

VI - Analiza opisowa i jej podstawowe elementy

Sphinx Classroom

pakiet Sphinx przeznaczony do nauki metod i techniki badań ankietowych.



Sphinx
ClassRoom

Spis treści

1. Rozplanowanie obliczeń
2. Różne rodzaje analiz opisowych
 - 2.1 Odpowiedzi na pojedyncze pytania zamknięte
 - 2.2 Odpowiedzi na pytania zamknięte wielokrotne
 - 2.3 Odpowiedzi na pytania zamknięte uporządkowane
 - 2.4 Odpowiedzi na pytania cyfrowe otwarte
 - 2.5 Parametry położenia i rozrzutu danych
 - 2.6 Tabele grup pytań
3. Wykresy do analizy opisowej i przetwarzania elementarnego



1. Rozplanowanie obliczeń

- Plan obliczania (lub obróbki), to predefiniowana lista analiz do wykonania, zaspokajająca potrzeby badania. Zdecydowanie zaleca się, aby plan obliczeń został jasno określony przed rozpoczęciem fazy aktywnej analizy.
- W dużych agencjach badawczych szef projektu definiuje swój plan obliczeń i przekazuje wyniki do swojego działu "liczenia".
- Plan analizy może być już zaprojektowany zgodnie z požądanym przez kierownika projektu scenariuszem prezentacji wyników.
- Obejmuje ona zazwyczaj po kolei:
 - "Płaską" interpretację pytań, które opisują populację, z którą przeprowadzono wywiad
 - "Płaskie" opracowanie kluczowych pytań ankietowych (wszystkie formaty), zgodnie z planem tematycznym.
 - Analizy "krzyżowe" według jednego lub więcej pytań profilowych (lub tożsamościowych), takich jak płeć, wiek, itp. w zależności od celów badania (patrz sekwencja 9).
 - Możliwe analizy wieloczynnikowe (patrz sekwencja 12)
- Niektóre programy, takie jak Sphinx iQ, umożliwiają skonfigurowanie planu obliczeń, a następnie uruchomienie (i uzyskanie) wszystkich pożądanych analiz.



2. Różne rodzaje analiz opisowych

- W przypadku sortowania płaskiego ważna jest obróbka tylko jednego pytania lub zmiennej w tym samym czasie. Jest ona również nazywana analizą jednoczynnikową lub opisową. Na początku należy skupić się na opisie wyników. Rodzaje rezultatów różnią się w zależności od charakteru analizowanych pytań: pojedyncze zamknięte, wielokrotne, uporządkowane, skalowe, cyfrowe, kodowe lub tekstowe (zob. sekwencja 2).

Pytanie zamknięte

pojedyncze

Wielokrotny wybór

uporządkowane

Skala

Wizerunek

☐	☐	☒	☐	☐	☐
☐	☒	☐	☒	☒	☐
3	2		1		
				☒	
- +					
					

Pytanie otwarte

Liczbowe

Tekstowe

Kodowe

1 2 3

Bla bla, bla blabla

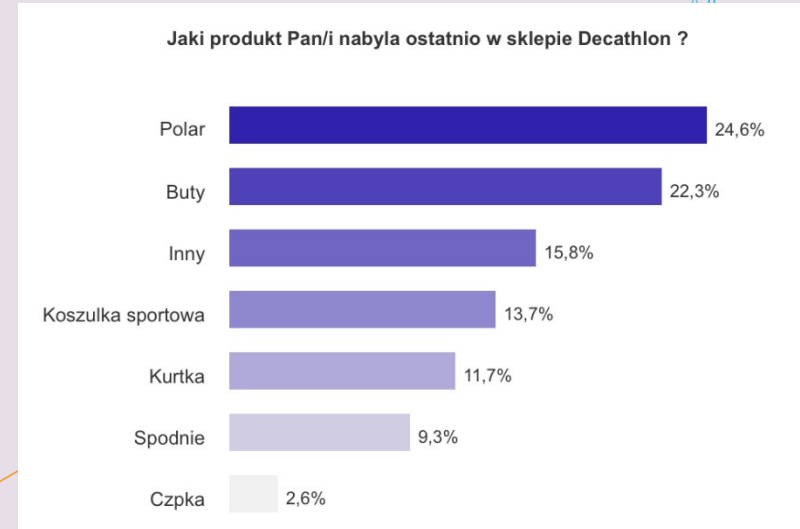
ABC 123



Sphinx
ClassRoom

2. Różne rodzaje analiz opisowych (ciąg dalszy)

- **2.1 Odpowiedzi na pojedyncze pytania zamknięte**
- Bardzo proste. Sprowadza się do przedstawienia tabeli zawierającej liczbę uzyskanych odpowiedzi i częstość (= procent) dla każdej proponowanej modalności.
- Odpowiedzi będą zazwyczaj uszeregowane w porządku malejącym według częstości, chyba że modalności są uporządkowane,
- Jedyną ważną opcją sprawozdawczą jest uwzględnienie lub nieuwzględnienie braku odpowiedzi. Brak odpowiedzi jest często wykluczany ze sprawozdawczości wyników, a mimo to podaje się, w różnych intencjach i celach, wskaźnik odpowiedzi na pytanie.

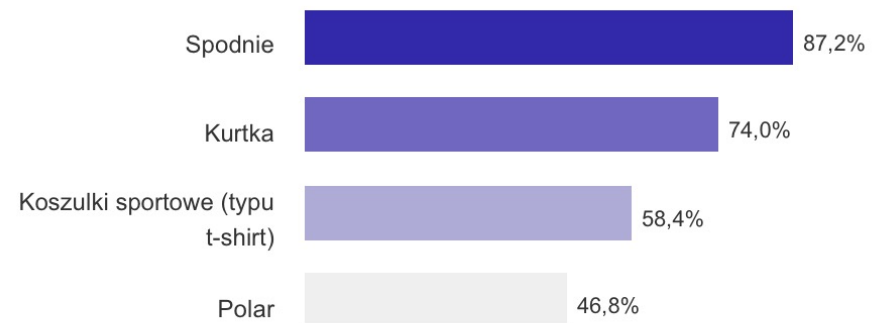


2. Różne rodzaje analiz opisowych (ciąg dalszy)

- **2.2 Odpowiedzi na pytania zamknięte wielokrotne**
- Poniższe zestawienie przedstawia tabela, w której z powodu wielu odpowiedzi, suma procentów jest większa niż 100.
- Lepiej jest obliczać częstotliwości **na podstawie obserwacji** niż cytatów.
- Odpowiedzi są na ogół uszeregowane w porządku malejącym według częstotliwości występowania.
- W poniższym przykładzie odpowiednim wynikiem jest, że 74% respondentów kupiło kurtkę dziecięcą w ciągu ostatnich 12 miesięcy.
- *Fakt, że ta modalność stanowi 27,8% (liczba = 574) wszystkich 2067 cytowań (liczba totalna) nie ma znaczenia.*

Ktore z ponizszych produktow kupowal Pan/Pani dla swoich dzieci w ostatnim roku?

	N	%
Spodnie	677	32,8%
Kurtka	574	27,8%
Koszulki sportowe (typu t-shirt)	453	21,9%
Polar	363	17,6%
SUMA	2 067	100,0%



2. Różne rodzaje analiz opisowych (ciąg dalszy)

- **2.3 Odpowiedzi na pytania zamknięte uporządkowane**
- Oprócz ich łącznej częstotliwości we wszystkich kategoriach, można również wyświetlić tylko odpowiedzi dla danej kategorii.
- Można także chcieć obliczać "średnią rangę" lub jej **ważność** wskazującą średnią pozycję danego kryterium w rankingu, wśród wszystkich badanych osób.
- Im większa ważność (lub średnia ranga), tym wyższa pozycja w rankingu. Im niższa ważność, tym niższa pozycja w rankingu.
- W naszym przykładzie solidność była cytowana rzadziej niż estetyka, ale jej średnia ranga (1,2) jest nieco wyższa. W związku z tym jest ona średnio ujmując ważniejsza.

Analiza opisowa i zabiegi elementarne

Sposród ponizszych kryteriow prosze wskazac 3 NAJWAZNIEJSZE dla Pana/Pani przy zakupie SPODNI dla swoich dzieci:

Efektywne odpowiedzi: 765

Odsetek odpowiedzi: 98,1%

	N	%	ZNACZENIE
Dopasownaie (ładny kroj)	483	63,1%	1,4
Estetyka	444	58,0%	1,1
Solidnosc / wytrzymałosc	434	56,7%	1,2
Mile w dotyku	252	32,9%	0,6
Niska cena	227	29,7%	0,5
Naturalne składniki	157	20,5%	0,5
Odpornosc UV	127	16,6%	0,4
Mozliwosc odpiecia nogawek	92	12,0%	0,2
Inne	21	2,7%	0,0
SUMA	765		

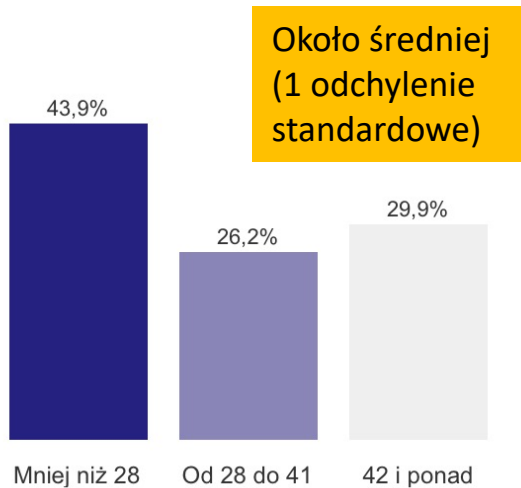


2. Różne rodzaje analiz opisowych (ciąg dalszy)

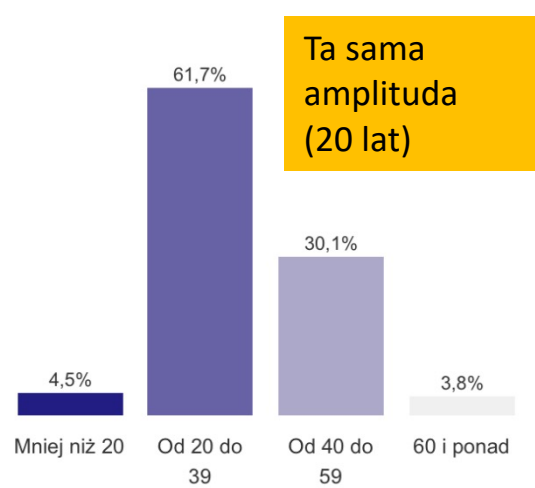
• 2.4 Odpowiedzi na pytania cyfrowe otwarte

- Możliwe jest również wybranie przedstawienia w postaci klasyfikacji odpowiedzi.
- Klasyfikacji można dokonać na różne sposoby: klasy o takich samych wartościach, amplitudach lub liczebnościach, klasy obliczone na podstawie średniej i odchylenia standardowego lub klasy ustalone przez badacza (najczęściej stosowana metoda).
- Odpowiedzialny za badania decyduje o pogrupowaniu osób według kategorii, mający sens dla niego i dla niego samego jak również dla prowadzonego badania.

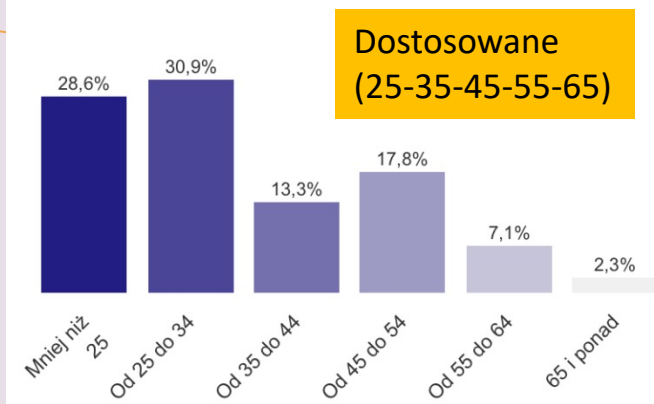
Ile masz lat?



Ile masz lat?



Ile masz lat?



2. Różne rodzaje analiz opisowych (ciąg dalszy)

• 2.5 Parametry położenia i rozrzutu

- Możemy również skupić się na głównych parametrach położenia i rozrzutu (średnia i odchylenie standardowe w ogólności) oraz ich graficznej reprezentacji.
- **Parametry położenia:** umiejscowienie populacji w odniesieniu do badanego zjawiska
- tryb: wartość, w której stan liczbowy jest największy ;
- mediana: wartość dzieląca siłę roboczą na dwie równe części, niedostatecznie wykorzystywana; jest jednak dobrym wskaźnikiem, mniej wrażliwym na wartości skrajne;
- średnia: "wartość typowa", uzyskana przez równomierne rozłożenie wszystkich wartości na wszystkie osoby.
- **Parametry rozrzutu:** ocena niejednorodności badanego zjawiska.
- zakres: wartość maksymalna - wartość minimalna
- odchylenie standardowe: średnie odchylenie od średniej dla wszystkich osobników; jeśli jest niskie, osobnicy formułują podobne odpowiedzi, jeśli jest wysokie ($>0,5$ średniej) zmienność jest duża w populacji badanej.



