



Fabian Czichos

Estymacja liczby kawalerek na wynajem
w Poznaniu z wykorzystaniem źródeł internetowych i metod capture-recapture

Estimation of the number of studio apartments in Poland using online data sources and capture-recapture methods

Praca licencjacka

Promotor: dr Maciej Beręsewicz

Pracę przyjęto dnia:

Podpis promotora

Kierunek: Informatyka i Ekonometria

Poznań 2021

Spis treści

Wstęp	1
1 Prawne i ekonomiczne aspekty najmu w Polsce	3
1.1 Definicje najmu	3
1.1.1 Polskie akty prawne, statystyka publiczna oraz portale internetowe . .	3
1.2 Definicja najmu okazjonalnego	4
1.3 Ekonomiczny wymiar najmu	5
1.4 Źródła danych o najmie kawalerek	6
1.5 OLX oraz OtoDom jako źródło danych o najmie kawalerek	8
1.5.1 Charakterystyka	8
1.6 Podsumowanie	9
2 Metody estymacji wielkości populacji trudnych do zbadania	11
2.1 Populacja trudna do zbadania	11
2.2 Metody estymacji wielkości populacji	14
2.3 Model log-liniowy	15
2.4 Mierniki oceny jakości dopasowania modelu	17
2.5 Wyznaczenie przedziałów ufności dla wielkości populacji	19
2.6 Implementacja modelu w programie statystycznym R	19
2.7 Podsumowanie	20
3 Estymacja wielkości rynku kawalerek w Poznaniu	21
3.1 Wykorzystane źródła danych	21
3.1.1 Opis zbioru danych z OLX	21
3.1.2 Opis zbioru danych z OtoDom	22
3.2 Proces przygotowania danych	22

3.3	Weryfikacja założeń modelu	25
3.4	Estymacja parametrów modelu log-liniowego oraz jego ocena	25
3.4.1	Konstrukcja modelu log liniowego	25
3.4.2	Wyniki estymacji parametrów modelu log-liniowego	27
3.5	Ocena modelu log-liniowego	27
3.6	Szacunek wielkości populacji	27
Podsumowanie		30
Literatura		33
Spis tabel		34
Spis rysunków		35
Spis programów w języku R		36
A	Załączniki	37
A.1	Procedura dodawania ogłoszeń na Otodom	37
A.2	Procedura dodawania ogłoszeń na OLX	42
A.3	Dane wykorzystane w modelach	42

Wstęp

W ciągu ostatniej dekady rynek najmu mieszkań w Polsce dynamicznie się rozwinął. W latach 2010-2018 podwoiła się liczba mieszkań przeznaczonych na wynajem (Pilcicka 2018). Szczególnie prężnie rozwija się rynek najmu kawalerek (Pilcicka 2018). Powodem takiej sytuacji jest przybywanie gospodarstw jednoosobowych, rozwodników oraz imigrantów zarobkowych ze Wschodu Europy (głównie obywateli Ukrainy). Celem tej pracy jest zbadanie skali najmu kawalerek na przykładzie m. Poznań.

Na potrzeby pracy przyjęto, że najem kawalerki to oddanie najemcy kawalerki, tj. najmniejszej odmiany lokalu mieszkalnego, składającego się z jednego pokoju (z aneksem kuchennym) do użytku, który zobowiązuje się płacić wynajmującemu czynsz.

Głównym celem badawczym pracy licencjackiej jest oszacowanie liczby kawalerek oferowanych na wynajem w Poznaniu w trzecim kwartale 2020 roku. Aby zrealizować ten cel wykorzystano dane pochodzące z dwóch portali internetowych: OLX oraz OtoDom oraz metodę *capture-recapture*¹. Metoda ta wykorzystywana jest do oszacowania wielkości populacji w przypadku niemożliwości obserwacji wszystkich jej jednostek oraz gdy dostępne źródła danych pokrywają tylko jej część.

Praca ma charakter teoretyczno-empiryczny i składa się z trzech rozdziałów. W pierwszym podano terminologię oraz aspekt prawny dotyczący najmu mieszkania. Przytoczono również definicję najmu kawalerek według polskiego prawa oraz publicznej statystyki. W tym rozdziale opisano także najem okazjonalny. Określono też źródła danych, które wykorzystano w badaniu.

W drugim rozdziale zdefiniowano populację trudną do zbadania oraz przedstawiono metodę estymacji wielkości takiej populacji. Przedstawiono również model log-liniowy wraz z założeniami. Na końcu rozdziału pokazano implementację modelu w pakiecie statystycznym R.

Na początku rozdziału trzeciego przedstawiono źródła danych, które wykorzystano w badaniu oraz proces ich oczyszczania. W badaniu skorzystano z danych pobranych przez OLX oraz

¹W pracy używane będzie angielskie pojęcie ponieważ w polskiej literaturze nie można znaleźć tłumaczenia, które oddawałoby istotę tej metody.

OtoDom². Później opisano konstrukcję modeli użytych do oszacowania wielkości populacji. Na końcu rozdziału zamieszczono wyniki estymacji dla trzech modeli – jednego bez zmiennych pomocniczych oraz dwóch z zmiennymi pomocniczymi.

²Dane te zostały udostępnione na potrzeby tego badania przez promotora dra Macieja Beręsewicza.

Rozdział 1

Prawne i ekonomiczne aspekty najmu w Polsce

1.1 Definicje najmu

1.1.1 Polskie akty prawne, statystyka publiczna oraz portale internetowe

W polskim prawie głównymi aktami prawnymi, które regulują podmiot najmu oraz stosunek między najemcą, a wynajmującym są:

- Ustawa z dnia 21 czerwca 2001 o ochronie praw lokatorów (Dz.U.2020.0.611);
- Kodeks cywilny.

W powyższych aktach prawnych pojawiają się takie definicje jak:

- **Umowa najmu** – w polskim prawie cywilnym umowa, w której wynajmujący zobowiązuje się oddać przedmiot najmu do używania najemcy, na czas oznaczony lub nieoznaczony, w zamian za wynagrodzenie w postaci czynszu płaconego przez najemcę.
- **Najemca** to osoba, która jest lokatorem danego lokum (mieszkania, domu) i najmuje dany lokal za korzyść materialną, odprowadzaną na rzecz wynajmującego
- **Wynajmujący** to właściciel mieszkania, który udostępnia je najmującemu za określoną korzyść materialną na podstawie specjalnie przygotowanej umowy.

W polskim prawie nie ma jasnej definicji najmu mieszkania. Jednakże na podstawie przepisu ogólnego artykułu 659 Kodeksu cywilnego o „Istocie umowy najmu”: *”Przez umowę najmu wynajmujący zobowiązuje się oddać najemcy rzecz do używania przez czas oznaczony lub nie ozna-*

czony, a najemca zobowiązuje się płacić wynajmującemu umówiony czynsz”, możemy stworzyć własną definicję najmu mieszkania: „oddanie najemcy lokalu mieszkalnego do używania, który zobowiązuje się płacić wynajmującemu czynsz”. W połączeniu z definicją kawalerki: „najmniejsza odmiana lokalu mieszkalnego, składającego się z jednego pokoju (z aneksem kuchennym)”, określenie najmu kawalerki brzmi: „oddanie najemcy kawalerki, tj. najmniejszej odmiany lokalu mieszkalnego, składającego się z jednego pokoju (z aneksem kuchennym) do użytku, który zobowiązuje się płacić wynajmującemu czynsz”

W statystyce publicznej również nie mamy jasno określonej definicji najmu. W słowniku pojęć Głównego Urzędu Statystycznego mamy jedynie definicję najmu socjalnego: „Najem lokalu nadającego się do zamieszkania ze względu na wyposażenie i stan techniczny, którego powierzchnia pokoi przypadająca na członka gospodarstwa domowego najemcy nie może być mniejsza niż 5 m², a w przypadku jednoosobowego gospodarstwa domowego 10 m², przy czym lokal ten może być o obniżonym standardzie”. Dopiero na poziomie międzynarodowej statystyki pojawia się wynajem w definicji opłaty na rzecz właściciela: „kwota za wynajem domu/mieszkania, płacona właścicielowi lub głównemu użytkownikowi, będąca jego zyskiem”.

Zarówno portal OLX jak i OtoDom nie podają swojej definicji najmu. Jednak wymienione portale mają własne wzory umów najmu. Umowy te spełniają wszystkie normy prawa polskiego, więc można uznać, iż te portale stosują się do prawnej definicji umowy najmu kawaleerek.

1.2 Definicja najmu okazjonalnego

Najem okazjonalny pojawił się w polskim systemie prawnym dzięki ustawie z 21 czerwca 2001 r. o ochronie praw lokatorów, mieszkaniowym zasobie gminy i o zmianie Kodeksu cywilnego. Zgodnie z zapisami tej regulacji umową najmu okazjonalnego lokalu jest umowa najmu lokalu mieszkalnego, którego właściciel, będący osobą fizyczną, nie prowadzi działalności gospodarczej w zakresie wynajmowania lokali, zawarta na czas oznaczony, nie dłuższy niż 10 lat. Umowa najmu okazjonalnego posiada identyczną moc prawną, co standardowa umowa najmu. Posiada jednak jedną, bardzo istotną zaletę dla właścicieli domów i mieszkań: możliwość szybkiej egzekucji komorniczej, opróżnienia lokalu i usunięcia nieuczciwego najemcy, gdyż do umowy najmu okazjonalnego lokalu załącza się:

1. oświadczenie najemcy w formie aktu notarialnego, w którym najemca poddał się egzekucji i zobowiązał się do opróżnienia i wydania lokalu używanego na podstawie umowy najmu okazjonalnego lokalu w terminie wskazanym w żądaniu, o którym mowa w art. 19d żądanie opróżnienia lokalu ust. 2;
2. wskazanie przez najemcę innego lokalu, w którym będzie mógł zamieszkać w przypadku wykonania egzekucji obowiązku opróżnienia lokalu;
3. oświadczenie właściciela lokalu lub osoby posiadającej tytuł prawny do lokalu, o którym mowa w pkt 2 lub ust. 3, o wyrażeniu zgody na zamieszkanie najemcy i osób z nim zamieszkujących w lokalu wskazanym w oświadczeniu; na żądanie wynajmującego załącza się oświadczenie z podpisem notarialnie poświadczonym.

W ten sposób, w przypadku wystąpienia jakichkolwiek problemów po wygaśnięciu lub rozwiązaniu umowy, właściciel mieszkania może wszcząć postępowanie egzekucyjne niemal natychmiast – bez konieczności wytaczania długotrwałych procesów sądowych oraz czekania na dostarczenie przez gminę lokalu tymczasowego (jak w przypadku zwykłej umowy najmu).

