

Profily úzkostných lidí: shluková analýza na základě fyziologických a behaviorálních ukazatelů¹

Teoretické zakotvení

Výzkum úzkostných poruch dlouhodobě ukazuje, že jedinci trpící úzkostí netvoří homogenní skupinu, ale liší se v profilech příznaků, spouštěčů i behaviorálních vzorců (Barlow, 2002). Identifikace přirozených podskupin v populaci může přispět k lepšímu porozumění etiologii úzkosti i k cílenější intervenci. V naší zprávě jsme se rozhodli prozkoumat skupinu úzkostných jedinců pomocí shlukové analýzy, která umožňuje identifikovat přirozené skupiny respondentů na základě podobnosti jejich profilů bez předem daných kategorií (Gentle et al., 1991). Vycházeli jsme přitom z předpokladu, který zdůrazňuje transdianogstické přístupy (Harvey et al., 2004) a to, že fyziologické ukazatele (např. srdeční frekvence a dechová frekvence) spolu s behaviorálními faktory (např. spánek, pohyb a příjem kofeinu) tvoří komplexní systém, jehož konfigurace se u různých jedinců liší. Cílem této zprávy je pomocí metody **k-means shlukové analýzy** identifikovat odlišné profily respondentů na základě vybraných fyziologických a behaviorálních proměnných a tyto profily psychologicky interpretovat.

Popis souboru

Výzkumný soubor tvořilo celkem **11 000 respondentů**. Soubor byl vyvážený z hlediska pohlaví: **33,9 % žen** (N = 3730), **33,2 % mužů** (N = 3657) a **32,9 % osob** identifikujících se jako jiné pohlaví (N = 3613). Průměrný věk respondentů byl **40,24 let** (SD = 13,24). Dataset byl získán z otevřeného zdroje Kaggle.com (Zhang, 2024) a je volně ke stažení. Je třeba poznamenat, že dataset pravděpodobně obsahuje **synteticky generovaná data**, což může vysvětlovat neobvykle vysokou vyváženost souboru a silné statistické efekty. Výsledky je proto třeba interpretovat primárně jako metodologickou demonstraci.

Metoda

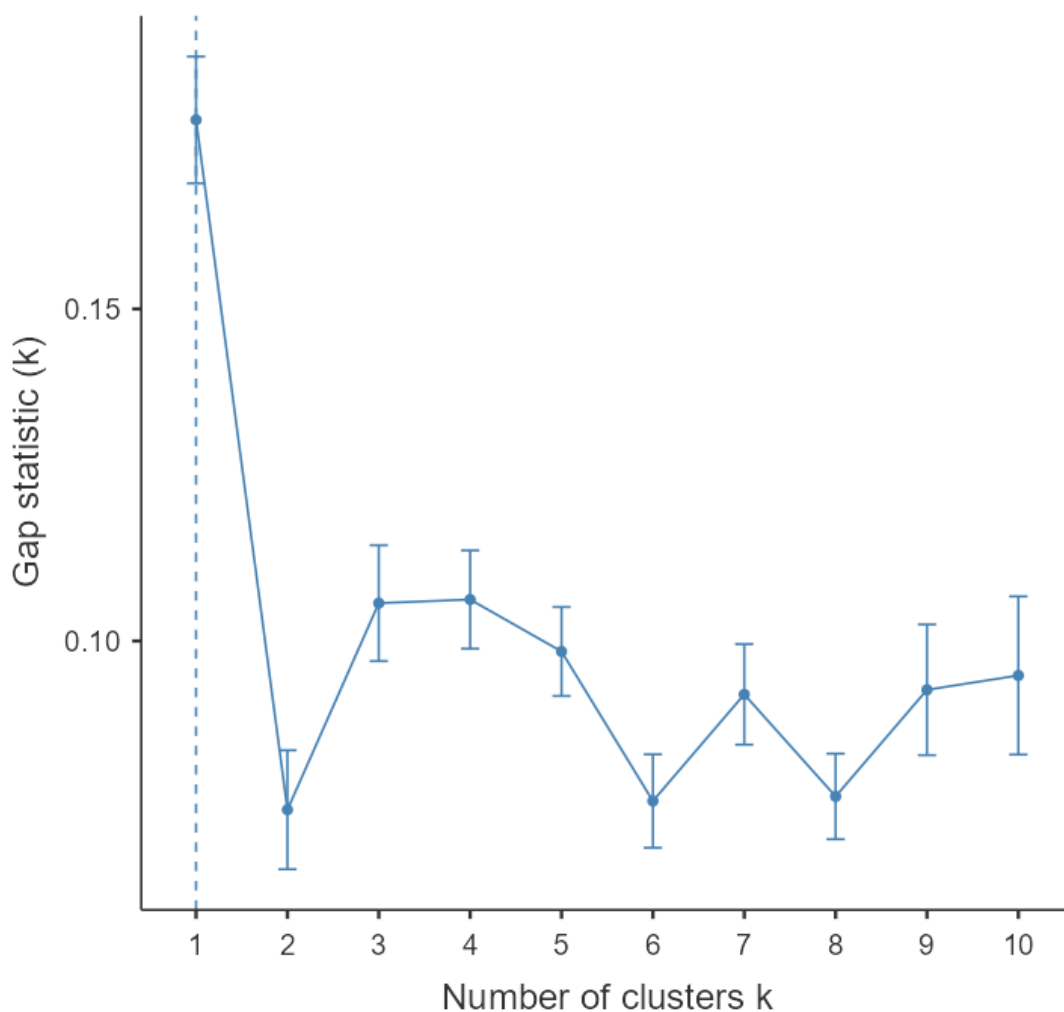
Pro identifikaci profilů respondentů byla zvolena metoda **k-means shlukové analýzy**. Vstupními proměnnými bylo sedm kontinuálních ukazatelů: počet hodin spánku, fyzická aktivita (hod/týden), příjem kofeinu (mg/den), míra stresu (škála 1–10), srdeční frekvence (bpm), dechová frekvence (dechy/min) a kvalita stravy (škála 1–10). Před analýzou byly všechny proměnné **standardizovány na z-skóry**, aby rozdílné měřítko proměnných

