

Dzień 1 - Regresja wielowymiarowa

Spis treści

Regresja wielowymiarowa

1

Regresja wielowymiarowa

Wersja pdf

Regresja - metoda statystyczna pozwalająca na badanie związku pomiędzy wielkościami danych. Celem regresji wielowymiarowej jest ilościowe ujęcie związków pomiędzy wieloma zmiennymi niezależnymi (objaśniającymi, czynnikami) a zmienną zależną (kryterialną, objaśnianą, odpowiedzią).

W ogólności badać współzależność zmiennych możemy na cztery różne sposoby:

- Zmienna zależna jednowymiarowa oraz jedna zmienna niezależna.
- Zmienna zależna jednowymiarowa oraz wiele zmiennych niezależnych.
- Zmienna zależna wielowymiarowa oraz jedna zmienna niezależna.
- Zmienna zależna wielowymiarowa oraz wiele zmiennych niezależnych.

Przykłady regresji wielowymiarowej:

- Wytrzymałość betonu zależy od składników użytych przy jego produkcji. Pytanie: W jakiej proporcji stosować te składniki, by wytrzymałość była największa?
- Cena mieszkania zależy od... Pytanie: jak udział poszczególnych elementów wpływa na to, aby cena rynkowa była najwyższa?
- Udzielenie kredytu zależy od Pytanie: jak udział poszczególnych elementów wpływa na decyzję o przyznaniu lub nie kredytu? *czy to na pewno regresja?*

W ujęcie "naukowym", badania statystyczne mają w ogólności wyjaśniać zależności pomiędzy różnymi cechami badanej populacji.

Cele badań w analizie regresji:

- Scharakteryzowanie relacji (między innymi jej zasięgu, kierunku i siły).
- Określenie modelu matematycznego, który w najbardziej wiarygodny sposób oddaje zachowanie się odpowiedzi (innymi słowy, znalezienie odpowiedniej funkcji, która może być później wykorzystana do predykcji).
- Określenie, które ze zmiennych objaśniających są ważne w analizie współzależności i uszeregowanie tych zmiennych ze względu na siłę wpływu na zmienną objaśnianą.
- Porównywanie różnych modeli dla jednej zmiennej objaśnianej, tzn. porównanie modeli, które składają się z różnych zestawów zmiennych objaśniających.
- Określenie interakcji zmiennych objaśniających oraz (przy dwukierunkowej zależności) określenie zależności zmiennych objaśniających od zmiennej objaśnianej.
- Oszacowanie punktowe wartości współczynników regresji (kierunek i siła współzależności oraz istotność statystyczna parametrów wprowadzonych do modelu).

Uzyskiwane wyniki należy zweryfikować pod kątem następujących kryteriów:

- Określenie logicznego związku pomiędzy zmiennymi, tzn. sprawdzenie czy uzyskane wyniki nie kolidują z naturą zjawiska.
- Sprawdzenie czy przyczyna poprzedza w czasie skutek.
- Analiza siły związku pomiędzy zmiennymi np. wysoka korelacja między zmiennymi, które w rzeczywistości nie oddziałują na siebie.

- Sprawdzenie czy otrzymany model sprawdza się w rzeczywistości.
- Spójności wyników.
- Określenie zgodności wyników z wiedzą teoretyczną oraz doświadczalną.
- Rozpatrzenie możliwości otrzymania badanego skutku, jako przejawu działania różnych przyczyn oraz możliwości wystąpienia kilku skutków jednej przyczyny.

Najczęściej stosowane funkcje w analizie regresji:

- funkcja liniowa $f(x) = ax + b$,
- funkcja wielomianowa, np. kwadratowa $f(x) = ax^2 + bx + c$,
- funkcja logarymiczna $f(x) = \ln x$,
- funkcja eksponencjalna $f(x) = e^{-x}$,
- funkcja logistyczna $f(x) = \frac{1}{1+e^{-x}}$.

Wybór metody współzależności wielu zmiennych:

Metoda analizy	Zmienna objaśniana	Zmienne objaśniające
Regresja wieloraka	ciągła	ciągłe (dopuszcza się także dyskretne)
Analiza wariancji	ciągła	jakościowe
Analiza kowariancji	ciągła	jakościowe (symboliczne) i ciągłe
Regresja Poissona	dyskretna	różne typy
Regresja logistyczna	dwuwartościowa	różne typy