

Analiza wielowymiarowa

Wprowadzenie

Dorota Celińska-Kopczyńska, Paweł Strawiński

Uniwersytet Warszawski

Zajęcia 1
8 października 2019

Plan zajęć I

- 1 Sprawy organizacyjne
- 2 Zasady zaliczenia
- 3 O czym będą te zajęcia?

Prowadzący

- dr Dorota Celińska-Kopczyńska
 - mail: dot@mimuw.edu.pl
 - strona: mimuw.edu.pl/~dot
 - dyżur: piątek 12:00, MIM, p. 1700, po umówieniu e-mailem
- dr hab. Paweł Strawiński
 - mail: pstrawinski@wne.uw.edu.pl
 - strona: coin.wne.uw.edu.pl/pstrawinski
 - dyżur: wtorek 17:15, WNE, p. 304, po umówieniu e-mailem

Kontakt mailowy

- W celu utrzymania porządku podczas kontaktu z Państwem prosimy o stosowanie formatu tytułu maili:
`AW_Nazwisko_skrót_sprawy`
- Niedopuszczalne jest rozpoczynanie zupełnie nowego tematu (np. wysłania propozycji tematu pracy zaliczeniowej) w wątku dotyczącym innej sprawy
- **Kontakt wyłącznie z wykorzystaniem adresów poczty uniwersyteckiej**

Forma zajęć

- Zajęcia w formie konwersatorium, 14 spotkań w semestrze zimowym
- Wymagana rejestracja w USOS
- Przyjście na inną grupę niż jest się zapisanym możliwe po wcześniejszym powiadomieniu mailowo prowadzących i uzyskaniu ich zgody (w zależności od liczby wolnych miejsc)
- Nie dopuszczamy do sytuacji, gdy liczba uczestników jest większa niż komputerów :)

Elementy zaliczenia

- 1 Obecność: obowiązkowa, dopuszczalne co najwyżej 3 nieobecności, powyżej NK
- 2 Raport z przeprowadzonego badania: 60%
- 3 30 minutowa prezentacja wygłoszona w trakcie ostatnich 3 zajęć: 15%
- 4 Prace domowe: 25%

Prace domowe

- 2 prace domowe (jedna techniczna, druga interpretacyjna) z ok. 3 tygodniowym terminem oddania
- Rozwiązania indywidualne – Państwo decydują, czy chcą rozwiązać
- Brak punktów z prac domowych nie uniemożliwia zaliczenia przedmiotu, ale nie można też uzyskać oceny końcowej wyższej niż 4 (db)
- Prace domowe stanowią element Państwa aktywności – nie poprawiamy ich w drugim terminie

Raporty i prezentacje

- Wykonanie badania w grupach. Preferujemy grupy 3 osobowe. Prosimy unikać grup składających się wyłącznie z osób z II roku studiów II stopnia
- W każdej grupie powinno powstać co najmniej 7 zespołów, ale co najwyżej 9
- Można dobrać się w zespół z osobami z grupy z innej godziny
- W razie problemów w zespole, prosimy nas informować, postaramy się znaleźć rozwiązanie

Raporty i prezentacje – tematyka

- Tematy do wyboru z poniżej listy, zgłaszane mailowo do obojga prowadzących:
 - 1 analiza korelacji i analiza ANOVA
 - 2 analiza korespondencji i analiza korelacji
 - 3 analiza kanoniczna
 - 4 analiza dyskryminacji
 - 5 metody segmentacji
 - 6 analiza czynnikowa, inne techniki redukcji wymiaru (★)
 - 7 analiza conjoint

UWAGA! W każdej grupie wszystkie tematy muszą zostać wyczerpane, o przydzieleniu tematu decyduje kolejność zgłoszeń.

Zgłoszenia tematów

- Zgłoszenie powinno zawierać: wybraną technikę badawczą, hipotezy/cel, opis źródła/bazy danych, propozycję literatury
- Zgłoszenie może być:
 - 1 **zakceptowane** (odpowiedź “akceptuję temat” od prowadzącego)
 - 2 **negocjowane** (prowadzący zwracają się o uszczegółowienie lub modyfikację)
 - 3 **odrzucone** (prowadzący podaje merytoryczne/formalne powody, dla których podany temat nie może zostać zrealizowany)

UWAGA! Zgłoszenia znacząco niepełne nie otrzymują statusu negocjowanych (nie ma np. rezerwacji techniki)

Wymagania dotyczące raportu

- Wyraźnie sformułowana hipoteza badawcza lub cel
- Krótki wstęp teoretyczny wraz z odniesieniami do istotnej literatury
- Literatura: co najmniej 3 artykuły angielskojęzyczne
- Opis i źródło użytych danych
- Sformułowany model
- Wyniki przeprowadzonej empirycznej analizy wraz z komentarzem, wynikającymi z nich wnioskami oraz odniesieniem do literatury