¹ Data a další informace o této zprávě jsou dostupné na adrese: <https://dostal.vyzkum-psychologie.cz/stat4/zprava.php?id=748>

neovlivnilo výsledky shlukování. Analýza byla provedena v softwaru **Jamovi** (modul snowCluster) s algoritmem Hartigan-Wong a 10 náhodnými startovními hodnotami.

Optimální počet shluků byl určen pomocí **Gap statistiky** (Tibshirani et al., 2001), která porovnává kvalitu shlukování s náhodným rozložením dat. Graf Gap statistiky (obrázek 1) ukázal nejvyšší hodnoty pro $K = 3$ a $K = 4$, přičemž rozdíl mezi nimi byl minimální. Pro další analýzu jsme zvolili **tří shluky**, jelikož se jedná o interpretačně úspornější řešení, které odpovídá didaktickým účelům této zprávy.

Obrázek 1: Graf Gap statistiky pro určení optimálního počtu shluků (popisky os v originálním jazyce softwaru)



Výsledky

Shluková analýza identifikovala tři odlišné skupiny respondentů. Složení a centroidy shluků zachycuje tabulka 1, profily proměnných ilustruje obrázek 2 a jejich následnou 2D vizualizaci nabízí obrázek 3.

Tabulka 1: Složení shluků a centroidy (z-skóry)

Proměnná	Shluk 1 (N = 2 060)	Shluk 2 (N = 4 392)	Shluk 3 (N = 4 548)
Spánek (hod)	-1,10	0,30	0,20
Fyzická aktivita	-0,67	0,21	0,10
Kofein (mg/den)	0,91	-0,27	-0,15
Stres (1–10)	0,80	-0,27	-0,10
Srdeční frekvence	0,61	-0,20	-0,09
Dechová frekvence	0,54	0,78	-1,00
Kvalita stravy	-0,72	0,24	0,09

