.pl.

Poniżej krótko scharakteryzowano 13 serów wskazanych w ankiecie, w tym 10 o najsilniejszym związku pomiędzy jego jakością a regionem wytworzenia, tzn. mającym certyfikat Chroniona Nazwa Pochodzenia (ChNP/PDO). Wszystkie etapy wytwórcze serów mających Chronioną Nazwę Pochodzenia muszą odbywać się na wskazanym w specyfikacji obszarze.

Cabrales – ser z przerostem niebieskiej pleśni, wytwarzany z mleka krowiego, owczego, koziego lub mieszanego w Asturii i w Hiszpanii (rys. 2).



Rys. 2. Ser Cabrales

Fig. 2. Cabrales cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Grana Padano – włoski ser, produkowany w okolicach miasta Brescia z mleka krowiego, zaliczany do serów dojrzewających, twardych, podpuszczkowych (rys. 3).



Rys. 3. Ser Grana Padano

Fig. 3. Grana Padano cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Allgäuer Emmentaler – niemiecki ser dojrzewający wytwarzany w okolicach Lindau nad Jeziorem Bodeńskim (rys. 4).



Rys. 4. Ser Allgäuer Emmentaler

Fig. 4. Allgäuer Emmentaler cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Camembert de Normandie – miękki ser podpuszczkowy. Nazwa sera pochodzi od wsi o tej samej nazwie w departamencie Orne, w Normandii, gdzie w 1791 r. opracowano ścisłą recepturę produkcji tego sera. Produkowany jest z niepasteryzowanego mleka krowiego (rys. 5).



Rys. 5. Ser Camembert de Normandie

Fig. 5. Camembert de Normandie cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Feta – grecki ser produkowany z mieszanki pasteryzowanego mleka owczego z mlekiem kozim (to drugie stanowi maksymalnie 30% mieszanki) lub tylko z mleka

owczego. Wytwarzany jest na terenie Grecji kontynentalnej oraz w departamencie Lesbos (rys. 6).



Rys. 6. Ser Feta

Fig. 6. Feta cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Bryndza Podhalańska – polski miękki ser podpuszczkowy z mleka owczego i krowiego wyrabiany na Podhalu (rys. 7).



Rys. 7. Ser Bryndza Podhalańska

Fig. 7. Bryndza Podhalańska cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Roquefort – pikantny ser owczy, niebieskawy, żyłkowany, wyrabiany we francuskim miasteczku Roquefort-sur-Soulzon (rys. 8).



Rys. 8. Ser Roquefort

Fig. 8. Roquefort cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Oscypek – polski dojrzewający, wędzony ser, przygotowywany z mleka owczego i krowiego, wyrabiany na Podhalu (rys. 9).



Rys. 9. Ser Oscypek

Fig. 9. Oscypek cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Parmigiano Reggiano – włoski ser dojrzewający z mleka krowiego wytwarzany w regionie Parmy (rys. 10).



Rys. 10. Ser Parmigiano Reggiano

Fig. 10. Parmigiano Reggiano cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Gorgonzola – podpuszczkowy, miękki, dojrzewający, niebieski ser pleśniowy pochodzący z północnych Włoch, z regionu Lombardia. Jego nazwa pochodzi od miejscowości Gorgonzola (rys. 11).



Rys. 11. Ser Gorgonzola

Fig. 11. Gorgonzola cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Pozostałe wskazane w ankiecie trzy sery miały certyfikat Chronione Oznaczenie Geograficzne. Znak ten mogą otrzymać produkty, których jeden z etapów wytwórczych, tj. produkcja, przetwarzanie lub przygotowanie, odbywa się na wskazanym obszarze geograficznym. Związek pomiędzy jakością tych serów a miejscem wytworzenia jest nieco słabszy w porównaniu z wcześniej wymienionymi dziesięcioma produktami. Poniżej przedstawiono krótką charakterystykę trzech serów z certyfikatem Chronione Oznaczenie Geograficzne wymienionych w ankiecie.

Tomme de Savoie – odmiana sera Tomme z Sabaudii we francuskich Alpach. Jest to łagodny, półtwardy ser z mleka krowiego o beżowym wnętrzu i grubej brązowo-szarej skórce (rys. 12).



Rys. 12. Ser Tomme de Savoie

Fig. 12. Tomme de Savoie cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Edam Holland – gatunek sera półtwardego, podpuszczkowego, dojrzewającego, produkowanego z mleka krowiego lub – rzadziej – koziego. Pochodzi z okolic holenderskiego małego miasta Edam położonego na północ od Amsterdamu (rys. 13).



Rys. 13. Ser Edam Holland

Fig. 13. Edam Holland cheese

Źródło/Source: <https://world.openfoodfacts.org>.

Gouda Holland – gatunek sera półtwardego, podpuszczkowego, dojrzewającego, produkowanego z mleka krowiego. Pochodzi z okolic holenderskiego miasta Gouda (rys. 14).



Rys. 14. Ser Gouda Holland

Fig. 14. Gouda Holland cheese

Źródło/Source: <https://agriculture.ec.europa.eu>.

Celem pracy było praktyczne wskazanie metodologii testu niezależności chi-kwadrat z wykorzystaniem pakietu programu *Statistica* na przykładzie wyników z ankiety dotyczącej europejskich serów regionalnych.

2. Przygotowanie danych statystycznych

Dokładne przedstawienie podstaw teoretycznych testu niezależności chi-kwadrat Pearsona można znaleźć w wielu publikacjach (Ostasiewicz, Rusnak i Siedlecka, 2006; Słowińska, 2019; Stanisław, 2006). Należy jednak zwrócić uwagę, że w tym nieparametrycznym teście rozkład badanej zbiorowości nie musi być rozkładem normalnym. Ponadto testy nieparametryczne dotyczą rozkładu zbiorowości według określonej cechy, a nie jej parametrów, takich jak średnia czy wariancja (Rajs, 2006). Z tego też względu analizować można, oprócz cech ilościowych, również cechy jakościowe. Można więc ocenić niezależność stochastyczną dwóch cech jakościowych lub dwóch cech ilościowych albo cechy ilościową i jakościową opisujące zbiorowość generalną (Ostasiewicz i in., 2006). W tym celu weryfikuje się odpowiednio sformułowaną hipotezę nieparametryczną za pomocą testu niezależności chi-kwadrat. Przyjmuje się hipotezę zerową orzekającą, że badane cechy są niezależne. Do zweryfikowania tej hipotezy należy dysponować odpowiednio dużą próbą (tzn. ponad 30 obserwacji/respondentów według Ostasiewicz i in., 2006). Wyniki przedstawia się w postaci tzw. tablicy dwudzielnej (dwudzielczej) mającej zastosowanie

tylko w przypadku analizy dwóch cech zmiennych, czyli przy teście niezależności chi-kwadrat. Do zbadania łącznie zależności trzech lub więcej zmiennych konieczne jest zastosowanie wielowymiarowych tablic kontyngencji oraz uogólnionej postaci testu niezależności chi-kwadrat (Słowińska, 2019). Jeśli na podstawie stosownych obliczeń odrzuci się hipotezę zerową, to przyjmuje się hipotezę alternatywną wskazującą na istnienie zależności pomiędzy badanymi cechami (zmiennymi).

Test niezależności (Pearsona) chi-kwadrat jest idealną techniką do analizy zależności pomiędzy dwiema zmiennymi o niezbyt znacznej liczbie kategorii czy grup. Przykładowo jedną zmienną może być płeć respondentów (2), drugą zaś częstość zakupów serów regionalnych (np. 4 różne częstości: codziennie, kilka razy w tygodniu, kilka razy w miesiącu, okazjonalnie). Gdyby druga zmienna miała np. kilkanaście możliwych wskazań, to w takich przypadkach zaleca się korzystanie z innego rodzaju testów, np. z testu U Manna-Whitneya.

Dane do analizy statystycznej przygotowano, wykorzystując wyniki ankiety konsumenckiej przeprowadzonej w 2016 roku, w której uczestniczyło 101 osób. Wyniki dotyczące każdego ankietowanego znajdowały się w 101 wierszach. Dane dotyczące poszczególnych wyróżników (zmiennych) były zatem zgrupowane w kolumnach. Ankieta składała się z kilku zasadniczych części, które ze względu na analizę chi-kwadrat (po zebraniu wszystkich odpowiedzi) podzielono na trzy kategorie – „zmiennie jakościowe (kategoryzujące)”, „wielokrotne odpowiedzi” oraz „wielokrotne dychotomie”.

2.1. Zmienne jakościowe (kategoryzujące)

Pierwszą kategorię uzyskanych odpowiedzi ankietowych stanowiły „zmiennie jakościowe (kategoryzujące)” grupowane rozłącznie. W analizowanej ankiecie były to następujące wyróżniki:

- jak często kupowany jest ser (jedna z 4 możliwych odpowiedzi: codziennie, kilka razy w tygodniu, kilka razy w miesiącu, okazjonalnie);
- gdzie głównie kupowany jest ser (jedna z 4 możliwych odpowiedzi: supermarket, hipermarket, targ, bezpośrednio od rolnika);
- jaki jest główny powód, dla którego ser jest kupowany (jedna z 6 możliwych odpowiedzi: smak, atrakcyjna cena, atrakcyjny wygląd opakowania, renoma/zaufanie/marka, wartość odżywcza, reklama);
- jaka jest główna postać kupowanego sera (jedna z 3 możliwych odpowiedzi: w kostce, w plastrach, tarty);
- jaka jest gramatura kupowanego sera (jedna z 4 możliwych odpowiedzi: do 100 g, 100-250 g, 250-500 g, powyżej 500 g);
- jaki jest główny rodzaj kupowanych serów (jedna z 4 możliwych odpowiedzi: dojrzewający [tzw. żółty], twarogowy [tzw. biały], pleśniowy [tzw. topiony];
- czy ankietowany ma jakiś ulubiony ser (możliwe odpowiedzi to: nie lub tak, jeśli tak, to jaki).

