**Autorzy:**

Roni Chikhmous 69684

Andrzej Drzystek 53667

Damian Gwóźdź 53320

Semestr zimowy 2015/2016

model zarażenia wirusem hiv w populacji

Raport z projektu zaliczeniowego z przedmiotu Modelowanie wieloagentowe [234900]

Raport dotyczy modelu zarażania się wirusem HIV w danej populacji. W symulacji zakłada się, że istnieje populacja osób, wśród których znajdują się agenci zdrowi, zarażeni wirusem HIV i tego nieświadomi oraz tacy, którzy są zarażeni oraz o tym wiedzę. Agenci mogą wiązać się w pary. Zakłada się, że wirus może być przenoszony wyłącznie drogą płciową. Co więcej, będąc w związku, agenci są sobie wierni. Aby uchronić się (i partnera/partnerkę) przed zarażeniem, agenci mogą stosować antykoncepcję. Przyjmuje się, że jej skuteczność wynosi 100%. Po zbadaniu się i uzyskaniu pozytywnego wyniku testu na obecność wirusa HIV, agent zawsze używa antykoncepcji (w szczególności również po ustaniu relacji z aktualnym partnerem/partnerką).

Omawiany problem został zasymulowany w czasie dyskretnym – w danym okresie (który może być utożsamiany z tygodniem) agenci:

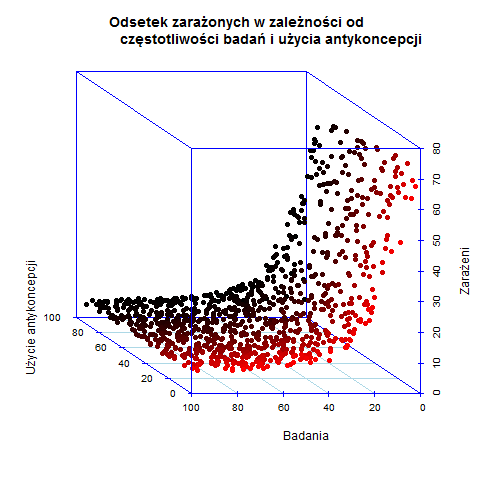
* współżyją ze sobą (zabezpieczając się bądź nie), co może skutkować zarażeniem się wirusa – dotyczy agentów będących aktualnie w związku;
* szukają partnera – jeśli nie są obecnie w związku; proces szukania pary przebiega dwuetapowo: agent z ustalonym prawdopodobieństwem spotyka potencjalnego partnera – jeśli to zdarzenie zakończy się sukcesem, to prawdopodobieństwo wejścia w związek danego agenta jest zestawiane z wartością losowaną z wybranego rozkładu;
* niezależnie od faktu bycia w związku mogą przebadać się na obecność wirusa; ponadto, po minięciu arbitralnie przyjętej liczby dwustu okresów pojawia się dodatkowa szansa przebadania się (odwzorowuje ona sytuację, w której po określonym czasie inkubacji choroby, osoba zaczyna odczuwać jej objawy, co może skłonić ją do wykonania testu).

Parametry, które przyjęto w modelu, to:

1. Początkowy odsetek agentów zarażonych: 5%.
2. Szansa na przekazanie wirusa partnerowi/partnerce: 50%.
3. W przypadku bycia samotnym agent ma 10% szansę na spotkanie potencjalnego partnera, natomiast prawdopodobieństwo wejścia w związek jest losowane z rozkładu normalnego N(0,2;0,1).
4. Długość trwania związku (w okresach) jest zmienną losową z rozkładu normalnego N(50;25).
5. Liczba agentów w modelu wynosiła 1000 osób.

Zarówno przeciętna szansa na używanie antykoncepcji, jak i częstotliwość badania się były losowane z rozkładu jednostajnego o przedziale <0;1> tysiąc razy. W każdej symulacji maksymalna liczba iteracji wynosiła trzy tysiące.

