

Klasifikace skupin pacientů akutní psychiatrické péče dle vybraných kritérií

1. Úvod

Cílem této analýzy je pomocí shlukové analýzy identifikovat přirozeně se vyskytující skupiny pacientů akutní psychiatrické péče na základě jejich klinického profilu, definovaného kombinací psychiatrické diagnózy, počtu předchozích hospitalizací a délky aktuální hospitalizace. Tento přístup umožňuje odhalit typické vzorce chování pacientů při využívání akutní péče a přispět tak k hlubšímu porozumění jejich potřebám.

Délka hospitalizace v akutní psychiatrické péči bývá ovlivněna řadou faktorů, včetně diagnózy, komorbidit a předchozí anamnézy pacienta (Romney, 1984). Studie ukazují, že hospitalizační trajektorie pacientů s psychickými poruchami jsou velmi heterogenní a že klasifikační přístupy založené na datech mohou přispět k efektivnějšímu plánování péče (Everitt et al., 2011). S rostoucím tlakem na efektivní využití lůžkových kapacit je proto klasifikace pacientů významným nástrojem nejen pro predikci délky hospitalizace, ale i pro alokaci zdrojů a individualizaci terapeutických strategií.

Výsledky takové klasifikace mohou mít praktický přínos pro optimalizaci péče, zefektivnění systému a cílení intervencí u specifických podskupin pacientů. V kontextu narůstajících nároků na psychiatrické služby je důležité znát strukturu hospitalizované populace nejen z pohledu diagnóz, ale i z hlediska opakovaného kontaktu se systémem a délky péče. V literatuře se opakovaně ukazuje, že proaktivní a integrovaná psychiatrická péče může zkrátit délku hospitalizace (Oldham, Chahal, & Lee, 2019). Identifikace pacientů s vyšším rizikem dlouhých pobytů – jak to umožňuje shluková analýza – tak může být klíčovým prvním krokem k cílené intervenci.

Shluková analýza jako explorativní vícerozměrná metoda nevyžaduje předem definované skupiny. Umožňuje identifikovat v datech vnitřně homogenní, ale mezi sebou odlišné shluky pozorování na základě zvolených proměnných. Tím je ideálním nástrojem pro popis heterogenní populace pacientů, jakou představují osoby hospitalizované na akutních psychiatrických odděleních.

2. Shluková analýza

Pro klasifikaci pacientů do skupin využíváme přístup shlukové analýzy (cluster analysis), což je soubor metod sloužících k identifikaci přirozeně se vyskytujících skupin (shluků) v datech. Shluková analýza je explorativní technika, která nevyžaduje předem definované kategorie – skupiny jsou vytvořeny čistě na základě vnitřní podobnosti pozorování podle vybraných proměnných.

V této práci kombinujeme dvě komplementární metody:

- hierarchickou shlukovou analýzu (HCA),
- K-Means shlukování.

Hierarchická shluková analýza slouží zejména k exploraci struktury dat a k určení vhodného počtu shluků. Využíváme tzv. aglomerativní přístup, kdy každé pozorování začíná jako samostatný shluk a v každém kroku jsou slučovány dvojice nejbližších shluků. Jako míru vzdálenosti používáme Eukleidovskou vzdálenost, přičemž k výběru shluků ke sloučení používáme Wardovu metodu, která minimalizuje nárůst celkové vnitroshlukové variability.

Metoda K-Means je iterativní algoritmus, jehož cílem je rozdělit pozorování do K předem stanovených shluků tak, aby pozorování uvnitř každého shluku byla si co nejpodobnější a mezi shluky co nejodlišnější. Každý shluk je reprezentován svým centroidem (průměrný bod ve vícerozměrném prostoru). Data jsou přiřazována k tomu centroidu, ke kterému mají nejmenší Eukleidovskou vzdálenost. Centroidy se opakovaně přepočítávají, dokud se přiřazení pozorování neustálí.

Tato kombinace metod nám umožní jednak spolehlivě stanovit počet významných skupin, a zároveň zajistit konzistentní a replikovatelné přiřazení jednotlivých pacientů ke konkrétním shlukům.

3. Data a proměnné

Analýza vychází z původního datového souboru obsahujícího **35 608 záznamů pacientů** hospitalizovaných v rámci **akutní psychiatrické péče**. Dataset zahrnoval široké spektrum demografických, klinických a organizačních proměnných, mimo jiné:

- **Počet hospitalizací** – diskrétní číselná proměnná udávající celkový počet epizod hospitalizace daného pacienta;
- **Délka hospitalizace (v dnech)** – spojitá číselná proměnná;
- **Diagnóza dle MKN-10** – kategoriální proměnná (textová);
- **Pohlaví, věková kategorie, obor péče, kategorie péče (dospělý/dítě), způsob ukončení hospitalizace, kraj** – další popisné proměnné, které slouží pro kontext a případné porovnání vytvořených skupin.