- jaki jest stosunek do nowych lub nieznanymi wcześniej serów (jedna z 4 możliwych odpowiedzi: niechętnie kupuję nowości; kupuję nowy produkt dopiero po wypróbowaniu go przez rodzinę, znajomych, bądź po degustacji; kupuję nowy produkt po dłuższym namyśle; lubię próbować nowe produkty);
- czy ankietowany zna regionalne sery (jedna z dwóch odpowiedzi: tak lub nie);
- czy ankietowany akceptuje większy wydatek za sery regionalne (jedna z dwóch odpowiedzi: tak, nie).

Wartość testu chi-kwadrat zależy od liczby obserwacji i liczby komórek w tabeli. Jeśli jakies liczebności teoretyczne będą poniżej 5, to wartość testu może być wysoce nieprecyzyjna. Jeśli z kolei liczebność próbki jest duża, to stosunkowo małe odchylenia częstości względnych od oczekiwań mogą okazać się istotne (Stanisz, 2006).

Po przeprowadzeniu wstępnej analizy chi-kwadrat zauważono, że liczebności w wielu przypadkach nie były większe od 5, dlatego zdecydowano się połączyć niektóre odpowiedzi w jedną kategorię. W pytaniu pierwszym połączono w jedną kategorię odpowiedzi „codziennie” i „kilka razy w tygodniu”. W drugim pytaniu połączono odpowiedzi „targ” i „bezpośrednio od rolnika”. W trzecim pytaniu połączono odpowiedzi „atrakcyjny wygląd opakowania” i „reklama”. W czwartym pytaniu połączono odpowiedzi „w plastrach” i „tarty”. W piątym pytaniu o gramaturę kupowanych serów cztery możliwe odpowiedzi połączono w dwie nowe kategorie „do 250 g” i „powyżej 250 g”. W szóstym pytaniu połączono ze sobą trzy kategorie „twarogowy”, „pleśniowy” i „topiony”. W siódmym pytaniu ze względu na bardzo rozproszone odpowiedzi wskazujące ulubiony ser zdecydowano się na pozostawienie dwóch kategorii: „nie” (respondent nie ma ulubionego sera) i „tak” (respondent ma ulubiony ser bez analizowania, jaki to ser). W ósmym pytaniu połączono w jedną kategorię dwie odpowiedzi: „niechętnie kupuję nowości” i „kupuję nowy produkt dopiero po wypróbowaniu go przez rodzinę”. Kategorie w pytaniach dziewiątym i dziesiątym pozostawiono bez zmian (odpowiedzi „tak” lub „nie”).

Przygotowując dane do analizy statystycznej, tekst odpowiedzi respondentów często kodowano w formie skróconej, np. odpowiedź „codziennie lub kilka razy w tygodniu” była kodowana jako „kilka razy w tygodniu”. Nie miało to oczywiście wpływu na sam wynik obliczeń.

Na rysunku 15 przedstawiono początkową tabelę jako przykładowe formatowanie danych w programie *Statistica* ukazujące 4 wybrane wyróżniki (po powyżej wskazanym łączeniu poszczególnych kategorii) dla pierwszych 12 respondentów (ze 101) w ramach kategorii danych „zmiennie jakościowe (kategoryzujące)”.

Po zebraniu wszystkich ankiet dokonano dla każdego respondenta osobno zliczenia, ile zna serów regionalnych, ile ich kupuje, ile serów byłby skłonny kupować po wyższej cenie oraz ile stosuje dodatków z serów przy przyrządzaniu wybranych dań lub potraw.

1 jak-często-kupuje	2 gdzie-kupuje	3 główna-cecha-sera	4 forma-sera
kilka razy w tygodniu	hipermarket	smak	w kostce/ tarty
kilka razy w miesiącu	supermarket	smak	w kostce/ tarty
kilka razy w tygodniu	hipermarket	smak	w plastrach
kilka razy w miesiącu	supermarket	smak	w plastrach
okazjonalnie	supermarket	atrakcyjną cenę	w kostce/ tarty
okazjonalnie	supermarket	wartość odżywczą	w kostce/ tarty
kilka razy w miesiącu	supermarket	renomu/ zaufanie/ marka	w plastrach
kilka razy w miesiącu	supermarket	smak	w kostce/ tarty
kilka razy w tygodniu	supermarket	atrakcyjną cenę	w plastrach
kilka razy w tygodniu	supermarket	smak	w kostce/ tarty
kilka razy w miesiącu	supermarket	smak	w kostce/ tarty
kilka razy w miesiącu	hipermarket	atrakcyjną cenę	w kostce/ tarty

Rys. 15. Przykładowe formatowanie danych w programie *Statistica* ukazujące 4 wybrane wyróżniki dla pierwszych 12 respondentów (ze 101) w ramach kategorii danych „zmiennie jakościowe (kategoryzujące)”

Fig. 15. Example of data formatting in *Statistica* program with 4 selected distinguishing features for the first 12 respondents (out of 101) within the data category “Qualitative Variables (Categorization)”

Źródło/Source: opracowanie własne/own study.

Do zmiennych jakościowych zaklasyfikowano również pytania dotyczące tzw. metryczki, czyli związane z osobą wypełniającą ankietę. Pytań takich było cztery i dotyczyły one:

- płci (uzyskano dwa wskazania: kobieta i mężczyzna);
- wieku (cztery kategorie: 18-25 lat, 25-35 lat, 35-50 lat, powyżej 50 lat);
- wykształcenia (cztery kategorie: podstawowe, zawodowe, średnie, wyższe);
- miejscowości zamieszkania (cztery kategorie: wieś, miasto do 50 tys. mieszkańców, miasto od 50 do 100 tys. mieszkańców, miasto powyżej 100 tys. mieszkańców).

W ramach wstępnego testu chi-kwadrat w odniesieniu do 4 zmiennych z metryczki wykazano, że w kilku przypadkach licznosci oczekiwane były mniejsze niż 5. Z tego też powodu dokonano połączenia kilku kategorii. W przypadku wieku podzielono respondentów na dwie kategorie wiekowe: „18-25 lat” i „powyżej 25 lat”. Wykształcenie ostatecznie również otrzymało dwie kategorie: „wyższe” i „nie wyższe” (tzn. podstawowe, zawodowe i średnie). Miejsce zamieszkania otrzymało również dwie kategorie: „wieś lub miasto do 100 tys. mieszkańców” oraz „miasto powyżej 100 tys. mieszkańców”.

Na rysunku 16 ukazano przykładowe formatowanie danych w programie *Statistica* ukazujące 4 „zmiennie jakościowe (kategoryzujące)” po zmianach związane z osobą ankietowaną (tzw. metryczka) dla pierwszych 12 respondentów (ze 101).

60 płeć	61 wiek	62 wykształcenie	63 miejscowość
mężczyzna	18-25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	18-25 lat	Wyższe	wieś lub miasto do 100 tys. mieszkańców
kobieta	18-25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	18-25 lat	Wyższe	wieś lub miasto do 100 tys. mieszkańców
mężczyzna	18-25 lat	nie wyższe	wieś lub miasto do 100 tys. mieszkańców
kobieta	18-25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	powyżej 25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	18-25 lat	nie wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	18-25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	18-25 lat	nie wyższe	wieś lub miasto do 100 tys. mieszkańców
kobieta	powyżej 25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców
kobieta	18-25 lat	Wyższe	miasto powyżej 100 tys. mieszkańców

Rys. 16. Przykładowe formatowanie danych w programie *Statistica* ukazujące 4 „zmiennne jakościowe (kategoryzujące)” związane z osobą ankietowaną (tzw. metryczka) dla pierwszych 12 respondentów (ze 101)

Fig. 16. Example of formatting data in *Statistica* program with 4 “qualitative variables (categorization variables)” for the first 12 respondents (out of 101)

Źródło/Source: opracowanie własne/own study.

2.2. Wielokrotne odpowiedzi

Drugą kategorię uzyskanych odpowiedzi ankietowych stanowiły „wielokrotne odpowiedzi” wskazujące obiekty (sery) z grupy wymienionych w ankiecie.

Respondenci mogli wskazać dowolną liczbę z 13 znanych im następujących europejskich serów regionalnych: Gorgonzola, Parmigiano Reggiano, Oscypek, Roquefort, Bryndza Podhalańska, Feta, Camembert de Normandie, Gouda Holland, Allgäuer Emmental, Edam Holland, Tomme de Savoie, Cabrales, Grana Padano.

W innym pytaniu respondenci mogli wskazać dowolną liczbę serów (z tej samej grupy 13), które kupują.

W kolejnym pytaniu można było wskazać z tej samej grupy 13 serów te, które byłyby kupowane przez respondentów nawet wtedy, kiedy ich cena byłaby wyższa od dotychczasowej (rys. 17).

Respondenci mogli wskazać także sposoby wykorzystania serów przy przyrządzaniu dań i potraw spośród 7 takich jak: surówki, sałatki, sosy, gorące przekąski, desery, zupy, drugie dania z ryb/drobieu/mięsa.