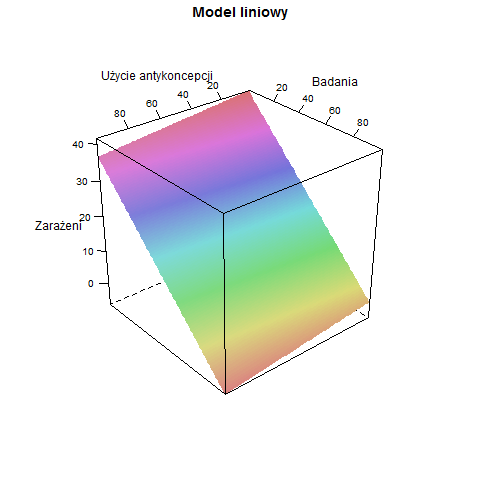
Z uśrednionych rezultatów wynika, że dla średnich wartości parametrów: użycia antykoncepcji i częstotliwości badania, średni odsetek zarażonych w populacji wyniósł 17% z odchyleniem standardowym na poziomie 2%. Uzyskane dane prezentuje wykres 1:



Wykres 1. Odsetek osób zarażonych wirusem HIV w zależności od częstotliwości badań i użycia antykoncepcji. Źródło: opracowanie własne.

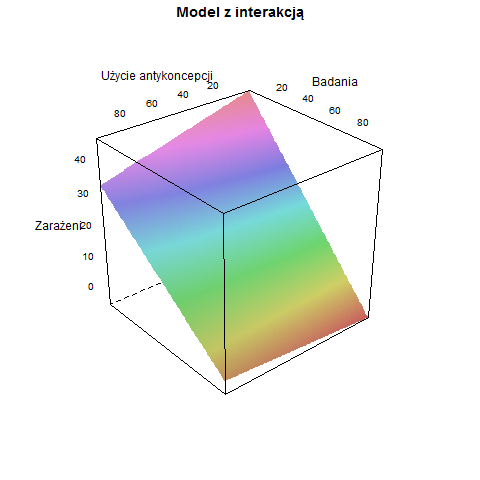
Jak można zauważyć, odsetek osób zarażonych wirusem dochodzi do 80% dla średniej wartości prawdopodobieństwa przebadania się wynoszącej 0. Dla wartości tego parametru z zakresu <0;0,2> obserwujemy gwałtowny spadek odsetku zarażonych do poziomu ok. 20%.

Po eksploracji uzyskanych danych podjęto próbę budowy metamodelu. Na początku wykorzystano w tym celu prosty model KMNK, za zmienne objaśniane przyjmując przeciętna szansa na używanie antykoncepcji, jak i częstotliwość badania. Wszystkie parametry (włącznie z wyrazem wolnym) były istotne statystycznie, skorygowany R2 wyniósł jednakże tylko 0.6. Na podstawie tego moelu uzyskano wykres 2, który zdecydowanie pokazuje rozbieżność pomiędzy wynikami symulacji a rezultatami uzyskanymi dla modelu liniowego:



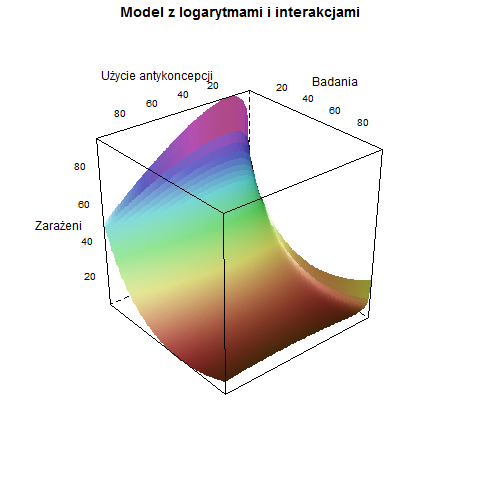
Wykres 2. Model liniowy (KMNK). Źródło: opracowanie własne.

W drugim modelu dodano interakcję między zmiennymi objaśnianymi. Ponownie, wszystkie zmienne okazały się być statystycznie istotne. Skorygowany R2 wzrósł do poziomu 0,607, niemniej jednak, model nadal nie odzwierciedlał relacji zarejestrowanych w danych z symulacji, co można zauważyć na wykresie 3:



Wykres 3. Model z interakcją. Źródło: opracowanie własne.

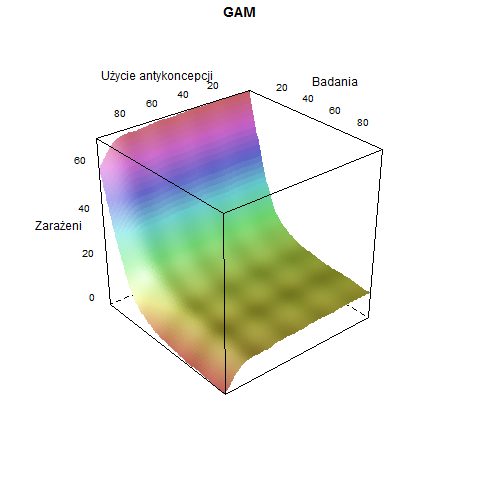
W następnym kroku do modelu dodano logarytmy zmiennych objaśnianych oraz interakcje między wszystkimi zmiennymi. Ponownie, wszystkie regresory okazały się być istotne statystycznie, a skorygowany R2 wyniósł tym razem aż 0,931. Otrzymane wyniki prezentuje wykres 4:



Wykres 4. Model z logarytmami i interakcjami. Źródło: opracowanie własne.