4. Čištění a výběr dat

Pro účely shlukové analýzy byl dataset důsledně pročištěn následujícím způsobem:

- Ze vstupního souboru byly vybrány pouze záznamy s jednoznačně identifikovatelnou diagnózou, tj. takové, jejichž kód odpovídal standardnímu formátu MKN-10 (např. F20, F31).

- Byly vyloučeny složené, neúplné nebo nejednoznačné diagnózy, případně záznamy s neidentifikovatelným kódem.
- Diagnózy byly zredukovány na úroveň hlavních kategorií F0–F9, přičemž záznamy mimo tuto škálu byly rovněž odstraněny.

Po tomto kroku zůstalo 14 224 validních záznamů.

Vzhledem k výpočetní a časové náročnosti hierarchické shlukové analýzy byla pro samotnou analýzu dále náhodně vybrána podmnožina 500 záznamů, která byla nadále použita pro klasifikaci.

5. Transformace proměnných pro analýzu

Pro potřeby shlukové analýzy byly jako vstupní proměnné zvoleny:

- **počet hospitalizací**
- **délka hospitalizace**

Tyto proměnné byly vybrány jako kvantitativní ukazatele využití akutní psychiatrické péče, které mohou odhalit rozdíly ve způsobu využívání systému mezi jednotlivými typy pacientů.

Vzhledem k výrazně rozdílnému měřítku a rozsahu hodnot obou proměnných byla provedena logaritmická transformace ($\log(x+1)$) – s cílem zmírnit vliv extrémních hodnot, které jsou běžné zejména u délky hospitalizace.

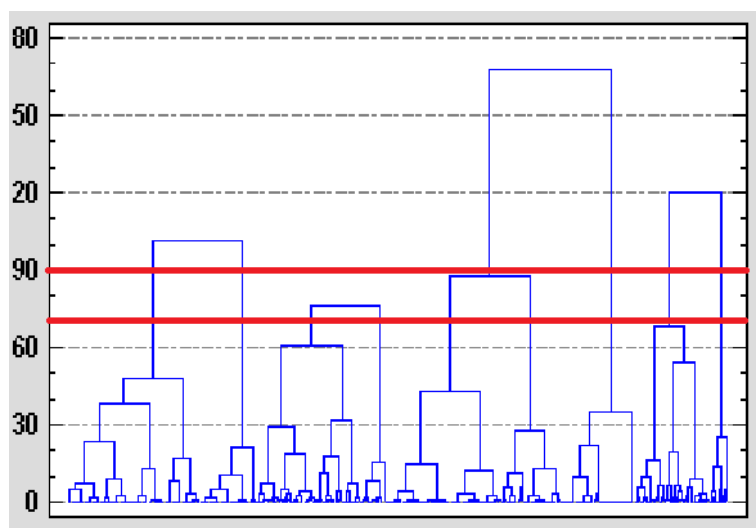
6. Hierarchická shluková analýza

Shluková analýza byla provedena ve dvou krocích. Nejprve byla aplikována hierarchická shluková analýza (HCA) s využitím Wardovy metody a Eukleidovské (čtvercové) vzdálenosti. Tento přístup slučuje pozorování do větších celků tak, aby v každém kroku minimalizoval nárůst vnitroshlukové variability. Vstupními proměnnými byly:

- logaritmicky transformovaný počet hospitalizací,
- logaritmicky transformovaná délka hospitalizace ve dnech.

Na základě výsledků hierarchické analýzy (zejména dendrogramu a změny vzdáleností) určíme vhodný počet shluků. Dendrogram (viz Obr. 1) zobrazuje, jak jsou jednotlivá pozorování postupně slučována na základě výpočtu vzdáleností mezi nimi. Na ose Y lze pozorovat míru (vzdálenost) slučování – čím vyšší hodnota, tím méně podobné jsou slučované skupiny. Červená horizontální čára znázorňuje námi zvolený řez dendrogramem, který rozdělil soubor do dobře interpretovatelných skupin pacientů. Podle výstupního dendrogramu a aglomerační vzdálenosti (scree plot) byl tak určen vhodný počet shluků. Nejvýraznější skok byl patrný mezi 4. a 5. slučovacím krokem, což indikovalo, že optimální počet skupin je čtyři.

Obr. 1: Dendrogram (Wardova metoda)



Hierarchický strom ukazující slučování pozorování. Červená čára vyznačuje řez, kterým byly identifikovány 4 klastry.

