

Analiza danych ankietowych

Sprawozdanie 1

Weronika Jaskiewicz

Weronika Pyrtak

Spis treści

1	Część I	2
1.1	Zadanie 1	2
2	Zadanie 2	12
3	Zadanie 6	12
4	Zadanie 7	12
5	Zadanie 9	13

1 Część I

1.1 Zadanie 1

W pewnej dużej firmie technologicznej przeprowadzono ankietę mającą na celu ocenę skuteczności programów szkoleniowych dla pracowników. Wzięło w niej udział 200 losowo wybranych osób (losowanie proste ze zwracaniem).

1.1.1 Zadanie 1.1

Wczytywanie danych z pliku ankieta.csv.

Powyższe dane zawierają 200 wierszy oraz 8 kolumn.

Następnie sprawdzono typy przyjmowanych zmiennych.

Tabela 1: Typy zmiennych

DZIAŁ	character
STAŻ	integer
CZY_KIER	character
PYT_1	integer
PYT_2	integer
PYT_3	integer
PŁEĆ	character
WIEK	integer

Zamieniono zmienne o typie *character* na typ *factor*.

Przeszukano zbiór pod względem braków danych.

Liczba wartości brakujących wynosi: 0

Sprawdzono, czy typy zmiennych zostały prawidłowo rozpoznane.

1. zmienne ilościowe (typ numeric)

Tabela 2: Zmienne ilościowe

STAŻ	2
PYT_1	4
PYT_2	5
PYT_3	6
WIEK	8

Liczba zmiennych ilościowych: 5

2. zmienne jakościowe (typ factor)

Tabela 3: Zmienne jakościowe (factor)

DZIAŁ	1
CZY_KIER	3
PŁEĆ	7

Liczba zmiennych jakościowych (typ factor): 3

1.1.2 zadanie 1.2

Utworzono zmienna “WIEK_KAT” przeprowadzając kategoryzację zmiennej “WIEK” korzystając z następujących przedziałów do 35 lat, między 36 a 45 lat, między 46 a 55 lat, powyżej 55 lat.

1.1.3 zadanie 1.3

Sporządzono tablice licznosci dla zmiennych: DZIAŁ, STAŻ, CZY_KIER, PŁEĆ, WIEK_KAT.

Tabela 4: Tablica ilości dla zmiennej DZIAŁ

	HR	IT	MK	PD
Ilość	31	26	45	98

Tabela 5: Tablica ilości dla zmiennej STAŻ

	<1 rok	1-2 lata	3+ lat
Ilość	41	140	19

Tabela 6: Tablica ilości dla zmiennej CZY_KIER

	TAK	NIE
Ilość	173	27

Tabela 7: Tablica ilości dla zmiennej PŁEĆ

	Kobieta	Mężczyzna
Ilość	71	129

Tabela 8: Tablica ilości dla zmiennej WIEK_KAT

	0-35	36-45	46-55	55+
Ilość	26	104	45	25

WNIOSKI:

Największą grupę pracowników stanowi dział PD (98 osób), a najmniejszą IT (26 osób). Działy HR i MK mają pośrednie wartości (odpowiednio 31 i 45 osób). Może to wskazywać na różne zapotrzebowanie na pracowników w poszczególnych działach.

Większość pracowników (140 osób) ma od 1 do 2 lat stażu. 41 osób pracuje krócej niż rok, a tylko 19 osób ma staż powyżej 3 lat. Wskazuje to na dużą rotację pracowników lub na to, że firma stosunkowo niedawno zatrudniła większość obecnej kadry.

Większość pracowników (173 osoby) pełni funkcje kierownicze, a jedynie 27 osób nie. Może to oznaczać, że w firmie jest dużo stanowisk kierowniczych lub że definicja „kierownika” obejmuje szerokie spektrum stanowisk.

W firmie przeważają mężczyźni (129 osób) nad kobietami (71 osób). Może to wynikać z charakteru działalności firmy lub preferencji rekrutacyjnych.

