



T-Mobile Polska S.A.
ul. Marynarska 12, 02-674 Warszawa

Metodologia Ekstrapolacji Wyników Analiz w T-Mobile Polska S.A.

1. Cel ekstrapolacji wyników

T-Mobile Polska S.A. dysponuje i analizuje jedynie dane dotyczące swoich własnych abonentów. Celem ekstrapolacji jest ustalenie ile osób w rzeczywistości przebywało w danym miejscu i czasie na podstawie ilości zarejestrowanych tam akcji kart SIM.

2. Określenie próby do badania

T-Mobile Polska S. A. rejestruje akcje, które zostały wykonane na danej karcie SIM. Przez akcję rozumiane są zdarzenia takie jak: wysyłanie i odbieranie wiadomości, przychodzące i wychodzące połączenia telefoniczne, korzystanie z danych komórkowych itp. Jeżeli w danym miejscu i czasie zostanie wykonana akcja to takie IMSI (ang. International Mobile Subscriber Identity), czyli unikatowy numer karty SIM jest traktowany jako obecny w analizowanym miejscu i czasie. T-Mobile Polska S.A. analizuje dane o IMSI, tylko tych abonentów, którzy wyrazili na to zgodę. Co więcej z powodu ochrony danych osobowych rekordy, dla których ilość IMSI nie przekracza 5 są odrzucane i nie są brane pod uwagę w analizie, gdyż w połączeniu z danymi CRM o abonentach pozwalałby na wskazanie konkretnej osoby. Liczbą szacującą naszą próbę jest ilość aktywnych numerów IMSI, dla których abonenci wyrazili zgodę na badania i analizy geolokalizacyjne.

3. Określenie wielkości populacji

Aby móc określić wielkość populacji postanowiono sprawdzić ilość kart SIM na rynku polskim. Po podzieleniu ilości kart SIM przez penetrację rynku kart SIM (wynoszącą 147% według Urzędu Komunikacji Elektronicznej) otrzymana została liczba tylko o ok. 1% większa od ilości Polaków podanej przez Główny Urząd Statystyczny (GUS). Postanowiono zatem oprzeć się na danych podanych przez GUS. Ilość wszystkich Polaków według GUS wynosi 38 437 239 osób. Z uwagi na charakter analizowanych miejsc przez T-Mobile Polska S.A. (tj. imprezy masowe, przejazdy na danym odcinku,

przejazdy przez dany punkt) postanowiono odrzucić dwie grupy wiekowe, uznane za "niemobilne" czyli osoby wieku 0 – 6 lat oraz osoby 85+. Po odjęciu tych grup nasza populacja została oszacowana na **34 933 522** osób.

4. Współczynnik ekstrapolacyjny

Dzięki podzieleniu wielkości populacji na wielkość próby otrzymaliśmy współczynnik ekstrapolacyjny. W ten sposób wiemy ilu osobom w rzeczywistości odpowiada jedno IMSI, które wykonało akcję w określonym miejscu i czasie.

$$\text{współczynnik ekstrapolacyjny} = \frac{\text{wielkość populacji}}{\text{wielkość próby}}$$

5. Metoda ekstrapolacji

Suma numerów IMSI, które wykonują akcje jest mnożona przez współczynnik ekstrapolacyjny. Ten iloczyn jest dzielony przez wielkość populacji. W taki sposób otrzymujemy estymowaną frakcję, populacji, która w danym momencie i czasie znajduje się w analizowanym obszarze.

$$\text{frakcja} = \frac{\text{ilość IMSI} \cdot \text{współczynnik ekstrapolacyjny}}{\text{wielkość populacji}}$$

Dla tak określonej frakcji określony zostaje przedział ufności, czyli zakres dla którego jesteśmy w stanie powiedzieć, że z 90% prawdopodobieństwem frakcja populacji zawiera się między dolną a górną granicą przedziału ufności. Do estymowania przedziału ufności zastosowano metodę Wilson – score zgodnie z poniższym wzorem:

$$\frac{1}{1 + \frac{1}{n} z^2} \left[\hat{p} + \frac{1}{2n} z^2 \pm z \sqrt{\frac{1}{n} \hat{p}(1 - \hat{p}) + \frac{1}{4n^2} z^2} \right]$$

gdzie:

z – 95. percentyl wystandaryzowanego układu normalnego

n – wielkość populacji

p – frakcja (obliczona zgodnie ze wzorem powyżej)

Np. Jeżeli frakcja wyniesie 5,364%, to używając metody Wilson-score możemy obliczyć,

że z 90% prawdopodobieństwem realna wielkość tej frakcji mieści się między 4,965%



a 5,985%. Po przemnożeniu tych wyników przez wielkość populacji otrzymamy przedział dla faktycznej ilości osób znajdujących się w danym miejscu i czasie.