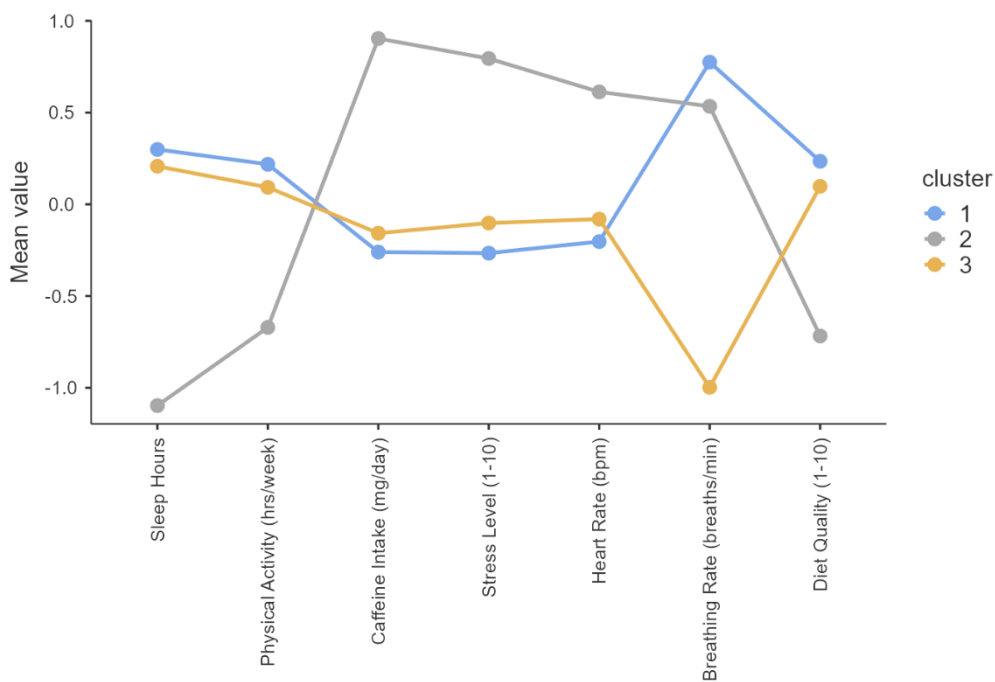
Poznámka: Hodnoty jsou standardizované centroidy shluků (z-skóry). Kladné hodnoty = nad průměrem souboru, záporné = pod průměrem.

Shluk 1: Riziková (N = 2 060; 18,7 %). Jedná se o nejmenší skupinu, vyznačující se kumulací negativních faktorů: výrazně nedostatečným spánkem ($z = -1,10$), nízkou fyzickou aktivitou, vysokým příjmem kofeinu, nejvyšším stresem a zvýšenou srdeční frekvencí, přičemž kvalita stravy je nejhorší ze všech skupin. Tito jedinci vykazují profil konzistentní s chronickým stresovým zatížením.

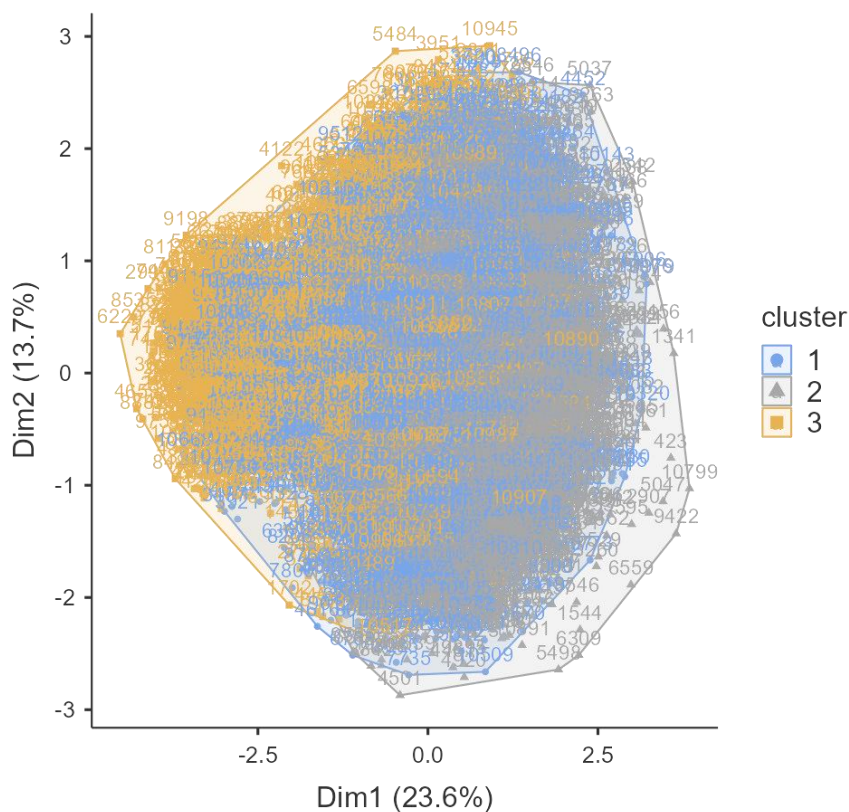
Shluk 2: Fyziologicky napjatí (N = 4 392; 39,9 %). Skupina se vyznačuje relativně průměrnými hodnotami spánku, aktivity i stravy a nízkým stresem. Nicméně vykazuje výrazně nejvyšší dechovou frekvenci ($z = 0,78$). Tento profil může naznačovat přítomnost chronického fyziologického napětí nebo úzkostné reaktivity navzdory tomu, že tito jedinci stres subjektivně nepocítují jako vysoký.