11 ile-zna	12 ile-kupuje	13 ile-kupuje-drożej	14 ile-dodatków
3	3	2	1
7	4	7	6
8	5	6	4
10	4	5	6
9	5	2	2
3	2	1	2
6	4	4	2
3	3	2	4
3	2	1	2
10	5	5	3
4	2	1	1
4	3	3	2

Rys. 17. Przykładowe formatowanie danych w programie *Statistica* ukazujące odpowiedzi ankietowanych na 4 pytania, gdzie można było wskazać dowolną liczbę spośród wszystkich 13 serów. Ukazane wyniki (dla pierwszych 12 respondentów ze 101) zaliczone zostały do kategorii danych „wielokrotne odpowiedzi”

Fig. 17. Example of data formatting in *Statistica* program, which shows respondents' answers to 4 questions where any number of 13 cheeses could be specified. The results (for the first 12 of the 101 respondents) were classified in the data category “multiple responses”

Źródło/Source: opracowanie własne/own study.

2.3. Wielokrotne dychotomie

Trzecią kategorię uzyskanych odpowiedzi ankietowych stanowiły „wielokrotne dychotomie” wskazujące np. na znajomość określonego sera, gdzie 0 oznaczało brak znajomości, zaś 1 – znajomość określonego sera. Sposób takiej prezentacji wyników przygotowano po zebraniu wszystkich ankiet w odniesieniu do wybranych pytań, na które można było udzielić wielokrotnych odpowiedzi. Dotyczyło to pytań nie tylko o znajomość serów, ale również trzech innych – odnoszących się do zakupu określonych serów, możliwości ich zakupu po wyższych cenach, jak również sposobów wykorzystania serów przy przygotowywaniu dań i posiłków. Kodowanie takich odpowiedzi dotyczących osobno każdego sera lub dodatku przeprowadzono w postaci dwóch liczb: 0 (np. brak znajomości sera, niekupowanie sera, niestosowanie określonego dodatku) lub 1 (np. znajomość sera, stosowanie wymienionego dodatku). Przykład takiego kodowania przedstawiono na rys. 18.

12 zna- Cabr ales	13 zna- Gran a-Pa dano	14 zna- Tom me-d e-Sa voie	15 zna- Eda m-Ho lland	16 zna- Allga uer- Emm ental er	17 zna- Goud a-Holl and	18 zna- Cam embe rt-de- Norm andie	19 zna- Feta	20 zna- Bryn dza- Podh alans ka	21 zna- Roq uefo rt	22 zna- Osc ypek	23 zna- Parm igian o-Re ggian o	24 zna- Gorg onzol a
0	0	0	0	0	0	0	0	0	0	1	1	1
0	0	0	0	1	1	0	1	0	1	1	1	1
0	0	0	0	1	1	0	1	1	1	1	1	1
1	0	0	0	1	1	0	1	1	1	1	1	1
0	1	0	1	1	1	1	1	0	0	1	1	1
0	0	0	0	0	0	0	1	0	0	1	0	1
0	0	0	0	0	1	1	1	1	0	1	0	1
0	0	0	0	0	0	0	1	0	0	1	0	1
0	0	0	0	0	0	0	1	1	0	1	0	0
0	1	0	1	1	1	1	1	1	1	1	1	1
0	0	0	1	0	1	0	1	0	0	0	1	0
0	0	0	0	0	0	0	1	1	0	1	0	1

Rys. 18. Przykładowe formatowanie danych w programie *Statistica* ukazujące deklarowaną znajomość wymienionych 13 serów w ramach kategorii danych „wielokrotne dychotomie” dla pierwszych 12 respondentów (ze 101)

Fig. 18. Example of the formatting of data in the *Statistica* program to represent the stated familiarity with the 13 cheeses in the data category “multiple dichotomies” for the first 12 respondents (out of 101)

Źródło/Source: opracowanie własne/own study.

3. Analiza statystyczna chi-kwadrat

Celem przeprowadzenia analizy statystycznej testu niezależności chi-kwadrat Pearsona oraz NW za pomocą programu *Statistica* wybrano następującą analizę: zakładka Statistica → Statystyki podstawowe i tabele → Tabele wielodzzielcze → ok → Określ tabelę [w pierwszej liście za pomocą klawisza ctrl wskazano np. pierwsze 10 zmiennych, zaś w drugiej liście 4 zmienne związane z metryczką] → ok → ok → zakładka Opcje → w kolumnie „Oblicz tabele” zaznaczono: Podświetl licznosci > 4, Licznosci oczekiwane, Licznosci resztowe [ewentualnie: Procenty z całości, Procenty w wierszach, Procenty w kolumnach], zaś w kolumnie „Statystyki dla tabel dwudzzielczych” zaznaczono: Chi-kwadrat Pearsona i NW [ewentualnie: Dokł. Fishera, Yatesa, McNemara (2x2), Fi (tabela 2x2), V i C Craméra], Współczynnik korelacji rang Spearmana → zakładka Więcej → Dokładne tabele dwudzzielcze → wszystkie → ok.

3.1. Analiza zmiennych jakościowych (kategoryzujących)

Po przeprowadzonej analizie możliwe było wskazanie określonych tabel, które zostały wygenerowane w odpowiednich raportach programu. Na rysunku 19 ukazano przykładowy raport odnoszący się do licznosci obserwowanych, licznosci oczeki-

wanych oraz statystyk chi-kwadrat (i ewentualnie innych – wcześniej wskazanych) dotyczących zależności pomiędzy częstością zakupów serów regionalnych a płcią respondentów.

	pleć kobieta	pleć mężczyzna	Wiersz Razem
jak-często-kupuje			
kilka razy w tygodniu	37	11	48
kilka razy w miesiącu	25	10	35
okazjonalnie	9	9	18
Ogół	71	30	101

Licznosci obserwowane/ Observed counts

	pleć kobieta	pleć mężczyzna	Wiersz Razem
jak-często-kupuje			
kilka razy w tygodniu	33,74257	14,25743	48,0000
kilka razy w miesiącu	24,60396	10,39604	35,0000
okazjonalnie	12,65347	5,34653	18,0000
Ogół	71,00000	30,00000	101,0000

Licznosci oczekiwane/ Expected counts

statystyka	Chi-kwadr.	df	p
Chi² Pearsona	4,631566	df=2	p=,09869
Chi ² NW	4,376114	df=2	p=,11213
Fi	,2141427		
Wsp. kontyngencji	,2093954		
V Craméra	,2141427		
R rang Spearmana	,1875167	t=1,8995	p=,06041

Statystyka chi-kwadrat Pearsona oraz chi-kwadrat największej wiarygodności (NW) – wartość, stopień swobody (*df*) oraz poziom istotności (*p*) jak również wartości innych współczynników (wcześniej zaznaczonych), w tym R rang Spearmana/ Pearson's chi-squared statistic and the chi-square of highest reliability (NW) – value, degree of freedom (*df*) and significance level (*p*) as well as values of other coefficients (previously marked) including Spearman rank R

Rys. 19. Przykładowe raporty wygenerowane przez program *Statistica* dotyczące licznosci obserwowanych, licznosci oczekiwanych oraz statystyk chi-kwadrat i innych dotyczących zależności pomiędzy częstością zakupów serów regionalnych a płcią respondentów

Fig. 19. Examples of reports produced by *Statistica* program on the observed and expected counts, chi-square statistics and others on the relationship between the frequency of purchases of regional cheese and the gender of respondents

Źródło/Source: opracowanie własne/own study.

W tabeli licznosci obserwowanych podana jest liczba osób (w zależności od płci) kupujących sery regionalne z różną częstością w określonym czasie. W tabeli licznosci oczekiwanych podane są wartości teoretyczne ukazujące, ile osób powinno

kupować sery, gdyby absolutnie nie było żadnej zależności między płcią a częstością dokonywanych zakupów. Wartość taka dla kobiet kupujących kilka razy w tygodniu sery regionalne wyniosła 33,74, co po zaokrągleniu wskazuje na 34 kobiety (według ankiet było ich 37). Im większe różnice pomiędzy licznosciami obserwowanymi a oczekiwanymi, tym większa jest szansa, że test chi-kwadrat wskaże na hipotezę alternatywną, tzn. wykaże istnienie statystycznie zależności pomiędzy badanymi wyróżnikami. W podanym przykładzie różnic takich nie stwierdzono. Wartości chi-kwadrat wynosiły 4,63 (Pearsona) oraz 4,37 (NW – największej wiarygodności) przy stopniach swobody (df) wynoszących dwa. Wartości poziomu istotności (p) były większe od granicznej wartości 0,05 i wynosiły 0,0987 (chi-kwadrat Pearsona) oraz 0,1121 (chi-kwadrat największej wiarygodności), co wskazuje na brak zależności pomiędzy częstością zakupów serów regionalnych a płcią respondentów. W zależności od potrzeb możliwa jest analiza innych współczynników wyliczonych w programie i umieszczonych w raporcie, jak np. wartość R rang Spearmana omówiona w dalszej części artykułu.

Oprócz powyższej analizy chi-kwadrat badającej istnienie zależności (lub jej brak) pomiędzy częstością zakupów serów regionalnych a płcią przeprowadzono inne podobne analizy pomiędzy 4 wyróżnikami metryczki (płeć, wiek, wykształcenie, miejsce zamieszkania) a 10 pytaniami ankiety, co łącznie dało $4 \times 10 = 40$ analiz zależności. Pytania dotyczyły: częstości zakupów (3 wybrane do analizy), miejsca zakupów (3), głównej cechy serów (5), formy kupowanych serów (2), ich gramatury (2), rodzaju serów (3), istnienia ulubionego sera (2), relacji ankietowanych do nowych serów (3), znajomości serów regionalnych (2) oraz akceptacji kupowania ulubionych serów po wyższych cenach (2). Po przeprowadzeniu wszystkich tych testów chi-kwadrat w żadnym przypadku nie wykazano zależności pomiędzy analizowanymi pytaniami a którymś z czterech wyróżników metryczki.