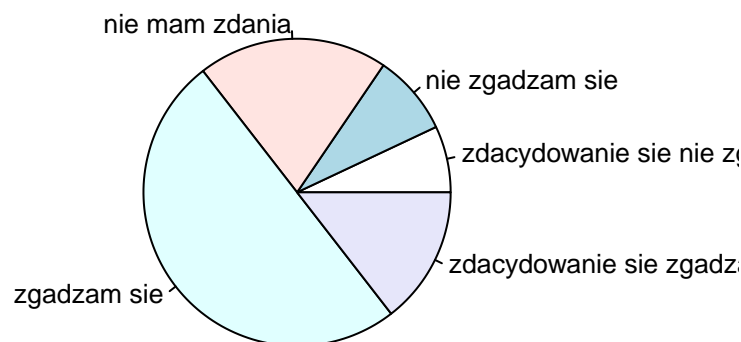
Najwięcej pracowników jest w grupie wiekowej 36-45 lat (104 osoby). Pozostałe grupy wiekowe mają znacznie mniejszą reprezentację.

5 zmiennych ilościowych (np. staż, wiek, odpowiedzi na pytania PYT_1–PYT_3). 4 zmienne jakościowe, np. dział, płeć, kategoria wieku. Brak wartości brakujących wskazuje na kompletność i dobrą jakość danych.

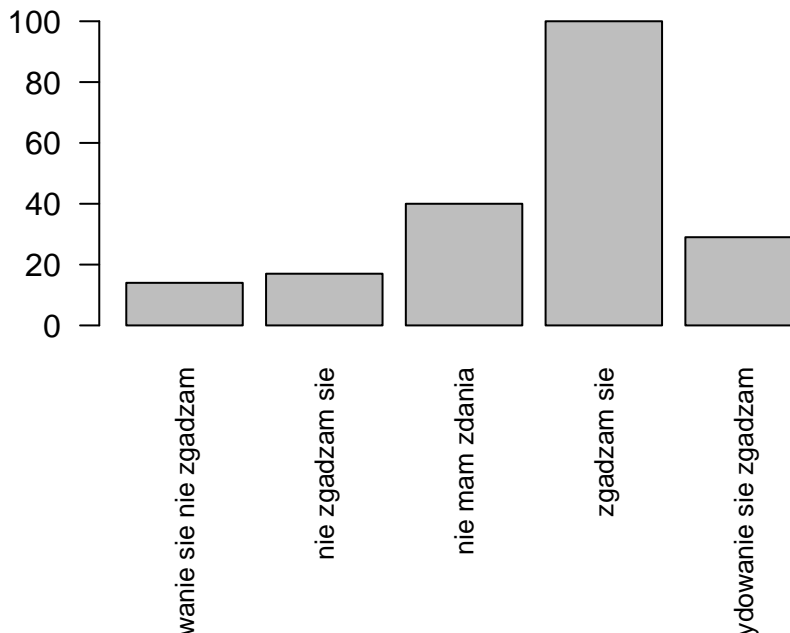
1.1.4 Zadanie 1.4

Sporządzono wykresy kołowe oraz wykresy słupkowe dla zmiennych: PYT_1 oraz PYT_2.

Odpowiedź na pierwsze pytanie



Odpowiedź na pierwsze pytanie

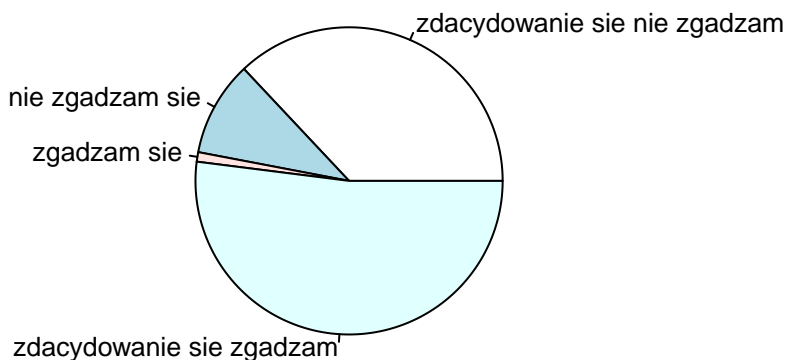


Największa liczba respondentów zaznaczyła odpowiedź “zgadzam się”, co wskazuje, że ogólna ocena wsparcia i materiałów dostarczanych przez firmę jest pozytywna. Istnieje także pewna grupa osób, które “zdecydowanie się zgadzają”, co podkreśla, że część pracowników uważa wsparcie za bardzo dobre.