2. Różne rodzaje analiz opisowych (ciąg dalszy)

2.6 Tabele grup pytań

- Kolejne pytania mogą być przedstawione w formie tabeli (patrz sekwencja 2) z identycznymi odpowiedziami.
- Do prezentacji wyników często preferuje się bardziej syntetyczną prezentację w jednej tabeli, niż serię kilku tabel dla każdego pytania.
- Jeśli chodzi o pytania skalowe, mogą być one przedstawione w formie liczbowej (średnia), zgodnie ze wskaźnikiem (patrz sekwencja 2) przypisaną do nich, domyślnie lub dostosowaną.

Precyzując bardziej, jaki rodzaj produktów przeznaczonych dla hikingu i/lub trekkingu masz zamiar nabyć w przeciągu następnych 12 miesięcy					
	TAK, DEFINITYWNE	TAK, PRAWDOPODOBNE	NIE, PRAWDOPODOBNE NIE	NIE, DEFINITYWNE NIE	CAŁKOWITY
ubrania	29,4%	59,9%	7,3%	3,4%	100,0%
obuwie	29,2%	47,7%			
plecak	18,2%	36,2%			
kijki	9,6%	20,6%			
zastaw do gotowania i sztućce	7,5%	17,7%			
termos/ butelkę do wody	12,2%	25,3%			
sprzęt do oświetlenia/ energii (dynamo, lampa główna, ładowarki)	11,3%	30,3%			
SUMA	17,1%	34,6%			

Precyzując bardziej, jaki rodzaj produktów przeznaczonych dla hikingu i/lub trekkingu masz zamiar nabyć w przeciągu następnych 12 miesięcy

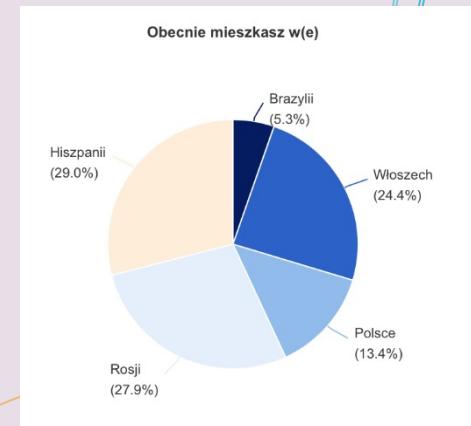
	SREDNIA
ubrania	7,4
obuwie	6,7
plecak	5,1
kijki	3,3
zastaw do gotowania i sztućce	2,9
termos/ butelkę do wody	
sprzęt do oświetlenia/ energii (dynamo, lampa główna, ładowarki)	
SUMA	

N.B. Wskaźnik zastosowany tutaj:
0 dla nie, definitywnie nie
następnie 3; 7 i 10 dla tak,
definitywnie

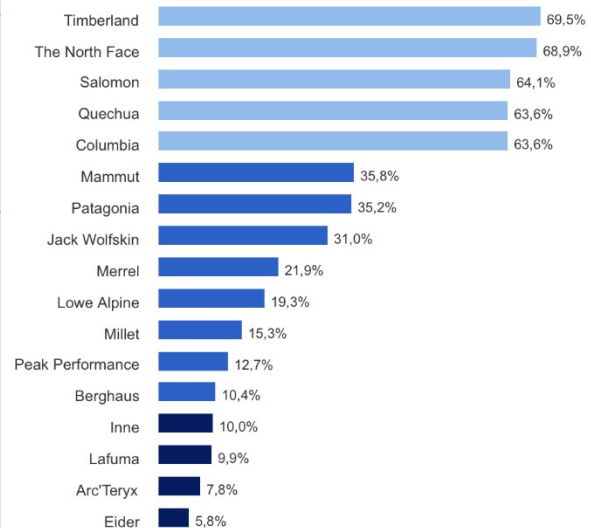
ou

3. Wykresy do analizy opisowej i przetwarzania elementarnego

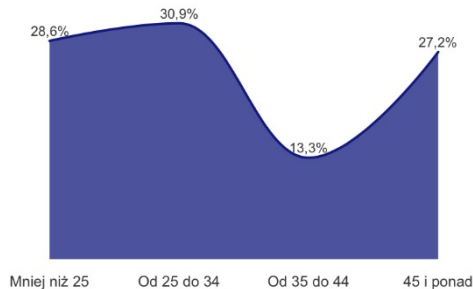
- Aby przedstawić wyniki analizy opisowej kierownik badania ma do dyspozycji kilka rodzajów wykresów:
- Sektory:** odpowiednie dla mniej niż 6 modalności - i do wizualizacji rozkładu wyników.
- Paski poziome:** dla więcej niż 4 modalności - aby wyraźnie zidentyfikować najczęściej używane modalności.
- Pionowe słupki** (histogramy): dla więcej niż 3 modalności - i modalności rosnących lub malejących (np. skale)
- Krzywe** lub **obszary**: to samo co pionowe słupki - interesujące do wizualizacji rozkładu zmiennej liczbowej



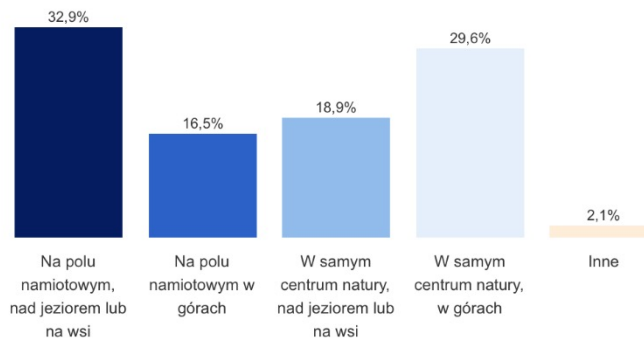
Spośród tych sprecyzowanych marek, które oferują produkty do hikingu i/lub trekkingu, który znasz nawet jeśli tylko z nazwy?



Ile masz lat?



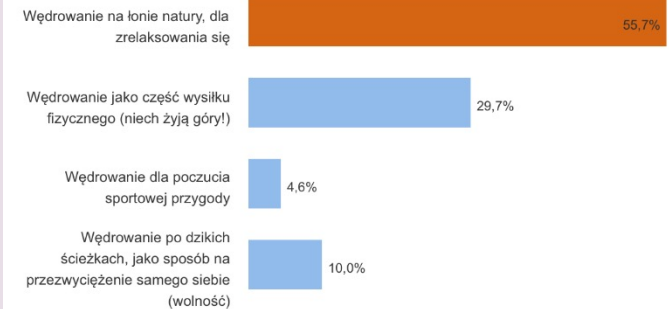
Gdzie najczęściej rozbijasz swój namiot?



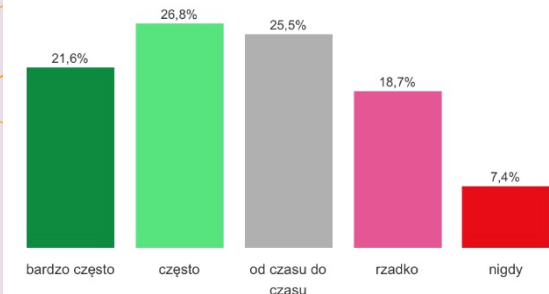
3. Wykresy do analizy opisowej i przetwarzania elementarnego (c.d.)

- Wybór kolorów w grafice może ułatwić jej komunikatywność lub dać dobre wrażenie o jakości wykonanej pracy. Można na przykład:
- Używać znaczących kolorów: zielony dla pozytywnych, czerwony dla negatywnych, biały dla neutralnych.
- Należy posługiwać się tabelą graficzną instytucji sporządzającej opracowanie.
- Podkreślanie wyników innym kolorem

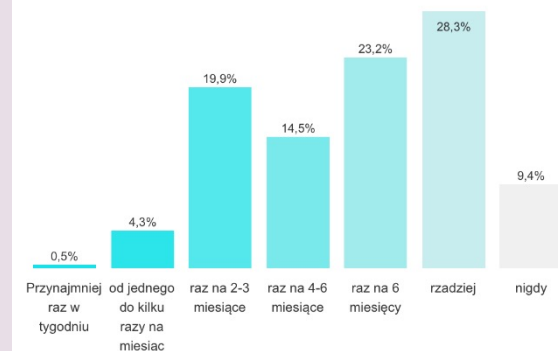
Czy możesz stwierdzić, jaki rodzaj hikingu najczęściej uprawiasz?



Decydujesz się na trekking w górach, aby przeżyć przygodę

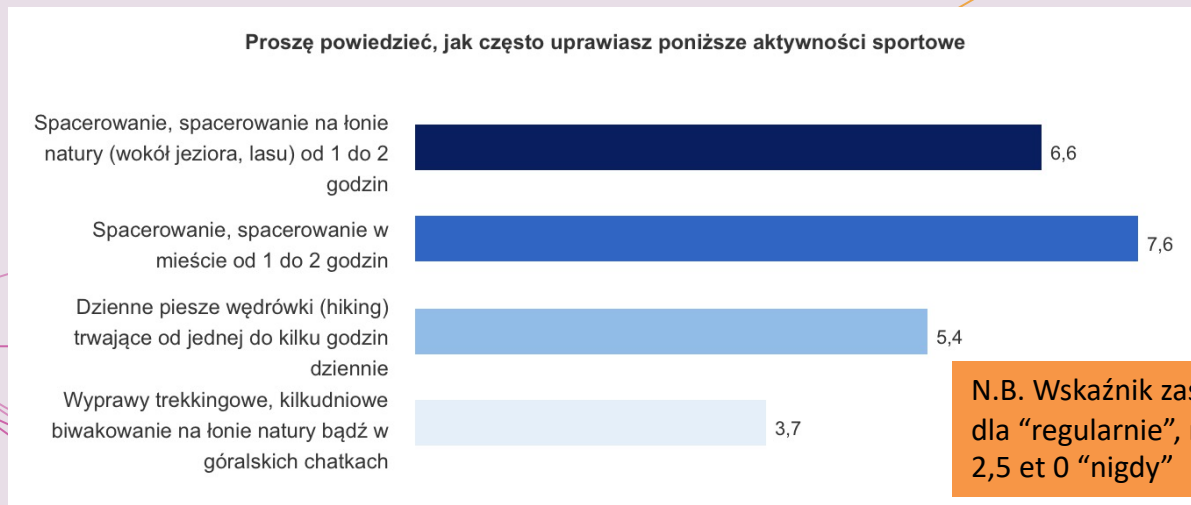


Jak często kupujesz produkty z Decathlonu



3. Wykresy do analizy opisowej i przetwarzania elementarnego (c.d.)

- Dla tabeli grup pytań
- Z tabeli można zrezygnować prezentując jedynie wykres ze słupkami zawierającymi procenty. Ważne jest, aby kolory słupków podążały za semantyczną ewolucją odpowiedzi, n.p.: od czerwonego dla "wcale nie jestem zadowolony" do zielonego dla "bardzo zadowolony". Obok wykresu znajduje się legenda.
- W przypadku pytań ze skalą, odpowiedzi można również przedstawić w formie liczbowej, podając dla każdego pytania "średni wynik", obliczony zgodnie z przypisanym wskaźnikiem (zob. pkt 2.6).
- Preferowana jest ocena w skali 10, która jest łatwiejsza do zrozumienia i manipulacji. Oczywiście komunikowania jest zastosowany wskaźnik.



N.B. Wskaźnik zastosowany tutaj: 10 dla "regularnie", następnie 7,5 – 5 – 2,5 et 0 "nigdy"



VII - Analiza krzyżowa (lub dwuzmienna)

Sphinx Classroom

pakiet Sphinx przeznaczony do nauki metod i techniki badań ankietowych.



Sphinx
ClassRoom

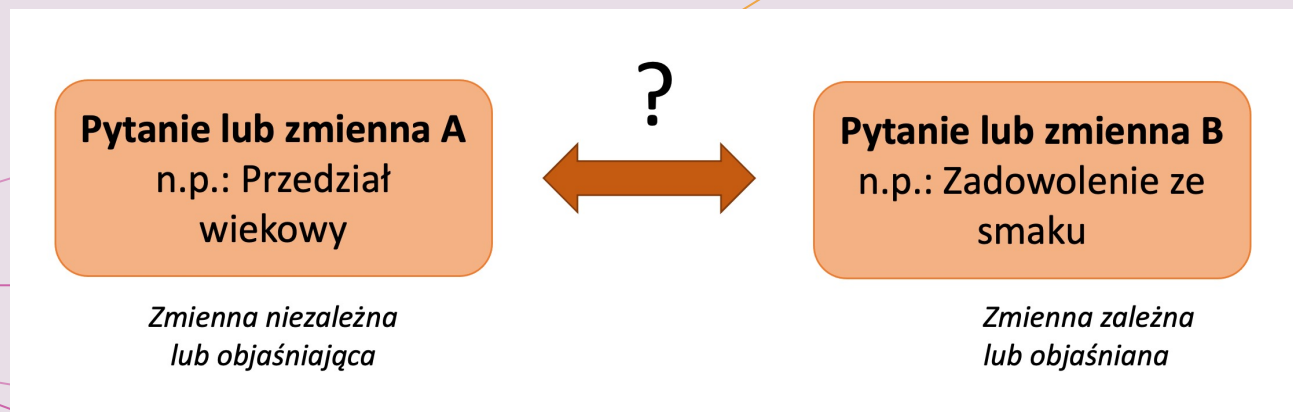
Spis treści

1. Zasady analizy krzyżowej
2. Odtworzenia wynikające z analizy krzyżowej
3. Przypadek zmiennych cyfrowych
4. Przypadek z pytaniami skalowymi
5. Reprezentacje graficzne



1. Zasady analizy krzyżowej

- Cel ogólny = sprawdzenie, czy zaobserwowane wyniki (sortowanie płaskie - analiza opisowa) różnią się w zależności od pewnych cech respondentów.
- Analizy krzyżowe (lub analizy dwuzmienne) polegają na jednoczesnym opracowywaniu dwóch pytań w celu określenia powiązań między odpowiedziami udzielonymi na jedno i drugie pytanie.
- Jest to określane z jednej strony jako zmienna niezależna lub objaśniająca, a z drugiej jako zmienna zależna lub objaśniana. Stanowią one dwa warunki wniosku:
 - JEŚLI wiek jest duży, TO prawdopodobnie satysfakcja jest duża.
 - JEŚLI zawód to "pracownik umysłowy", TO najbardziej prawdopodobnym miejscem zamieszkania jest "hotel".
 - I tak dalej
- Następnie bada się statystyczne powiązanie lub związek między A i B (należy pamiętać, że nie jest to wpływ A na B, ale raczej ich powiązanie).