3.2. Analiza wielokrotnych odpowiedzi

W następnej kolejności zbadano, czy istnieją zależności pomiędzy liczbą: znanych serów, kupowanych, potencjalnie kupowanych serów po wyższej cenie oraz liczbą potraw z wykorzystaniem serów a 4 parametrami metryczki. W sumie przeanalizowano w tej serii 16 zależności (4×4) i wykazano 4 zależności, które były istotne. Jedną z nich była zależność pomiędzy liczbą znanych serów regionalnych a wykształceniem respondentów. Na rysunku 20 przedstawiono 4 raporty uzyskane w ramach tej analizy. Respondenci deklarowali znajomość od 1 do 10 serów regionalnych (lub 12). Mogli wskazać dowolną liczbę znanych im serów z 13 wymienionych w ankiecie. Nikt nie wskazał znajomości 11 serów.

Zależność pomiędzy liczbą znanych respondentom serów regionalnych a ich wykształceniem była istotna (wartość $p = 0,0361$ dla chi-kwadrat Pearsona oraz $p = 0,0294$ dla chi-kwadrat NW). Analizując licznosci resztowe (tj. różnicę pomiędzy licznosciami obserwowanymi a oczekiwanymi), można stwierdzić, że osoby

ile-zna	wykształcenie Wyższe	wykształcenie nie wyższe	Wiersz Razem
1	4	4	8
2	3	5	8
3	13	4	17
4	13	2	15
5	7	1	8
6	14	2	16
7	8	4	12
8	5	1	6
9	4	0	4
10	6	0	6
12	0	1	1
Ogół	77	24	101

Liczności obserwowane/ Observed counts

ile-zna	wykształcenie Wyższe	wykształcenie nie wyższe	Wiersz Razem
1	6,09901	1,90099	8,0000
2	6,09901	1,90099	8,0000
3	12,96040	4,03960	17,0000
4	11,43564	3,56436	15,0000
5	6,09901	1,90099	8,0000
6	12,19802	3,80198	16,0000
7	9,14851	2,85149	12,0000
8	4,57426	1,42574	6,0000
9	3,04950	0,95050	4,0000
10	4,57426	1,42574	6,0000
12	0,76238	0,23762	1,0000
Ogół	77,00000	24,00000	101,0000

Liczności oczekiwane/ Expected counts

ile-zna	wykształcenie Wyższe	wykształcenie nie wyższe	Wiersz Razem
1	-2,09901	2,09901	0,00
2	-3,09901	3,09901	0,00
3	0,03960	-0,03960	0,00
4	1,56436	-1,56436	0,00
5	0,90099	-0,90099	0,00
6	1,80198	-1,80198	0,00
7	-1,14851	1,14851	0,00
8	0,42574	-0,42574	0,00
9	0,95050	-0,95050	0,00
10	1,42574	-1,42574	0,00
12	-0,76238	0,76238	0,00
Ogół	0,00000	0,00000	0,00

Liczności resztowe (tj. obserwowane – oczekiwane)/ Residual counts (i.e. observed – expected)

statystyka	Chi-kwadr.	df	p
Chi ² Pearsona	19,34697	df=10	p=,03607
Chi ² NW	19,98787	df=10	p=,02937
F _i	,4376690		
Wsp. kontyngencji	,4009488		
V Craméra	,4376690		
R rang Spearmana	-,223157	t=-2,278	p=,02489

Statystyka chi-kwadrat Pearsona oraz chi-kwadrat największej wiarygodności (NW) – wartość, stopień swobody (*df*) oraz poziom istotności (*p*) jak również wartości innych współczynników/ Pearson's chi-square and the chi-square of highest reliability (NW) – value, degree of freedom (*df*) and significance level (*p*) as well as values of other coefficients

Rys. 20. Raporty wygenerowane przez program *Statistica* odnoszące się do obserwowanych, oczekiwanych i resztowych licznosci oraz statystyk chi-kwadrat i innych dotyczących zależności pomiędzy liczbą serów regionalnych znanych respondentom a ich wykształceniem

Fig. 20. *Statistica* program reports on observed, expected and residual counts as well as chi-square statistics and other statistics on the relationship between the quantity of regional cheeses known to respondents and their formation.

Źródło/Source: opracowanie własne/own study.

niemające wykształcenia wyższego najczęściej wskazywały na znajomość jednego lub dwóch serów regionalnych. Osoby z wykształceniem wyższym przeważnie deklarowały znajomość 4, 5 lub 6 serów. Przy innych deklarowanych licznosciach zależność od wykształcenia była zmienna (np. przy 7 na korzyść osób niemających wykształcenia wyższego, ale przy 10 zauważono odwrotną zależność).

Na rysunku 21 przedstawiono licznosci resztowe oraz poziom istotności dla testów chi-kwadrat zarówno Pearsona, jak i największej wiarygodności odnoszące się do zależności pomiędzy liczbą kupowanych różnych rodzajów serów regionalnych a płcią, wiekiem i wykształceniem respondentów. Wartość poziomu istotności dla testów chi-kwadrat największej wiarygodności we wszystkich trzech przypadkach

była mniejsza od 0,05, dla chi-kwadrat Pearsona zaś była mniejsza w odniesieniu do płci i wykształcenia.

W pozostałych 12 przypadkach tej serii pytań nie stwierdzono istotnej zależności pomiędzy analizowanym czynnikiem a jednym z parametrów metryczki.

ile-kupuje	pleć kobieta	pleć mężczyzna	Wiersz Razem
1	1,01980	-1,01980	0,00
2	-4,08911	4,08911	0,00
3	-1,95050	1,95050	0,00
4	2,97030	-2,97030	0,00
5	2,45545	-2,45545	0,00
6	0,29703	-0,29703	0,00
7	-0,70297	0,70297	0,00
Ogół	0,00000	0,00000	0,00

Licznosci resztowe dla zależności „ilość kupowanych serów – pleć”/ Residual counts for the relationship “number of cheeses purchased – gender”

ile-kupuje	wiek 18-25 lat	wiek powyżej 25 lat	Wiersz Razem
1	0,95050	-0,95050	0,00
2	-3,27723	3,27723	0,00
3	2,37624	-2,37624	0,00
4	-2,42574	2,42574	0,00
5	2,86139	-2,86139	0,00
6	-0,74257	0,74257	0,00
7	0,25743	-0,25743	0,00
Ogół	0,00000	0,00000	0,00

Licznosci resztowe dla zależności „ilość kupowanych serów – wiek”/ Residual counts for the relationship “the number of cheeses purchased – age”

ile-kupuje	wykształcenie Wyższe	wykształcenie nie wyższe	Wiersz Razem
1	2,58416	2,58416	0,00
2	-2,87129	2,87129	0,00
3	2,03960	-2,03960	0,00
4	2,37624	-2,37624	0,00
5	1,56436	-1,56436	0,00
6	0,23762	-0,23762	0,00
7	-0,76238	0,76238	0,00
Ogół	0,00000	0,00000	0,00

Licznosci resztowe dla zależności „ilość kupowanych serów – wykształcenie”/ Residual counts for the relationship “number of cheeses purchased – education”

Zależność ilości kupowanych serów od: Dependence of the amount of cheese purchased on:	Poziom istotności <i>p</i> (chi-kwadrat Pearsona)/ Significance level <i>p</i> (chi-square Pearsona)	Poziom istotności <i>p</i> (chi-kwadrat NW)/ Significance level <i>p</i> (chi-square NW)
płci/ gender	0,0452	0,0132
wieku/ age	0,0438	0,0338
wykształcenia/ education	0,0673	0,0285

Rys. 21. Raporty wygenerowane przez program *Statistica* odnoszące się do resztowych licznosci oraz poziomu istotności chi-kwadrat (Pearsona i NW) dotyczących zależności pomiędzy liczbą serów regionalnych kupowanych przez respondentów a ich płcią, wiekiem i wykształceniem

Fig. 21. *Statistica* program reports on the residual counts and significance of the chi-square (Pearson and NW) on the relationship between the quantity of regional cheeses purchased by respondents and their sex, age and education

Źródło/Source: opracowanie własne/own study.

3.3. Analiza wielokrotnych dychotomii

Trzecią grupą analizowanych danych były wielokrotne dychotomie odnoszące się do pytań o wskazanie spośród 13 wymienionych serów regionalnych tych, które są znane respondentom, tych, które są przez nich kupowane, oraz tych, które byłyby wciąż kupowane pomimo wyższej ceny. Dodatkowo ankietowani w ramach tej serii pytań mieli wskazać dania i potrawy, przy tworzeniu których wykorzystują sery regionalne.