Według badań Polskiej Federacji Rynku Nieruchomości (Hardek 2014), umowę najmu okazjonalnego podpisuje jedynie 5% właścicieli mieszkań, ponieważ ten rodzaj najmu budzi poważne obawy najemców ze względu na:

- **dodatkowe koszty** – sporządzenie aktu notarialnego wymaga m.in. zapłaty za usługi notariusza, a poszukujący mieszkania na wynajem niechętnie godzą się na dodatkowe koszty
- **niewystarczającej wiedzy** – już sam zwrot „łatwiejsza eksmisja” powoduje, że najemcy boją się umów najmu okazjonalnego.
- **czasochłonnym procedur** – dopięcie wszelkich formalności i dostarczenie dodatkowych dokumentów dołączanych do umowy najmu okazjonalnego to dodatkowy wysiłek i czas, który muszą poświęcić najemcy

1.3 Ekonomiczny wymiar najmu

Rynek najmu w Polsce jest niewielki, gdyż wartość usług generowana przez nieruchomości komercyjne z dochodów czynszowych dla ich właścicieli wynosił w 2018 roku 0,7% PKB (Łaszek i in. 2019). Jest to wartość mniejsza niż w rozwiniętych krajach UE, gdzie zasób nieruchomości

komercyjnych na wynajem jest znacznie większy (Pilcicka 2018). Jednak polski rynek dynamicznie się rozwija. Według raportu Amron w 2018 roku co dwudziesty Polak mieszkał w wynajmowanym mieszkaniu, a w 2010 roku ta wartość była dwa razy mniejsza (Pilcicka 2018). Wpływ na taką sytuację ma postrzeganie zakupu nieruchomości pod wynajem jako inwestycję w bezpieczne aktywa z dobrą stopą zwrotu. Narodowy Bank Polski wyliczył, że jeżeli zakup mieszkania jest w pełni finansowany gotówką to roczna stopa zwrotu wynosi 4%. Jest to wartość lepsza od nominalnych stóp procentowych lokat bankowych czy obligacji skarbu państwa. Niskie koszty kredytów mieszkaniowych również przyczyniają się do tego, że zamożniejsze gospodarstwa kupują mieszkania pod wynajem, gdyż nawet przy wkładzie własnym wynoszącym zaledwie 20% stopa zwrotu oscyluje wokół 3%, co powoduje, że jest to lepsza inwestycja niż zakup obligacji. Innym czynnikiem, który sprawia, że rynek najmu w Polsce rośnie jest zwiększająca się liczba imigrantów zarobkowych. Imigranci zdecydowanie częściej decydują się na najem niż zakup własnego mieszkania (Andrews i Sánchez 2011).

1.4 Źródła danych o najmie kawalerek

W celu oszacowania liczby kawalerek przeznaczonych na najem w Poznaniu należy zgromadzić dane, które mogą pochodzić z różnych źródeł. Źródła dzielimy na statystyczne oraz niestatystyczne. Dane niestatystyczne są gromadzone do celów niestatystycznych jednak mogą być wykorzystywane na potrzeby statystyki publicznej, a dane statystyczne są gromadzone bezpośrednio w celu tworzenia statystyk opisujących stan i rozwój określonych populacji, gdy żadne inne odpowiednie dane nie są dostępne (Nordbotten 2010; A. Wallgren i B. Wallgren 2014).

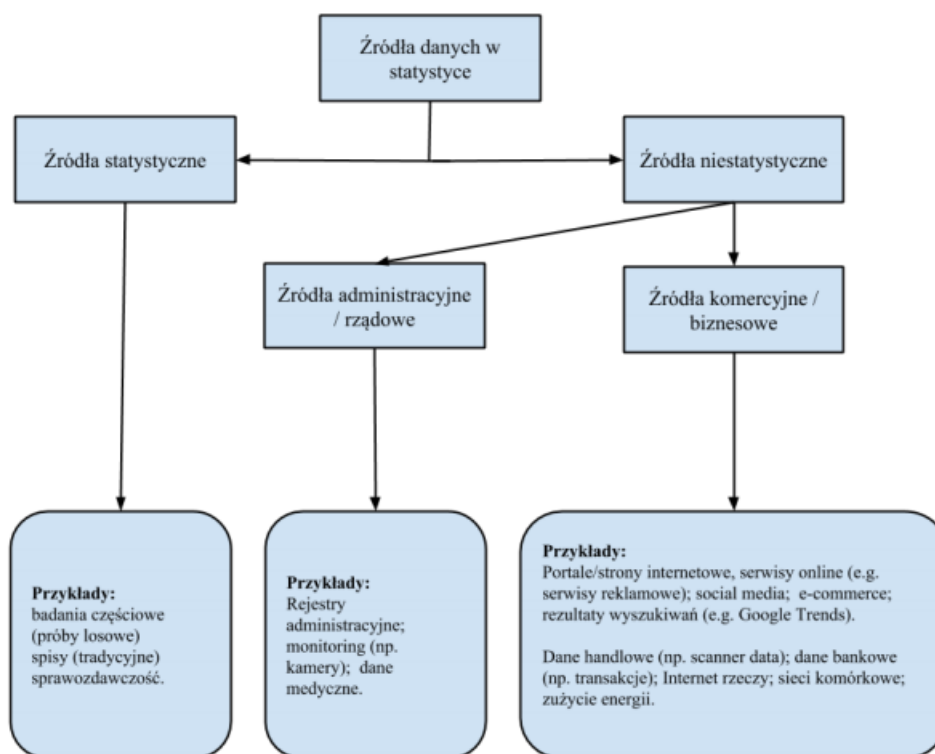
Klasycznymi źródłami danych w statystyce publicznej są spisy powszechne, które mogą być przeprowadzone na całej zbiorowości, bądź na wylosowanej próbie jako badania reprezentacyjne. Problemem jest okres, który dzieli kolejne badania (np. 10 lat), który w obecnych czasach i gospodarkach opartych na informacji jest zbyt długim odstępem czasowym. Innym źródłem są badania reprezentacyjne prowadzone w określonym celu badawczym np. BAEL – Badanie Aktywności Ekonomicznej Ludności. Takie badania są prowadzone z dużo większą częstotliwością niż spisy powszechne np. miesięczną. Niestety informacje pozyskiwane z tych badań agregowane są bardzo często na poziomie całego kraju czy co najwyżej województw (Beręsewicz i Szymkowiak 2015).

Oprócz klasycznych źródeł danych istnieją też nowe, nieklasyczne źródła danych. Dzieli się je na źródła stworzone na potrzeby sektora publicznego oraz sektora prywatnego. Pierwszą grupę stanowią głównie rejestry administracyjne (m.in. PESEL, REGON, PIT/CIT, ZUS, RCIWN) oraz inne dane ewidencyjne (m.in. pochodzące z fotoradarów, satelitów, pomiaru ruchu samochodowego czy giełdy), które mają za zadanie osiągnięcie celów określonych w odpowiednich ustawach. Natomiast druga grupa związana jest z działalnością prywatną (profit i non-profit), których przykładem mogą być dane transakcyjne sieci handlowych, transakcje bankowe, sieci komórkowe i Internet Rzeczy (m.in. GPS, smartfony), portale internetowe (m.in. Facebook, Twitter, Pracuj.pl, OtoDom) czy szeroko rozumiany e-commerce (np. Allegro, Ceneo). Nowe źródła danych pojawiają się głównie pod nazwą big data, co ma ujmować ich charakter przez wskazanie na: duży wolumen danych (ang. *volume*), dużą zmienność i dynamikę ich powstawania (ang. *velocity*) oraz dużą różnorodność oraz ich nieustrukturyzowanie (ang. *variety*) (Beręsewicz i Szymkowiak 2015).

Zaletą statystycznych źródeł informacji jest ich ustrukturyzowanie. Dane posiadają odpowiednie definicje jednostek oraz zmiennych. Statyczne informacje przechowywane są ponadto w relacyjnych bazach danych, skąd może je łatwo wyeksportować za pomocą języka SQL. Wadą klasycznych źródeł danych jest niski wolumen, liczony maksymalnie w milionach rekordów. Kolejną wadą statystycznych źródeł danych jest uzależnienie jakości danych od metody pozyskania informacji, np. źle napisane pytania w kwestionariuszu mogą doprowadzić do tego, że badana grupa opacznie je zrozumie.

Z kolei zaletą nowych źródeł danych jest obniżanie kosztów otrzymania danych oraz ich utrzymania. Technologie Big Data takie jak Hadoop czy Azure potrafią znacząco zredukować koszty przechowywania ogromnej ilości danych. Pomimo tego, że implementacja tych technologii może wydawać się droga, w dłuższym okresie pozwoli ona zaoszczędzić pieniądze. Duża ilość informacji oraz szczegółowa danych pozwala dokładniej badać interesujące zjawisko co w połączeniu z analityką czasu rzeczywistego pozwala podejmować szybsze oraz dokładniejsze decyzje

Ilość oraz szczegółowość informacji jest również wadą nowych źródeł informacji. Oczyszczenie takich danych jest żmudne oraz czasochłonne. Należy pamiętać również o problemie korelacji. Powszechną techniką używaną do analizy Big Data jest obliczanie korelacji, czyli związku pomiędzy dwoma zmiennymi losowymi. Jednakże częstym błędem jest przyjmowanie, że zmienna nawet silnie skorelowana z inną jest jej skutkiem. Przykładem może być korelacja



Rysunek 1.1. Klasyfikacja źródeł danych w statystyce

Źródło: Beręsewicz (2016).

liczby zgonów osób bezdomnych w Indiach a poziomem spożycia lodów w Stanach Zjednoczonych. Zmienną, której nie uwzględniamy w analizie jest średnia temperatura lata na półkuli północnej. Jej zwiększenie powoduje więcej zgonów spowodowanych upałem i oczywiście zwiększenie spożycia lodów. Kolejną wadą Big Data są kwestie związane z bezpieczeństwem oraz prywatnością danych. Skandal "Facebook–Cambridge Analytica" (Criddle 2020) pokazuje, że informacje o użytkownikach mogą zostać wykorzystane do zmanipulowania ich. Z kolei historia wycieku danych z Yahoo (Fiegerman 2016) uświadamia, że dane użytkowników nie są bezpiecznie przechowywane.

1.5 OLX oraz OtoDom jako źródło danych o najmie kawalerek

1.5.1 Charakterystyka

W celu oszacowania liczby kawalerek przeznaczonych na wynajem wykorzystane zostaną niestatystyczne źródła informacji, a dokładniej specjalistyczne strony internetowe:

OtoDom – internetowy serwis ogłoszeniowy dotyczący nieruchomości działający w Polsce od 2006 roku. OtoDom umożliwia użytkownikom przeglądanie i zamieszczanie ogłoszeń sprzedaży i wynajmu nieruchomości: mieszkań, domów, pokoi, działek, lokali użytkowych, hal, magazynów oraz garaży. Zgodnie z przeprowadzonym badaniem przez PBS w lutym 2016 OtoDom został uznany za najskuteczniejszy portal ogłoszeniowy – 46% respondentów uważa, że to właśnie z tego serwisu otrzymują największą ilość zapytań kończących się sprzedażą nieruchomości (PBS 2017). W OtoDom nie ma możliwości dodania bezpłatnego ogłoszenia. Warto jednak zaznaczyć, że opłata za dodanie podstawowego ogłoszenia wynosi zaledwie 2,46 zł. OtoDom oprócz użytkowników prywatnych współpracuje zarówno z pośrednikami oraz deweloperami. Warto zwrócić uwagę na fakt iż, każde ogłoszenie umieszczone w OtoDom pojawia się również na stronie OLX, co pozwala dotrzeć do większej liczby odbiorców.