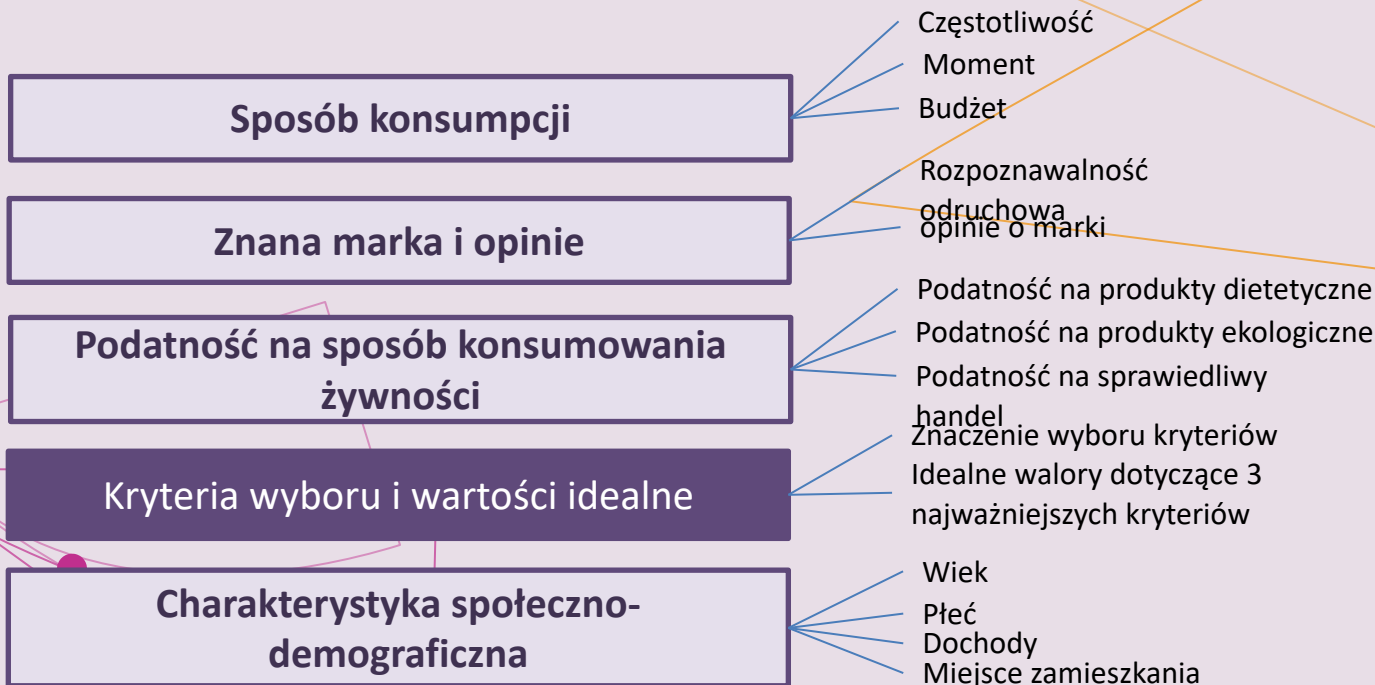


Analiza krzyżowa (lub dwuzmiennowa)



1. Zasady analizy krzyżowej

- **Zmienna niezależna/wyjaśniająca czy zmienna zależna/wyjaśniająca?**
- Generalnie, kwestionariusz został skonstruowany z myślą o analizach, które będą przeprowadzone.
- Każde pytanie w ankiecie ma wówczas status zmiennej objaśniającej lub zmiennej objaśnianej, lub obu, w zależności od **celów badania i pytań badawczych**.
- W przypadku badania w Equatorii (patrz sekwencje 1 i 2), głównymi zmiennymi objaśnianymi są pytania o renomę, opinie, a przede wszystkim **kryteria wyboru i wartości idealne**. Są one "sercem" ankiety lub pytaniami kluczowymi.
- Sposoby konsumpcji i cechy społeczno-demograficzne są zmiennymi objaśniającymi, ponieważ pozwolą określić, czy wyniki badania różnią się w zależności od wieku, płci, bardziej lub mniej regularnej konsumpcji itp.



2. Odtworzenia wynikające z analizy krzyżowej

W wierszu, zmienna objaśniająca

Treść dwóch pytań wzajemnie powiązanych

W kolumnie zmienna objaśniana

Countries x Shops

SHOPS

COUNTRY	SHOPS											
	H_ET_M		C_ET_A		DECATHLON		ZARA		BENETTON		CAŁKOWITY	
	N	%	N	%	N	%	N	%	N	%	N	%
Francja	17	28%	3	5%	47	77%	7	11%	0	0%	61	
Finlandia	37	97%	0	0%	0	0%	0	0%	1	3%	38	
Niemcy	38	81%	23	49%	0	0%	3	6%	1	2%	47	
Węgry	35	48%	47	64%	20	27%	15	21%	1	1%	73	
Włochy	8	16%	0	0%	4	8%	5	10%	41	80%	51	
Polska	13	50%	4	15%	5	19%	22	85%	1	4%	26	
SUMA	148	50%	77	26%	76	26%	52	18%	45	15%	296	

Wartość pracowników odpowiadająca dwóm warunkom (wiersz i kolumna) oraz % w rzędzie.

Analiza krzyżowa (lub dwuzmiennowa)

inx
ClassRoom

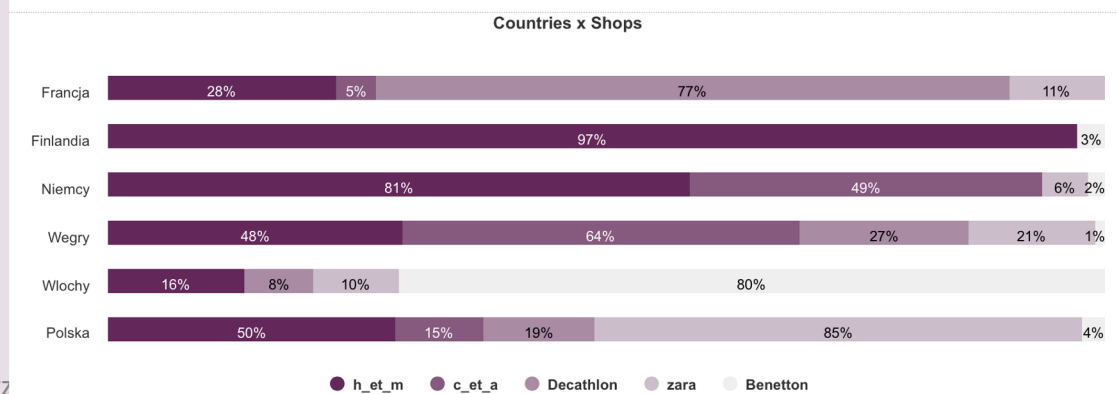
2. Odtworzenia wynikające z analizy krzyżowej

- **Kilka zasad:**
- Zmienna objaśniająca jest przeważnie umieszczona w szeregu, aby zapewnić lepszą czytelność od lewej do prawej.
- Wartości procentowe są wyświetlane w celu ułatwienia porównania między kategoriami. W razie potrzeby można je usunąć.
- Procenty edytowane są w szeregu, aby zaobserwować rozkład kategorii zmiennej objaśnianej w zależności od kategorii zmiennej objaśniającej.

- Powyższa tabela wykazuje, iż 77% francuskich konsumentów kupuje w Decathlonie w porównaniu z 26% respondentów ogółem.

- Wykres potwierdza ideę, szczególnie wysokiego udziału francuskich konsumentów w kupnie artykułów Decathlonu i że jest szczególnie wysoki.

Countries x Shops												
COUNTRY	SHOPS											
	H_ET_M		C_ET_A		DECATHLON		ZARA		BENETTON		CAŁKOWITY	
	N	%	N	%	N	%	N	%	N	%	N	%
Francja	17	28%	3	5%	47	77%	7	11%	0	0%	61	
Finlandia	37	97%	0	0%	0	0%	0	0%	1	3%	38	
Niemcy	38	81%	23	49%	0	0%	3	6%	1	2%	47	
Węgry	35	48%	47	64%	20	27%	15	21%	1	1%	73	
Włochy	8	16%	0	0%	4	8%	5	10%	41	80%	51	
Polska	13	50%	4	15%	5	19%	22	85%	1	4%	26	
SUMA	148	50%	77	26%	76	26%	52	18%	45	15%	296	



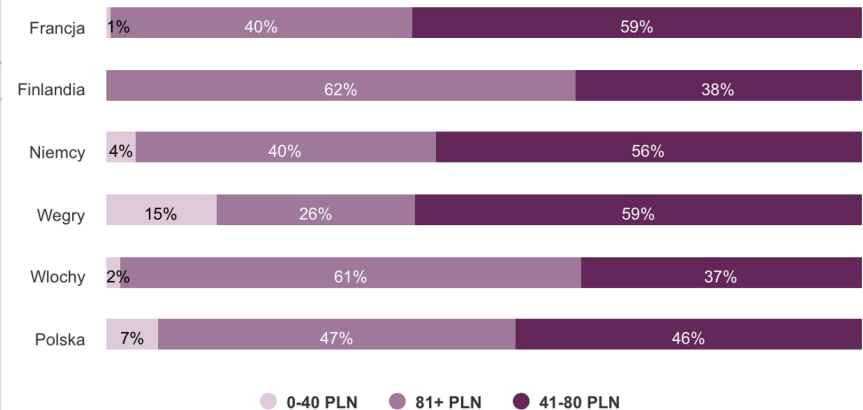
3. Przypadek zmiennych cyfrowych

- Wybierając zmienną cyfrową w analizie krzyżowej należy dokonać pewnych wyborów.
- Może być traktowana jako prawdziwa zmienna cyfrowa. W tym przypadku otrzymujemy **średnią** w zestawieniu krzyżowym.
- Można preferować **klasyfikacje**
- Następnie należy wybrać, jaka **klasyfikacja** będzie pożądana: n klasa na podstawie średniej (lub na podstawie odchylenia standardowego), albo n klasa o tej samej wielkości lub klasy i dostosowanych (patrz sekwencja N).
- Po dokonaniu klasyfikacji otrzymujemy tabelę lub wykres podobny do tabeli krzyżowej z dwóch pytań zamkniętych.

Age x Budget for pants

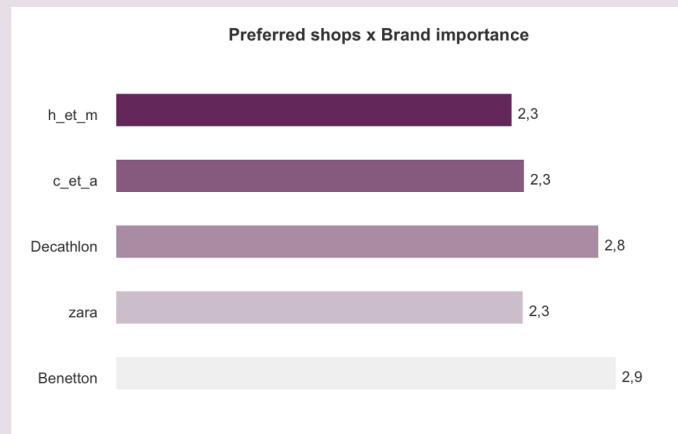
	ILE NAJCZĘŚCIEJ WYDAJE PAN/PANI NA ZAKUP SPODNI DLA SWOICH DZIECI?		
Francia	20,5	Francia	20,5
Finlandia	23,3	Finlandia	23,3
Niemcy	20,6	Niemcy	20,6
Węgry	16,7	Węgry	16,7
Włochy	24,2	Włochy	24,2
Polska	20,3	Polska	20,3
SUMA	21,2		

Age x Budget for pants

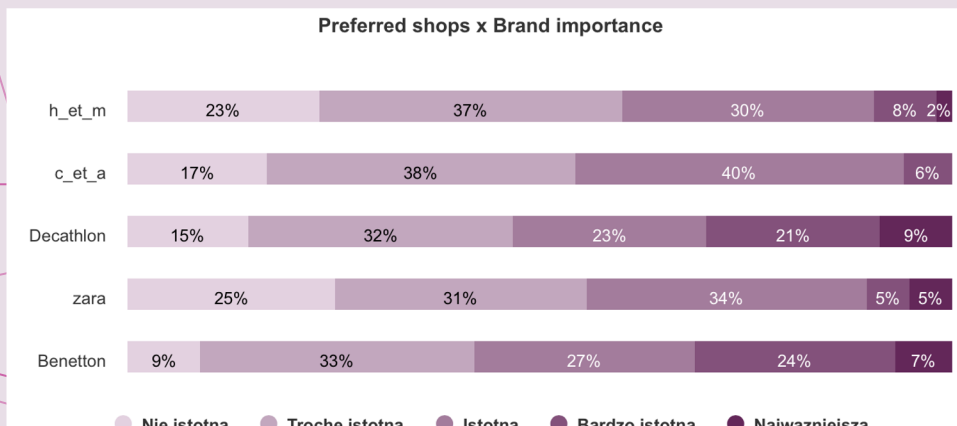


4. Przypadek z pytaniami skalowymi

- W przypadku pytań skalowych uwzględnionych w analizie krzyżowej również należy dokonać wyboru.
- Mogą one zostać przedstawione:
 - W postaci **cyfrowej** (średnia), zgodnie z przypisaną do nich **skalą** (por. sekwencja 2), domyślną lub własną
 - Można też potraktować je jako pojedyncze pytania **zamknięte**, przedstawiając rozkład odpowiedzi na wszystkich poziomach pytania.