Deklarowana znajomość 13 serów była porównana z 4 wyróżnikami metryczki, co dało w sumie $4 \times 13 = 52$ analizy chi-kwadrat. Ostatecznie uzyskano 8 istotnych zależności. Jedną z nich była udowodniona zależność pomiędzy deklarowaną znajomością sera Edam Holland a miejscem zamieszkania. Na rysunku 22 przedstawiono dla tego przypadku liczności obserwowane, oczekiwane, resztowe, jak również statystykę chi-kwadrat wraz z innymi statystykami uzupełniającymi.

	mięscowość miasto powyżej 100 tys. mieszkańców	mięscowość wieś lub miasto do 100 tys. mieszkańców	Wiersz Razem
zna-Edam-Holland			
0	33	24	57
1	38	6	44
Ogół	71	30	101

Licznosci obserwowane/ Observed counts

	mięscowość miasto powyżej 100 tys. mieszkańców	mięscowość wieś lub miasto do 100 tys. mieszkańców	Wiersz Razem
zna-Edam-Holland			
0	40,06931	16,93069	57,00000
1	30,93069	13,06931	44,00000
Ogół	71,00000	30,00000	101,00000

Licznosci oczekiwane/ Expected counts

	mięscowość miasto powyżej 100 tys. mieszkańców	mięscowość wieś lub miasto do 100 tys. mieszkańców	Wiersz Razem
zna-Edam-Holland			
0	-7,06931	7,06931	0,00
1	7,06931	-7,06931	0,00
Ogół	0,00000	0,00000	0,00

Licznosci resztowe (tj. obserwowane – oczekiwane)/ Residual counts (i.e. observed – expected)

statystyka	Chi-kwadr	df	p
Chi ² Pearsona	9,638527	df=1	p=.00191
Chi ² NW	10,23917	df=1	p=.00137
Chi ² Yatesa	8,323311	df=1	p=.00391
dokł. Fishera, 1-stronny			p=.00157
2-stronny			p=.00210
Chi ² McNemara (A/D)	17,33333	df=1	p=.00003
(B/C)	2,725806	df=1	p=.09874
Fi dla tabel 2 x 2	-.308919		
Korel. tetrachoryczne	-.510620		
Wsp. kontyngencji	,2951564		
R rang Spearmana	-.308919	t=-3,232	p=.00167

Statystyka chi-kwadrat Pearsona oraz chi-kwadrat największej wiarygodności (NW) – wartość, stopień swobody (*df*) oraz poziom istotności (*p*), jak również wartości innych współczynników/ Pearson's chi-square and the chi-square of highest reliability (NW) – value, degree of freedom (*df*) and significance level (*p*) as well as values of other coefficients

Rys. 22. Raporty wygenerowane przez program *Statistica* odnoszące się do obserwowanych, oczekiwanych i resztowych licznosci oraz statystyk chi-kwadrat i innych dotyczących zależności pomiędzy deklarowaną znajomością sera Edam Holland a miejscem zamieszkania respondentów

Fig. 22. *Statistica* program reports on observed, expected and residual counts as well as chi-square statistics and other statistics on the relationship between the reported knowledge of Edam Holland and the respondents' place of residence

Źródło/Source: opracowanie własne/own study.

Na podstawie uzyskanych danych można stwierdzić, że udowodniono zależność pomiędzy deklarowaną znajomością sera Edam Holland a miejscem zamieszkania (poziom istotności $p = 0,0019$ dla testu chi-kwadrat Pearsona oraz $p = 0,0014$ dla testu chi-kwadrat NW). Opierając się na tabeli licznosci resztowych, ustalono, że ser Edam Holland jest istotnie częściej znany wśród mieszkańców miast powyżej 100 tys. osób.

Przeprowadzona w podobny sposób analiza wykazała także inne zależności pomiędzy znajomością podanych serów a parametrami metryczki, co ukazano w kompaktowy sposób w tab. 1.

Tabela 1. Istotne zależności pomiędzy deklarowaną znajomością serów regionalnych a parametrami metryczki respondentów – udowodnione na podstawie testu chi-kwadrat

Table 1. Significant relationships between the declared knowledge of regional cheeses and the parameters of the respondents' specification – proved on the basis of the chi-square test

Deklarowana znajomość sera/ Declared knowledge of cheese	Zależny parametr metryczki/ Dependent parameter of the imprint	Kto zna lub gdzie lepiej znają wskazany ser/ Who knows or where they know better the indicated cheese	Poziom istotności (p) chi-kwadrat Pearsona/ Significance level (p) Pearson's chi-square	Poziom istotności (p) chi-kwadrat NW/ Significance level (p) chi-squared NW
Grana Padano	płeć/ gender	mężczyźni	0,0376	0,0446
Tomme de Savoie	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0280	0,0261
Edam Holland	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0019	0,0014
Allgäuer Emmentaler	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0466	0,0350
Gouda Holland	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0056	0,0054
Camembert de Normandie	płeć/ gender	kobiety/ women	0,0177	0,0169
Camembert de Normandie	wykształcenie/ education	wyższe/ higher	0,0416	0,0403
Parmigiano Reggiano	wykształcenie/ education	wyższe/ higher	0,0152	0,0111

Źródło/Source: opracowanie własne/own study.

Dane z tab. 1 ukazują, że znajomość serów Tomme de Savoie, Edam Holland, Allgäuer Emmentaler oraz Gouda Holland jest istotnie większa wśród mieszkańców miast powyżej 100 tys. Osoby z wykształceniem wyższym w sposób statystycznie istotny częściej deklarują znajomość serów Camembert de Normandie oraz Parmigiano Reggiano. Biorąc pod uwagę płeć, stwierdzono, że w istotny sposób ser Camembert de Normandie jest lepiej znany wśród kobiet, Grana Padano zaś wśród mężczyzn.

W podobny sposób przeanalizowano 52 zależności dotyczące deklarowanych zakupów wskazanych 13 serów z 4 parametrami metryczki respondentów. Istotne zależności stwierdzono tylko w 4 przypadkach, co ukazano w tab. 2.

Wykazano, że respondenci z miast powyżej 100 tys. mieszkańców częściej dokonują zakupów serów Edam Holland oraz Allgäuer Emmentaler. Z drugiej strony, mieszkańcy wsi i miast do 100 tys. częściej kupują ser Feta. Oprócz tego stwierdzono, że młodszy respondenci – w wieku od 18 do 25 lat – częściej kupują ser Gorgonzola niż starsi konsumenci.

Tabela 2. Istotne zależności pomiędzy deklarowanym dokonywaniem zakupów serów regionalnych a parametrami metryczki respondentów – udowodnione na podstawie testu chi-kwadrat

Table 2. Significant relationships between the declared purchases of regional cheeses and the parameters of the respondents' specification – proved on the basis of the chi-square test

Deklarowana znajomość sera/ Declared knowledge of cheese	Zależny parametr metryczki/ Dependent parameter of the imprint	Kto zna lub gdzie lepiej znają wskazany ser/ Who knows or where they know better the indicated cheese	Poziom istotności (<i>p</i>) chi-kwadrat Pearsona/ Significance level (<i>p</i>) Pearson's chi-square	Poziom istotności (<i>p</i>) chi-kwadrat NW/ Significance level (<i>p</i>) chi-squared NW
Edam Holland	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0135	0,0085
Allgäuer Emmentaler	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0303	0,0062
Feta	miejsowość/ town	wieś lub miasto do 100 tys./ city of over 100 thousand	0,0125	0,0100
Gorgonzola	wiek/ age	18-25 lat/ years	0,0307	0,0150

Źródło/Source: opracowanie własne/own study.

Następne 52 analizy zostały przeprowadzone w celu ustalenia, jaka grupa konsumentów byłaby skłonna do dalszych zakupów serów regionalnych, ale już po wyższej cenie. W sposób jednoznaczny, a więc na podstawie obu testów chi-kwadrat, tj. zarówno Pearsona, jak i NW, wykazano istnienie jednej takiej zależności. Respondenci ze wsi lub mniejszych miast (do 100 tys. mieszkańców) częściej deklarowali dalsze zakupy sera Oscypka pomimo jego wyższej ceny. W pozostałych trzech przypadkach poziom istotności (*p*) testu chi-kwadrat Pearsona był większy niż 0,05, ale w teście chi-kwadrat NW wartość ta była poniżej 0,05. Od osób przeprowadzających test chi-kwadrat zależy, czy chcą takie wyniki uwzględnić. Wszystkie cztery zależności ukazano w tab. 3.

Ostatnią grupą wyników w ramach wielokrotnych dychotomii była analiza stosowania serów jako dodatków do siedmiu różnych dań w zależności od 4 parametrów metryczki (przeanalizowano więc $4 \times 7 = 28$ zależności). W dwóch przypadkach zależności takie zostały udowodnione przez obydwa testy chi-kwadrat (tj. Pearsona oraz NW), także w dwóch tylko przez test największej wiarygodności NW, zaś w jednym przypadku – przez test Pearsona. Wszystkie pięć zależności przedstawiono w tab. 4.

Tabela 3. Istotne zależności pomiędzy deklarowanym dokonywaniem zakupów droższych serów regionalnych a parametrami metryczki respondentów – udowodnione na podstawie przynajmniej jednego z testów chi-kwadrat (Pearsona lub największej wiarygodności NW)

Table 3. Significant dependencies between the declared making more expensive purchases of regional cheeses and the parameters of the respondents' specification – proven on the basis of at least one of the chi-square tests (Pearson or the highest reliability NW)

Deklarowana znajomość sera/ Declared knowledge of cheese	Zależny parametr metryczki/ Dependent parameter of the imprint	Kto zna lub gdzie lepiej znają wskazany ser/ Who knows or where they know better the indicated cheese	Poziom istotności (p) chi-kwadrat Pearsona/ Significance level (p) Pearson's chi-square	Poziom istotności (p) chi-kwadrat NW/ Significance level (p) chi-squared NW
Edam Holland	wiek/ age	18-25 lat/ years	0,1064	0,0371
Allgäuer Emmentaler	pleć/ gender	kobieta/ woman	0,0554	0,0149
Allgäuer Emmentaler	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,0554	0,0149
Oscypek	miejsowość/ town	wieś lub miasto do 100 tys./ city of over 100 thousand	0,0025	0,0026

Źródło/Source: opracowanie własne/own study.

