

Konferencja entuzjastów analizy danych z użyciem programu R

WZUR 3.0

Zapraszamy wszystkich entuzjastów analizy danych z użyciem programu statystycznego R na kolejne spotkanie.

Cel

Celem spotkania jest zaprezentowanie wybranych metod analizy danych medycznych z przykładami dla danych rzeczywistych, przygotowanych przez organizatorów. Najlepiej z przykładami z użyciem programu R.

Spotkanie będzie miało charakter kursu/warsztatów. Liczymy na uczestnictwo młodych osób chcących się czegoś nauczyć jak i osób z praktycznym doświadczeniem, tak statystyków jak i lekarzy czy epidemiologów. Podczas warsztatów przedstawione będą różne techniki analizy przykładowych danych (patrz sekcja Dane), zaprezentowane metody zostaną opisane w przeglądowej pracy, która zostanie wydrukowana w anglojęzycznym czasopiśmie.

Chcielibyśmy, by podczas tego spotkania (tego i jeżeli się powiedzie, kolejnych) uczestnicy pogłębiali swoją wiedzę dotyczącą analizy danych biologicznych, medycznych, epidemiologicznych i pokrewnych.

Dane

Na potrzeby konferencji została zebrana baza danych dotycząca zmienności zachorowalności na grypę w poszczególnych województwach. Baza została przygotowana w oparciu o meldunki epidemiologiczne z ostatnich 10 lat dotyczące zachorowań na grypę publikowane przez PZH.

Dane znaleźć można na stronie <http://biecek.pl/WZUR3/>.

Szerzej ten zbiór danych jest opisany na końcu tego komunikatu.

Konferencja

Konferencja złożona będzie z pięciu lub sześciu referatów. Każdy referat będzie trwał od 60 do 90 minut. Podczas każdego referatu przedstawiona będzie inna metoda analizy danych. Każdy referat składać się będzie z prezentacji podstaw matematycznych omawianej metody, informacji o dostępności danej metody w programie statystycznym R oraz z przykładowej analizy ww. zbioru danych.

Po zakończeniu konferencji planujemy napisać wspólną pracę porównującą metody przedstawione podczas konferencji. Zakres jak i sposób opublikowania tej pracy będzie ustalony przez autorów referatów. Najprawdopodobniej będzie to przeglądowa praca opublikowana w anglojęzycznym czasopiśmie.

Referaty

Poszukujemy pojedynczych osób jak i całych zespołów badawczych, które chciały by wygłosić referat podczas konferencji.

Dostępne dane mogą być analizowane na różne sposoby z użyciem rozmaitych metod, takich jak szeregi czasowe, analiza skupień, analiza przestrzenna, modelowanie wielowymiarowe, modelowanie danych sezonowych i wiele innych.

Szukamy entuzjastów. Tak studentów, pracowników naukowych czy pracowników prywatnych jednostek, którzy poświęcą trochę czasu i wezmą czynny udział w analizach. Prelegenci mogą liczyć na zwrot kosztów podróży.

Zeszłoroczne materiały pokonferencyjne zostały upublicznione na stronie <http://www.biecek.pl/WZUR/>. Relacja z tej konferencji pojawiła się w drugim numerze *The R Journal*, <http://journal.r-project.org/>.

Gdzie i kiedy

WZUR 3.0 planowany jest na 24-25 września. Odbędzie się we Wrocławiu w budynku Instytutu Matematycznego IMPAN, Kopernika 18, Wrocław.

Osoby chętne do wygłoszenia referatu proszone są o możliwie szybkie skontaktowanie się z organizatorami. Zgłoszenia referatów przyjmowane będą do 15 czerwca lub do wyczerpania miejsc.

Opłata

Nie jest planowana opłata za udział w konferencji. Liczba miejsc jest jednak ograniczona do 40. Pierwszeństwo przy rejestracji mają prelegenci. Zachęcamy do wcześniejszego zgłaszania się.

Serdecznie zapraszamy i zachęcamy do zgłaszania się.

Przemysław Biecek (MIM UW), Przemyslaw.Biecek@gmail.com

Marta Zalewska (WUM), Zalewska.Marta@gmail.com

Konrad Furmańczyk (SGGW), konfur@wp.pl

Stanisław Jaworski (SGGW), stanislaw.jaworski@ekonometria.info

Szczegółowy opis danych

Dane o zapadalności na grypę pochodzą z meldunków sprawozdawczych z poszczególnych województw (zgodnie z drukiem MZ-55: Dwutygodniowy, tygodniowy meldunek o zachorowaniach i podejrzeniach zachorowań na grypę) publikowanych na stronach internetowych. Można je znaleźć wchodząc w kolejne zakładki PZH -Państwowego Zakładu Higieny: - Meldunki epidemiologiczne- Grypa - Grypa i podejrzenia zachorowań na grypę – kolejne lata- raporty Rejestrowane są zachorowania grypopodobne lub ostre zakażenia dróg oddechowych ogółem i w grupach wiekowych – zgodnie z definicją grypy umieszczoną na stronie (grypa str 25):

http://www.pzh.gov.pl/oldpage/epimeld/inne/Def_PL2_Rob_1g.pdf

Zebrane dane dotyczą okresu 2003-2010 (aktualnie dostępne są tylko lata 2003, 2004, 2009, 2010, pełna baza pojawi się wkrótce)

Kolejne kolumny w bazie danych opisują:

1. województwo, nazwa województwa którego dotyczy dany raport,
2. liczba.ogol - liczba zachorowań w danym okresie czasu w danym województwie.

Objaśnienia znaków umownych: „Kreska (-) zjawisko nie wystąpiło ("zero"); Zero (0,0) zjawisko istniało, jednakże w ilościach mniejszych od liczb, które mogły być wyrażone uwidocznionymi w tablicy znakami cyfrowymi; Kropka (.) zupełny brak informacji albo brak informacji wiarygodnej; lks (x) wypełnienie pozycji, ze względu na układ tablicy, jest niemożliwe lub niecelowe ("nie dotyczy)"

3. zapadalność.ogol – frakcja zachorowań unormowana do wielkości województwa, unormowana do 100 tys mieszkańców

$$\text{zapadalność} = (\text{liczba zachorowań} / \text{narażona populacja}) * 100\ 000$$

4. wzrost/spadek.ogol – wzrost spadek zapadalności w stosunku do poprzedniego okresu, wartość dodatnia oznacza wzrost zapadalności

5. liczba0_4 Liczba zachorowań i zapadalność w grupie wiekowej 0-4 lata
6. zapadalność0_4
7. liczba5_14 Liczba zachorowań i zapadalność w grupie wiekowej 5-14 lat
8. zapadalność5_14
9. liczba15_64 Liczba zachorowań i zapadalność w grupie wiekowej 15-64 lat
10. zapadalność15_64
11. liczba65pow Liczba zachorowań i zapadalność w grupie wiekowej powyżej 65 lat
12. zapadalność65pow – dzień rozpoczęcia raportu.