

Wojciech Niemirow
Uniwersytet Warszawski

Ankiety zawierające pytania drażliwe: model nieparametryczny, estymatory największej wiarygodności i algorytm EM

Chcemy zbadać, jaki jest procent studentów, którzy użyli kiedyś niedozwolonej substancji psychoaktywnej. Przeprowadzamy badanie ankietowe, ale nie zadamy pytania wprost, bo nie spodziewamy się szczerych odpowiedzi. Żeby respondenci czuli się bezpiecznie, dzielimy ich na dwie grupy. Osobom z **pierwszej** grupy zadajemy dwa pytania i prosimy o podanie **sumy** odpowiedzi:

Pytanie 1) *Ile przeciętnie godzin śpisz dziennie w dni robocze (wliczając w to połówki godziny)? Zapamiętaj swoją liczbę, ale jej nie ujawniaj.*

Pytanie 2) *Czy zdarzyło Ci się kiedyś użyć jakiejś niedozwolonej substancji psychoaktywnej (marihuana, kokaina, amfetamina czy inne). Jeśli tak — przypisz temu pytaniu 1, jeśli nie, przypisz temu pytaniu 0. Zapamiętaj swoją liczbę i nie ujawniaj jej.*

Do swojej liczby przypisanej do pytania 1 dodaj swoją liczbę przypisaną do pytania 2. Suma wynosi: ???

Osobom z **drugiej** grupy zadajemy te same dwa pytania, ale prosimy o **różnicę** odpowiedzi:

Od swojej liczby przypisanej do pytania 1 odejmij swoją liczbę przypisaną do pytania 2. Różnica wynosi: ???

Porównując średnie odpowiedzi w 2 grupach, możemy wyestymować interesującą nas wielkość, rozpoznać procent respondentów, którzy odpowiedzieli „tak” na pytanie 2). To zdroworozsądkowe podejście faktycznie sprowadza się do zastosowania metody momentów.

Okazuje się jednak, że zastosowanie metody największej wiarygodności w tej samej sytuacji prowadzi do estymatorów wyraźnie lepszych, niż metoda momentów. O tym opowiem w zapowiedzianym referacie.

Tradycyjnie, dzielono respondentów grupę „suma” i grupę „kontrolną”, w której zdawano tylko pytanie 1) (maskujące). Pomysł skonstruowania ankiety w podziale na grupy „suma” i „różnica” pochodzi od Barbary Kowalczyk.