Shluk 3: Vyvážení (N = 4 548; 41,3 %). Jedná se o největší skupinu s průměrnými hodnotami většiny proměnných a výrazně nejnižší dechovou frekvencí ($z = -1,00$). Tito jedinci vykazují fyziologicky i behaviorálně nejklidnější profil.

Obrázek 2: Profily tří shluků na standardizovaných proměnných (popisky os v originálním jazyce softwaru)



Obrázek 3: Vizualizace shluků ve dvourozměrném prostoru po redukci pomocí PCA (popisky os v originálním jazyce softwaru)



Obrázek 3, který zobrazuje rozmístění respondentů ve dvourozměrném prostoru, vznikl redukcí sedmi původních proměnných pomocí **analýzy hlavních komponent (PCA)**. Dim1 (23,6 % vysvětlené variance) a Dim2 (13,7 % vysvětlené variance) dohromady zachycují přibližně 37 % celkové variability dat. Překryv shluků v grafu proto neodráží skutečnou kvalitu shlukování — ve vícerozměrném prostoru jsou shluky odděleny lépe, než jak je patrné z tohoto 2D zobrazení. Graf slouží jako doplňková ilustrace prostorového rozložení skupin.

Interpretace výsledků

Shluková analýza odhalila tři kvalitativně odlišné profily respondentů. Nejvýraznějším zjištěním je, že skupiny se od sebe liší především v **konfiguraci** proměnných, nikoli jen v jejich celkové výši. Shluk 1 kumuluje více rizikových faktorů současně (málo spánku, vysoký stres, hodně kofeinu, špatná strava), zatímco Shluk 2 vykazuje fyziologické napětí (vysoká dechová frekvence) přesto, že subjektivně vnímaná hladina stresu je nízká. Tento rozdíl je psychologicky zajímavý: naznačuje, že fyziologická reaktivita nemusí být nutně doprovázena vědomým prožitkem stresu. Toto zjištění je v souladu s koncepty alexithymie nebo disociace tělesného a psychického prožívání (Taylor et al., 1997).

Diskuse a závěr

Výsledky ukazují, že respondenti tvoří jednotnou skupinu, ale přirozeně se člení do odlišných profilů. Nalezené profily jsou v souladu s transdianogstickými modely úzkosti, které zdůrazňují různé cesty k rozvoji úzkostné symptomatiky (Harvey et al., 2004). Praktickým důsledkem by mohla být **individualizace intervencí** — např. jedinci ze Shluku 1 by mohli profitovat zejména ze spánkové hygieny a redukce kofeinu, zatímco jedinci ze Shluku 2 spíše z technik tělesného zklidnění (např. dechová cvičení).

Je třeba upozornit na několik limitů. Metoda k-means předpokládá kulovité shluky stejné velikosti, což nemusí odpovídat realitě. Počet shluků byl zvolen na základě Gap statistiky, avšak hranice mezi shluky jsou plynulé, jak dokládá překryv v cluster plotu (obrázek 3). Zásadním limitem je také již zmíněná pravděpodobná syntetická povaha dat, která omezuje zobecnitelnost závěrů na reálnou populaci. Budoucí výzkum by měl tyto analýzy replikovat na validovaných klinických datech.

Literatura

- Barlow, D. H. (2002). *Anxiety and its disorders: The nature and treatment of anxiety and panic, 2nd ed* (s. xvi, 704). The Guilford Press.
- Gentle, J. E., Kaufman, L., & Rousseuw, P. J. (1991). Finding Groups in Data: An Introduction to Cluster Analysis. *Biometrics*, 47(2), 788.
<https://doi.org/10.2307/2532178>
- Harvey, A., Watkins, E., Mansell, W., & Shafran, R. (2004). *Cognitive Behavioural Processes across Psychological Disorders: A transdiagnostic approach to research and treatment*. Oxford University Press.
<https://doi.org/10.1093/med:psych/9780198528883.001.0001>
- Taylor, G. J., Bagby, R. M., Parker, J. D. A., & Grotstein, J. (1997). *Disorders of Affect Regulation: Alexithymia in Medical and Psychiatric Illness* (1. vyd.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511526831>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(2), 411–423.
- Zhang, N. (2024). *Social Anxiety Dataset* [Data set]. Kaggle.
<https://www.kaggle.com/datasets/natezhang123/social-anxiety-dataset>