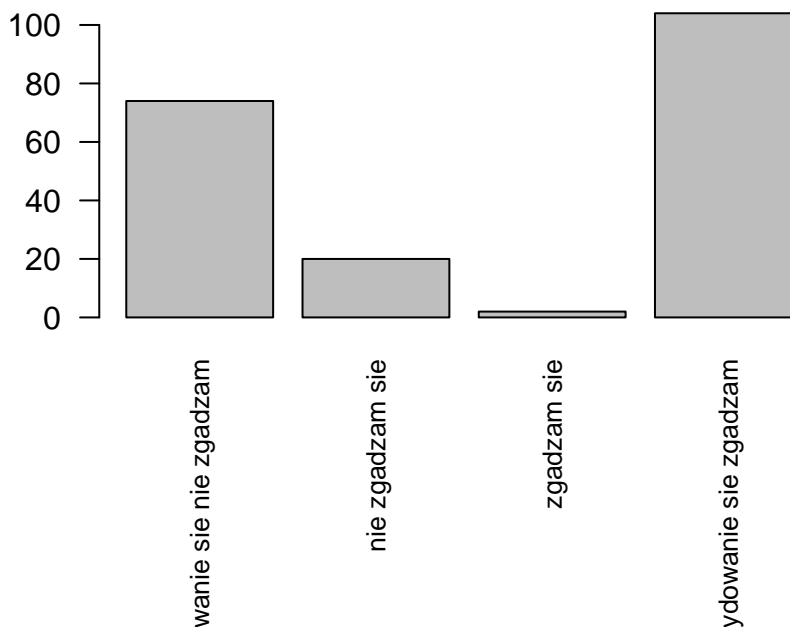
Znaczna część pracowników wybrała odpowiedź “nie mam zdania”, co może sugerować, że nie mieli oni okazji skorzystać ze wsparcia lub materiały nie są dla nich wystarczająco widoczne.

Pewna część respondentów zaznaczyła opcje “nie zgadzam się” i “zdecydowanie się nie zgadzam”, ale ich liczba jest stosunkowo mała w porównaniu do grupy zadowolonych pracowników. Może to wskazywać na pewne niedociągnięcia w dostępie do materiałów lub ich jakości, jednak nie jest to powszechny problem.

Odpowiedź na drugie pytanie



Odpowiedź na drugie pytanie



Największa część respondentów zaznaczyła odpowiedź “zdecydowanie się zgadzam”, co wskazuje, że większość pracowników uważa szkolenia oferowane przez firmę za dobrze dopasowane do ich potrzeb. Istnieje także spora grupa osób, które wybrały opcję “zgadzam się”, co dodatkowo potwierdza ogólnie pozytywne nastawienie wobec polityki szkoleniowej firmy.

Stosunkowo niewielka część respondentów wybrała odpowiedź “nie zgadzam się” oraz “zdecydowanie się nie zgadzam”, co sugeruje, że niektóre osoby mogą mieć trudności z dostępem do odpowiednich szkoleń lub nie uważają ich za skuteczne w kontekście swojego rozwoju zawodowego.

Pomimo że liczba negatywnych odpowiedzi jest niewielka, warto przeanalizować, czy istnieją konkretne obszary, w których szkolenia mogą być bardziej dostosowane do indywidualnych potrzeb

1.1.5 zadanie 1.5

Sporządzono tablice wielozmiennicze dla par zmiennych: PYT_1 i DZIAŁ, PYT_1 i STAŻ, PYT_1 i CZY_KIER, PYT_1 i PŁEĆ oraz PYT_1 i WIEK_KAT.