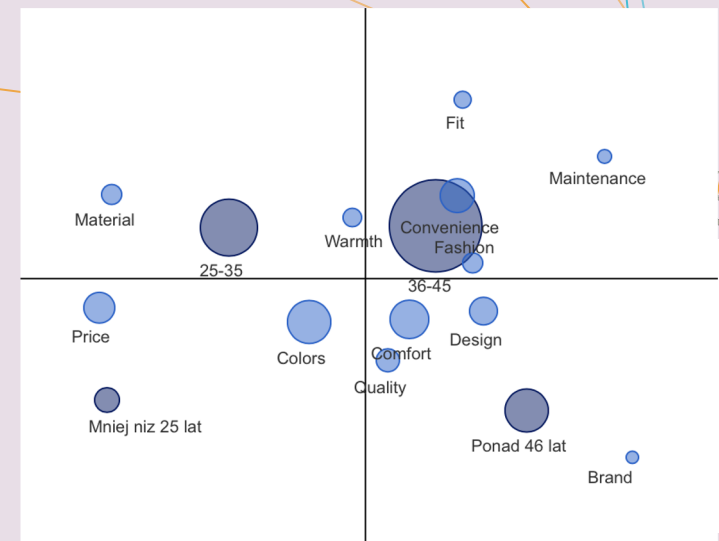
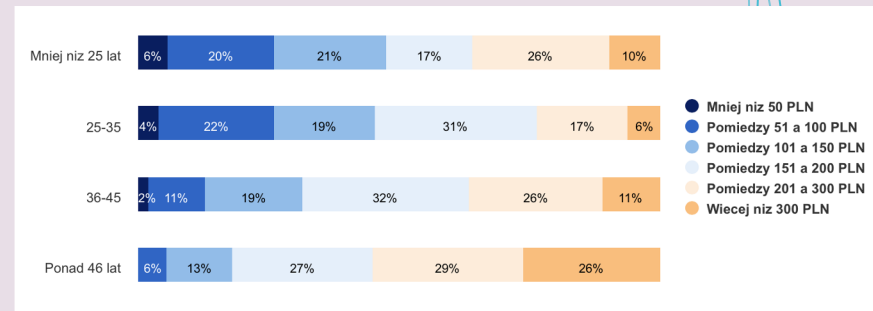
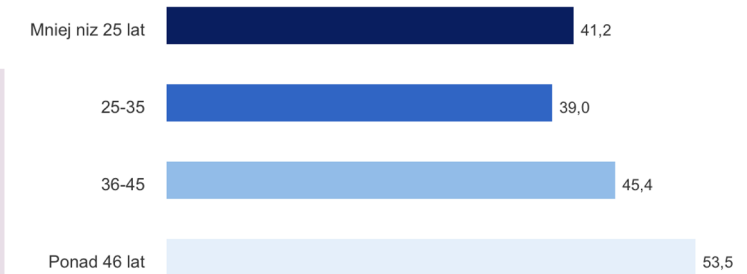


N.B. Skala zastosowana tutaj: 1 dla wcale nie jest ważne, a następnie 2; 3; 4 i 5 dla bardzo ważne



5. Reprezentacje graficzne

- Graficzna reprezentacja przetwarzania krzyżowego jest delikatna, ponieważ ilość informacji w zestawieniu krzyżowym może być duża.
- Dla krzyżowania zmiennej liczbowej lub skali przetwarzanej w postaci liczbowej, skuteczny jest **wykres słupkowy**.
- Dla skrzyżowania odpowiedzi na dwa pytania zamknięte,
 - Poniżej dwudziestu lub więcej kombinacji (liczba modalności w wierszach x liczba modalności w kolumnach), wykres słupkowy może pozostać czytelny.
 - Poza tym, zamiast tego zostanie użyta mapa czynnikowa (szczegóły w sekwencji 10).



Analiza krzyżowa (lub dwuzmienna)

VIII - Testy statystyczne dla analizy dwuzmiennej

Sphinx Classroom

pakiet Sphinx przeznaczony do nauki metod i techniki badań ankietowych.



Sphinx
ClassRoom

Spis treści

1. Kolejne etapy badania
2. Różne rodzaje analizy krzyżowej i odpowiadające im testy
3. Test chi-kwadrat
4. Analiza wariancji i test Fishera F
5. Regresja i korelacja



1. Kolejne etapy badania

- W przypadku analiz dwuzmiennych można stosować testy statystyczne aby stwierdzić obecność lub brak ważnego **powiązania** lub korelacji między dwoma zmiennymi lub dwoma kategoriami.
- Analiza krzyżowa (lub dwuzmienna) wymaga przeprowadzenia testu w trzech etapach:
 1. **Czy istnieje** znacząca korelacja między tymi zmiennymi?
=> Prowadzi to do konkluzji, iż wyniki pozwalają nam ustalić prawdopodobieństwa, które są wiarygodne.
 2. Jeśli tak, to czy związek ten jest przeciętny, **silny** czy bardzo silny?
=> Czy prawdopodobieństwo naszego wniosku jest umiarkowane czy wysokie?
 3. Jeśli tak, to **jak** to się przekłada na rzeczywistość?
=> Jest to najważniejszy etap, ponieważ chodzi o wskazanie, które cechy są powiązane: wiek/zadowolenie; lokalizacja/hotel, itp. Są to najbardziej **użyteczne** informacje uzyskane z tego typu analizy.



2. Różne rodzaje analizy krzyżowej i odpowiadające im testy

- Ta seria badań jest zależna od charakteru pytań - zmiennych - do krzyżowania.
- Stosuje się 3 metody badania krzyżowego w zależności od rodzaju zarejestrowanych odpowiedzi: zamknięte (zmienne nominalne) lub cyfrowe otwarte.

Przypadek	Przykład	Badanie, które należy zastosować	Test statystyczne potwierdzające zależność
Zamknięte x zamknięte	Płeć x uprawiany sport	Sortowanie krzyżowe	Chi-kwadrat
Zamknięty x cyfrowe	Płeć x Cel na dopłatę	Analiza wariancji	Test Fischer
Cyfrowe x Cyfrowe	Wiek x Cel na dopłatę	Regresja liniowa	Współczynnik korelacji

3. Test chi-kwadrat

• 3.1 Ogólne zasady testu

- Jest to test niezależności, w celu sprawdzenia, czy rozkład warunkowy zmiennej objaśniającej jest funkcją rozkładu warunków zmiennej objaśnianej.
- Punktem wyjścia jest tabela krzyżowa lub tabela kontyngencyjna.
- Na przykład, ankieta dotycząca zakupów odzieży dziecięcej przeprowadzona wśród konsumentów europejskich. Konfrontuje się dwa następujące pytania: płeć respondenta i ostatni produkt dla dzieci kupiony w Decathlonie. Chce się dowiedzieć, czy rodzaj kupowanego produktu różni się w zależności od płci respondenta.
- Z tabelą, z liczbami lub z wykresem nie można wnioskować "naukowo" bez użycia testu statystycznego.

Gender x last product bought at D

JEZELI TAK, JAKI PRODUKT PAN/I NABYŁA OSTATNIO W SKLEPIE DECATHLON ?	PLEC		
	KOBIETA	MEZCZYZNA	CAŁKOWITY
Kurtka	26	19	45
Polar	72	21	93
Koszulka sportowa	39	14	53
Spodnie	23	13	36
Buty	71	13	84
Czпка	3	7	10
Inny	48	13	61
SUMA	282	100	382

3. Test chi-kwadrat (c.d.)

- Test chi kwadrat polega na porównaniu między:
 - Liczebnościami rzeczywistymi (zgodnie z wynikami badania)...
 - Oraz liczebnościami teoretycznymi (które powinny być otrzymane w przypadku braku jakiegokolwiek związku między dwoma zmiennymi).

Gender x last product bought at D			
JEZELI TAK, JAKI PRODUKT PANI/ NABYŁA OSTATNIO W SKLEPIE DECATHLON ?	PLEC		
	KOBIETA	MEZCZYZNA	CAŁKOWITY
Kurtka	26	19	45
Polar	72	21	93
Koszulka sportowa	39	14	53
Spodnie	23	13	36
Buty	71	13	84
Czпка	3	7	10
Inny	48	13	61
SUMA	282	100	382

Liczebności rzeczywiste

≠

Gender x last product bought at D			
JEZELI TAK, JAKI PRODUKT PANI/ NABYŁA OSTATNIO W SKLEPIE DECATHLON ?	PLEC		
	KOBIETA	MEZCZYZNA	CAŁKOWITY
Kurtka	33,2	11,8	
Polar	68,7	24,3	
Koszulka sportowa	39,1	13,9	
Spodnie	26,6	9,4	
Buty	62,0	22,0	
Czпка	7,4	2,6	
Inny	45,0	16,0	
SUMA			

Liczebności teoretyczne

- Obliczenie liczebności teoretycznych: 73,82 % próby (n=282) stanowią kobiety (n=382). Gdyby nie istniała zależność między płcią, a rodzajem kupowanego produktu, 73,82% próby stanowiłyby kobiety w każdej kategorii produktowej.
- Np. 73,82% kobiet kupiłoby kurtkę, tj. 73,82% z 45 = 33,2. To samo dla 100 chłopców $\Leftrightarrow$ 26,18% próby; teoretycznie powinni oni reprezentować 26,18% kategorii kupujących kurtki, czyli 26,18% z 45 = 11,8, i tak dalej.



3. Test chi-kwadrat (c.d.)

- Test chi kwadrat = suma różnic między liczebnościami rzeczywistymi i teoretycznymi.
- Jeżeli różnice są bardzo małe, to zbliżamy się do przypadku, w którym nie ma związku między tymi dwoma zmiennymi. Im większe różnice, tym silniejszy związek między tymi dwoma zmiennymi.
- Wzór obliczeniowy:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}.$$

O_i – wartość mierzona,

E_i – wartość teoretyczna (oczekiwana) wynikająca z hipotezy odpowiadająca wartości mierzonej,

n – liczba pomiarów.

- W przykładzie ta formuła obliczeniowa jest zastosowana 14 razy (2 wiersze x 7 kolumn).
- Obliczone $\chi^2 = (33,2 - 26)^2 / 33,2 + (11,8 - 19)^2 / 11,8 + (68,7 - 72)^2 / 68,7 + \dots + (16 - 13)^2 / 16 =$
- $1,56 + 4,39 + 0,15 + \dots + 0,56 = \mathbf{24,12}$

3. Test chi-kwadrat (c.d.)

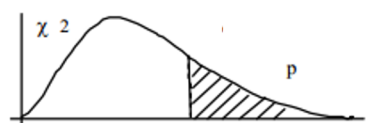
3.2. Pierwszy etap wnioskowania: zatwierdzenie lub odrzucenie związku

Aby ustalić, czy różnica ta jest znacząca, obliczony chi kwadrat należy porównać z wartością wzorcową, którą należy przekroczyć (uzyskaną zgodnie z prawami probabilistycznymi), czyli z teoretycznym chi kwadratem.

- Uzyskuje się to w tabeli według dwóch kryteriów (tabela dostępna np. pod [tym linkiem](#)):

- p = poziom błędu, domyślnie 0,05, co oznacza tylko 5% szans na wyciągnięcie błędnego wniosku;
- Stopnie swobody: (liczba wierszy - 1) x (liczba kolumn - 1). W naszym przypadku stopnie swobody wynoszą $(7 - 1) \times (2 - 1) = 6$.
- Tutaj wartość teoretycznego chi kwadrat, przy $p = 0,05$ i stopniach swobody = 6, wynosi **12,59**.