OLX – serwis ogłoszeniowy działający w Polsce, który pozwala na tworzenie oraz przeglądanie ofert (w tym także ofert najmu mieszkań). Serwis ogłoszeniowy pod marką OLX działa w Polsce od 14 kwietnia 2014 roku. Wywodzi się z serwisu tablica.pl, który wtedy należał do Grupy Allegro. Użytkownicy serwisu OLX mają do dyspozycji 15 głównych kategorii: Motoryzacja, Nieruchomości, Praca, Dom i Ogród, Elektronika, Usługi i Firmy, Moda, Rolnictwo, Zwierzęta, Dla Dzieci, Sport i Hobby, Muzyka i Edukacja, Ślub i Wesele, a także Oddam za darmo oraz Zamienię. Korzystanie z serwisu OLX jest możliwe z poziomu bezpłatnej aplikacji mobilnej oraz strony internetowej OLX.pl. Według badania Gemius/PBI w lutym 2020 serwis OLX.pl zanotował 13,4 mln użytkowników (dane PBI/Gemius), którzy wygenerowali 1,5 mld odsłon, co czyni OLX jedną z najpopularniejszych stron internetowych w Polsce (Gemius/PBI 2020).

W większości kategorii zamieszczanie ogłoszeń jest bezpłatne. Jednym z wyjątków są ogłoszenia dotyczące nieruchomości. W tej kategorii można mieć tylko jedno bezpłatne ogłoszenie na 30 dni. Za każde kolejne ogłoszenie trzeba zapłacić 30 złotych. OLX skupia się bardziej na prywatnych właścicielach niż pośrednikach nieruchomości czy deweloperach, pojedynczy użytkownik może zamieścić maksymalnie 50 ogłoszeń najmu. W załączniku A.2 i A.1 przedstawiono odpowiednio procedurę dodania ogłoszenia na portalu OLX oraz OtoDom.

1.6 Podsumowanie

W rozdziale pierwszym wyjaśniono czym jest najem kawalerek według polskiego prawa oraz publicznej statystyki. Dodatkowo wyjaśniono czym jest najem okazjonalny. W dalszej części

rozdziału omówiono ekonomiczny wymiar najmu w Polsce. Przedstawiono również źródła danych, na podstawie których można oszacować liczbę kawalerek. W kolejnym rozdziale zostanie opisana metoda, za pomocą której oszacowano wielkość populacji kawalerek oraz przedstawiona implementacja tej metody w programie statystycznym R (R Core Team [2019](#)).

Rozdział 2

Metody estymacji wielkości populacji trudnych do zbadania

2.1 Populacja trudna do zbadania

Wszystkie istniejące populacje mogą wydawać się trudne do zbadania jednak niektóre z nich ze względu na różne czynniki stanowią większe wyzwanie w badaniach. Populacje te mogą być rzadkie, ukryte, trudne do odnalezienia bądź kontaktu, lub trudne w kooperacji z ankieterem (Tourangeau i in. [2014](#)). Populacje trudne do zbadania możemy podzielić na:

1. Trudne w wylosowaniu (ang. *hard-to-sample*)

W tego typu populacjach nie istnieje właściwy operat losowy, z którego można byłoby wylosować jednostki do próby, wykorzystując odpowiedni schemat do pobrania. Rozwiązaniem tego problemu może być zastosowanie wstępnej rozmowy rekrutacyjnej (ang. *screening interview*), lecz jest to zależne od kosztu wstępnej rozmowy kwalifikacyjnej w stosunku do głównej wywiadu (ang. *main interview*). Rozważmy sytuację, w której przedstawiciele rzadkiej grupy stanowią tylko 5% całej populacji. Oznacza to, że ankie-terzy będą musieli wykonać 20 wywiadów wstępnych przed 1 wywiadem właściwym. Jeśli jeden wywiad wstępny kosztuje 1/20 wywiadu głównego to połączone koszty badań przesiewowych oraz badania głównego są tylko dwa razy większe od kosztu samego badania głównego. Jeżeli jednak jeden wywiad wstępny kosztuje połowę wywiadu głównego wtedy łączony koszt badań przesiewowych oraz badania głównego jest 11 razy większy od kosztu samego wywiadu głównego (Tourangeau i in. [2014](#)).

2. Trudne w identyfikacji (ang. *hard to identify*)

W niektórych przypadkach, zwłaszcza stygmatyzowanych grup społecznych, członkowie populacji mogą nie chcieć udostępnić swoich danych z obawy przed ujawnieniem nielegalnego lub wstydlwego statusu społecznego, co uniemożliwia identyfikację jednostek należących do takich populacji. Przykład: narkomani, alkoholicy, mniejszości (LGBT, wyznawcy określonych religii czy ideologii) (Bereśewicz i in. 2019).

Metoda kuli śnieżnej lub coraz częściej wykorzystywana metoda doboru sterowania przez respondenta (ang. *respondent-driven sampling*, RDS) są metodami, które redukują problem identyfikacji członków rzadkich lub stygmatyzowanych grup. Metoda kuli śnieżnej polega na tym, że ankietowana osoba rekrutuje inne osoby z danej społeczności, a te inne osoby rekrutują jeszcze innych członków tej grupy itd. By metodę śnieżnej kuli lub RDS można było wykorzystać potrzebne są trzy założenia:

- respondent zna innego członka badanej populacji,
- wszystkie jednostki z populacji są połączone w jedną zwartą sieć,
- respondent potrafi dokładnie wskazać rozmiar swojej osobistej sieci,

jeżeli te założenia są spełnione, to członkowie populacji, ukrywają np. swoją orientację tylko przed osobami spoza tej społeczności

3. Trudne w dotarciu (ang. *hard to reach*)

Trudność w nawiązaniu kontaktu związana jest przede wszystkim z mobilnością tego typu populacji. Są przynajmniej 4 typy mobilnych społeczności, które mogą być ciężkie do zlokalizowane:

- tradycyjne nomadyczne plemiona (Beduini w krajach arabskich),
- wędrownie mniejszości (Romowie),
- osoby które są tymczasowo mobilne lub przesiedlone (np. imigranci, bezdomni czy uchodźcy),
- osoby które są mobilne na pewnym etapie życia (studenci).

Niekiedy też występują problemy z skontaktowaniem się, gdyż część osób celowo się izoluje, np. krajach rozwiniętych buduje się coraz więcej zamkniętych osiedli. Szacuje się, że w połowie lat 90-tych 40% budowanych w Ameryce osiedli było osiedlami zamkniętymi.

4. Trudne do rekrutacji do badania (ang. *hard to persuade*)

To, że osoba została wybrana do badania nie oznacza, że wywiad zostanie przeprowadzony. Respondent może odmówić udziału w badaniu i odmowa może być spowodowana

różnymi czynnikami. Część społeczeństwa, która ma niezbyt przychylne zdanie o rządzie może nie chcieć brać udziału w badaniach organizowanych przez rządowe instytucje. Poziom aktywności społecznej uczestnika badania ma również wpływ na to czy wyrazi zgodę na wywiad. Osoby, które głosują w wyborach lub są wolontariuszami o wiele chętniej biorą udział w badaniach. Co ciekawe osoby, które regularnie głosują częściej wyrażają zgodę na badanie od tych niegłosujących nawet, jeżeli temat badania nie dotyczy polityki. Temat ankiety, sponsor, oraz jej długość mają niebagatelne znaczenie czy ktoś weźmie udział w badaniu czy nie. Trzeba też pamiętać, że tą samą grupę ludzi można łatwo namówić do jednej ankiety, a może być ciężko ich namówić do drugiej ankiety.

5. Trudne do przeprowadzenia badania (ang. *hard to interview*)

Przeprowadzenie badania może być mocno utrudnione, gdy:

- osoba badana należy do specjalnej grupy (np. więźniowie lub małe dzieci), które muszą mieć wyraźną zgodę od opiekuna, rodzica bądź strażnika na przeprowadzenie wywiadu.
- jest upośledzona umysłowa, bądź fizycznie,
- mogą nie znać języka w jakim został napisany kwestionariusz.

Jednak nawet w tych grupach przeprowadzenie wywiadu nie jest niemożliwe. Dzieci bardzo często biorą udział w badaniach (np. testy szkolne). Rozwiązanie problemu braku znajomości języka jest bardzo proste. Wystarczy na podstawie danych dotyczących ilości imigrantów przetłumaczyć ankietę na odpowiednie języki. W Kalifornii ze względu na dużą liczbę osób nie posługujących się angielskim California Health Interview Survey (CHIS) przeprowadzana jest w aż pięciu językach: angielskim, hiszpańskim, chińskim, koreańskim oraz wietnamskim. Wywiad z osobą niepełnosprawną fizyczną w odpowiednich warunkach też nie sprawia większego problemu. Osoby głuche mogą wypełnić kwestionariusz lub wywiad z nimi może być przeprowadzany w języku migowym. Z kolei trudność przeprowadzenia wywiadu z osobą niepełnosprawną umysłową zależy od stopnia upośledzenia. Wywiad osoby z downem bądź autyzmem będzie trudny w przeprowadzeniu, ale nie niemożliwy. Z kolei mózgowe porażenie dziecięce uniemożliwia przeprowadzenie wywiadu.

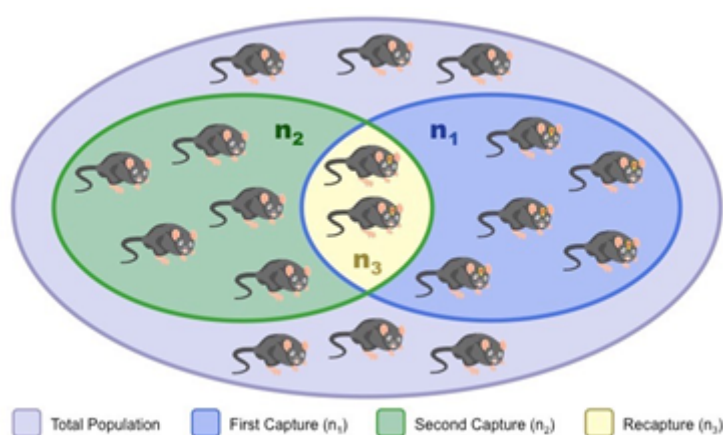
Trudność w zbadaniu określonych populacji może być pochodną wielu czynników. Tak też jest w przypadku populacji kawalerek przeznaczonych na wynajem w Poznaniu, której jednostki trudno wylosować do próby (brak pełnego operatu losowania oraz komplekso-

wych źródeł danych statystycznych). Trudno również zidentyfikować członków tej grupy, gdyż ogłoszenie dotyczące tej samej kawalerki może być na wielu portalach internetowych. Kolejną trudnością jest mobilność tej populacji. Wynajmujący bez większego trudu może usunąć ogłoszenie z jednego portalu i wstawić je na drugie. Czynniki te powodują, że estymacja liczebności tego typu populacji, zwłaszcza z uwzględnieniem dodatkowych przekrojów, jest niezwykle złożonym zadaniem.