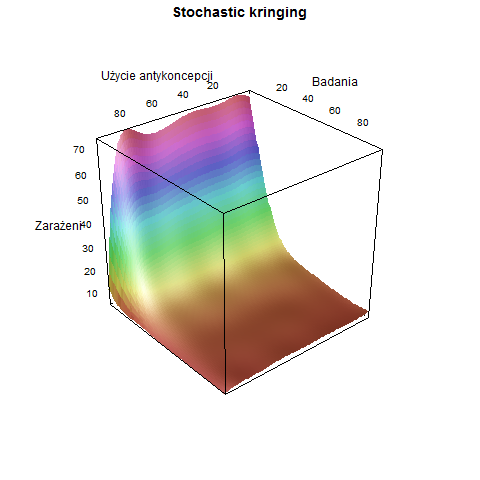
Jak można zauważyć, tym razem wyniki z modelu oraz symulacji są do siebie zdecydowanie bardziej zbliżone. Uwagę zwraca szczególnie fragment, w którym częstotliwość badań jest bardzo niska, częstotliwość stosowania koncepcji bardzo wysoka i dużo niższy niż w poprzednich modelach odsetek zarażonych (na poziomie 50-60% dla przeciętnej szansy na stosowanie antykoncepcji rzędu 80-100%).

Zbadano również analogiczny model dla kwadratów zmiennych zamiast logarytmów, lecz ze względu na bardzo podobne wyniki i ograniczoną objętość raportu nie jest on dokładnie opisywany. Do danych próbowano również dopasować model GAM. Obie zmienne były istotne statystycznie, skorygowany R2 kształtował się na poziomie 94%. Wyniki przedstawia wykres 5. Warto zwrócić uwagę na bardzo mały spadek liczby zarażonych wraz ze wzrostem częstotliwości stosowania antykoncepcji i niebadaniu się. Dopiero w przypadku, gdy antykoncepcja stosowana jest niemalże zawsze, odsetek osób zarażonych spada dość znaczenie (co jest także obserwowane dla różnych częstotliwości badania na obecność wirusa HIV).



Wykres 5. Model GAM. Źródło: opracowanie własne.

Na koniec wykorzystano metodę *stochastic kringing*. Wyniki przedstawia wykres 6:



Wykres 6. Model stochastic kringing. Źródło: opracowanie własne.

Jak można odnotować, dla tego modelu, podobnie jak w przypadku modelu GAM, w przypadku stosowania antykoncepcji w prawie każdym przypadku spada znacznie odsetek osób zarażonych. Ponadto, jeżeli przeciętna szansa na wykonanie badania wynosi ok 50% lub więcej, odsetek osób zarażonych wirusem HIV dla całej populacji wynosi nie więcej niż 10%. Jest to lekki spadek w porównaniu z modelem GAM.

Podsumowując, na podstawie przeprowadzonej symulacji (dla ustalonych parametrów) oraz analizy wyników można stwierdzić, że:

1. Mole nieliniowe odzwierciedlają dane w sposób dużo lepszy niż liniowe.
2. W przypadku szansy na wykonanie badania wynoszącej przynajmniej 50% odsetek osób zarażonych w populacji wynosi ok. 10% lub mniej.
3. W przypadku, gdy szansa na zbadanie się jest mniejsza niż 50%, ale większa niż 20%, następuje stopniowy wzrost odsetka zarażonych (do poziomu ok. 30-35%).
4. Gdy szansa na badanie wynosi mniej niż 20%, następuje gwałtowny wzrost liczby osób zarażonych, osiągający maksimum na poziomie ok. 80%.
5. W przypadku, gdy szansa na użycie antykoncepcji jest większa lub równa ok. 95%, odsetek zarażonych wirusem HIV spada dla każdego poziomu częstotliwości wykonywania badań, przy czym im częstotliwość ta jest większa, tym wyższy spadek odsetka osób zarażonych.