Fragment tabeli rozkładu chi-kwadrat



n P	0.10	0.05	0.02	0.01
1	2,706	3,841	5,412	6,635
2	4,605	5,991	7,824	9,210
3	6,251	7,815	9,837	11,341
4	7,779	9,488	11,668	13,277
5	9,236	11,070	13,388	15,086
6	10,645	12,592	15,033	16,812
7	12,017	14,067	16,622	18,475
8	13,362	15,507	18,168	20,090
9	14,684	16,919	19,679	21,666
10	15,987	18,307	21,161	23,209

Jeżeli obliczony chi kwadrat jest większy niż teoretyczny chi kwadrat, można stwierdzić, że istnieje związek między dwoma zmiennymi (na odpowiednim poziomie błędu). W naszym przykładzie mamy $24,12 > 12,59$. Wtedy możemy powiedzieć, z błędem 5%, że produkt zakupiony w Decathlonie różni się **znacząco** w zależności od płci.

3. Test chi-kwadrat (c.d.)

- **3.3. Wniosek Etap 2: Ustalenie siły związku lub jego wiarygodności**
- Przy teoretycznym chi kwadrat wynoszącym 5%, mamy 95% wiarygodny wniosek. Podniesienie poziomu precyzji pozwala na określenie (poprzez obniżenie p poziomu błędu), czy nasz wniosek jest nadal ważny, a więc ustalić poziom jego wiarygodności, czyli siłę.
- Siła związku jest tym bardziej gwarantowana, im bardziej poziom ufności jest bliski 100%. Poniższa tabela ilustruje, jak teoretyczne chi kwadrat wzrasta wraz ze wzrostem wymaganego poziomu wiarygodności; możemy zatem zobaczyć, od którego poziomu zależność pozostaje ważna, nadal przy 6 stopniach swobody.

p = poziom błędu	Wiarygodność wniosku	Chi kwadrat teoretyczny	Obliczony Chi kwadrat	Wniosek
0,05	95 %	12,59	24,12	Potwierdzony związek
0,02	98 %	15,03		Potwierdzony związek
0,01	99 %	16,81		Potwierdzony związek
0,005	99,5 %	18,55		Potwierdzony związek
0,001	99,9 %	22,46		Potwierdzony związek

Z tabeli wynika, że zależność pozostaje ważna na poziomie 98% (lub $p=,02$), a nawet na poziomie 99,9% (lub $p=,001$), gdzie obliczony chi kwadrat (24,12) jest zawsze wyższy niż teoretyczny chi kwadrat (22,46). Jest to więc bardzo wiarygodny wniosek.



3. Test chi-kwadrat (c.d.)

3.4 Wniosek Etap 3: Szczegóły dotyczące relacji

Następnie należy przyjrzeć się szczegółom relacji, aby ustalić, jak przekłada się ona na rzeczywistość. Należy się odnieść do chi kwadratów na jednostkę (lub chi kwadraty na parę), obliczonych dla każdej jednostki w tabeli, biorąc pod uwagę każdą kombinację.

N.p. dla "mężczyzna" i "kurtka", mamy obliczenie pary i uproszczoną tabelę z 2 x 2 modalnościami dla płci i zakupionego produktu. Na tym pojedynczym skrzyżowaniu 2 x 2 obliczany jest chi kwadrat. W tym przykładzie wynosi on 6,76 wobec teoretycznego chi kwadratu 3,84 (przy $p=0,05$). Następnie zatwierdzana jest zgodność między modalnościami "człowiek" i "kurtka". Nawet przy $p=0,01$, ponieważ teoretyczne χ^2 wynosi 6,64.

	Mężczyzna	Inny
Kurtka	19	26
Inny	81	256

Chi2=6,76 p=0,009

JEZELI TAK, JAKI PRODUKT PAN/I NABYŁA OSTATNIO W SKLEPIE DECATHLON ?	PLEC					
	KOBIETA			MEZCZYZNA		
	N	%	TEST STATYSTY...	N	%	TEST STATYSTY...
Kurtka	26	58%	- VS	19	42%	+ VS
Polar	72	77%		21	23%	
Koszulka sportowa	39	74%		14	26%	
Spodnie	23	64%		13	36%	
Buty	71	85%	+ S	13	15%	- S
Czпка	3	30%	- VS	7	70%	+ VS
Inny	48	79%		13	21%	

■ Poniżej reprezentowane elementy ■ Nad reprezentacyjne elementy

W tabelach krzyżowych Sphinx, pary modalności, dla których korelacja jest znacząca, mogą być zidentyfikowane, na niebiesko dla nadreprezentacji i na różowo dla podreprezentacji. VS dla bardzo istotnych (przy $p=0,01$) lub S dla istotnych (przy $p=0,05$). Tutaj w Decathlonie kobiety kupują buty, a mężczyźni kurtki i czapki.

4. Analiza wariancji i test Fishera F

4.1 Ogólne zasady analizy i badań

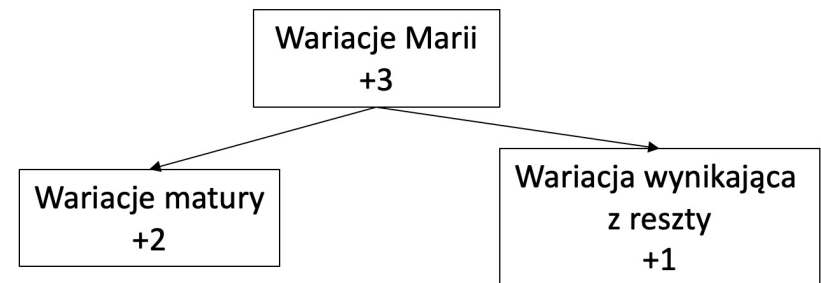
Analiza wariancji zajmuje się wariacjami zmiennej liczbowej (objaśnianej) w zależności od kategorii nominalnej zmiennej objaśniającej (lub pytania zamkniętego). Polega ona na rozłożeniu wariancji (wszystkich wariantów) zmiennej liczbowej na dwie części :

- Wariancja "międzygrupowa", mierzy różnice pomiędzy kategoriami pytań zamkniętych; wariancja ta wynika ze zmiennej objaśniającej, "czynnika", w słownictwie analizy wariancji.
- Wariancja "wewnątrzgrupowa", określa różnice w każdej kategorii pytania zamkniętego. Wariancja ta wynika z innych czynników, zwanych "resztami", które składają się ze wszystkich czynników innych niż zmienna objaśniająca.

Studentka Maria zdała maturę ścisłą, a jej ocena jest równa 14/20. Średnia klasy wynosi 11/20. Średnia klasy dla uczniów klasy maturalnej ścisłej wynosi 13/20. W tym przypadku rozkładamy wariancję Marii od średniej (+3) na dwie części:

- wariancja ze względu na maturę: (+2), ponieważ średnia z matury ścisłych jest o 2 punkty wyższa niż z całości;
- wariancja rezydualna, wynikająca z reszty: (+1), ponieważ Maria ma o jeden punkt wyższą niż średnia z maturzystów ścisłych.

- Maria zdała maturę ścisłą (S)
- Ocena od Marii
- Średnia ocena w klasie
- Średnia ocena maturzysty ścisłego (S)



4. Analiza wariancji i test Fishera F (c.d.)

• 4.2 Tabela analizy wariancji

- Za pomocą tabeli analizy wariancji oblicza się dwie wariancje: wariancję ze względu na czynnik oraz tzw. wariancję "resztową", a następnie porównuje się je z F od Fishera.

Wariancja	suma kwadratowa	Stopnie swobody	Średni kwadrat	F odFisher
Wynikająca z czynnika (zmienna objaśniająca)	SCf	$k - 1$	$CMf = SCf / k - 1$	CMf / CMr
Resztowy (wynikający z reszty)	SCr	$n - k$	$CMr = SCr / n - k$	

k = liczba modalności czynnika (zmiennnej objaśniającej); n = liczba obserwatorów.

Wymagane formuły:

$$SCf = \sum ((\sum x_{.j})^2 / n_i) - (\sum x_{ij})^2 / n$$

$\sum x_{.j}$ = suma j-tej modalności (kolumna)

$$SCr = \sum (x_{ij})^2 - \sum ((\sum x_{.j})^2 / n_i)$$

4. Analiza wariancji i test Fishera F (c.d.)

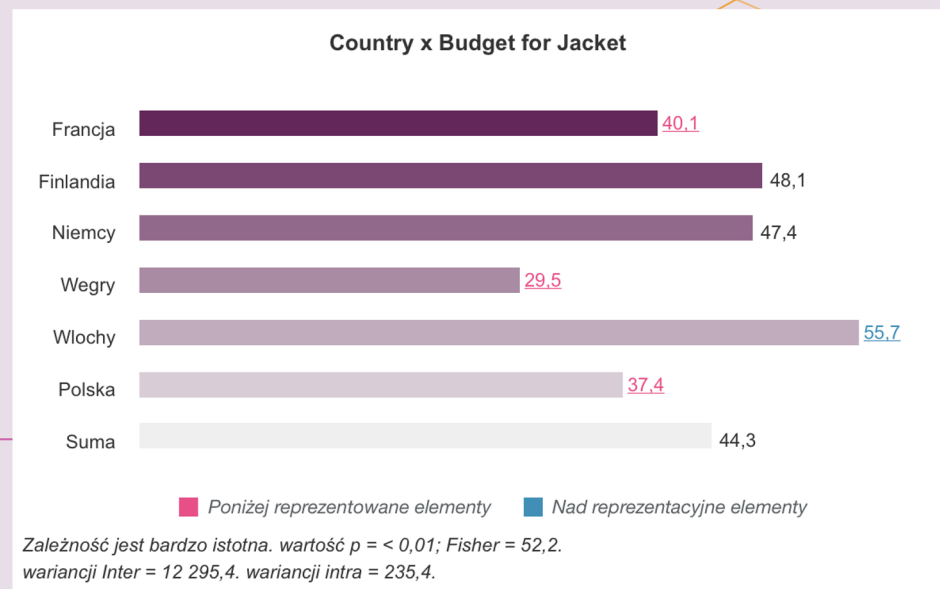
• 4.3 Test F Fishera

- Po zakończeniu tabeli analizy wariancji otrzymujemy "wyliczoną" wartość F Fishera, która jest stosunkiem wariancji ze względu na czynnik do wariancji resztowej.
- Zgodnie z tą samą zasadą, jak w przypadku innych testów statystycznych, uzyskaną wartość należy następnie porównać z wartością wzorcową, którą należy przekroczyć (uzyskaną zgodnie z prawami probabilistycznymi), czyli teoretyczną wartością F
- Teoretyczna wartość F znajduje się w odpowiedniej tabeli statystycznej (test Fishera - patrz [ten link](#)). Teoretyczne F uzyskuje się jako funkcję poziomu błędu (zwykle 0,05); z $V1$ lub $df1 = k-1$ i $V2$ lub $df2 = n-k$.
- W przypadku analizy wariancji i testu F wnioski są również wyciągane w trzech etapach:
 1. Czy istnieje związek między tymi dwoma zmiennymi? Tak, jeśli obliczone F jest większe niż teoretyczne F - dla poziomu błędu 5%.
 2. Jeśli TAK, to czy związek ten jest przeciętny, silny czy bardzo silny? Jeśli obliczone F jest domyślnie wyższe, gdy $p = 5\%$, to czy zawsze jest wyższe dla $p = 1\%$, potem $p = 0.5\%$, potem $p = 0.1\%$? Jeśli zależność jest zawsze potwierdzona, gdy p jest bliskie 0, jest tym bardziej wiarygodna i dlatego można ją zakwalifikować jako "silną" lub "bardzo silną".
 3. Jeśli TAK, to jak to się przekłada na rzeczywistość? Jakie rodzaje matur otrzymałyby średnie silniejsze lub słabsze w porównaniu do ogólnej średniej? Następnie przyglądamy się statystycznie różnym średnim według testu t-Sphinx (porównanie średnich).

4. Analiza wariancji i test Fishera F (c.d.)

4.4 Wyniki oprogramowania

- Wynik testu Fishera oraz poziom błędu są wyświetlane.
- Średnie statystycznie różniące się zaznaczone są kolorami: niebieskim (wyższe) lub różowym (niższe).
- W poniższym przykładzie budżet na zakup kurtek przez rodziców różni się znacznie w zależności od kraju.
- Konkretnie, średni budżet jest wyższy we Włoszech (średnia w kolorze niebieskim), podczas gdy niższy jest we Francji, w Polsce i na Węgrzech (średnia w kolorze różowym).



Testy statystyczne dla analizy dwuzmiennej

5. Regresja i korelacja

• Za pomocą tej techniki badamy związki między dwoma zmiennymi cyfrowymi. Należy określić, czy zmienna objaśniana y zmienia się jako funkcja innej zmiennej objaśniającej x .

•5.1 Linia regresji

• Szukamy linii współrzędnych $y = ax + b$, która opisuje jak y ulega zmianie jako funkcja x .

• Wzory potrzebne do obliczeń: $a = \text{Cov}(x,y)/V(x)$

• $\mathbf{a} = 1/n (\sum x_i y_i) - \bar{x} \bar{y} / 1/n (\sum x_i^2) - \bar{x}^2$ i $\mathbf{b} = \bar{y} - a \bar{x}$

• a jest nachyleniem (lub współczynnikiem sterowalności), b jest stałą (lub punktem przecięcia)

•5.2 Współczynnik korelacji

• Współczynnik korelacji r wskazuje siłę i związek. Mieści się w przedziale od -1 do +1. Blisko 1 (lub -1) wskazuje na bardzo silną korelację; blisko 0 odpowiada zerowej korelacji.