Czego nie proponować

- Badanie ma być ćwiczeniem z wykorzystaniem “realnych” danych, tak jak w życiu zawodowym
- Ważne jest jak Państwo sobie radzą i reagują na pojawiające się problemy, a nie spełnienie wszystkich założeń techniki
- (Realność danych) Prosimy nie korzystać ze zbiorów danych pochodzących z podręczników, tutoriali, stron wykładowców
- (Antyplagiat) Prosimy nie korzystać z Diagnozy Społecznej (badanie przerwane) ani powtarzać analiz przeprowadzanych podczas zajęć

Wymagania dotyczące raportu – strona techniczna

- Raport w formacie pdf
- Raport powinien być wyczerpujący – forma artykułu, a nie logu z programu
- Oprogramowanie dowolne (ostrożnie z SaaS) – ważna jest poprawność uzyskanych wyników i ich opis
- Prosimy również zadbać o schludność raportu (brak literówek, poprawność odwołań do literatury, estetykę tabel i opisów...)
- 30 000 znaków (ok. 18 stron bez obrazków), przekroczenie limitu należy zgłosić i uzasadnić wykładowcom
- Nie czytamy prac anonimowych i bez ponumerowanych stron

Prezentacje

- Prezentacje mają za zadanie zasugerować Państwu, co należałoby poprawić przed oddaniem końcowego raportu
- **Nie oceniamy zawartości merytorycznej prezentacji**, jedynie technikę prezentacji
- Pokazanie wyników badania zapewnia więcej komentarzy – warto je pokazać nawet, jeśli wiedzą Państwo, że są nie do końca prawidłowe lub błędne
- Od Państwa zależy podział zadań w grupie (**nie każda osoba musi prezentować**) – grupa oceniana jest jako całość
- Jeśli zasugerowane merytoryczne poprawki zostaną uwzględnione w końcowym raporcie wcześniejsze błędy/uchybieńia nie mają znaczenia
- Prezentacji nie poprawiamy w drugim terminie (publiczność!)

Prezentacje – na co zwracać uwagę

- Prezentacja != odczyt
- Staramy się mówić do sali (rozumiemy, że np. przy wynikach wygodniej jest spojrzeć na rzutnik)
- Generalnie *less is more* na slajdach, ale umiar jest jeszcze lepszy
- Nie tracimy czasu na to, co publiczność wie – technikę omawiamy tylko, jeśli istotnie wykracza poza zakres omawiany podczas zajęć
- Prezentacja powinna trwać ok. 20 min (z tolerancją 4 min.)

Na co zwrócić uwagę przy opisie badania?

- Czy temat jest ważny?
- Czy temat jest osadzony w teorii ekonomii lub czy jest to badanie interdyscyplinarne (tło teoretyczne)?
- Czy hipotezy są weryfikowalne i poprawnie sformułowane?
- Opis bazy danych
- Szacowanie parametrów modelu
- Diagnostyka, określenie ogólności wyników, ograniczenia badania
- Weryfikacja hipotez (o ile są)
- Interpretacja wyników, wnioski, odniesienie do literatury

Terminarz

- 1 Zgłoszenie składu osobowego grup i tematów prac zaliczeniowych: **6 listopada 20:00 - 26 listopada 20:00**
- 2 Przesłanie wybranej bazy danych do obojga prowadzących:
6 listopada 20:00 - 30 listopada 20:00 (jeśli baza jest ogólnodostępna wystarczy link i wskazanie zmiennych)
- 3 Prezentacje w kolejności tematów zgodnej z omówieniem ich na zajęciach: **7, 14, 21 stycznia**
- 4 Przesłanie finalnej wersji raportu oraz dostarczenie papierowej wersji pracy na WNE: **nie później niż w 7 dniu po terminie prezentacji**

Kilka słów o analizie wielowymiarowej

- Analiza wielowymiarowa to zbiór metod i technik analizy danych zawierających informacje o wielu obiektach opisanych jednocześnie za pomocą wielu zmiennych
- Jej celem jest redukcja dużego zbioru danych, uproszczenie ich struktury oraz zapewnienie przejrzystej interpretacji wyników
- Analizę wielowymiarową można również zastosować do sortowania, grupowania, skupiania obiektów wykazujących podobne cechy
- Dodatkowo za jej pomocą można badać zależności pomiędzy zmiennymi, ich siłę powiązań oraz wyciągać wnioski
- Jest to również warsztat narzędzi przydatnych podczas analizy zmiennych jakościowych

Omawiane techniki analizy

- Powtórzenie statystyki oraz testy (nie)parametryczne
- Analiza korelacji
- Analiza wariancji i kowariancji (ANOVA)
- Analiza kanoniczna
- Analiza korespondencji
- Analiza dyskryminacji
- Metody grupowania i segmentacji
- Analiza czynnikowa, inne techniki redukcji wymiaru (★)
- Analiza conjoint