2.2 Metody estymacji wielkości populacji

Istnieje wiele metody szacowania wielkości nieznanej populacji. Wśród najbardziej znanych można wymienić metody mnożnikowe czy metodę estymacji syntetycznych. Jednak ze względu na dostępność źródeł danych i spełnienie założeń do oszacowania liczby kawalerek na wynajem została wykorzystana metoda *capture-recapture*

Grupa metod pod nazwą *capture-recapture* wywodzi się z nauk przyrodniczych, jednak często stosować je również do szacowania wielkości populacji ludzkich. W przypadku zwierząt typowym sposobem jest umieszczenie siatek, które służą do wielokrotnego wylapywania jednostek danej populacji. Każdy osobnik jest znakowany i następnie wypuszczany. W każdej kolejnej próbie znakuje się osobniki złapane po raz pierwszy i notuje się jednostki, które wcześniej były już złapane. Na koniec analizuje się historię złowień. Ten system nazywany jest *mark-recapture*, *tag-recapture*, lub *multiple-record system* (Pawlukiewicz 2019).



Rysunek 2.1. Poglądowa wizualizacja metody *capture-recapture* na przykładzie populacji zwierzęcych

Źródło: Rysunek na podstawie Pawlukiewicz (2019)

Metoda *capture-recapture* opiera się na wykorzystaniu dwóch lub więcej źródeł danych wobec, których istnieją następujące założenia (Bereświcz i in. 2019):

1. definicje populacji we wszystkich źródłach są takie same (przykładowo: populacja kawalerów w Poznaniu),
2. populacja jest zamknięta (tj. stała w danym czasie),
3. źródła danych są niezależne (tj. wystąpienie jednostki w jednym źródle danych nie wpływa na pojawienie się tej samej jednostki w drugim źródle danych),
4. brak błędów pokrycia i duplikatów (tj. jeden rekord w zbiorze danych oznacza jedną kawalerkę),
5. brak błędów łączenia między źródłami (tj. łączenie następuje po identyfikatorze),
6. prawdopodobieństwa włączenia do co najmniej jednego z rejestrów powinny być jednorodne.

Spełnienie tych założeń jest kluczowe w kontekście możliwości stosowania omawianych metod zarówno w przypadku dwóch, jak i wielu źródeł. Niespełnienie tych założeń będzie skutkowało obciążonymi szacunkami wielkości populacji.

W literaturze można odnaleźć wiele przykładów wykorzystania metody *capture-recapture*. (Bereświcz i in. 2019) podjęli się próby oszacowania liczby cudzoziemców przebywających na terytorium Polski w roku 2015 oraz 2016 na podstawie danych z Rejestru PESEL, Systemu Pobyt (Urząd ds. Cudzoziemców) oraz Centralnego Rejestru Ubezpieczonych (ZUS). Z kolei (Gerritse i in. 2015) oszacowali liczbę Polaków oraz osób urodzonych na Bliskim Wschodzie, a przebywających w Holandii w roku 2009 oraz 2011. (Zhang 2008) w swojej pracy oszacował liczbę nielegalnie przebywających cudzoziemców w Norwegii. Autorowi nie jest znana żadna praca badawcza, która wykorzystywałaby metody *capture-recapture* do szacowania populacji na rynku nieruchomości

2.3 Model log-liniowy

Na potrzeby estymacji liczby kawalerek z uwzględnieniem dodatkowych zmiennych zdecydowano się wykorzystać metodę *capture-recapture* bazującą na modelach log-liniowych.

Właściwie skonstruowana tablica kontyngencji, w której wykorzystuje się informacje z dwóch bądź lub większej liczby źródeł danych jest niezbędna do prawidłowego zastosowania

modeli log-liniowych w estymacji. W tablicy 2.1 przedstawiono sytuację, gdy wykorzystujemy dwa niezależne źródła danych A i B.

Tablica 2.1. Przypadek dwóch źródeł - tablica kontyngencji 2×2

	Źródło B			\sum
		Tak (1)	Nie (0)	
Źródło A	Tak (1)	n_{11}	n_{10}	n_{1+}
	Nie (0)	n_{01}	n_{00}	n_{0+}
\sum		n_{+1}	n_{+0}	n

Źródło: Beresewicz i in. (2019)

W powyższej tabeli Tak (1) oznacza, że dana jednostka występuje w odpowiednim źródle, a Nie (0), że jednostka w tym źródle nie występuje. Przykładowo, n_{11} oznacza liczbę jednostek, które występują zarówno w źródle A i B. Kluczową kwestią jest zatem oszacowanie liczebności n_{00} , tj. liczby jednostek, które nie występują zarówno w źródle A jak i B, gdyż całkowitą wielkość populacji otrzymujemy po sumowaniu wartości tablicy 2.1 po wcześniejszej estymacji liczebności n_{00} .

Aby móc oszacować n_{00} możemy zastosować model logarytmiczno-liniowy do niekompletnej tablicy kontyngencji. Dla Tablicy 2.1 pełen model log-liniowy [AB]¹ został przedstawiony we wzorze (2.1)

$$\log(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}, \quad (2.1)$$

gdzie: m_{ij} – oczekiwaną liczebność w komórce i,j, a λ_j^B – wpływ źródła B. Ponieważ jednak komórka $m_{00} = m_{Nie,Nie}$ nie jest obserwowana, model [AB] ma o jeden parametr za dużo i nie może być zatem oszacowany. W takim przypadku należy rozważyć model niezależności [A][B] postaci:

$$\log(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B \quad (2.2)$$

Który ma tylko trzy parametry do oszacowania. Po dopasowaniu tego modelu do danych możemy użyć oszacowanych parametrów do wyznaczenia liczebności brakującej komórki ('Nie', 'Nie'), a następnie wyznaczyć liczebność populacji poddanej analizie. Wzór na oszacowanie wartości komórki to $n_{00} = \exp(\mu)$.

¹Jest to tzw. notacja nawiasowa, która w przypadku modeli log-liniowych jest często stosowana.

W oszacowaniu wielkości populacji możliwe jest wykorzystanie zmiennych pomocniczych takich jak dzielnica w jakiej znajduje się kawalerka lub czy ogłoszenie jest prywatne bądź biznesowe. Celem jest z jednej strony obejście jednego z założeń metody *capture-recapture* (o stałej stopie pokrycia przez źródło w populacji) i uwzględnienie faktu heterogeniczności przynależności poszczególnych jednostek do źródeł (Bereśewicz i in. 2019).

W poniższej tablicy kontyngencji oprócz przynależności do dwóch źródeł A i B badamy też dodatkową zmienną X (czy ogłoszenie jest prywatne lub biznesowe). Model log-liniowy [AX][BX] dopasowany do takiej tablicy jest postaci:

$$\log(m_{ij}) = \mu + \lambda_i^A + \lambda_j^B + \lambda_x^X + \lambda_{ix}^{AX} + \lambda_{jx}^{BX}, \quad (2.3)$$

gdzie λ_{ix}^{AX} oraz λ_{jx}^{BX} oznaczają efekty interakcji pomiędzy zmienną pomocniczą X i źródłami danych A oraz B odpowiednio.

Tablica 2.2. Przypadek dwóch źródeł A i B oraz jednej zmiennej pomocniczej X

	Zmienna X								
	X_1			X_2					
	Źródło B			Źródło B					
Źródło A	Tak (1)	n_{111}	Nie (0)	n_{101}	Tak (1)	n_{110}	Nie (0)	n_{100}	Σ
	Nie (0)	n_{011}	n_{001}	n_{010}	n_{010}	n_{000}	n_{0++}		
	Σ	n_{+11}	n_{+01}	n_{+10}	n_{+00}	\mathbf{n}			

Źródło: Bereśewicz i in. (2019)

W przypadku dwóch źródeł danych oraz jednej zmiennej pomocniczej przyjmującej dwa warianty X_1 – ogłoszenie prywatne, X_2 – ogłoszenie biznesowe. Mamy 6 znanych wartości oraz dwie: n_{001} i n_{000} których nie znamy. Po dopasowaniu modelu do danych brakujące wartości komórek może obliczyć ze wzorów: $n_{000} = \exp(\mu)$ i $n_{001} = \exp(\mu + \lambda_{x1}^X)$.

2.4 Mierniki oceny jakości dopasowania modelu

- **Iloraz wiarygodności**

Celem analizy logarytmiczno-liniowej jest wybór modelu o jak najprostszej postaci, który jest jednocześnie modelem dobrze dopasowanym do danych (Brzezińska 2015). Niestety w analizie log-liniowej nie ma jednoznacznych kryteriów pozwalających na wybór najlep-

szego modelu. Jedną z miar pozwalających na ocenę dopasowania modelu do danych jest iloraz wiarygodności.

$$G^2 = 2 \sum \sum n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right). \quad (2.4)$$

Wysoka wartość współczynnika G^2 oznacza, że model nie jest dopasowany do danych i powinien zostać odrzucony, gdyż w nieprawidłowy sposób odwzorowuje zależność między zmiennymi. Model akceptowalny to taki, którego liczebności oczekiwane nie odbiegają od liczebności teoretycznych. W przypadku $G^2 = 0$ dopasowanie modelu do danych jest idealne (sytuacja ta występuje jedynie w przypadku modelu pełnego). Iloraz wiarygodności ma rozkład chi-kwadrat i niestety może być stosowany tylko dla niewielkich prób ($n < 1500$). Dla większych prób musimy użyć kryteria AIC lub BIC.

- **Kryterium AIC**

Kryterium informacyjne Akaike (Akaike Information Criteria, AIC) służy do porównania większej liczby badanych modeli.

$$AIC = G^2 - df, \quad (2.5)$$

gdzie G^2 – iloraz wiarygodności badanego modelu, df – liczba stopni swobody. Przewagą AIC nad ilorazem wiarygodności jest uwzględnienie złożoności zjawiska poprzez liczbę parametrów w modelu.

- **Kryterium BIC**

$$BIC = G^2 - df \times \log(n), \quad (2.6)$$

gdzie n – liczebność tablicy kontyngencji. Kryterium BIC uwzględnia wielkość próby i w porównaniu z kryterium AIC jest znacznie częściej wykorzystywane w celu wyboru najlepszego modelu. Im mniejsza wartość kryterium AIC lub BIC tym model jest lepiej dopasowany do danych. Zarówno kryterium AIC jak i BIC są miarami relatywnymi, co oznacza, że można stosować je jedynie do porównania kilku modeli ze sobą. W przypadku pojedynczego modelu współczynniki nie posiadają interpretacji, a ich wartości nie mówią nic o jakości modelu.