• $r = \text{Cov}(x,y) / \text{odchylenie standardowe}(x) \times \text{odchylenie standardowe}(y)$

• Aby oszacować, czy korelacja jest istotna, sprawdzamy, czy linia regresji ma znaczenie i czy można użyć jej współrzędnych, aby dobrze przewidzieć y jako funkcję x .

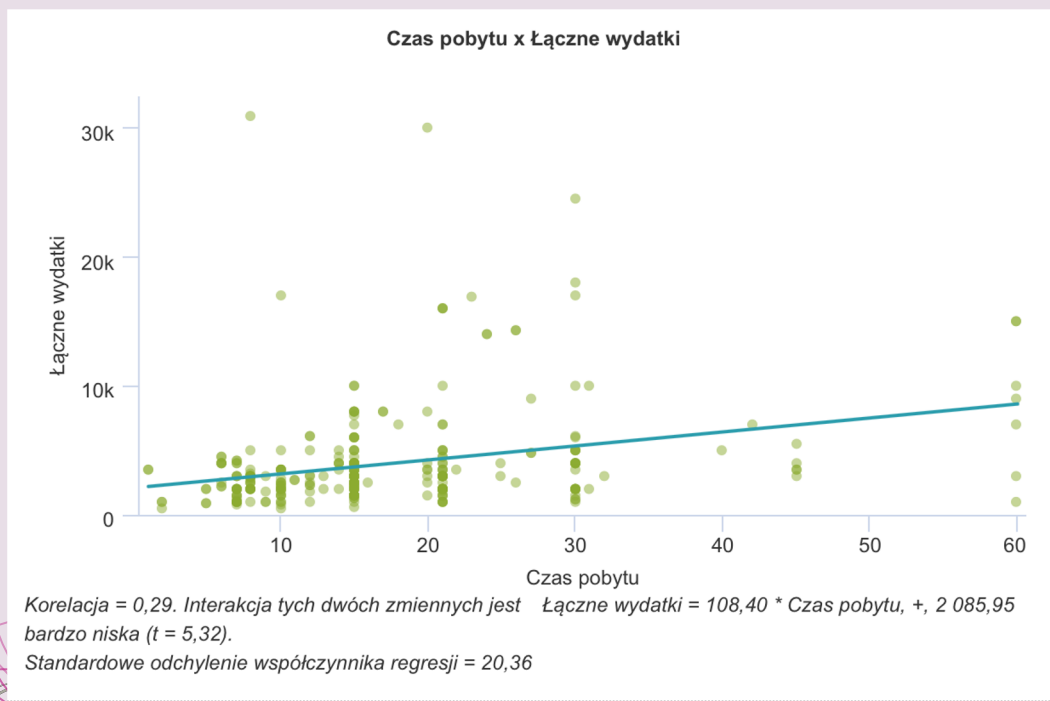
• Należy sprawdzić, czy współczynnik $a \neq 0$. Zna się jego wartość i odchylenie standardowe. Dzięki odchyleniu standardowemu a możemy ocenić p , czyli prawdopodobieństwo, że jest ono różne od 0. Jest to równoznaczne z określeniem prawdopodobieństwa błędu korelacji między x i y .



5. Regresja i korelacja (c.d.)

•5.3 Wnioski z korelacji i regresji

- 1. Czy współczynnik korelacji jest znacznie różny od 0? Jeśli tak, to istnieje związek między x i y
- 2. Jeśli tak, to jak silna jest ta korelacja? W naukach społecznych i humanistycznych, $|r| < 0.25$ jest słaby, $|r|$ między 0.25 a 0.5: średni; $|r|$ między 0.5 a 0.75: mocny; $|r|$ większy niż 0.75: bardzo mocny.
- 3. Linia regresji wskazuje, jak zachowuje się y jako funkcja x . Wiarygodność prognozy zależy od siły współczynnika korelacji.



W przykładzie:

Wydatki ogółem = $108,4 \times \text{długość pobytu} + 2\,086$

Linia regresji wskazuje, że wydatki rosną wraz z długością pobytu. Normalnie pobyt kosztuje 2000 jednostek monetarnych i dodaje się 108 jednostek monetarnych na dzień, aby uzyskać oszacowanie całkowitych wydatków.

Współczynnik korelacji wynosi 0,29. Prognoza wskazuje na niewielką korelację (ale różną od 0), czyli prognoza nie będzie zbyt wiarygodna.



Sphinx
ClassRoom

IX - Analiza danych tekstowych

Sphinx Classroom

pakiet Sphinx przeznaczony do nauki metod i techniki badań ankietowych.



Sphinx
ClassRoom

Spis treści

1. Zalety zbierania danych tekstowych w ankiecie
2. Przykłady służących do zbierania tekstów
3. Wykorzystanie odpowiedzi tekstowych
 - 3.1 Analiza leksykalna
 - 3.2 Kodowanie
 - 3.3 Ekstrakcja verbatimów



1. Zalety zbierania danych tekstowych w ankiecie

- Pytania otwarte pozwalają na zapisanie bardziej spontanicznych **odpowiedzi**.
- Pozwalają one na zmianę rodzaju odpowiedzi, czyli zapewniają bardziej **zróżnicowany** protokół.
- Jeśli respondent jest zaangażowany i/lub chętny, można zebrać bardzo **ciekawe**, czasem nieoczekiwane informacje, np. na temat uczuć, lub motywacji. W ten sposób wprowadzamy trochę **"jakości"** w podejściu kwestionariusza będącym ogólnie ilościowym.
- Verbatim (wyrazy wypowiedziane przez respondentów) pozwalają **"ożywić"** wyniki, które czasami są zbyt liczbowe.



2. Przykłady służących do zbierania tekstów

- **Komentarze/sugestie na zakończenie kwestionariusza:** Czy są jakieś uwagi, krytyki oraz propozycje na ten lub inny temat?
- **Inne, proszę określić:** Jaka inna(e) nowość(i) mogłaby(y) być interesująca(e)?
- **Zdanie do uzupełnienia:** Dla ciebie dawanie prezentów świątecznych jest...
- **Trzy słowa:** Podaj trzy słowa (lub trzy przymiotniki) określające Paris St Germain.
- **Słotwarzyszenie idei:** Jeśli powiem PSG, myślisz o... Jakby PSG byłby filmem, to byłoby to...
- **Dalsze działania w związku z pytaniem zamkniętym:** Jesteś niezadowolony z..., wyjaśnij dlaczego?
- **Narracja:**
 - Kiedy ostatnio robiłeś to lub kupowałeś tamto... jak poszło? Jakie to było uczucie?
 - Na przykład, czy pamiętasz ostatni film, który widziałeś w kinie i który Ci się podobał? W kilku słowach, dlaczego podobał Ci się ten film?

Wskazówki:

- Należy zaplanować liczbę wierszy odpowiadającą spodziewanej ilości odpowiedzi.
- W razie potrzeby, szczególnie w ważnych pytaniach otwartych, nie wahać się przed określeniem obowiązkowej odpowiedzi.
- Pytania otwarte nie stanowią więcej niż 10-15% ankiety

3. Wykorzystanie odpowiedzi tekstowych

- Aby analizować verbatimy, istnieją zasadniczo trzy podejścia:
 - **Synteza intelektualna:** na podstawie prawie całkowitej lektury verbatimów i swojej wiedzy na ten temat, analizant przeprowadza syntezę i identyfikuje n głównych tematów tekstu. Tradycja analizy jakościowej.
 - **Kodowanie:** czytanie verbatimów i kodowanie każdej odpowiedzi zgodnie z siatką kodową (lub code-book) wstępnie zdefiniowaną przez analityka. Analiza treści.
 - **Analiza leksykalna:** lokalizuje się słowa i zwroty cytowane w całym zbiorze, a następnie oddzielamy zawartości tematyczne.
- Dla drugiego i trzeciego podejścia zaproponowane zostaną tabele i/lub wykresy w celu poparcia wniosków.
- We wszystkich trzech przypadkach wyniki zostaną uzupełnione przez **verbatimy dobrane** tak, aby zilustrować każdą kategorię tematyczną.



3.1. Kodowanie

- Analiza treści poprzez kodowanie "ręczne" polega na przeczytaniu wszystkich istotnych verbatimów oraz zakodowaniu ich, zgodnie z mniej lub bardziej opracowanym schematem kodowym. Schemat ten, w zależności od pytań badawczych, został wcześniej ustalony przez analityka.
- Programy komputerowe często oferują specyficzne środowisko, gdzie wybiera się zmienną do zakodowania, definiuje się tabelę kodową, a następnie konsultuje verbatimy. Na koniec sporządza się kodowanie.
- Zmienna(-e) w ten sposób utworzona(-e) jest(są) następnie dostępna(-e) do analizy opisowej lub krzyżowej.

KODYFIKACJA - Analizowana zmienna: review (REVIEW)

“ Po raz kolejny moi goście korzystali z gościnności hotelu Sheraton w Warszawie. Szczególnie ujęło ich profesjonalne, bardzo osobiste i serdeczne (nie bezosobowe) podejście ze strony obsługi. serdecznie dziękuję Pani Ilonie z Guest Service Center za ogromną pomoc i cenne wskazówki dla moich gości. Z pewnością do Pani wróci życzliwość od ludzi..Byli bardzo z hotelu i obsługi zadowoleni.

 Zmienne podpisu

◀ 2 ▶ Niekodowana obserwacja ▶

1 obs. zakodowane / 197 nagrania wypełnione

Pokaż wyciągi ☐

NOWA KATEGORIA ⓘ

Category #1

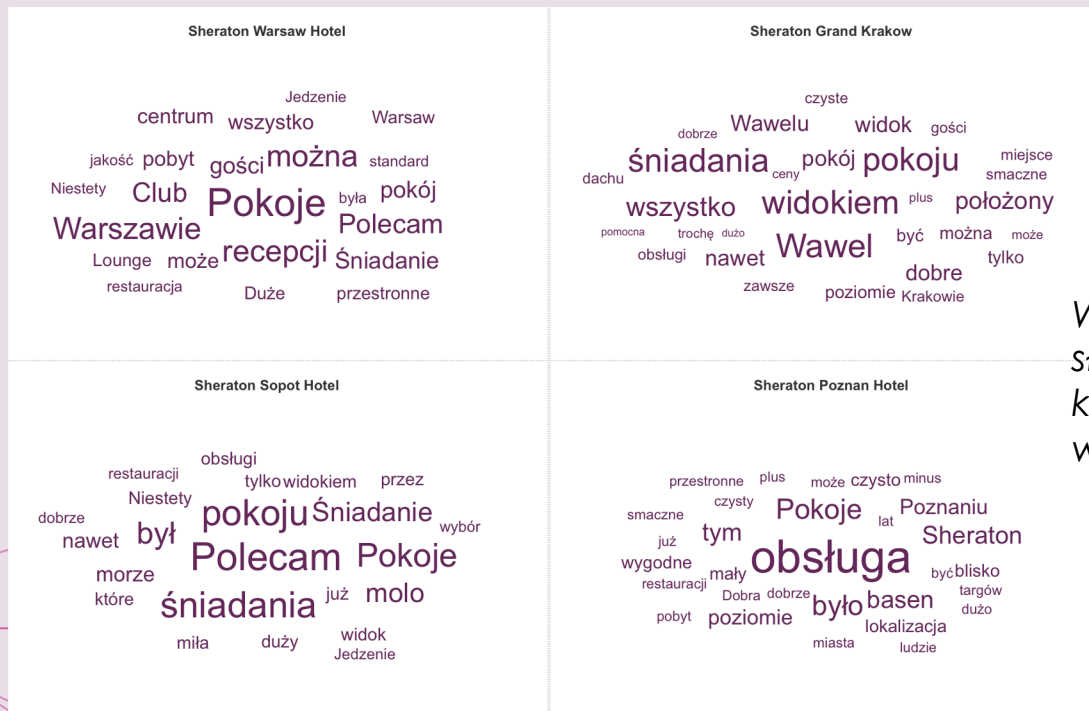
2 obs. zakodowane ✎ 🗑

☐ Restauracja 1 obs. zakodowane ☐ Śniadania 1 obs. zakodowane ☐ Rezerwacja 1 obs. zakodowane ☐ Wystrój 1 obs. zakodowane ☒ Obsługa 1 obs. zakodowane

+ Dodaj motyw

3.2 Analiza leksykalna

- **a. Chmury słów**
- Analiza leksykalna polega na odtworzeniu najczęściej używanych słów lub wyrażen cellem szybkiego wychwycenia ich treści w całym korpusie lub według kategorii respondentów.
- Listy słów kluczowych mogą być wyświetlane jako lista, tabela lub chmura słów (tag cloud).



W tym przykładzie, zauważamy, że słowa kluczowe używane przez klientów różnią się w zależności od wybranego hotelu.

