

Wydajne wnioskowanie statystyczne dla danych funkcjonalnych w złożonych układach czynnikowych

Analiza danych funkcjonalnych (FDA, ang. *functional data analysis*) stała się znaczącym narzędziem w statystyce, szczególnie skutecznym, gdy dane są gęsto próbkowane. Analizuje ona dane składające się z lub mogące być reprezentowane przez funkcje, krzywe lub powierzchnie. Rozwój bardzo precyzyjnych instrumentów pomiarowych pozwala na obserwowanie takich danych w wielu dziedzinach nauki jak biologia, chemometria, ekonomia, medycyna czy meteorologia. Przykładowo, gdy pacjenci noszą urządzenie, które automatycznie mierzy ciśnienie krwi i tętno w regularnych odstępach czasu przez 24 godziny, jako wynik otrzymujemy trajektorie wartości tych zmiennych, które mogą być modelowane jako dane funkcjonalne. W przypadku zwiększenia częstotliwości obserwacji powtarzanych w czasie, analiza danych funkcjonalnych cechuje się znaczną przewagą nad tradycyjnymi metodami statystycznej analizy danych, ponieważ nie zakłada konkretnych struktur zależności ani nie wymaga równomiernie rozmieszczonych punktów czasowych, w których mierzone są cechy. Ta elastyczność może prowadzić do lepszych wniosków i umożliwia ekstrakcję dodatkowych informacji z danych.

Pomimo postępów metodologicznych nadal istnieje potrzeba konstrukcji wydajnych metod analizy danych funkcjonalnych, aby poradzić sobie z wyzwaniami, takimi jak brakujące dane, brak normalności rozkładu i heteroskedastyczność funkcji kowariancji. Projekt ma na celu opracowanie nowych metod statystycznych zarówno dla jednowymiarowej, jak i wielowymiarowej funkcjonalnej analizy wariancji lub kowariancji z powtarzanymi pomiarami, które nie są ograniczone założeniami tradycyjnych modeli. Metody te zostaną skonstruowane tak, aby zapewniać wiarygodne wyniki nawet w trudnych przypadkach takich jak małe rozmiary prób i częściowo lub całkowicie brakujące obserwacje. Będziemy również rozważać obecność czynników kowariancyjnych w różnych układach czynnikowych oraz weryfikować złożone hipotezy statystyczne obejmujące efekty główne i interakcyjne zarówno w modelach krzyżowych, jak i hierarchicznych, bez przyjmowania konkretnego rozkładu prawdopodobieństwa. Konstrukcja testów statystycznych w tych modelach opierać się będzie zarówno na tradycyjnych, jak i nowatorskich estymatorach, takich jak estymatory funkcji średniej i kowariancji z próby, a także oszacowania bazujące na falkach. Szczególną uwagę poświęcimy metodom imputacyjnym i statystycznym, które umożliwiają efektywne wnioskowanie w obecności braków danych. Takie braki mogą mieć różną charakterystykę jak i przyczyny. Pierwszym przykładem jest brak całej trajektorii, co może się pojawić w przypadku, gdy pacjent miał mierzone ciśnienie krwi i tętno przez cały dzień przed rozpoczęciem leczenia, ale po jego zakończeniu nie dokonał tego pomiaru ponownie. Drugi przykład skutkuje tak zwanymi częściowo obserwowanymi danymi funkcjonalnymi, gdy pacjent wyłącza urządzenie pomiarowe na kilka godzin, co powoduje powstanie braku pomiaru przez znaczną część dnia.

W projekcie przewidziano kompleksowe badanie teoretycznych własności skonstruowanych procedur statystycznych oraz ich skuteczności w praktyce. Planujemy zastosowanie intensywnych badań symulacyjnych w celu oceny ich własności skończenie-próbkowych. Wyniki będą integrowane z praktycznymi narzędziami statystycznymi w postaci wydajnego oprogramowania, które znajdzie zastosowanie w rzeczywistych problemach badawczych, zwłaszcza w medycynie. Przykładowe zastosowania obejmują analizę danych biomedycznych i biostatystycznych, gdzie braki danych są powszechne, a uwzględnienie zmiennych kowariancyjnych i struktury funkcjonalnej powtarzalnych pomiarów jest kluczowe dla poprawnego wnioskowania.

Współpraca w ramach projektu ma na celu nie tylko rozwój metod analizy danych funkcjonalnych, skupiając się na testach statystycznych i obszarach ufności, ale także przygotowanie wytycznych dotyczących stosowania tych metod do problemów biostatystycznych. To kompleksowe podejście ma na celu zniwelowanie luki między zaawansowaną teorią statystyczną a zastosowaniami w świecie rzeczywistym, co czyni je znaczącym krokiem naprzód w dziedzinie analizy danych funkcjonalnych.