Tabela 4. Istotne zależności pomiędzy deklarowanym stosowaniem serów jako dodatków do dań a parametrami metryczki respondentów – udowodnione na podstawie przynajmniej jednego z testów chi-kwadrat (Pearsona lub największej wiarygodności NW)

Table 4. Significant relationships between the declared use of cheeses as additives to dishes and the parameters of the respondents' record – proven on the basis of at least one of the chi-square tests (Pearson or the highest reliability NW)

Danie/ potrawa/ Dish/ meal	Zależny parametr metryczki/ Dependent parameter of the imprint	Kto zna lub gdzie lepiej znają wskazany ser/ Who knows or where they know better the indicated cheese	Poziom istotności (p) chi-kwadrat Pearsona/ Significance level (p) Pearson's chi-square	Poziom istotności (p) chi-kwadrat NW/ Significance level (p) chi-squared NW
zupa/ soup	wykształcenie/ education	nie wyższe/ not higher	0,0406	0,0524
deser/ dessert	pleć/ gender	kobieta/ woman	0,1006	0,0362
deser/ dessert	miejsowość/ town	miasto ponad 100 tys./ city of over 100 thousand	0,1006	0,0362
sałatka/ salad	wykształcenie/ education	wyższe/ higher	0,0014	0,0012
surówka/ salad	wykształcenie/ education	wyższe/ higher	0,0014	0,0012

Źródło/Source: opracowanie własne/own study.

Na podstawie testu chi-kwadrat Pearsona wykazano, że osoby niemające wykształcenia wyższego częściej wykorzystują sery regionalne jako składnik zup. Z drugiej strony, osoby po studiach znacznie częściej dodają sery regionalne jako składnik sałatek i surówek. Ser jest częściej wykorzystywany jako składnik deserów przez respondentów mieszkających w większych miastach, jak również przez kobiety, ale zależność tę wykazano tylko przy zastosowaniu testu chi-kwadrat największej wiarygodności NW.

3.4. Korelacje rang Spearmana

Współczynnik korelacji rang Spearmana jest wykorzystywany do opisu siły korelacji dwóch cech, wtedy gdy są one mierzalne, badana zbiorowość jest nieliczna oraz mają charakter jakościowy i istnieje możliwość ich uporządkowania.

Po przeprowadzeniu testów chi-kwadrat, chcąc uzupełnić uzyskane wyniki, zdecydowano się na obliczenie nieparametrycznych korelacji rang Spearmana pomiędzy wybranymi zmiennymi. Z pierwszej kategorii uzyskanych odpowiedzi ankietowych będących „zmiennymi jakościowymi (kategoryzującymi)” wybrano wszystkie dziesięć zmiennych oraz dodatkowo dołączono cztery zmienne z metryczki ankietowanych, co łącznie stanowiło 14 zmiennych.

Uzyskane współczynniki określiły siłę i kierunek tych korelacji. Znak współczynnika korelacji informuje o kierunku korelacji. Dodatni współczynnik korelacji wskazuje na prostą zależność między analizowanymi zmiennymi, czyli że są one o współzależności dodatniej, zaś ujemny współczynnik informuje o korelacji ujemnej. Im bardziej współczynnik korelacji jest bliższy jedności, tym zależność korelacyjna jest silniejsza. Zero jako uzyskana wartość współczynnika świadczy o braku związku korelacyjnego między badanymi zmiennymi.

Przyjęto następujące kryterium siły danych związków korelacyjnych: $<0,2$ – korelacja słaba; $0,2-0,4$ – korelacja niska; $0,4-0,6$ – korelacja umiarkowana; $0,6-0,8$ – korelacja wysoka; $0,8-0,9$ – korelacja bardzo wysoka; $0,9-1,0$ – prawie pełna zależność. W niniejszej pracy został założony poziom istotności $\alpha = 0,05$ (Stanisz, 2006, 2007).

Mając przygotowany plik z danymi z ankiety, wykorzystano program *Statistica 13.3*, aby przeprowadzić obliczenia współczynników korelacji rang Spearmana w parach dla wszystkich 10 zmiennych jakościowych (kategoryzujących) i 4 związanych z metryczką. W tym celu wybrano odpowiednią analizę: zakładka Statystyka → Statystyki nieparametryczne → Korelacje (Spearmana, tau Kendalla, gamma) → Zmienne (tu: wskazanie odpowiednich 14 zmiennych z wykorzystaniem klawisza ctrl) → ok → Kontynuuj z bieżącym wyborem → R Spearmana (zostawiając domyślne wskazania „macierzy kwadratowej”).

Analizując otrzymaną macierz (rys. 23), wyraźnie można zauważyć, że zdecydowana większość analizowanych par zmiennych jest ze sobą skorelowana słabo ($R < 0,2$) lub nisko ($R = 0,2-0,4$). Biorąc pod uwagę tylko pary: pierwsza zmienna z grupy 10 analizowanych, druga zaś – z 4 zmiennych, z metryczki uzyskano tylko słabe korelacje ($R < 0,2$). Niskie korelacje ($R = 0,2-0,4$) stwierdzono w 5 przypadkach, ale w odniesieniu do 10 głównych zmiennych.

Zmienna	jak-często-kupuje	gdzie-kupuje	główna-cecha-sera	forma-sera	gramatura	rodzaj	ulubiony	relacja-do-nowych	regionalne	wiekszy-wydatek	pleć	wiek	wykształcenie	miejscowość
jak-często-kupuje	1,000	-0,082	0,246	0,098	0,030	0,032	-0,034	0,092	-0,146	-0,148	0,188	-0,072	-0,018	-0,133
gdzie-kupuje	-0,082	1,000	-0,107	0,107	0,281	-0,020	0,046	-0,069	0,010	-0,052	-0,068	0,077	0,104	0,058
główna-cecha-sera	0,246	-0,107	1,000	-0,026	-0,077	-0,093	-0,070	0,196	-0,007	-0,123	0,035	-0,126	0,130	0,058
forma-sera	0,098	0,107	-0,026	1,000	0,065	-0,204	-0,136	-0,165	-0,158	0,025	0,042	0,004	0,101	0,042
gramatura	0,030	0,281	-0,077	0,065	1,000	0,173	0,158	0,020	0,113	-0,148	0,166	0,162	0,190	0,003
rodzaj	0,032	-0,020	-0,093	-0,204	0,173	1,000	0,088	-0,008	0,038	-0,065	-0,012	-0,157	-0,023	-0,101
ulubiony	-0,034	0,046	-0,070	-0,136	0,158	0,088	1,000	-0,083	0,016	-0,166	-0,076	-0,037	-0,078	-0,020
relacja-do-nowych	0,092	-0,069	0,196	-0,165	0,020	-0,008	-0,083	1,000	0,253	0,118	0,070	0,047	0,111	-0,018
regionalne	-0,146	0,010	-0,007	-0,158	0,113	0,038	0,016	0,253	1,000	0,234	0,041	-0,134	0,001	-0,119
wiekszy-wydatek	-0,148	-0,052	-0,123	0,025	-0,148	-0,065	-0,166	0,118	0,234	1,000	0,005	-0,060	0,071	0,005
pleć	0,188	-0,068	0,035	0,042	0,166	-0,012	-0,076	0,070	0,041	0,005	1,000	-0,036	0,248	0,004
wiek	-0,072	0,077	-0,126	0,004	0,162	-0,157	-0,037	0,047	-0,134	-0,060	-0,036	1,000	-0,009	0,063
wykształcenie	-0,018	0,104	0,130	0,101	0,190	-0,023	-0,078	0,111	0,001	0,071	0,248	-0,009	1,000	0,197
miejscowość	-0,133	0,058	0,058	0,042	0,003	-0,101	-0,020	-0,018	-0,119	0,005	0,004	0,063	0,197	1,000

Rys. 23. Współczynniki korelacji rang R Spearmana w postaci macierzy kwadratowej dla 10 zmiennych jakościowych (kategoryzujących) i 4 zmiennych z metryczki ankiety o preferencjach regionalnych serów europejskich przeprowadzonej wśród 101 respondentów

Fig. 23. Spearman's R-Rang square matrix correlation coefficient for 10 quality variables (categorization variables) and 4 variables from the metric of the survey on regional preferences for European cheese among 101 respondents

Źródło/Source: opracowanie własne/own study.

W analogiczny sposób utworzono macierz dla danych z grupy „wielokrotnych odpowiedzi” (4 zmienne) wraz z parametrami metryczki (4 zmienne), co ukazano na rys. 24. Korelację umiarkowaną stwierdzono dla 3 par: „ile zna serów – ile kupuje” ($R = 0,504$), „ile zna serów – do ilu dań dodaje sery” ($R = 0,420$), „ile kupuje serów – ile serów kupi drożej” ($R = 0,400$).