2.5 Wyznaczenie przedziałów ufności dla wielkości populacji

W pracy badawczej na potrzeby oceny jakości oszacowań liczby kawalerek na wynajem dokonano konstrukcji 95% przedziałów ufności bazujących na parametrycznej metodzie bootstrap. W celu utworzenia przedziałów ufności w pierwszej kolejności dokonuje się estymacji parametrów modelu log-liniowego oraz wielkości badanej populacji. Mając oszacowane parametry modelu oraz liczebności brakujących komórek wyznaczane są prawdopodobieństwa teoretyczne przynależności dla wszystkich komórek w tablicy kontyngencji. Następnie losowana jest próba z rozkładu wielomianowego przy uwzględnieniu oszacowanych prawdopodobieństw, która w dalszym etapie jest korygowana, tak aby odpowiadała strukturze obserwowanych danych. Wówczas dokonuje się dopasowania odpowiedniego modelu log-liniowego do kompletnej tablicy kontyngencji i uzyskuje pierwsze oszacowanie bootstrapowe liczebności populacji trudnej do zbadania. Procedurę tą przeprowadza się wielokrotnie, wyznaczając wariancję, a następnie przedział ufności dla liczebności populacji (Bereśewicz i in. 2019). Program 2.6 przedstawia wyznaczenie 95% przedziałów ufności w pakiecie statystycznym R.

2.6 Implementacja modelu w programie statystycznym R

Model log-liniowy został zaimplementowany w języku programowania R poprzez wykorzystanie pakietu `stats`. Pakiet ten zawiera wiele funkcji statystycznych, w tym również funkcję **glm (Generalized Linear Models)**, która została użyta do estymacji wielkości badanej populacji.

```
glm(formula, family, data)
```

1

Program 2.1. Funkcja `glm` do szacowania uogólnionych modeli liniowych w języku R

Parametry powyższej funkcji są następujące:

- `formula` — symboliczny zapis modelu
- `family` — rozkład prawdopodobieństwa, który został użyty w modelu
- `data` — zbiór danych, które zawiera zmienne wykorzystane w modelu

```
model <- glm(formula = n ~ zrodlo1 + zrodlo2 + zmiennaPomocnicza, family= 1  
poisson(), data=dane)
```

Program 2.2. Wykorzystanie funkcji `glm` do oszacowania modelu log-liniowego

W formule funkcji podany jest model, który powinien najlepiej dopasowywać się do danych. Rozkładem prawdopodobieństwa jakiego użyjemy w modelu jest rozkład Poissona, który jest

dyskretnym rozkładem prawdopodobieństwa wyrażający prawdopodobieństwo szeregu wydarzeń mających miejsce w określonym czasie, gdy te wydarzenia występują ze znaną średnią częstotliwością i w sposób niezależny od czasu jaki upłynął od ostatniego zajścia takiego zdarzenia. Dane, do których będziemy dopasowywać model pochodzą z `data.frame` o nazwie `dane`.

Wynikiem tej funkcji jest obiekt 'model', z którego za pomocą funkcji `exp` oraz `coef` jesteśmy w stanie oszacować wielkość nieznanej populacji.

```
exp(coef((model)[1]))
```

1

Program 2.3. Wyznaczenie wielkości populacji w przypadku dwóch źródeł

```
n_0 <- exp(coef(model)[1])
n_obs <- count0$n
n_all <- c(n_obs, n_0)
N <- sum(n_all)
pr <- n_all/N
dane_sim <- count0
set.seed(123)
B <- 1000
bootstrap_samples <- rmultinom(n = B, size = N, prob = pr)
bootstrap_samples <- bootstrap_samples[1:3,]
wynik <- list()
for (i in 1:B) {
  dane_sim$n <- bootstrap_samples[,i]
  model_a <- glm(n ~ zdroj1 + zdroj2, data = dane_sim, family = poisson())
  wynik[[i]] <- exp(coef(model_a)[1])
}
bootstrap_oszacowania <- colSums(bootstrap_samples) + unlist(wynik)
c(mean(bootstrap_oszacowania), sd(bootstrap_oszacowania))
```

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

Program 2.4. Wyznaczenie 95% przedziałów ufności za pomocą metody bootstrap

2.7 Podsumowanie

W tym rozdziale zdefiniowano populację trudną do zbadania oraz przedstawiono dowody dlaczego populacja kawalerek na wynajem również zalicza się do takich populacji. Następnie opisano metodę estymacji wielkości populacji *capture-recapture* bazującą na modelach logliniowym. W następnym rozdziale metoda ta zostanie wykorzystana do estymacji nieznanej populacji kawalerek na wynajem w Poznaniu.

Rozdział 3

Estymacja wielkości rynku kawalerek w Poznaniu

3.1 Wykorzystane źródła danych

Dane, które wykorzystano do oszacowania liczby kawalerek na wynajem w Poznaniu, zostały pozyskane z dwóch prywatnych, nieklasycznych źródeł danych.

- OLX – dane o ogłoszeniach nieruchomości, które zostały opublikowane na portalu. Zakres czasowy zbioru danych to 2019 kw. 1-2020 kw.3.
- OtoDom – dane o ogłoszeniach nieruchomości, które zostały opublikowane na tymże portalu. Zakres czasowy zbioru danych to 2019 kw. 1-2020 kw.3.

3.1.1 Opis zbioru danych z OLX

Zbiór danych z OLX składa się z ogłoszeń opublikowanych na portalu w okresie od stycznia 2019 do września 2020 roku. Nieprzetworzony zbiór składa się z 85 tys rekordów oraz 30 kolumn. Z powodu tak dużej liczby zmiennych, omówione zostaną tylko najważniejsze kolumny:

- `Id` – identyfikator ogłoszenia,
- `Description` – opis nieruchomości,
- `User_id` – id osoby publikującej,
- `Phone_numbers` – numer telefonu osoby publikującej ogłoszenie,
- `Price` – cena najmu,
- `m` – metraż,

- `Rooms` – liczba pokoi,
- `Advertiser_type` – informuje czy ogłoszenie osoby prywatnej lub agencji,
- `District_id` – id dzielnicy w jakiej znajduje się mieszkanie,
- `Created_at` – dokładny czas stworzenia ogłoszenia,
- `Last_refresh` – data, kiedy ogłoszenie zostało ostatni raz odświeżone,
- `External_id` – id tego samego ogłoszenia na serwisie partnerskim,
- `External_partner_code` – nazwa serwisu partnerskiego.

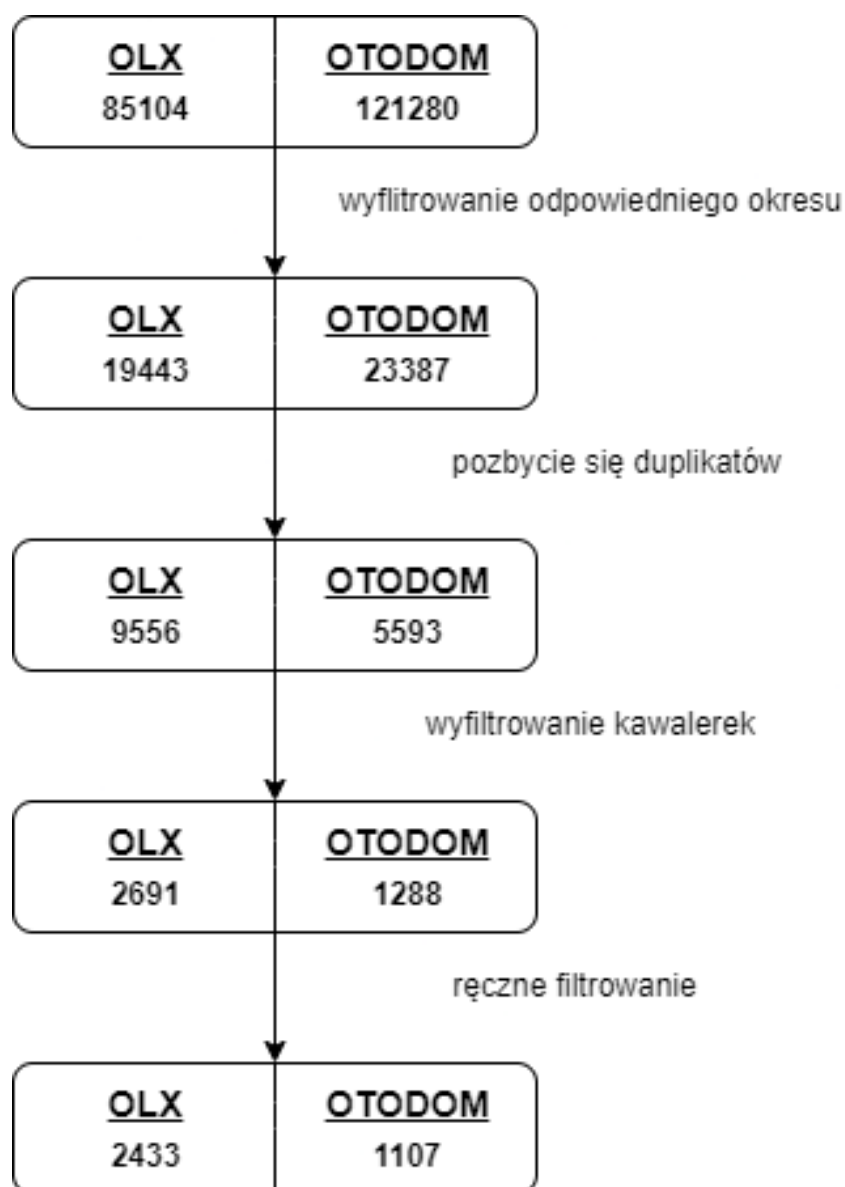
3.1.2 Opis zbioru danych z OtoDom

Zbiór danych z OtoDom składa się z ogłoszeń opublikowanych na tymże portalu w okresie od stycznia 2019 do września 2020 roku. Nieprzetworzony zbiór składa się z ponad 121 tys. rekordów oraz 88 kolumn. Z powodu tak dużej liczby zmiennych, omówione zostaną tylko najważniejsze kolumny:

- `ID` – identyfikator ogłoszenia,
- `Description` – opis nieruchomości,
- `SellerInfo.ID` – id osoby publikującej,
- `SellerInfo.Contacts.Number` – numer telefonu osoby publikującej ogłoszenie,
- `Price` – cena najmu,
- `Area` – metraż,
- `FlatDetails.RoomsNum` – liczba pokoi,
- `Individual` – informuje czy ogłoszenie osoby prywatnej bądź firmy,
- `District` – id dzielnicy w jakiej znajduje się mieszkanie,
- `ActivationDate` - dokładny czas stworzenia ogłoszenia,
- `ModificationDate` – data, kiedy ogłoszenie zostało ostatni raz zmodyfikowane,
- `RemoteId` – id tego samego ogłoszenia na serwisie partnerskim.

3.2 Proces przygotowania danych

Procedura oczyszczania danych zarówno z OLX jak i z OtoDom składała się z kilku etapów. Wykres 3.1 przedstawia poszczególne etapy odsiewania zbędnych danych oraz informuje ile rekordów zostało w zbiorze danych po każdym kroku oczyszczania danych. Z początkowych 121



Rysunek 3.1. Etapy oczyszczania danych

tysięcy rekordów w OtoDom oraz 85 tysięcy w OLX finalnie w zbiorze danych pozostały tylko 2433 rekordy w OLX oraz 1107 w OtoDom.