Tabela 9: Tabela wielozmiennicza dla zmiennych PYT 1 i DZIAŁ

	HR	IT	MK	PD
zdecydowanie się nie zgadzam	2	0	3	9
nie zgadzam się	2	2	3	10
nie mam zdania	5	4	14	17
zgadzam się	19	15	15	51
zdecydowanie się zgadzam	3	5	10	11

Najbardziej pozytywne opinie pochodzą z działów HR i PD – większość respondentów z tych działów wybrała opcje “zgadzam się” i “zdecydowanie się zgadzam”. Dział MK (Marketing) jest bardziej podzielony – występuje większy odsetek osób, które nie mają zdania, co może sugerować, że dla tej grupy pytanie nie było jednoznaczne lub temat ich mniej dotyczył. Dział IT wykazuje najmniejszy poziom negatywnych odpowiedzi (“zdecydowanie się nie zgadzam” = 0), ale też stosunkowo niewiele osób udzieliło odpowiedzi skrajnie pozytywnej.

Tabela 10: Tabela wielodzielcza dla zmiennych PYT 1 i STAŻ

	<1 rok	1-2 lata	3+ lat
zdacydowanie się nie zgadzam	5	5	4
nie zgadzam się	6	10	1
nie mam zdania	8	26	6
zgadzam się	19	75	6
zdacydowanie się zgadzam	3	24	2

Najbardziej pozytywne odpowiedzi (zgadzam się i zdecydowanie się zgadzam) pochodzą od osób z doświadczeniem 1-2 lata, co sugeruje, że osoby na tym etapie kariery widzą największą wartość badanego zagadnienia. Osoby z najmniejszym stażem (<1 rok) są bardziej podzielone, częściej nie mają zdania lub udzielają odpowiedzi negatywnych. Osoby z największym stażem (3+ lata) rzadko wybierają odpowiedzi skrajne, co może oznaczać, że ich ocena sytuacji jest bardziej neutralna.

Pracownicy z 1-2 latami stażu są najbardziej pozytywnie nastawieni do badanego zagadnienia. Osoby z krótszym stażem mogą wymagać większej ilości informacji lub wsparcia, aby mogły bardziej świadomie ocenić sytuację.

Tabela 11: Tabela wielodzielcza dla zmiennych PYT 1 i CZY_KIER

	TAK	NIE
zdacydowanie się nie zgadzam	10	4
nie zgadzam się	14	3
nie mam zdania	34	6
zgadzam się	88	12
zdacydowanie się zgadzam	27	2

Osoby na stanowiskach kierowniczych częściej wybierają opcję “zgadzam się” i “zdecydowanie się zgadzam”, co może świadczyć o ich większym zadowoleniu. Osoby niepełniące funkcji kierowniczych są bardziej podzielone – widoczny jest większy odsetek neutralnych i negatywnych odpowiedzi.

Pracownicy na stanowiskach kierowniczych mają bardziej pozytywne podejście do badanego aspektu, natomiast pracownicy bez funkcji kierowniczych mogą czuć się mniej związani z tematem lub mieć inne doświadczenia.

Mężczyźni są bardziej podzieleni, częściej wybierają opcje neutralne i negatywne. Kobiety częściej wybierają pozytywne odpowiedzi, co może świadczyć o lepszym dopasowaniu badanego zagadnienia do ich oczekiwań.

Tabela 12: Tabela wielodzielcza dla zmiennych PYT 1 i PŁEĆ

	Kobieta	Mężczyzna
zdacydowanie się nie zgadzam	3	11
nie zgadzam się	7	10
nie mam zdania	14	26
zgadzam się	36	64
zdacydowanie się zgadzam	11	18

Istnieją różnice w postrzeganiu badanego tematu w zależności od płci – warto byłoby zbadać, co wpływa na większą satysfakcję kobiet.

Tabela 13: Tabela wielodzielcza dla zmiennych PYT 1 i WIEK KAT

	0-35	36-45	46-55	55+
zdacydowanie się nie zgadzam	1	11	2	0
nie zgadzam się	6	7	1	3
nie mam zdania	3	24	5	8
zgadzam się	13	50	25	12
zdacydowanie się zgadzam	3	12	12	2

Najbardziej pozytywne odpowiedzi pochodzą od osób w wieku 36-45 lat oraz 46-55 lat – te grupy częściej wybierają opcje “zgadzam się” i “zdacydowanie się zgadzam”. Najmłodsza grupa (0-35 lat) oraz najstarsza grupa (55+) mają wyższy odsetek odpowiedzi neutralnych lub negatywnych.