Zmienna	ile-zna	ile-kupuje	ile-kupuje-drożej	ile-dodatki-ów	pleć	wiek	wykształcenie	miejscowość
ile-zna	1,000	0,504	0,305	0,420	-0,019	-0,090	-0,223	-0,197
ile-kupuje	0,504	1,000	0,400	0,333	-0,101	-0,042	-0,210	-0,097
ile-kupuje-drożej	0,305	0,400	1,000	0,246	-0,176	0,019	-0,121	-0,005
ile-dodatki-ów	0,420	0,333	0,246	1,000	-0,098	0,140	-0,217	0,116
pleć	-0,019	-0,101	-0,176	-0,098	1,000	-0,036	0,248	0,004
wiek	-0,090	-0,042	0,019	0,140	-0,036	1,000	-0,009	0,063
wykształcenie	-0,223	-0,210	-0,121	-0,217	0,248	-0,009	1,000	0,197
miejscowość	-0,197	-0,097	-0,005	0,116	0,004	0,063	0,197	1,000

Rys. 24. Współczynniki korelacji rang R Spearmana w postaci macierzy kwadratowej dla 4 zmiennych z grupy „wielokrotnych odpowiedzi” i 4 zmiennych z metryczki ankiety o preferencjach regionalnych serów europejskich przeprowadzonej wśród 101 respondentów

Fig. 24. Spearman's R-Rang square matrix correlation coefficient for 4 variables from the "multiple responses" group and 4 variables from the metric of the survey on regional preferences for European cheese among 101 respondents

Źródło/Source: opracowanie własne/own study.

Kolejna macierz (rys. 25) ukazuje wyłącznie korelacje pomiędzy wyróżnikami dotyczącymi znajomości 13 serów regionalnych. Najwyższy współczynnik Spearmana zaobserwowano pomiędzy znajomością sera Gouda Holland i sera Edam Holland ($R = 0,523$, korelacja umiarkowana).

Zmienna	zna-Cabrales	zna-Grana-Padano	zna-Tomme-de-Savoie	zna-Edam-Holland	zna-Allgauer-Emmentaler	zna-Gouda-Holland	zna-Camembert-de-Normandie	zna-Feta	zna-Bryndza-Podhalanska	zna-Roquefort	zna-Oscypek	zna-Parmigiano-Reggiano	zna-Gorgonzola
zna-Cabrales	1,000	-0,047	-0,014	-0,088	0,184	0,091	-0,103	0,056	0,103	0,249	0,038	0,129	0,112
zna-Grana-Padano	-0,047	1,000	0,120	0,113	0,302	-0,042	-0,066	0,138	0,066	0,188	-0,053	0,386	0,155
zna-Tomme-de-Savoie	-0,014	0,120	1,000	0,018	-0,077	0,130	0,138	0,079	0,004	0,149	0,055	0,183	0,159
zna-Edam-Holland	-0,088	0,113	0,018	1,000	0,285	0,523	0,373	0,209	0,106	0,110	-0,020	0,101	0,217
zna-Allgauer-Emmentaler	0,184	0,302	-0,077	0,285	1,000	0,212	0,102	0,081	0,276	0,329	0,068	0,212	0,131
zna-Gouda-Holland	0,091	-0,042	0,130	0,523	0,212	1,000	0,425	0,283	0,212	0,194	0,064	0,054	0,220
zna-Camembert-de-Normandie	-0,103	-0,066	0,138	0,373	0,102	0,425	1,000	0,296	0,268	0,160	0,337	0,100	0,232
zna-Feta	0,056	0,138	0,079	0,209	0,081	0,283	0,296	1,000	0,030	0,224	0,063	0,097	0,360
zna-Bryndza-Podhalanska	0,103	0,066	0,004	0,106	0,276	0,212	0,268	0,030	1,000	0,184	0,255	0,023	0,047
zna-Roquefort	0,249	0,188	0,149	0,110	0,329	0,194	0,160	0,224	0,184	1,000	0,154	0,221	0,390
zna-Oscypek	0,038	-0,053	0,055	-0,020	0,068	0,064	0,337	0,063	0,255	0,154	1,000	0,054	0,285
zna-Parmigiano-Reggiano	0,129	0,386	0,183	0,101	0,212	0,054	0,100	0,097	0,023	0,221	0,054	1,000	0,291
zna-Gorgonzola	0,112	0,155	0,159	0,217	0,131	0,220	0,232	0,360	0,047	0,390	0,285	0,291	1,000

Rys. 25. Współczynniki korelacji rang R Spearmana w postaci macierzy kwadratowej dla 13 zmiennych z grupy „wielokrotnych dychotomii” dotyczących znajomości wskazanych serów regionalnych, na podstawie ankiety przeprowadzonej wśród 101 respondentów

Fig. 25. R-Rang correlation coefficients of Spearman square matrix for 13 variables from the group of “multiple dichotomies” for familiarity with the indicated regional cheeses, based on a survey of 101 respondents

Źródło/Source: opracowanie własne/own study.

Zmienna	kupuje-Grana-Padano	kupuje-Tomme-de-Savoie	kupuje-Edam-Holland	kupuje-Allgauer-Emmentaler	kupuje-Gouda-Holland	kupuje-Camembert-de-Normandie	kupuje-Feta	kupuje-Bryndza-Podhalanska	kupuje-Roquefort	kupuje-Oscypek	kupuje-Parmigiano-Reggiano	kupuje-Gorgonzola
kupuje-Grana Padano	1,000	-0,031	0,047	-0,104	0,011	0,054	0,105	-0,109	0,150	-0,095	0,055	0,036
kupuje-Tomme-de-Savoie	-0,031	1,000	-0,060	-0,033	0,147	0,170	0,079	-0,035	-0,017	0,103	0,230	0,215
kupuje-Edam-Holland	0,047	-0,060	1,000	0,174	0,358	0,207	-0,118	0,076	0,026	-0,183	0,044	-0,106
kupuje-Allgauer-Emmentaler	-0,104	-0,033	0,174	1,000	0,131	-0,044	-0,078	0,203	-0,058	-0,255	0,401	0,105
kupuje-Gouda-Holland	0,011	0,147	0,358	0,131	1,000	0,134	-0,028	-0,033	0,006	-0,193	0,054	0,017
kupuje-Camembert-de-Normandie	0,054	0,170	0,207	-0,044	0,134	1,000	0,141	0,012	0,164	-0,028	0,117	-0,037
kupuje-Feta	0,105	0,079	-0,118	-0,078	-0,028	0,141	1,000	-0,310	0,019	0,200	0,010	0,104
kupuje-Bryndza-Podhalanska	-0,109	-0,035	0,076	0,203	-0,033	0,012	-0,310	1,000	-0,061	-0,021	-0,065	-0,080
kupuje-Roquefort	0,150	-0,017	0,026	-0,058	0,006	0,164	0,019	-0,061	1,000	0,064	-0,076	-0,081
kupuje-Oscypek	-0,095	0,103	-0,183	-0,255	-0,193	-0,028	0,200	-0,021	0,064	1,000	-0,096	0,169
kupuje-Parmigiano-Reggiano	0,055	0,230	0,044	0,401	0,054	0,117	0,010	-0,065	-0,076	-0,096	1,000	0,223
kupuje-Gorgonzola	0,036	0,215	-0,106	0,105	0,017	-0,037	0,104	-0,080	-0,081	0,169	0,223	1,000

Rys. 26. Współczynniki korelacji rang R Spearmana w postaci macierzy kwadratowej dla 13 zmiennych z grupy „wielokrotnych dychotomii” dotyczących kupowania wskazanych serów regionalnych, na podstawie ankiety przeprowadzonej wśród 101 respondentów

Fig. 26. Spearman’s R-Rang square matrix correlation coefficient for 13 variables from the group “multiple dichotomies” for the purchase of the indicated regional cheeses, based on a survey of 101 respondents

Źródło/Source: opracowanie własne/own study.

W podobny sposób obliczona macierz (rys. 26) badająca korelacje pomiędzy wyróżnikami związanymi z kupnem serów regionalnych wskazała najwyższą korelację dla tej samej pary serów, tj. Gouda Holland oraz Edam Holland ($R = 0,358$, korelacja niska).

Kolejna macierz odnosząca się do deklaracji zakupów poszczególnych serów regionalnych nawet po wyższych cenach ukazana jest na rys. 27. Wynika z niej, że najwyższa korelacja została wykazana dla pary serów Gouda Holland oraz Allgäuer Emmentaler ($R = 0,423$, korelacja umiarkowana).

Zmienna	drozej-kupi-Cabrales	drozej-kupi-Grana-Padano	drozej-kupi-Tomme-de-Savoie	drozej-kupi-Edam-Holland	drozej-kupi-Allgauer-Emmentaler	drozej-kupi-Gouda-Holland	drozej-kupi-Camembert-de-Normandie	drozej-kupi-Feta	drozej-kupi-Bryndza-Podhalanska	drozej-kupi-Roquefort	drozej-kupi-Oscypek	drozej-kupi-Parmigiano-Reggiano	drozej-kupi-Gorgonzola
drozej-kupi-Cabrales	1,000	-0,025	-0,014	-0,027	0,341	-0,031	-0,045	-0,086	-0,045	-0,025	-0,083	-0,042	-0,042
drozej-kupi-Grana-Padano	-0,025	1,000	-0,036	-0,069	0,081	-0,079	-0,001	-0,047	-0,113	0,291	-0,122	0,012	-0,105
drozej-kupi-Tomme-de-Savoie	-0,014	-0,036	1,000	0,241	-0,042	-0,044	0,126	0,021	0,126	-0,036	0,027	-0,060	-0,059
drozej-kupi-Edam-Holland	-0,027	-0,069	0,241	1,000	0,064	0,325	-0,019	-0,156	-0,019	-0,069	-0,067	-0,005	-0,114
drozej-kupi-Allgauer-Emmentaler	0,341	0,081	-0,042	0,064	1,000	0,423	-0,132	-0,030	0,162	0,392	-0,093	0,289	0,084
drozej-kupi-Gouda-Holland	-0,031	-0,079	-0,044	0,325	0,423	1,000	-0,141	0,082	0,045	0,215	-0,046	0,161	0,163
drozej-kupi-Camembert-de-Normandie	-0,045	-0,001	0,126	-0,019	-0,132	-0,141	1,000	0,094	0,081	-0,001	0,005	-0,041	0,035
drozej-kupi-Feta	-0,086	-0,047	0,021	-0,156	-0,030	0,082	0,094	1,000	-0,013	0,122	-0,019	0,040	0,147
drozej-kupi-Bryndza-Podhalanska	-0,045	-0,113	0,126	-0,019	0,162	0,045	0,081	-0,013	1,000	0,111	0,059	-0,041	-0,039
drozej-kupi-Roquefort	-0,025	0,291	-0,036	-0,069	0,392	0,215	-0,001	0,122	0,111	1,000	-0,037	0,248	0,131
drozej-kupi-Oscypek	-0,083	-0,122	0,027	-0,067	-0,093	-0,046	0,005	-0,019	0,059	-0,037	1,000	-0,293	0,108
drozej-kupi-Parmigiano-Reggiano	-0,042	0,012	-0,060	-0,005	0,289	0,161	-0,041	0,040	-0,041	0,248	-0,293	1,000	-0,020
drozej-kupi-Gorgonzola	-0,042	-0,105	-0,059	-0,114	0,084	0,163	0,035	0,147	-0,039	0,131	0,108	-0,020	1,000