1. Ujednolicenie okresu

Zarówno zbiór danych z OLX jak i z OtoDom zawiera mnóstwo informacji o ogłoszeniach, które nie były aktywne w trzecim kwartale 2020 roku. Część ogłoszeń, które zawierają te zbiory danych zostało opublikowane jeszcze w 2016 roku. By pozbyć się niepotrzebnych danych zostawiono tylko te rekordy, których wartość w kolumnie `last_refresh` (w OtoDom `ModificationDate`) znajduje się w przedziale od 1 lipca 2020 roku do 30 września 2020 roku. Zmienne `last_refresh` oraz `ModificationDate` zostały wy-

brane do filtrowania, gdyż informują o tym czy w danym okresie ogłoszenie było aktualne, a nie czy w tym okresie zostało opublikowane. Pozwoli to lepiej oddać rzeczywistość. Często w badanym zbiorze danych występowały ogłoszenia, które były opublikowane np. na początku 2020 roku, a przez to, że nie było chętnych na dane lokum ogłoszenie były aktywne w badanym trzecim kwartale 2020 roku.

2. Usunięcie duplikatów

By pozbyć się ze zbioru danych duplikatów wykonano następujące czynności. Utworzono sztuczną kolumnę `point`, która zawiera wynik odejmowania daty z kolumny `last_refresh` od daty „2021-01-01”. Kolejnym krokiem jest pogrupowanie ogłoszeń po ID i wyfiltrowanie tych ogłoszeń z każdej grupy, których wartość w kolumnie `point` jest najmniejsza.

3. Wybranie kawalerek

Na obecnym etapie w zbiorze danych mamy ogłoszenia dotyczące zarówno mieszkań, kawalerek czy pokoi. Badanie dotyczy tylko kawalerek, dlatego wybrano tylko te rekordy w których opisie ogłoszenia znajduje się słowo klucz „kawaler”. Dodatkowo jeszcze wyfiltrowano tylko te ogłoszenia, które w kolumnie `rooms` miały wartość „1”. W badaniu również nie brano pod uwagę najmu krótkoterminowego oraz najmu z odstępnym, dlatego wydobyto tylko te rekordy, które nie zawierają słów-kluczy: „odstępne” oraz „krótko”. Dodatkowo pozbyto się ogłoszeń, które dotyczą najmu kawalerek, których metraż jest mniejszy niż $15m^2$ lub większy niż $50m^2$ oraz wynajem kosztuje mniej niż 900 zł miesięcznie.

4. Ręczne filtrowanie

Powyższe kroki zapewniają wysoką jakość danych, ale nie gwarantują 100% pewności, że wszystkie przefiltrowane rekordy są poprawne. Nierzadko wynajmujący publikował kilka ogłoszeń, które dotyczyły tej samej kawalerki. Ogłoszenia te miały różne ID, co powodowało, że algorytm użyty do oczyszczania danych nie traktował ich jako duplikatów. Dlatego ręcznie przefiltrowano dane, które zostały po tych wszystkich krokach. W wyniku ręcznego filtrowania pozbyto się 178 rekordów z OLX oraz 139 z OtoDom.

3.3 Weryfikacja założeń modelu

W związku z użyciem w badaniu metody capture-recapture z wykorzystaniem dwóch źródeł danych spełniono założenia, które są niezbędne do poprawnego oszacowania wielkości nieznannej populacji:

- Definicje populacji we wszystkich rozważanych źródłach są takie same – określono populację kawalerek na wynajem w Poznaniu w trzecim kwartale 2020 roku. Każde z wykorzystanych źródeł zostało ograniczone do tej populacji.
- Populacja jest zamknięta – wielkość populacji na dzień 30.09.2020 roku była stała oraz aktualna.
- Źródła danych są niezależne – pojawienie się jednostki w jednym źródle nie zależy od przynależności tej jednostki do drugiego źródła. Na potrzeby tego badania założono, że tak jest (nie można tego zweryfikować na podstawie dostępnych źródeł).
- Brak błędów nadreprezentacji i duplikatów – dane zostały pozbawione duplikatów oraz pozbawione nadreprezentacji w procesie oczyszczania danych.
- Każdą jednostkę będzie można zidentyfikować i połączyć między źródłami bez błędów – dane zostały ze sobą połączone za pomocą identyfikatora ID oraz external_ID.

3.4 Estymacja parametrów modelu log-liniowego oraz jego ocena

3.4.1 Konstrukcja modelu log liniowego

Na podstawie zbiorów danych z OLX oraz OtoDom został stworzony połączony zbiór danych który, zawiera takie zmienne:

- OtoDom – zmienna informuje czy ogłoszenie zostało opublikowane na portalu OtoDom
- OLX – zmienna informuje czy ogłoszenie zostało opublikowane na portalu OLX
- Ogłoszeniodawca – informuje czy oferta została opublikowana przez osobę prywatną lub agencję nieruchomości
- floor_cut – kolumna, która informuje w jakim przedziale metrażowym znajdują się kawalerka. Przedziały wyznaczono metodą ekspercką na podstawie rozkładu zmiennej.

Jeżeli ogłoszenie zostało opublikowane na portalu OtoDom, to w kolumnie OtoDom pojawi się wartość 1, w przeciwnym wypadku 0. Podobnie w przypadku OLX.

Dzięki temu połączonemu zbiorowi danych możliwe jest powstanie tablic kontyngencji, które są nie zbędne do stworzenia modelu. Dla każdego modelu została przygotowana osobna tablica. Poniżej zaprezentowano dwie tablice kontyngencji. Tablica A.1, która została wykorzystana do estymacji modelu 2 ze względu na wielkość oraz złożoność została zamieszczona w Aneksie.

Tablica 3.1. Tablica kontyngencji do modelu 0

OtoDom	OLX	n
1	1	541
1	0	542
0	1	1774

Źródło: Opracowanie własne.

Tablica 3.2. Tablica kontyngencji do modelu 1

OtoDom	OLX	Ogłoszeniodawca	n
1	1	biznesowy	170
1	1	prywatny	371
1	0	biznesowy	516
1	0	prywatny	26
0	1	biznesowy	437
0	1	prywatny	1337

Źródło: Opracowanie własne.

Skonstruowano trzy modele, które pozwolą określić wielkość populacji kawalerek na wynajem w Poznaniu:

- Model 0 – model ten został stworzony tylko na podstawie przynależności ogłoszeń do źródła,

```
model_0 <- glm(formula = n ~ OtoDom + OLX, family = poisson(), data = dane0)
```

Program 3.1. Model 0 – dwa źródła danych

- Model 1 – w tym modelu oprócz źródła w jakim pojawiło się ogłoszenie jest jedna zmienna pomocnicza: Ogłoszeniodawca (zgodnie ze wzorem (2.3))

```
model_1 <- glm(formula = n ~ OtoDom + OLX + Ogłoszeniodawca, family =  
poisson(), data=dane1)
```

Program 3.2. Model 1 – dwa źródła danych oraz jedna zmienna pomocnicza

- Model 2 – zmiennymi objaśniającymi w tym modelu jest: źródło w jakim pojawiło się ogłoszenie, Ogłoszeniodawca oraz floor_cut.

```
model_2 <- glm(formula = n ~ OtoDom + OLX + Ogłoszeniodawca + floor_  
cut, family = poisson(), data=dane2)
```

Program 3.3. Model 2 – dwa źródła danych oraz dwie zmienne pomocnicze

3.4.2 Wyniki estymacji parametrów modelu log-liniowego

Tabela 3.3 przedstawia wartości parametrów poszczególnych modeli, błąd standardowy parametrów oraz wartość p. Kolumna 'Oszacowanie parametrów' przedstawia wyniki estymacji parametrów poszczególnych modeli. Kolumna 'Błąd standardowy' pokazuje błąd standardowy oszacowanych parametrów. Z kolei kolumna 'p-value' informuje czy dany parametr był istotny statystycznie.

3.5 Ocena modelu log-liniowego

Tabela 3.4 zawiera poszczególne miary jakości dopasowania modeli do danych. Występują duże różnice w miarach pomiędzy modelem bez zmiennych pomocniczych, a modelami z zmiennymi pomocniczymi. Z kolei wartości miar dopasowania pomiędzy modelem pierwszym, a drugim są zbliżone. Model 0 okazał się najlepiej dopasowany do danych na co wskazują najniższe wartości kryteriów informacyjnych AIC i BIC. Model 2, który zawiera dwie zmienne pomocnicze okazał się najgorzej dopasowany do danych, co pokazują bardzo wysokie wartości AIC i BIC.

3.6 Szacunek wielkości populacji

Tabela 3.5 zawiera wyniki oszacowań oraz 95% przedziały ufności dla oszacowanej wielkości populacji kawalerek na wynajem w 3 kwartale 2020 roku w Poznaniu dla modelu 0 (bez zmiennej pomocniczej), modelu 1 (z jedną zmienną pomocniczą) oraz modelu 2 (z dwoma zmiennymi pomocniczymi). Według modelu bez zmiennych pomocniczych szacunek liczba kawalerek na wynajem wynosi 4644. Kolejne dwa modele oszacowały podobną wielkość populacji kawalerek na wynajem.

Tablica 3.3. Wartości parametrów modelu zerowego

Model 0			
Zmienna	Oszacowanie parametrów	Błąd standardowy	p-value
(Intercept)	7.482	0.065	<2e-16
OtoDom	-1.187	0.049	<2e-16
OLX	-0.001	0.060	0.976
Model 1			
Zmienna	wartość parametru	Błąd standardowy	p-value
(Intercept)	6.549	0.069	<2e-16
OtoDom	-1.187	0.049	<2e-16
OLX	-0.001	0.060	0.976
Ogłoszeniodawca	0.434	0.038	<2e-16
Model 2			
Zmienna	wartość parametru	Błąd standardowy	p-value
(Intercept)	6.549	0.069	<2e-16
OtoDom	-1.187	0.049	<2e-16
OLX	-0.001	0.060	0.976
Ogłoszeniodawca	0.434	0.038	<2e-16
floor_cut[20,25)	0.865	0.090	<2e-16
floor_cut[25,30)	1.434	0.084	<2e-16
floor_cut[30,35)	1.784	0.082	<2e-16
floor_cut[35,40)	0.792	0.091	<2e-16
floor_cut[40,50]	-0.248	0.114	0.030

Źródło: Opracowanie własne.

Tablica 3.4. Wybrane miary jakości modeli log-liniowych

model	Dewiancja	G^2	AIC	BIC
0	1.9273e-13	-12.79	31.584	28.87986
1	966.05	-505.46	1018.90	1018.09
2	1021.40	-608.50	1235.00	1249.25

Źródło: Opracowanie własne.

Na podstawie modelu zerowego 95% przedział ufności wynosi od 4519 do 4769 kawalerek na wynajem. Podobne przedziały powstały na podstawie modelu pierwszego i drugiego. Wyniki tej pracy badawczej można porównywać tylko z danymi pochodzącymi z Narodowych Spisów Powszechnych. Niestety ostatnie informacje o liczbie kawalerek pochodzą ze spisu sprzed 10 lat, co oznacza, że dane z 2011 roku, które mówiły o 7491 kawalerkach w Poznaniu są nie aktualne. Oszacowana wielkość populacji wskazuje na dużą dominację portali OLX oraz OtoDom

Tablica 3.5. Szacunek wielkości populacji wraz z 95% przedziałem ufności

model	\hat{N}	Przedział ufności
0	4644	(4519, 4769)
1	4637	(4506, 4768)
2	4644	(4500, 4776)

Źródło: Opracowanie własne.

na poznańskim rynku nieruchomości. Około 75% ofert wynajmu kawalerek zostało opublikowanych w trzecim kwartale 2020 roku na tych dwóch portalach.