Pracownicy w średnim wieku są najbardziej pozytywnie nastawieni do badanego zagadnienia. Osoby młodsze i starsze mogą mieć inne oczekiwania lub mniej doświadczenia w tym obszarze.

1.1.6 zadanie 1.6

Sporządzono tablicę wielodzielczą dla pary zmiennych: PYT_2 i PYT_3.

Tabela 14: Tabela wielodzielcza dla zmiennych PYT 1 i PYT 2

	zdacydowanie się nie zgadzam	nie zgadzam się	zgadzam się	zdacydowanie się zgadzam
zdacydowanie się nie zgadzam	13	0	1	
nie zgadzam się	16	0	0	
nie mam zdania	39	0	1	
zgadzam się	3	17	0	
zdacydowanie się zgadzam	3	3	0	

Dane obejmują zarówno zmienne ilościowe (np. STAŻ, WIEK, PYT_1), jak i jakościowe (np. DZIAŁ, PŁEĆ, CZY_KIER). Brak wartości brakujących sugeruje, że zestaw danych jest kompletny i dobrze przygotowany do analizy.

W zbiorze znajduje się 5 zmiennych ilościowych i 4 zmienne jakościowe (faktory).

Najwięcej osób w próbie należy do działu PD (98 osób), najmniej do IT (26 osób).

Większość badanych ma staż pracy 1-2 lata (140 osób), co może wskazywać na młodą kadrę pracowników.

Większość respondentów pełni funkcję kierowniczą (173 TAK vs. 27 NIE).

Przewaga mężczyzn w próbie (129 mężczyzn vs. 71 kobiet).

Największa grupa wiekowa to 36-45 lat (104 osoby).

PYT_1 a DZIAŁ: Najwięcej pozytywnych odpowiedzi (“zgadzam się” lub “zdecydowanie się zgadzam”) udzieliły osoby z działu PD.

PYT_1 a STAŻ: Wśród osób z najkrótszym stażem (<1 rok) więcej było odpowiedzi neutralnych lub negatywnych. Może to wynikać z braku doświadczenia lub innego spojrzenia na temat badania.

PYT_1 a CZY_KIER: Kierownicy częściej zgadzali się z twierdzeniem zawartym w PYT_1 niż osoby niepełniące funkcji kierowniczych.

PYT_1 a PŁEĆ: Mężczyźni częściej wyrażali zdecydowaną opinię niż kobiety, które częściej wybierały opcję “nie mam zdania”.

PYT_1 a WIEK_KAT: Najwięcej osób w wieku 36-45 lat zgadzało się z PYT_1, natomiast najmniej zdecydowanych odpowiedzi było w grupie 0-35 lat.

PYT_1 a PYT_2: Istnieje silna zależność – osoby, które nie zgadzały się z PYT_1, często miały podobne odpowiedzi na PYT_2, a ci, którzy zgadzali się z PYT_1, także zgadzali się z PYT_2.

Występują różnice w opiniach w zależności od działu, stażu pracy, płci, funkcji kierowniczej i wieku. Kierownicy oraz osoby z dłuższym stażem mają bardziej zdecydowane opinie. Istnieje silna korelacja między odpowiedziami na PYT_1 i PYT_2. Możliwe różnice w podejściu do badanej kwestii między kobietami a mężczyznami oraz młodszymi a starszymi pracownikami.