Rys. 27. Współczynniki korelacji rang R Spearmana w postaci macierzy kwadratowej dla 13 zmiennych z grupy „wielokrotnych dychotomii” dotyczących deklaracji kupowania wskazanych serów regionalnych nawet po wyższych cenach, na podstawie ankiety przeprowadzonej wśród 101 respondentów

Fig. 27. Spearman's R-Rang square matrix correlation coefficient for 13 variables from the group “multiple dichotomies” indicating that they also buy the indicated regional cheese at higher prices, based on a survey of 101 respondents

Źródło/Source: opracowanie własne/own study.

Ostatnia macierz ukazana jest na rys. 28. Przeanalizowano w tym przypadku deklarację użycia serów do 7 różnych dań lub potraw. Dodatkowo umieszczono 4 wyróżniki z metryczki respondentów. Uzyskane korelacje były co najwyżej niskie. Najwyższą bezwzględną wartość $R = -0,319$ stwierdzono dla dwóch par powiązanych z wykształceniem, tj. stosowaniem serów jako składnika do surówek, jak i do sałatek.

Przedstawione macierze ukazujące współczynniki korelacji rang R Spearmana mogą być przydatne jako uzupełnienie wyników uzyskanych z testów chi-kwadrat. Taka całościowa analiza zastosowana łącznie może przyczynić się do lepszej interpretacji wyników badań ankietowych.

Zmienna	dodatek k-zupa	dodatek -drugie -danie- -miesne	dodatek k-deser	dodatek -goraca -przekaska	dodatek -sos	dodatek -sałatka	dodatek -surowka	pleć	wiek	wykszt acenie	miejsco wość
dodatek-zupa	1,000	0,045	0,235	-0,019	-0,045	-0,248	-0,248	0,074	0,055	0,204	-0,045
dodatek-drugie-danie-miesne	0,045	1,000	-0,026	-0,013	-0,267	0,136	0,136	-0,013	0,089	0,001	0,041
dodatek-deser	0,235	-0,026	1,000	-0,043	-0,093	-0,101	-0,101	-0,163	-0,148	0,057	-0,163
dodatek-goraca-przekaska	-0,019	-0,013	-0,043	1,000	-0,159	-0,223	-0,223	-0,067	-0,009	-0,095	-0,111
dodatek-sos	-0,045	-0,267	-0,093	-0,159	1,000	-0,142	-0,142	-0,009	-0,073	0,042	0,176
dodatek-salatka	-0,248	0,136	-0,101	-0,223	-0,142	1,000	1,000	-0,045	0,141	-0,319	0,129
dodatek-surowka	-0,248	0,136	-0,101	-0,223	-0,142	1,000	1,000	-0,045	0,141	-0,319	0,129
pleć	0,074	-0,013	-0,163	-0,067	-0,009	-0,045	-0,045	1,000	-0,036	0,248	0,004
wiek	0,055	0,089	-0,148	-0,009	-0,073	0,141	0,141	-0,036	1,000	-0,009	0,063
wykształcenie	0,204	0,001	0,057	-0,095	0,042	-0,319	-0,319	0,248	-0,009	1,000	0,197
miejscowość	-0,045	0,041	-0,163	-0,111	0,176	0,129	0,129	0,004	0,063	0,197	1,000

Rys. 28. Współczynniki korelacji rang R Spearmana w postaci macierzy kwadratowej dla 7 zmiennych z grupy „wielokrotnych dychotomii” dotyczących stosowania wskazanych serów regionalnych jako składników dań lub potraw, na podstawie ankiety przeprowadzonej wśród 101 respondentów

Fig. 28. Spearman's R-Rang correlation coefficient square matrix for 7 variables from the group “multiple dichotomies” for the use of regional cheese as a food or food ingredient, based on a survey of 101 respondents

Źródło/Source: opracowanie własne/own study.

4. Zakończenie

Zaproponowana w pracy metodologia testu niezależności chi-kwadrat z wykorzystaniem pakietu programu *Statistica* wykazała swoją przydatność w analizie badań ankietowych. Wydaje się, że wraz z klasyczną analizą chi-kwadrat Pearsona i/lub największej wiarygodności (NW) warto zebrane dane ankietowe poddać również badaniom korelacji rang Spearmana.

Na podstawie przeprowadzonych 240 podwójnych testów chi-kwadrat pomiędzy 60 zmiennymi a 4 wyróżnikami tzw. metryczki respondenta wykazano w 25 przypadkach istotne zależności – przynajmniej według jednego z testów. W 18 przypadkach obydwa testy chi-kwadrat (zarówno Pearsona, jak i największej wiarygodności) wykazały istnienie takich statystycznie istotnych zależności. Od prowadzących analizę zależy, czy istnienie tych różnic będą uzależniali od jednego z tych testów czy od obydwu.

Najwięcej udowodnionych przez oba testy zależności stwierdzono w ramach danych sformatowanych jako wielokrotne dychotomie, a w szczególności w odniesieniu do deklarowanej przez respondentów znajomości serów regionalnych. Udowodniono, że respondenci z dużych miast powyżej 100 tys. mieszkańców znają lepiej 4 sery (w porównaniu z mieszkańcami mniejszych miejscowości), tj.: włoski Grana Padano, niemiecki Allgäuer Emmentaler oraz holenderskie Gouda i Edam. Osoby z wykształceniem wyższym lepiej znają od innych dwa sery: włoski Parmigiano Reggiano oraz francuski Camembert de Normandie. Kobiety znają lepiej niż męż-

czyżni ser Camembert de Normandie, natomiast mężczyźni lepiej rozpoznają Grana Padano.

W dużych miastach znacznie częściej kupuje się sery Allgäuer Emmentaler i Edam Holland, w mniejszych miejscowościach zaś grecką Fetę. Młodzi ludzie w wieku 18-25 lat chętniej niż starsze osoby kupują włoską Gorgonzolę. Deklarację dalszych zakupów, nawet po wyższej cenie, odnotowano w przypadku Oscypka, (również wśród mieszkańców mniejszych miejscowości).

Sery regionalne jako dodatki do dań i potraw są powiązane głównie z wykształceniem. Osoby po studiach znacznie częściej wykorzystują sery do sałatek i surówek. Deklarowana liczba rodzajów kupowanych serów regionalnych jest istotnie powiązana z płcią i wiekiem respondentów, znajomość serów zaś – z wykształceniem. Wszystkie opisane powyżej istotne zależności udowodniono w ramach obu testów chi-kwadrat.

Analiza korelacji rang Spearmana pomiędzy wybranymi zmiennymi wykazała co najwyżej umiarkowaną korelację pomiędzy kilkoma parami zmiennych.

Na podstawie uzyskanych wyników można potwierdzić przydatność testu niezależności chi-kwadrat oraz analizy korelacji rang Spearmana w badaniach ankietowych.

Literatura

- Bishop, Y. M. M., Fienberg, S. E. i Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press.
- Fienberg, S. E. (1977). *The analysis of cross-classified categorical data*. Cambridge: The MIT Press.
- Gadzała, K. i Lesiów, T. (2019). Wiedza konsumentów na temat aktualnych trendów żywieniowych. *Nauki Inżynierskie i Technologie*, 2(33), 6-43.
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka. Elementy teorii i zadania*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Oziembłowski, M. i Salejda, A. M. (2021). Wybrane systemy żywności regionalnej i tradycyjnej w kontekście turystyki. In T. Lesiów (red.), *Determinanty jakości pracy przewodniczej*. Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.
- Rajs, R. (2006). *Ocena testów sprawdzających wiedzę studenta metodami testowania hipotez statystycznych. Modelowanie metodą (chi kwadrat)*. Pobrane 26 września 2022 z <http://www.kis.pwshchelm.pl/publikacje/V/Rajs.pdf>
- Słowińska, M. (2019). Wykorzystanie testu chi-kwadrat w badaniach preferencji żywieniowych konsumentów. *Nauki Inżynierskie i Technologie*, 1(32), 24-38.
- Stanisz, A. (2006). *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 1. Statystyki podstawowe*. Kraków: StatSoft Polska Sp. z o.o.
- Statsoft.pl. (2022). Strona internetowa firmy StatSoft Polska Sp. z o.o.

Internet

- <https://agriculture.ec.europa.eu>
<https://world.openfoodfacts.org>
serwis.polskasmakuje.pl