3.2 Analiza leksykalna (c.d.)

- **b. Grupy tematyczne**
- Analiza leksykalna umożliwia również tworzenie grup tematycznych na podstawie **leksykonu**. Leksykon oznacza zbiór słów cytowanych w korpusie (zbiór odpowiedzi). Jest on ogólnie klasyfikowany według występujących słów.
- Następnie można określić **tematy** i powiązać słowa kluczowe z każdym tematem (patrz zrzut ekranu poniżej).
- Na końcu tej operacji można utworzyć **nową zmienną**, która rejestruje temat(y) obecny(e) w każdej odpowiedzi i staje się dostępna w badaniu analiz opisowych, krzyżowych lub nawet wielu zmiennych.

Analiza tematyczna : review - 197 nagrania wypelnione

Importuj Eksportuj Zapisz

Słowa i wyrażenia

Wyraz	Occ ↓	Obs	Kot
Nie	247	102	N
bardzo	195	108	N
się	171	83	N
hotelu	162	100	N
jest	149	82	N
Hotel	149	115	N
ale	89	59	N
jak	74	53	N
Pokoje	71	68	N
	56	50	N
	55	45	N
przy	54	44	N
tym	52	35	N
było	52	33	N
pokoju	50	38	N
Sheraton	49	37	N
można	48	39	N
mnie	40	31	N
pokój	39	32	N
tego	37	31	N
jednak	37	24	N
był	35	26	N

Struktura tematyczna

Kategorie tematyczne

NOWA KATEGORIA

Nowa kategoria 1

Usługa

obsługa obsługi obsługa

Śniadanie

śniadania Śniadanie śniadaniami

Lokalizacja

lokalizacja lokalizacji lokalizacje

Widok

widokiem widok widoki

+ Dodaj motyw

Verbatim - 197 nagrania(e)

Tryb wyświetlania Zdania

Pokaż orientację

n° 1 W Warszawie z reguły zatrzymujemy się z mężem w tym hotelu.

n° 1 Parking bezpłatny obok.

n° 1 Świetna restauracja hotelowa In Azia.

n° 1 Super śniadania.

n° 1 Niezwykle pomocna uprzejma obsługa.

n° 1 Podoba mi się wystrój pokoju.

n° 1 Jedynie miejsca przy umywalce mogłoby być nieco więcej.

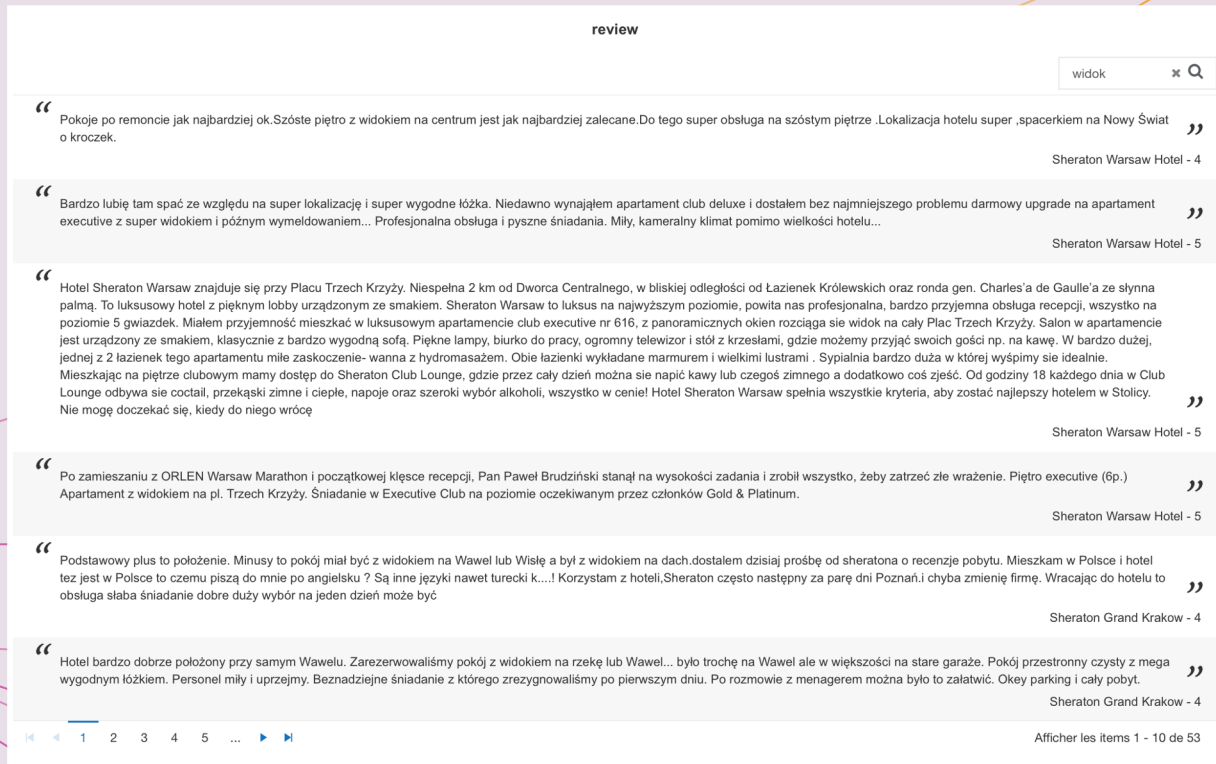
n° 1 Nie rezerwuję telefonicznie-ponieważ centrala mieści się w Londynie z tego co pamiętam i wiecznie są potem problemy z tą rezerwacją.

n° 1 I to jest moim zdaniem jeden poważny minus-reszta podoba mi się :

Verbatim

3.3 Rejestracja verbatimów

- Ekstrakcja verbatimów może być dokonywana na podstawie zapytań:
 - Verbatimy zawierające określone słowo kluczowe (jak na poniższym zrzucie ekranu z słowem "**widok**")
 - Według profilu respondenta (mężczyźni, kobiety, osoby powyżej określonego wieku itp.)
- Często możliwe jest również uzupełnienie tekstu o jedną lub więcej zmiennych " podpis", które wskazują np.
 - kto (wiek, płeć, itp.),
 - lub jakiego obiektu dotyczy (w tym przykładzie: odwiedzony hotel)



review

widok x Q

“ Pokoje po remoncie jak najbardziej ok.Szóste piętro z widokiem na centrum jest jak najbardziej zalecane.Do tego super obsługa na szóstym piętrze .Lokalizacja hotelu super ,spacerkiem na Nowy Świat o kroczeek. ”

Sheraton Warsaw Hotel - 4

“ Bardzo lubię tam spać ze względu na super lokalizację i super wygodne łóżka. Niedawno wynajęłam apartament club deluxe i dostałam bez najmniejszego problemu darmowy upgrade na apartament executive z super widokiem i późnym wymeldowaniem... Profesjonalna obsługa i pyszne śniadania. Miły, kameralny klimat pomimo wielkości hotelu... ”

Sheraton Warsaw Hotel - 5

“ Hotel Sheraton Warsaw znajduje się przy Placu Trzech Krzyży. Niespełna 2 km od Dworca Centralnego, w bliskiej odległości od Łazienek Królewskich oraz ronda gen. Charles'a de Gaulle'a ze słynna palmą. To luksusowy hotel z pięknym lobby urządzonym ze smakiem. Sheraton Warsaw to luksus na najwyższym poziomie, powita nas profesjonalna, bardzo przyjemna obsługa recepcji, wszystko na poziomie 5 gwiazdek. Miałem przyjemność mieszkać w luksusowym apartamencie club executive nr 616, z panoramicznych okien rozciąga się widok na cały Plac Trzech Krzyży. Salon w apartamencie jest urządzony ze smakiem, klasycznie z bardzo wygodną sofą. Piękne lampy, biurko do pracy, ogromny telewizor i stół z krzesłami, gdzie możemy przyjąć swoich gości np. na kawę. W bardzo dużej, jednej z 2 łazienek tego apartamentu mile zaskoczenie- wanna z hydromasażem. Obie łazienki wykładane marmurem i wielkimi lustrami . Sypialnia bardzo duża w której wyśpiemy się idealnie. Mieszkając na piętrze klubowym mamy dostęp do Sheraton Club Lounge, gdzie przez cały dzień można się napić kawy lub czegoś zimnego a dodatkowo coś zjeść. Od godziny 18 każdego dnia w Club Lounge odbywa się coctail, przekąski zimne i ciepłe, napoje oraz szeroki wybór alkoholi, wszystko w cenie! Hotel Sheraton Warsaw spełnia wszystkie kryteria, aby zostać najlepszym hotelem w Stolicy. Nie mogę doczekać się, kiedy do niego wrócę ”

Sheraton Warsaw Hotel - 5

“ Po zamieszaniu z ORLEN Warsaw Marathon i początkowej kłęsce recepcji, Pan Paweł Brudziński stanął na wysokości zadania i zrobił wszystko, żeby zatrzeć złe wrażenie. Piętro executive (6p.) Apartament z widokiem na pl. Trzech Krzyży. Śniadanie w Executive Club na poziomie oczekiwanym przez członków Gold & Platinum. ”

Sheraton Warsaw Hotel - 5

“ Podstawowy plus to położenie. Minusy to pokój miał być z widokiem na Wawel lub Wisłę a był z widokiem na dach.dostałem dzisiaj prośbę od sheratona o recenzje pobytu. Mieszkam w Polsce i hotel też jest w Polsce to czemu piszą do mnie po angielsku ? Są inne języki nawet turecki k....! Korzystam z hoteli,Sheraton często następny za parę dni Poznań.i chyba zmienię firmę. Wracając do hotelu to obsługa słaba śniadanie dobre duży wybór na jeden dzień może być ”

Sheraton Grand Krakow - 4

“ Hotel bardzo dobrze położony przy samym Wawelu. Zarezerwowaliśmy pokój z widokiem na rzekę lub Wawel... było trochę na Wawel ale w większości na stare garaże. Pokój przestronny czysty z mega wygodnym łóżkiem. Personel miły i uprzejmy. Beznadziejne śniadanie z którego zrezygnowaliśmy po pierwszym dniu. Po rozmowie z menagerem można było to załatwić. Okey parking i cały pobyt. ”

Sheraton Grand Krakow - 4

Afficher les items 1 - 10 de 53



X - Podstawy analizy wielu zmiennych

Sphinx Classroom

pakiet Sphinx przeznaczony do nauki metod i techniki badań ankietowych.



Sphinx
ClassRoom

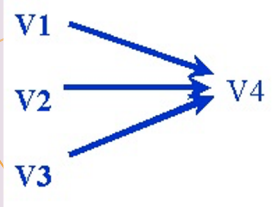
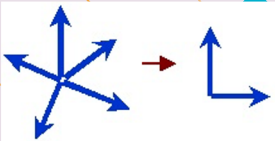
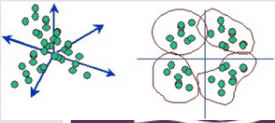
Spis treści

1. Cele i korzyści wynikające z analiz wielu zmiennych
2. Analiza regresji wielorakiej
3. Analiza czynnikowa korespondencji
4. Automatyczna klasyfikacja / typologia



1. Cele i korzyści wynikające z analiz wielu zmiennych

- Analizy wielowariantowe zasadniczo spełniają trzy różne cele, które przedstawiono w poniższej tabeli. Następnie przedstawi się podgląd trzech technik: po jednej dla każdej z głównych kategorii.

Cele	Zasady	Techniki wielu zmiennych	Przykłady
Wyjaśnienie	Podobnie jak w przypadku analiz dwuzmiennych lub sortowania krzyżowego, chodzi o to, aby zrozumieć, jak na daną zmienną może wpływać seria innych zmiennych: które z nich mają największy wpływ, które nie mają wpływu, itd.	- Regresja wieloraka. - Macierz ważności-wydajności - Drzewa decyzyjne	
Synteza	Celem jest uproszczenie danych poprzez zidentyfikowanie zmiennych lub modalności, które są podobne i które mierzą to samo. Te grupy "skorelowanych" zmiennych mogą być zidentyfikowane, aby otrzymać wyniki bardziej syntetyczne oraz prostsze.	- Analiza czynnikowa składowych głównych - Analiza czynnikowa korespondencji	
Klasyfikacja	Metody klasyfikacji skupiają się nie na zmiennych, ale na osobach. W celu uproszczenia, chodzi o identyfikację grup respondentów o podobnych zachowaniach.	- Analizy typologiczne lub klasyfikacje	

2. Analiza regresji wielorakiej

- Jest to rozszerzenie "prostej" analizy regresji opisanej w sekwencji 10. Zamiast analizować dwie zmienne, badane będą 3, 4 lub więcej zmiennych. Zmienne mają być cyfrowe (otwarte cyfrowe lub skalowe).
- Celem jest zrozumienie, w jaki sposób zmienna **y** jest wyjaśniana w funkcji kilku potencjalnie objaśniających zmiennych **x1, x2 i x3 itd.**
- Wyniki analizy regresji wielorakiej wskazują przede wszystkim:
 - Które zmienne potencjalne są istotnie skorelowane ze zmienną objaśnianą, a które nie są
 - Współczynnik korelacji wielokrotnej modelu, będące jego ogólną jakością (bliski) oraz jego możliwością dobrego przewidywania Y w funkcji X-ów.
 - Stała, a w szczególności wartość współczynników regresji dla każdej z zastosowanych zmiennych objaśniających.
 - Możemy zatem wnioskować o względnej wadze każdej zmiennej objaśniającej w jej wpływie na zmienną objaśnianą.