Následně byla provedena druhá fáze pomocí metody **K-Means**, kde byl počet klastrů fixován na čtyři. Tato metoda potvrdila stabilitu řešení a umožnila jednoznačné přiřazení pacientů do skupin. Na základě dendrogramu se jako možné jevily obě varianty – rozdělení do 4 nebo 5 klastrů. Nakonec bylo zvoleno čtyř shlukové řešení, které lépe odpovídalo výraznějšímu skoku ve vzdálenosti slučování a poskytlo interpretovatelné skupiny z hlediska klinické relevance.

Po rozdělení dat do čtyř shluků bylo dosaženo velmi dobrých výsledků z hlediska vnitřní konzistence i odlišení skupin:

- **Vysvětlená variabilita:** 83,2 %
- **Calinski-Harabasz index:** 816,3 (*vysoká hodnota potvrzuje dobré oddělení skupin*)

7. Výsledky – metoda K-Means

Po předchozí hierarchické shlukové analýze byla aplikována metoda **K-Means**, která umožňuje přiřazení každého pozorování k jednomu z předem stanovených shluků na základě minimalizace vzdálenosti k centroidu. Analýza proběhla na standardizovaných datech pro dvě proměnné jako v předchozí analýze.

Počet klastrů byl nastaven na **čtyři** ($K = 4$), v souladu s výsledky dendrogramu. Výsledky klasifikace potvrzují strukturu zjištěnou hierarchickou analýzou.

V následující tabulce (Tabulka 1) jsou uvedeny výsledky klasifikace, včetně počtu případů v jednotlivých klastrech, hodnot centroidů (logaritmické hodnoty) a jejich zpětný převod na přirozené jednotky. Je zřejmé, že jednotlivé klastry se liší jak ve frekvenci hospitalizací, tak v průměrné délce pobytu, což potvrzuje jejich praktickou významnost.

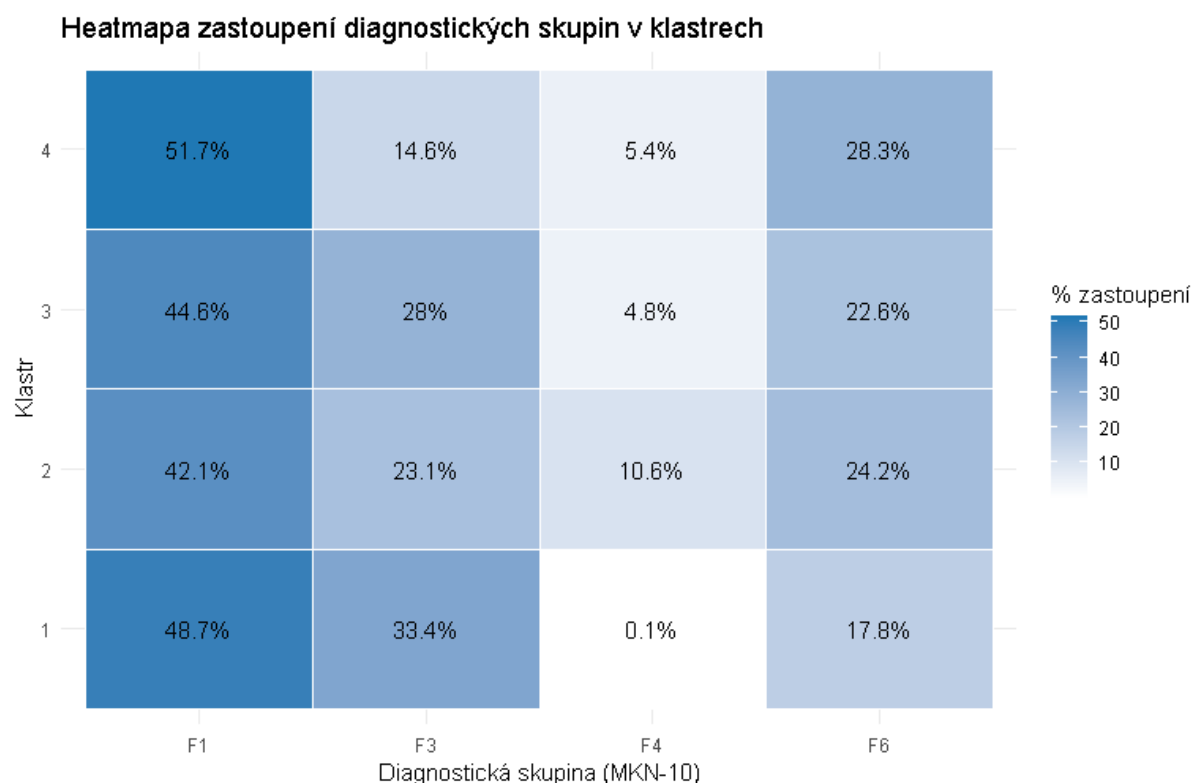
Tabulka 1: *Souhrn výsledků klasifikace (K-Means)*

