摘要

在因子設計實驗中所使用的效應編碼，常是由分析者所主觀選定的，但其未必最能解釋因子對於反應變數的影響。為了解決此問題，Huang (2020) 提出一個方法，其在全因子設計下，藉由數據依序估計出解釋變異量最大的多個編碼函數，稱之為本質編碼。但因為越後面求得之本質編碼的解釋變異量越小，解釋能力越差，故 Huang (2020) 提出利用在反應變數的總變異量中，本質編碼之累積解釋變異量所佔的比例，來挑選出解釋變異量較大之本質編碼的方法。但是此方法需分析者主觀地選定一個累積解釋變異比例之閾值，而針對不同類型的資料，其所適合之閾值可能皆不相同，並且此方法所決定之本質編碼個數亦缺乏數學上嚴謹的定義。針對上述缺點，在本論文中，我們先對重要本質編碼，藉由本質效應給予明確的數學定義。並透過變異數分析和序列式假設檢定的想法，提出決定重要本質編碼個數的方法。此外我們亦將呈現，將變異數分析應用在本質編碼估計上時，其檢定統計量會產生的一些問題與性質。我們並利用模擬資料，觀察本文所提出決定重要本質編碼個數之方法的各種表現和性質。

關鍵字：效應編碼、本質效應，特徵分解、最大解釋變異準則、序列檢定

Abstract

Effect codings in factorial design are commonly determined by analysts, but such subjective choices may not fully capture the most significant impacts of factors on the response. To address this issue, Huang (2020) proposed a data-driven approach to estimate multiple coding functions for each factor, referred to as essence codings. These codings are formulated to sequentially maximize the explained variation in the response. As a consequence of the sequential nature in the estimation process, the later an essence coding is estimated, the smaller the variation it can explain, leading to a weaker ability in explaining the response. Huang (2020) suggested a method to retain essence codings with high-enough explained variations based on their cumulative explained variation. This method, however, requires the assignment of a threshold, the determination of which depends on the data context and might lack objectivity. Furthermore, there is no rigorous mathematical definition for the number of essence codings retained. To resolve these issues, we start by presenting a mathematical definition for important essence coding based on the relevant essence effects. Subsequently, we propose an ANOVA-based approach that utilizes a sequential testing procedure to objectively determine the number of important essence codings. We further explore the issues that arise from directly applying ANOVA to estimated essence codings, which are random vectors unlike the fixed ones in the typical usage of ANOVA. Through a data simulation study, we evaluate the performance and examine various properties of our method for determining the number of important essence codings.

Keywords: Effect coding; Essence effect; Eigen-decomposition; Maximum explained variation criterion; Sequential testing