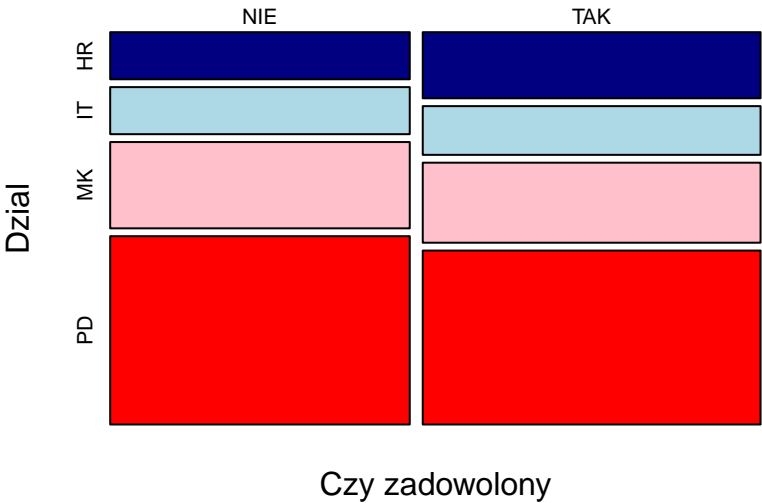
1.1.7 zadanie 1.7

Utworzono zmienną CZY_ZADOW na podstawie zmiennej PYT_2 i „aczkolwiek” kategorie “nie zgadzam się” i “zdecydowanie się nie zgadzam” oraz “zgadzam się” i “zdecydowanie się zgadzam”

1.1.8 zadanie 1.8

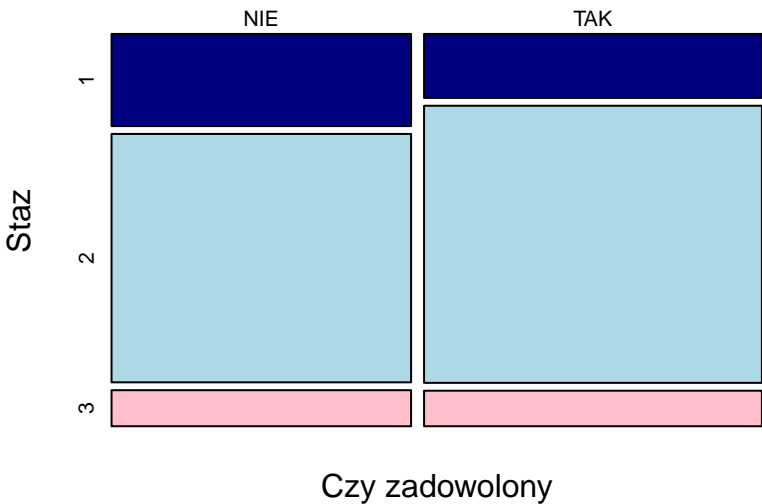
Sporządzono wykresy mozaikowe odpowiadające parom zmiennych CZY_ZADOW i DZIAŁ, CZY_ZADOW i STAŻ, CZY_ZADOW i CZY_KIER, CZY_ZADOW i PŁEĆ oraz CZY_ZADOW i WIEK_KAT.