Podsumowanie

Głównym celem badawczym niniejszej pracy licencjackiej było oszacowanie liczby kawalerek oferowanych na wynajem w Poznaniu w trzecim kwartale 2020 roku. W tym celu zastosowano model logarytmiczno-liniowy. Estymowanym parametrem jest teoretyczna wielkość populacji ogłoszeń najmu kawalerek, które nie znalazły się w żadnym badanym zbiorze.

Do oszacowania modelu wykorzystano dwa niezależne źródła danych: portale OLX oraz OtoDom. Z tych zbiorów danych wybrano tylko te rekordy, które spełniały definicję ogłoszenia najmu kawalerki przedstawionej w pierwszym rozdziale. Dane były aktualne na dzień 30 września 2020 roku.

W części praktycznej oszacowano trzy modele 1) model zerowy, który uwzględniał tylko źródła danych, 2) model pierwszy, który brał pod uwagę jedną zmienną pomocniczą oraz 3) model drugi, z dwiema zmiennymi pomocniczymi.

Szacowana liczba oferowanych kawalerek na wynajem w Poznaniu wynosiła dla poszczególnych modeli:

1. model zerowy – 4644 ogłoszeń ; 95% przedział ufności (4519, 4769),
2. model pierwszy – 4637 ogłoszeń ; 95% przedział ufności (4506, 4768),
3. model drugi – 4644 ogłoszeń ; 95% przedział ufności (4500, 4776).

Wszystkie trzy modele szacują podobną wielkość populacji ogłoszeń najmu w Poznaniu w trzecim kwartale 2020 roku. Niestety, obecnie nie jest możliwe zweryfikowanie tych liczb, gdyż ostatnie dane dotyczące liczebności kawalerek w Poznaniu pochodzą z 2011 roku. Wtedy w stolicy wielkopolski było 7491 kawalerek co wskazywałoby, że ponad 60% jednoizbowych lokali było przeznaczonych na wynajem. Jednak są to dane przestarzałe. Aktualne dane będą dostępne dopiero po przeprowadzeniu Narodowego Spisu Powszechnego 2021.

W latach 2010-2020 liczba mieszkań w Poznaniu zwiększyła się o 35 000 mieszkań do 271 688, a kawalerki stają się coraz bardziej pożądanym typem mieszkania. W tym okresie studenci

ze względu na naukę zdalną spowodowaną pandemią koronawirusa rezygnowali z wynajmu mieszkań, w tym kawalerek. Dlatego można przypuszczać, iż oszacowana populacja ofert jest blisko rzeczywistej liczby ogłoszeń.

Ograniczeniem tego modelu jest jakość dostępnych danych. Ogłoszenia zarówno z OLX jak i z OtoDom posiadały mnóstwo błędów, takich jak różniące się parametry lokalu pomiędzy opisem a kolumnami, które zawierały cenę oraz metraż mieszkania. Szczególnie ten problem występował w ofertach z OLX, gdzie sprzedający specjalnie zaniżali w parametrach kwotę najmu, by więcej osób odwiedziło ich ogłoszenie. Wynajmujący dodawali równocześnie kilka ofert dotyczących tej samej kawalerki. Często przy tym lekko zmieniali cenę oraz metraż co uniemożliwiało automatyczne usunięcie duplikatów i wymagało ręcznego filtrowania danych.

Kolejnym ograniczeniem modelu jest wątpliwa niezależność źródeł danych, gdyż jak przedstawiono w pierwszym rozdziale, ogłoszenia opublikowane w OtoDom automatycznie zamieszczane są również na OLX. Wyjątkiem od tej reguły są oferty deweloperów lub agencji nieruchomości. W takim przypadku oferta zostanie opublikowana tylko na portalu OtoDom. Jest to jednak wyjątek, który często ma miejsce, ponieważ tylko połowa ogłoszeń, które zostały zamieszczone w OtoDom znalazły się również na OLX.

Należy jednak podkreślić, że zgodnie z wiedzą autora i promotora jest to pierwsze badanie w Polsce, którego celem jest zastosowanie metod *capture-recapture* do oszacowania liczby kawalerek oferowanych na wynajem w Poznaniu. Dotychczas nie pojawiły się podobne szacunki w źródłach statystycznych (np. GUS) czy niestatystycznych (np. raporty branżowe podawały wyłącznie dane pochodzące z portali internetowych, nie brano pod uwagę kawalerek prezentowanych poza tymi źródłami).

Bibliografia

- Andrews, D. & Sánchez, A. C. (2011). Drivers of homeownership rates in selected OECD countries.
- Bereśewicz, M., Gudaszewski, G. & Szymkowiak, M. (2019). Estymacja liczby cudzoziemców w Polsce z wykorzystaniem metody capture-recapture. *Wiadomości Statystyczne*, 64(10).
- Beręśewicz, M. (2016). *Internet data sources for real estate market statistics*. Poznan University of Economics; Business. Retrieved from [http://www. wbc ...](http://www.wbc...)
- Beręśewicz, M. & Szymkowiak, M. (2015). Big data w statystyce publicznej–nadzieje, osiągnięcia, wyzwania i zagrożenia. *Ekonometria*, 2, 9–22.
- Brzezińska, J. (2015). *Analiza logarytmiczno-liniowa: teoria i zastosowania z wykorzystaniem programu R*. Wydawnictwo CH Beck.
- Criddle, C. (2020). Facebook sued over Cambridge Analytica data scandal.
- Fiegerman, S. (2016). Yahoo says 500 million accounts stolen.
- Gemius/PBI. (2020). Wyniki badania Gemius/PBI za luty 2020.
- Gerritse, S. C., van der Heijden, P. G. & Bakker, B. F. (2015). Sensitivity of population size estimation for violating parametric assumptions in log-linear models. *Journal of official statistics*, 31(3), 357–379.
- Hardek, L. A. (2014). Konsumenci niechętni umowom najmu okazjonalnego - podsumowanie.
- Łaszek, J., Olszewski, K., Waszczuk, J., Zaczek, M. & Augustyniak, H. (2019). Raport o sytuacji na rynku nieruchomości mieszkaniowych i komercyjnych w Polsce w 2018 r.
- Nordbotten, S. (2010). The Use of Administrative Data in Official Statistics – Past, Present, and Future – With Special Reference to the Nordic Countries. *Journal of official statistics*.
- Pawlukiewicz, K. (2019). Estymacja liczby cudzoziemców nielegalnie przebywających w Polsce z wykorzystaniem hierarchicznego modelu Gamma-Poissona.
- PBS. (2017). Badanie satysfakcji ze współpracy z serwisem Otodom.pl.
- Pilcicka, A. (2018). Rynek najmu w Polsce kwitnie.

- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Tourangeau, R., Edwards, B., Johnson, T. P., Wolter, K. M. & Bates, N. (2014). *Hard-to-survey populations*. Cambridge University Press.
- Wallgren, A. & Wallgren, B. (2014). *Register-based statistics: statistical methods for administrative data*. John Wiley & Sons.
- Zhang, L.-C. (2008). Developing methods for determining the number of unauthorized foreigners in Norway. *Statistics Norway*, www.ssb.no/a/english/publikasjoner/pdf/doc_20_811.

Spis tablic

2.1	Przypadek dwóch źródeł - tablica kontyngencji 2×2	16
2.2	Przypadek dwóch źródeł A i B oraz jednej zmiennej pomocniczej X	17
3.1	Tablica kontyngencji do modelu 0	26
3.2	Tablica kontyngencji do modelu 1	26
3.3	Wartości parametrów modelu zerowego	28
3.4	Wybrane miary jakości modeli log-liniowych	28
3.5	Szacunek wielkości populacji wraz z 95% przedziałem ufności	29
A.1	tablica kontyngencji do modelu 2	46

Spis rysunków

1.1	Klasyfikacja źródeł danych w statystyce	8
2.1	Poglądowa wizualizacja metody <i>capture-recapture</i> na przykładzie populacji zwierzęcych	14
3.1	Etapy oczyszczania danych	23

Spis programów

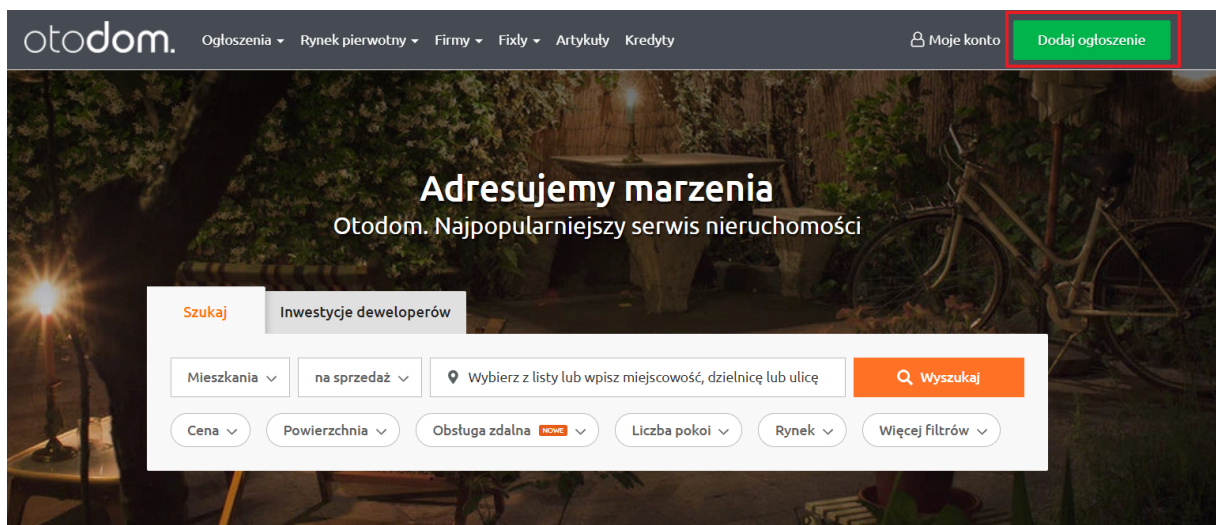
2.1	Funkcja glm do szacowania uogólnionych modeli liniowych w języku R	19
2.2	Wykorzystanie funkcji glm do oszacowania modelu log-liniowego	19
2.3	Wyznaczenie wielkości populacji w przypadku dwóch źródeł	20
2.4	Wyznaczenie 95% przedziałów ufności za pomocą metody bootstrap	20
3.1	Model 0 – dwa źródła danych	26
3.2	Model 1 – dwa źródła danych oraz jedna zmienna pomocnicza	27
3.3	Model 2 – dwa źródła danych oraz dwie zmienne pomocnicze	27

Dodatek A

Załączniki

A.1 Procedura dodawania ogłoszeń na Otodom

1. Wchodzimy na portal Otodom i klikamy przycisk "Dodaj ogłoszenie".
2. Najeżdżamy na zakładkę mieszkanie i klikamy "Wynajmę"
3. Uzupełniamy podstawowe informacje o mieszkaniu przeznaczonym na wynajem
4. Dodajemy multimedia
5. Podajemy lokalizację kawalerki
6. Podajemy informacje szczegółowe kawalerki
7. Zaznaczamy wyposażenie lokalu
8. Zaznaczamy wyposażenie lokalu
9. Zatwierdzamy ogłoszenie



otodom.
Ogłoszenia
Rynek pierwotny
Firmy
Fixly
Artykuły
Kredyty
Moje konto
Dodaj ogłoszenie

Dodaj ogłoszenie

Sprzedam

Wynajmę

mieszkania

domy

pokoje

dziatki

lokale użytkowe

hale i magazyny

garaże

Jesteś pośrednikiem lub deweloperem?
Zaloguj się

Nie masz jeszcze konta w Otodom?
Sprawdź naszą ofertę dla Klientów profesjonalnych

Wynajem Mieszkania zmień

Informacje podstawowe

Tytuł ogłoszenia ?

