Analiza Wielowymiarowa

Testy parametryczne i nieparametryczne

Maciej Nasiński, Paweł Strawiński

Uniwersytet Warszawski

Zajęcia 2
14 października 2021

Rodzaje skal pomiarowych

- Skala nominalna
- Skala porządkowa (rangowa)
- Skala przedziałowa
- Skala ilorazowa
- Skala absolutna

Skala Likerta

1. Nie zgadzam się
2. Raczej nie zgadzam się
3. Nie mam zdania
4. Raczej zgadzam się
5. Zgadzam się

Test t

- Najprostszym sposobem porównania średnich jest wykorzystanie testu opartego na statystyce o rozkładzie t-Studenta
- Niech zbiór $\mathbb{X}$ liczy n obserwacji, a zbiór $\mathbb{Y}$ m obserwacji
- Wówczas przy prawdziwej H_0 o równości średnich

$$t = \frac{\bar{X} - \bar{Y}}{\hat{var}} \sim t(n + m - 2)$$

- Gdzie $\hat{var}$ jest wariancją zmiennej w połączonych zbiorach
- Uwaga, gdy X lub Y nie ma rozkładu normalnego to rozkład statystyki testowej może różnić się od zakładanego

Test proporcji

- Test przeznaczony do weryfikowania hipotez o równości proporcji
- Niech zbiór $\mathbb{X}$ liczy n obserwacji, a zbiór $\mathbb{Y}$ m obserwacji
- Wówczas przy prawdziwej H_0 o równości proporcji

$$z = \frac{p_x - p_y}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n} + \frac{1}{m})}} \sim N(0,1)$$

- gdzie $\hat{p}$ jest udziałem sukcesów w połączonych zbiorach

Test F

- Najprostszym sposobem porównania wariancji jest wykorzystanie statystyki o rozkładzie F
- Niech zbiór $\mathbb{X}$ liczy n obserwacji, a zbiór $\mathbb{Y}$ m obserwacji
- Wówczas przy prawdziwej H_0 o równości wariancji

$$F = \frac{S_x^2}{S_y^2} \sim F(n-1, m-1)$$

- Ale rozkład tej statystyki jest czuły na spełnienie założenia o normalności rozkładu

Test Levena

- Levene (1960) zaproponował test równości wariancji odporny na brak normalności rozkładu analizowanej cechy
- Brown i Forsythe (1974) zaproponowali by w teście średnią zastąpić medianą która jest bardziej odporną miarą tendencji centralnej
- Ta poprawka jest istotna w przypadku skośnych rozkładów zmiennych
- Statystyka oparta jest o odchylenia wartości zmiennych od średnich w grupach

Statystyka testowa

- Niech X_{ij} będzie obserwacją j w grupie i
- Niech $Z_{ij} = \mid X_{ij} - \bar{X}_i \mid$, gdzie $\bar{X}_i$ jest średnią wartością zmiennej w grupie i.

$$W_0 = \frac{\frac{\sum_i n_i (Z_i - \bar{Z})^2}{(g-1)}}{\frac{\sum_i (Z_{ij} - \bar{Z}_i)^2}{\sum_i (n_i - 1)}}$$

- gdzie g to liczba grup, a n_i oznacza liczebność grupy i

Test Kruskala-Wallisa

- Test Kruskala-Wallisa jest uogólnieniem testu Manna-Whitneya na większą liczbę grup
- Test wykorzystuje rangowanie obserwacji
- Wzór statystyki testowej jest skomplikowany. Jeżeli nie występują obserwacje o identycznych rangach to niech n będzie liczbą obserwacji, n_j liczbą obserwacji z w zbiorze j, a R_j będzie sumą rang w j-tym zbiorze:

$$KW = \frac{12}{n(n+1)} \sum_{j=1}^{J} \frac{R_j^2}{n_j} - 3(n+1) \sim_a \chi^2(J-1)$$

- Test może być również traktowany jako niepraparametryczny odpowiednik jednoczynnikowej analizy wariancji

Test Kołmogorowa-Smirnowa

- Jest wykorzystywany do porównywania rozkładów jednowymiarowych cech statystycznych.
- Test ma dwie wersję:
  - dla jednej grupy, służy do weryfikacji hipotezy czy dana zmienna na określony rozkład. Ta wersja nazywana jest testem zgodności Kołmogorowa.
  - dla dwóch grup, służący do weryfikacji hipotezy czy rozkład zmiennej w dwóch grupach jest identyczny