2. Analiza regresji wielorakiej (c.d.)

Współczynnik korelacji wielokrotnej modelu = jakość modelu

Równanie linii regresji wielorakiej z współczynnikami i z stałą

Wielokrotna regresja liniowa

Równanie modelu

Wynik egzaminu statystycznego = $-3.25 + 0.41 \times \text{Liczba podjętych kursów statystycznych} + 0.33 \times \text{Ocena matury z matematyki} + 0.31 \times \text{Liczba godzin przygotowań do egzaminu}$

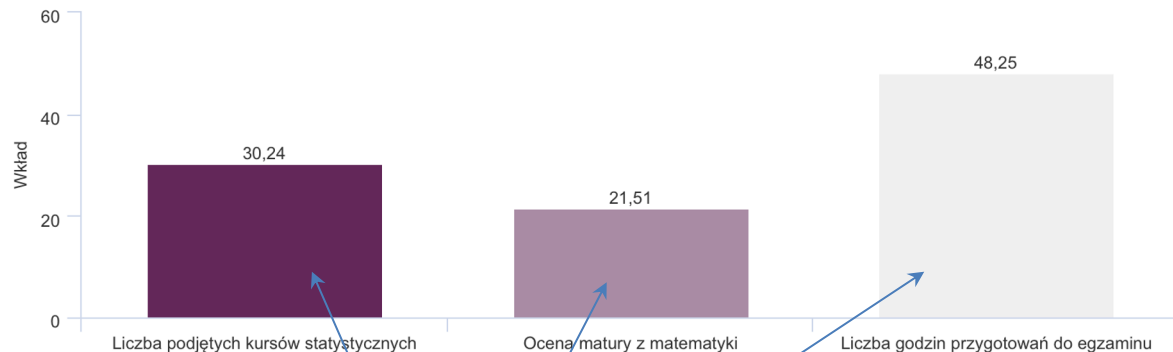
Model ten stanowi 81% variancji zmiennej zostanie wyjaśnione.

Wielokrotny Współczynnik korelacji $R = 0.9$

Współczynnik Fisher $C = 37.48$

P-value $R_p(r) = 0$

P-value $F_p(F) = 0$



Regresja liniowa polega na wytworzeniu równania do obliczania tak dokładnie, jak to możliwe, wartości zmiennej by została wyjaśniona. Możemy zatem określić:

1 / zmiennych, które nie mają żadnego wpływu (Ocena matury z matematyki w progu 5%).

2 / wkład innych zmiennych. Mierzy ich wpływ na zmienną objaśnianą.

[Szczegółowy wyniki](#)

W tym przykładzie, można bardzo dobrze przewidzieć y (ocena statystyczna) jako funkcję x_1 (liczba godzin zajęć), x_2 (liczba godzin przygotowań do egzaminu) i x_3 (ocena z matematyki na maturze). Wpływ korekty jest silniejszy niż wpływ frekwencji na zajęciach i poziomu matematyki na maturze.

Wagi relatywne każdej zmiennej objaśniającej

3. Analiza czynnikowa korespondencji

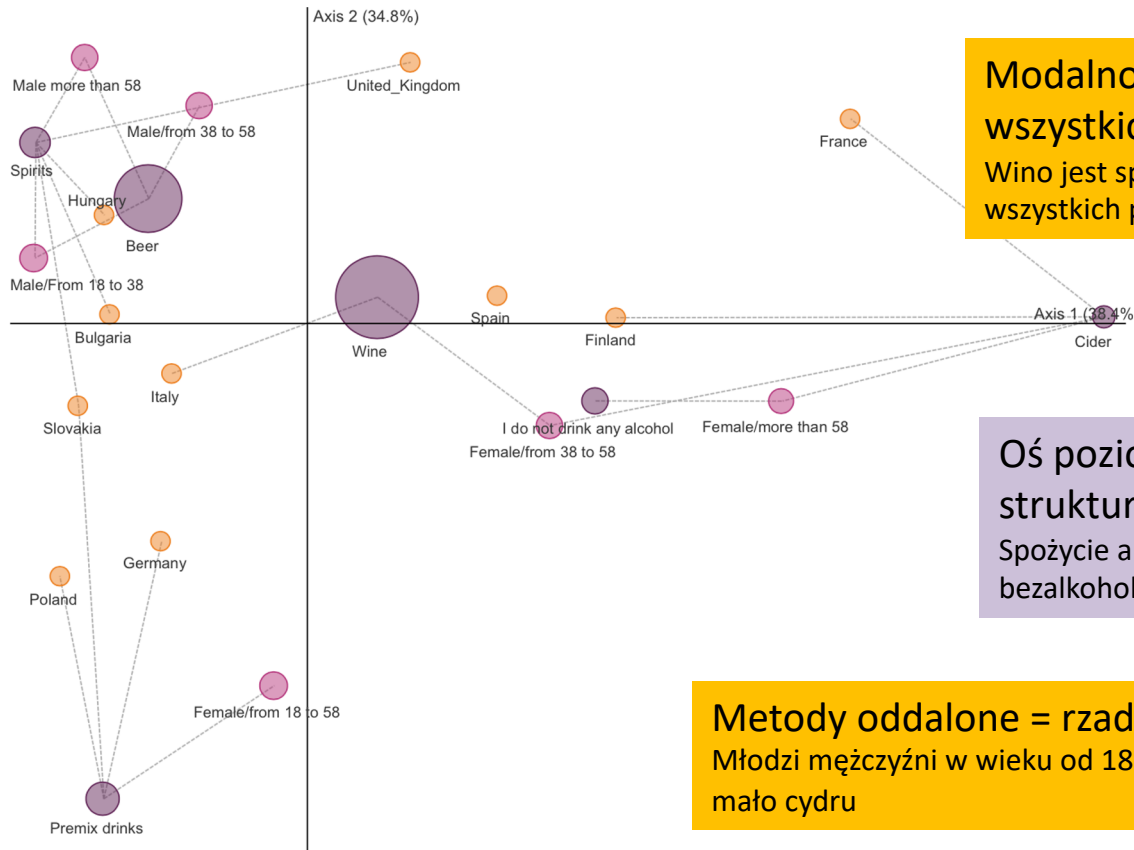
- Analiza czynnikowa korespondencji dotyczy odpowiedzi na pytania zamknięte (lub tzw. zmienne nominalne). Interesuje ją przede wszystkim **korespondencja** między **modalnościami** różnych badanych zmiennych.
- Proponuje ona kombinacje często powiązanych modalności ("**czynniki**"), które mogą być interpretowane jako duże wymiary strukturyzujące dla badanego tematu.
- Główną restytucją analizy czynnikowej korespondencji jest mapa, na której można wizualizować **skojarzenia** (i przeciwieństwa) między modalnościami zgodnie z ich **bliskością** (i odległością) na mapie.
- **Osie** poziome i pionowe - w zależności od umiejscowionych na nich modalności - mogą pozwolić na interpretację głównych wymiarów strukturyzujących badany przedmiot.
- Analiza czynnikowa korespondencji może być przeprowadzona na podstawie liczenia na płasko wielokrotnego pytania zamkniętego, sortowania krzyżowego lub analizy wielu zmiennych (gdzie wskazane są zmienne, które należy wziąć pod uwagę).

3. Analiza czynnikowa korespondencji (c.d.)

Bliskie modalności = często połączone.

Mężczyźni w wieku powyżej 58 lat często spożywają napoje spirytusowe

Kraj x Napoje alkoholowe x Wiek i płeć



Modalności w centrum = wspólne dla wszystkich

Wino jest spożywane we wszystkich krajach i we wszystkich profilach konsumentów.

Oś pozioma = główny wymiar strukturyzujący odpowiedzi

Spożycie alkoholu po lewej stronie, lekkie lub bezalkoholowe po prawej stronie

Metody oddalone = rzadko związane

Młodzi mężczyźni w wieku od 18 do 38 lat spożywają mało cydru

The correspondence map renders 73.3% of information, divided into 38.4% horizontally (F1) and 34.8% vertically (F2). The proximity or the distance between elements visualizes the associations over or under-represented.

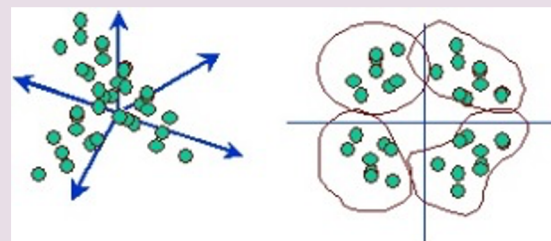
[Detailed results](#)



Sphinx
ClassRoom

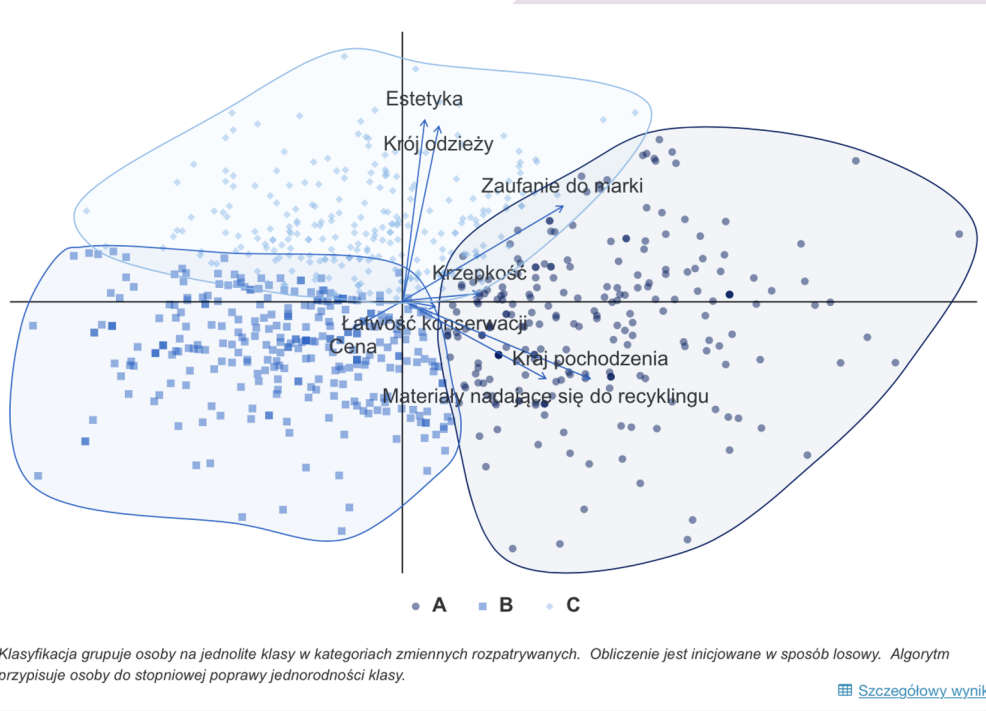
4. Automatyczna klasyfikacja / typologia

- Identyfikacja grup udzielających podobnych odpowiedzi w zakresie zdefiniowanych zmiennych (np. na potrzeby segmentacji w marketingu).
- Istnieje wiele metod, takich jak klasyfikacja hierarchiczna (rosnąca lub malejąca) lub metoda centrów mobilnych, którą zostanie omówiona.
- Metoda = proces kilkakrotnie inicjowany w sposób losowy, dający kolejno daje kilka rozwiązań, zachowując zawsze najlepsze (według kryteriów różnicowania i jednorodności grup).
- Następnie oprogramowanie (dzięki metodzie k-średnich) wskazuje zidentyfikowane przez siebie klasy, ich liczbę oraz wartości na użytych zmiennych. Pozwala to na ich interpretację.
- Następnie można zapisać klasę, do której należy każda osoba i porównać ją z innymi zmiennymi badania.



4. Automatyczna klasyfikacja / typologia

Przedstawienie klas na pierwszym planie czynnikowym zmiennych



- Ankieta: Junior Textile (zakupy odzieży dziecięcej) Klasyfikacja k-średnich przeprowadzona na zmiennych mierzących zgłaszane znaczenie czynników wpływających na zakup.

Cena: Łatwość konserwacji; Estetyka; Krój odzieży; Materiały nadające się do recyklingu; Krzepkość; Zaufanie do marki; Kraj pochodzenia.

Wyróżnia się trzy grupy:

- Klasa A: silna w zakresie materiałów nadających się do recyklingu, duża wytrzymałość, marka i kraj pochodzenia = **osoby racjonalne**
- Klasa B: szczególnie związana z ceną = **osoby ekonomiczne**.
- Klasa C: estetyka i krój ubrań: **osoby z stylami**.

Średnia wartość zmiennych dla każdej klasy.

(na zielono znacznie powyżej średniej, na czerwono poniżej - zgodnie z testem t na poziomie 0,05%)

Klasa	Cena	Łatwość konserwacji	Estetyka	Krój odzieży	Materiały / recyklingu	Krzepkość	Zaufanie do marki	Kraj pochodzenia
Racjonalny	4,33	6,33	5,92	6,05	4,85	7,05	5,65	6,25
Ekonomiczny	5,28	6,06	4,93	4,89	2,25	5,88	1,81	2,40
Stylamy	4,44	5,78	7,81	8,01	2,00	6,28	3,93	2,18
Całkowity	4,73	6,05	6,14	6,23	2,94	6,35	3,62	3,47