Kawalerka na Wynajem Wilda

24 znaków pozostało

Cena *

1000

zł

Czynsz - dodatkowo

500

zł

Kaucja

1500

zł

Powierzchnia *

30

m²

Liczba pokoi *

1

Typ ogłoszeniodawcy *

☒

właściciel nieruchomości

☐

pośrednik

Dodaj multimedia

Dodaj przynajmniej jedno zdjęcie, gdyż w przeciwnym razie Twoje ogłoszenie będzie wyświetlane pod koniec listy ogłoszeń.

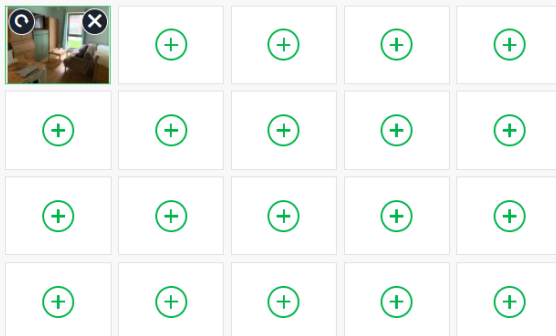


Możesz dodać do 20 zdjęć o wymiarach minimum 800x600 pikseli i wadze pliku nie większej niż 5MB, w formacie JPG, PNG lub GIF.

Aby dodać wiele zdjęć na raz, zaznaczaj je trzymając klawisz Ctrl na klawiaturze.

Dodaj multimedia

Dodaj przynajmniej jedno zdjęcie, gdyż w przeciwnym razie Twoje ogłoszenie będzie wyświetlane *pod koniec listy ogłoszeń*.



Możesz dodać do 20 zdjęć o wymiarach minimum 800x600 pikseli i wadze pliku nie większej niż 5MB, w formacie JPG, PNG lub GIF.

Aby dodać wiele zdjęć na raz, zaznaczaj je trzymając klawisz Ctrl na klawiaturze.

Pierwsza fotografia będzie wyświetlana jako główne zdjęcie nieruchomości.

Możesz zmienić kolejność zdjęć, przesuwając miniaturki z wciśniętym przyciskiem myszki.

Idealne wymiary zdjęć wynoszą 830x500 pikseli. Dowiedz się więcej.

film ?

wklej link

wirtualny spacer ?

wklej link

plan zagospodarowania ?

wklej link

Lokalizacja

miejsowość *

Poznań, wielkopolskie

dzielnica

Wilda

ulica lub osiedle

☐ lokalizacja dokładna

☒ lokalizacja przybliżona



Informacje szczegółowe

opis *

B *I* U

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

9665 znaków pozostało

Rodzaj zabudowy

blok

Piętro

parter

Liczba pięter

1

Materiał budynku

cegła

Okna

plastikowe

Ogrzewanie

miejskie

Rok budowy

2001

Stan wykończenia

do zamieszkania

Dostępne od

☒ Wynajmę również studentom

Wypożyczenie

☐ meble

☒ pralka

☐ zmywarka

☒ lodówka

☐ kuchenka

☒ piekarnik

☐ telewizor

Zabezpieczenia

☐ rolety antywłamaniowe

☐ drzwi / okna antywłamaniowe

☐ domofon / wideofon

☐ monitoring / ochrona

☐ system alarmowy

☐ teren zamknięty

Media

☒ internet

☐ telewizja kablowa

☐ telefon

Informacje dodatkowe

☐ balkon

☐ pom. użytkowe

☐ garaż/miejsce parkingowe

☐ piwnica

☐ ogródek

☐ taras

☐ winda

☐ dwupoziomowe

☐ oddzielna kuchnia

☐ klimatyzacja

☐ tylko dla niepalących

Obsługa zdalna

Korzystam z rozwiązań online wspierających transakcję (transmisja na żywo z nieruchomości, wideorozmowa, podpis elektroniczny itp.)

Obsługa zdalna

wybierz

Dane kontaktowe

Imię i nazwisko *

xyz

Numer telefonu

123 456 789

E-mail *

Xyz@zyx.pl

Dodając ogłoszenie akceptuję [Regulamin serwisu Otodom](#).

Przyjmuję do wiadomości, że Grupa OLX wykorzystuje moje dane osobowe zgodnie z [Polityką prywatności](#) oraz [Polityką dotyczącą plików cookie](#) i podobnych technologii. Grupa OLX wykorzystuje zautomatyzowane systemy i partnerów do analizowania, w jaki sposób korzystam z usług w celu zapewnienia odpowiedniej funkcjonalności produktu, treści, dostosowanych i opartych na zainteresowaniach reklam, jak również ochrony przed spamem, złośliwym oprogramowaniem i nieuprawnionym korzystaniem z naszych usług.



Wyrażam zgodę na używanie przez Grupę OLX sp. z o.o. środków komunikacji elektronicznej oraz telekomunikacyjnych urządzeń końcowych w celu przesyłania mi informacji handlowych oraz prowadzenia marketingu (np. newsletter, wiadomości SMS) przez Grupę OLX sp. z o.o., podmioty powiązane i partnerów biznesowych.

[Zobacz podgląd ogłoszenia](#)

Podgląd ogłoszenia

Edytuj

Zatwierdź

[Wróć do listy](#) > [Mieszkanie na wynajem](#) > [wielkopolskie](#) > [Poznań](#) > [Wilda](#) > [Kawalerka na Wynajem Wilda](#)

Kawalerka na Wynajem Wilda

[Poznań, Wilda](#)

1 000 zł



xyz

123

[Wyświetl numer](#)

Imię *

E-mail *

Numer telefonu

Administratorem Twoich danych osobowych jest Grupa OLX Sp. z o.o. [zobacz](#)

☐ Chcę otrzymać linki do podobnych ofert

[Wyślij wiadomość](#)

[Obserwuj](#)

Twoje ogłoszenia

Stan konta: 0 zł
Punkty na promowanie: 0 pkt.

Dodać

Uwaga

Proces kupna i sprzedaży przez [Przesyłki OLX](#) odbywa się na Twoim koncie OLX. Zignoruj linki wysyłane poza portalem, np. przez Whatsapp. Przeczytaj [więcej na naszym Blogu](#).

Ogłoszenia

Wiadomości

Płatności

Szukam pracy 94%

Ustawienia

📁 Twoje przesyłki

Dodaj ogłoszenie

Im więcej szczegółów, tym lepiej!

Tytuł ogłoszenia*

Kawalerka na Wynajem Wilda



Kategoria*



Wynajem

Nieruchomości / Mieszkania



Zdjęcia

Pierwsze zdjęcie będzie zdjęciem głównym. Przeciągaj zdjęcia na inne miejsca, aby zmienić ich kolejność.



A.2 Procedura dodawania ogłoszeń na OLX

1. Klikamy ikonę "Dodaj Ogłoszenie"
2. Podajemy tytuł ogłoszenia
3. Dodajemy opis kawalerki oraz cenę najmu
4. Podajemy dodatkowe informacje o kawalerce
5. Zamieszczamy dane kontaktowe
6. Klikamy ikonę "Dodaj ogłoszenie"

A.3 Dane wykorzystane w modelach

Opis*

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.



334/9000

Cena*

1000

zł

Do negocjacji



Dodatkowe informacje

Ogłoszenie prywatne czy firmowe?*

Prywatne

Firmowe

Poziom

Parter



Umeblowane*

Tak

Nie

Rodzaj zabudowy*

Blok



Powierzchnia*

m²

Liczba pokoi*

Kawalerka



Czynsz (dodatkowo)*

500



Dane kontaktowe

Lokalizacja*

Poznań, Wilda



Osoba kontaktowa*

XYZ

Adres e-mail

czitos100@gmail.com

Numer telefonu

123 456 678



Wyrażam zgodę na używanie przez Grupę OLX sp. z o.o. środków komunikacji elektronicznej oraz telekomunikacyjnych urządzeń końcowych w celu przysyłania mi informacji handlowych oraz prowadzenia marketingu (np. newsletter, wiadomości SMS) przez Grupę OLX sp. z o.o., podmioty powiązane i partnerów biznesowych.

[Podgląd ogłoszenia](#)

[Dodaj ogłoszenie](#)

Tablica A.1. tablica kontyngencji do modelu 2

OtoDom	OLX	Ogłoszeniodawca	floor_cut	n
otodom	olx	biznesowy	[15,20)	5
otodom	olx	biznesowy	[20,25)	22
otodom	olx	biznesowy	[25,30)	45
otodom	olx	biznesowy	[30,35)	55
otodom	olx	biznesowy	[35,40)	32
otodom	olx	biznesowy	[40,50)	11
otodom	olx	prywatny	[15,20)	18
otodom	olx	prywatny	[20,25)	59
otodom	olx	prywatny	[25,30)	96
otodom	olx	prywatny	[30,35)	142
otodom	olx	prywatny	[35,40)	48
otodom	olx	prywatny	[40,50)	8
otodom	0	biznesowy	[15,20)	26
otodom	0	biznesowy	[20,25)	71
otodom	0	biznesowy	[25,30)	132
otodom	0	biznesowy	[30,35)	177
otodom	0	biznesowy	[35,40)	78
otodom	0	biznesowy	[40,50)	32
otodom	0	prywatny	[15,20)	4
otodom	0	prywatny	[20,25)	2
otodom	0	prywatny	[25,30)	4
otodom	0	prywatny	[30,35)	12
otodom	0	prywatny	[35,40)	3
otodom	0	prywatny	[40,50)	1
0	olx	biznesowy	[15,20)	33
0	olx	biznesowy	[20,25)	73
0	olx	biznesowy	[25,30)	125
0	olx	biznesowy	[30,35)	125
0	olx	biznesowy	[35,40)	48
0	olx	biznesowy	[40,50)	33
0	olx	prywatny	[15,20)	87
0	olx	prywatny	[20,25)	184
0	olx	prywatny	[25,30)	324
0	olx	prywatny	[30,35)	519
0	olx	prywatny	[35,40)	173
0	olx	prywatny	[40,50)	50

Źródło: Opracowanie własne.