CZY ZADOWOLONY ~ DZIAL



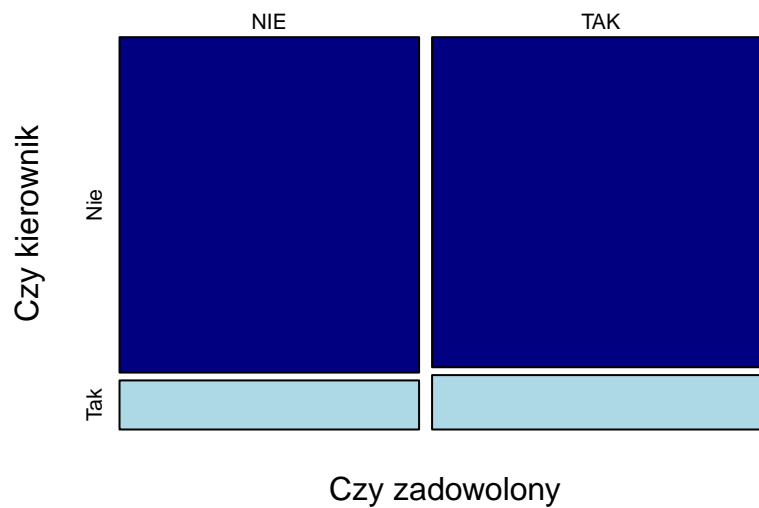
NULL

CZY ZADOWOLONY ~ STAZ



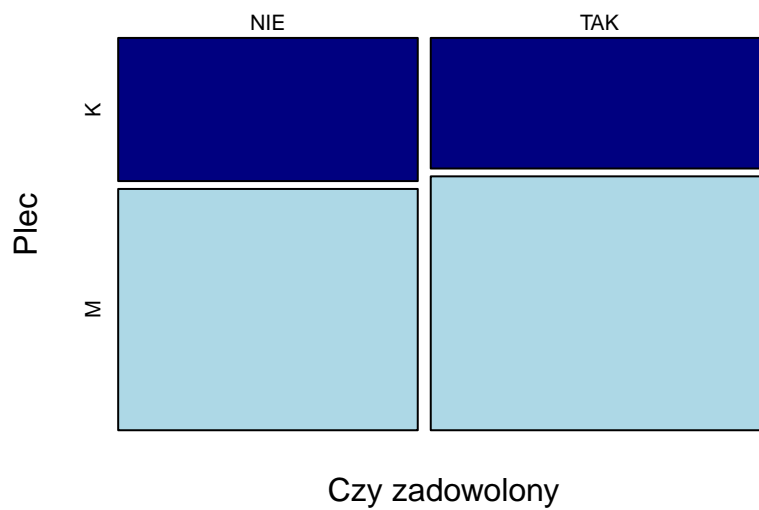
NULL

CZY ZADOWOLONY ~ CZY KIER



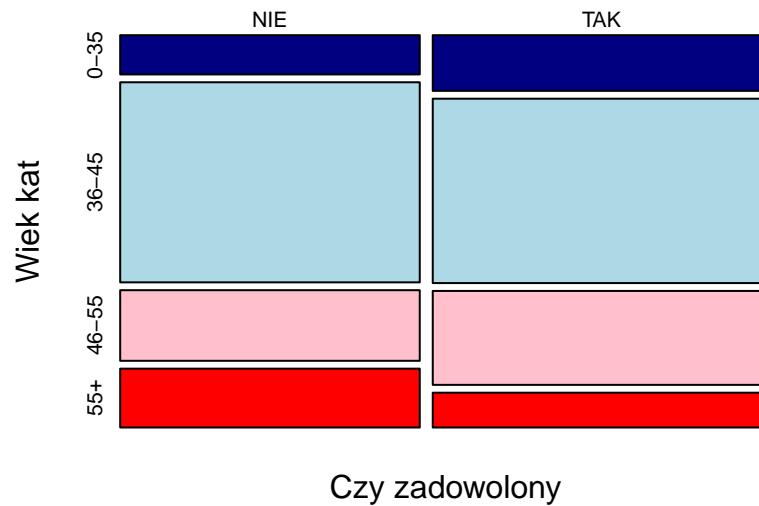
NULL

CZY ZADOWOLONY ~ PLEC



NULL

CZY ZADOWOLONY ~ WIEK KAT



NULL

2 Zadanie 2

3 Zadanie 6

Przedział ufności (Clopper-Pearson) dla 30 / 100 : 0.2124064 0.3998147

Przedział ufności (Clopper-Pearson) dla danych binarnych: 0.2212888 0.4103146

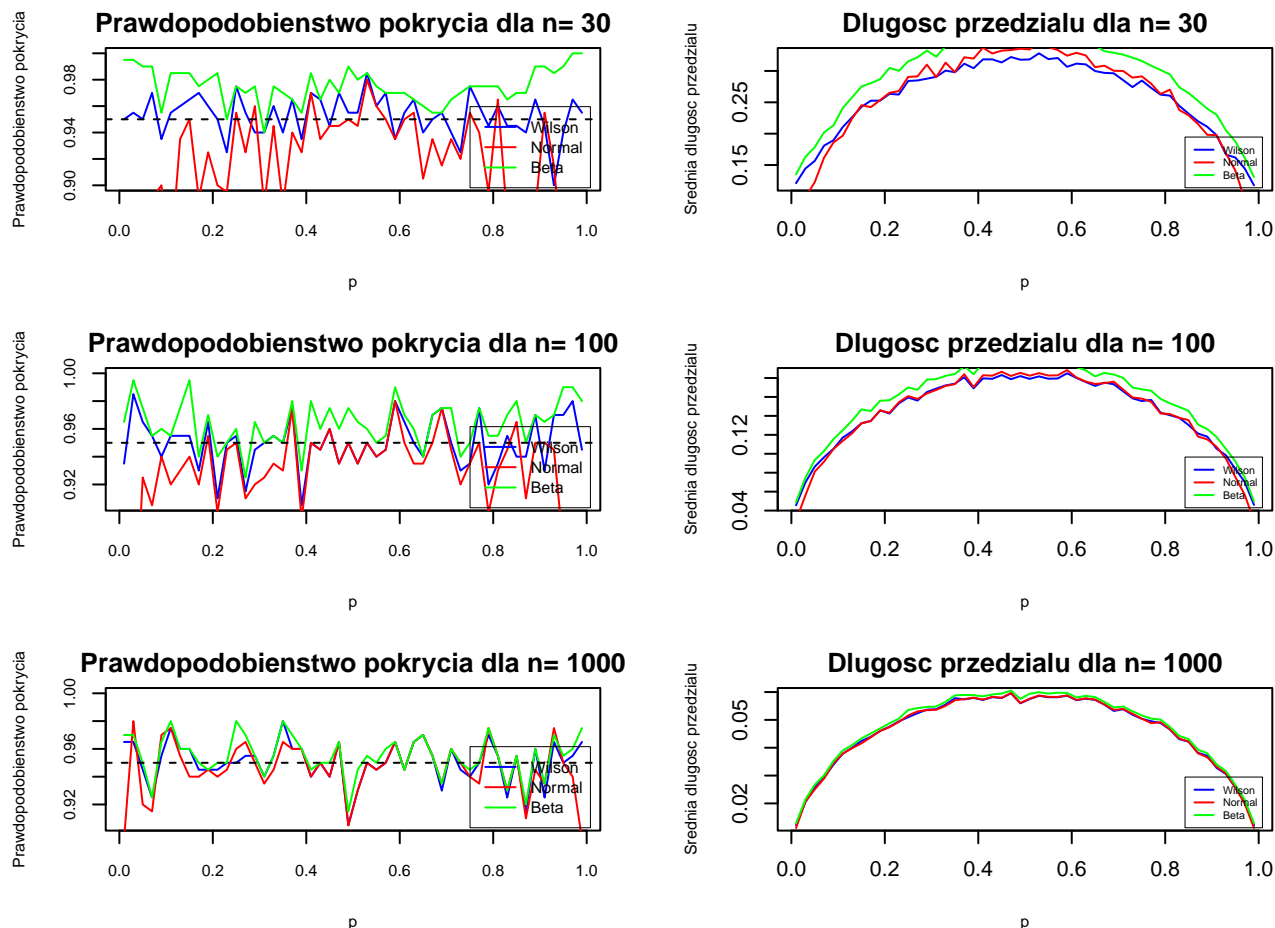
4 Zadanie 7

Przedział ufności Cloppera-Pearsona: 0.4905513 0.5102485

Przedział ufności Wald'a: 0.4906002 0.5101998

Przedział ufności Wilsona: 0.4906019 0.5101978

5 Zadanie 9



Dla małych próbek ($n=30$) metoda normalna wykazuje dużą niestabilność, z wyraźnymi spadkami poniżej poziomu pozostałych metod. W miarę wzrostu liczności próby ($n=100$, $n=1000$) różnice między metodami się zmniejszają, a wszystkie trzy metody zbliżają się do oczekiwanego poziomu pokrycia. Metoda Wilsona i Beta generalnie utrzymują lepsze prawdopodobieństwo pokrycia, szczególnie dla mniejszych próbek.

Wszystkie metody wykazują podobny kształt – długość przedziału jest największa dla $p=0.5$ i maleje dla wartości bliższych 0 i 1. Dla $n=30$ metoda normalna daje krótsze przedziały, ale kosztem gorszego prawdopodobieństwa pokrycia. Dla większych próbek długości przedziałów uzyskane różnymi metodami są bardzo podobne.

Przy większych n różnice między metodami stają się mniej istotne – wszystkie trzy metody dają podobne prawdopodobieństwo pokrycia i długości przedziałów. Dla małych próbek metoda Walda wydaje się najmniej stabilna i najgorzej dopasowana.

Metoda Walda nie sprawdza się dobrze dla małych prób – daje niestabilne pokrycie i krótsze przedziały kosztem wiarygodności. Metody Wilsona i Beta są bardziej niezawodne, szczególnie dla mniejszych wartości n . Dla dużych próbek wszystkie metody działają podobnie.