Klastr	Počet případů	Podíl [%]	log_pocet_hospitalizaci	log_delka_hospitalizace	Průměrný počet hospitalizací	Průměrná délka hospitalizace (dny)
1	143	28,6 %	1,10	3,76	~2,0	~42
2	98	19,6 %	2,00	4,54	~6,4	~93
3	188	37,6 %	0,84	2,14	~1,3	~7
4	71	14,2 %	3,13	5,77	~21,9	~318

Rozdíly mezi klastry jsou výrazné – zatímco například Klastr 1 sdružuje pacienty s velmi častými a dlouhými hospitalizacemi (v průměru 24,6 hospitalizací a 350 dní), Klastr 4 reprezentuje pacienty s krátkou, jednorázovou hospitalizací (v průměru 1,2 hospitalizace a 5 dní).

Pro lepší znázornění rozložení psychiatrických diagnóz v rámci jednotlivých klastrů byla vytvořena také heatmapa (Tabulka 2), která ukazuje procentuální zastoupení jednotlivých diagnostických skupin F1–F6. Diagnóza F1 (poruchy způsobené užíváním psychoaktivních látek) dominuje ve všech klastrech, přičemž nejvyšší zastoupení má ve skupinách s vyšší frekvencí hospitalizací.

Tabulka 2: Diagnózy podle klasifikačních skupin (v %)



8. Závěrečné zhodnocení

Shluková analýza ukázala, že na základě dvou klíčových proměnných – počtu hospitalizací a délky pobytu – je možné identifikovat čtyři relativně stabilní skupiny pacientů akutní psychiatrické péče. Každý klastr byl charakterizován odlišnými vzorci využívání péče:

Klastr 1 reprezentoval skupinu pacientů s velmi vysokým počtem hospitalizací i délkou pobytu (průměrně 24 hospitalizací a 350 dní). Tato skupina vykazuje nejvyšší zátěž systému a zároveň nejvyšší zastoupení diagnóz F1 (poruchy způsobené užíváním návykových látek).

Klastr 2 byl smíšený, s mírným zastoupením více typů diagnóz a střední zátěží.

Klastr 3 se nacházel mezi těmito extrémy a představoval pacienty se střední délkou hospitalizace a vyšší frekvencí opakovaných pobytů, což může odpovídat chronickým a těžším formám onemocnění.

Klastr 4 zahrnoval pacienty s krátkou a ojedinělou hospitalizací (průměrně 1,2 hospitalizace, 5 dní), převážně s diagnózami F3 a F6, tedy afektivními poruchami a poruchami osobnosti.

Významným zjištěním je dominance diagnóz F1 (závislosti) napříč všemi klastry, zejména však v klastru s nejvyšší zátěží. To potvrzuje silnou souvislost mezi užíváním návykových látek a opakovanými hospitalizacemi, což by mělo být zohledněno při plánování péče a intervencí.

9. Diskuze a doporučení

Tato analýza ukázala, že i pomocí relativně jednoduchých kvantitativních ukazatelů (počet hospitalizací, délka pobytu) lze v populaci pacientů akutní psychiatrické péče odhalit strukturované vzorce chování. Shluky, které analýza identifikovala, nejsou náhodné – dobře odpovídají různým klinickým typům pacientů, jak z hlediska průběhu onemocnění, tak i typologie diagnóz.

Zvláštní pozornost si zaslouží skupina pacientů s vysokou četností hospitalizací a dlouhodobým pobytem – typicky s diagnózami F1 (závislosti) a F6 (poruchy osobnosti). Tito pacienti představují významnou zátěž pro systém, a právě jejich identifikace může být prvním krokem k cílené intervenci. Naopak skupiny s krátkým, jednorázovým pobytem (např. afektivní poruchy) vyžadují jiný typ péče.

Klinická praxe by mohla využít podobnou segmentaci pro cílené plánování péče, např. při nastavování péče po propuštění nebo v rámci case managementu. Diagnostická skupina F1 (poruchy způsobené užíváním psychoaktivních látek) si zaslouží zvláštní pozornost – její dominance napříč klastry potvrzuje nezastupitelnost mezioborové spolupráce mezi adiktologií a psychiatrií.

Analýza byla provedena pouze na náhodném vzorku 500 pacientů kvůli výpočetní náročnosti, diagnózy byly zjednodušeny do hlavních kategorií F0–F9, což může zakrýt rozdíly uvnitř jednotlivých skupin a vstupní proměnné byly omezeny na kvantitativní ukazatele bez zohlednění dalších klinických a sociodemografických faktorů. V budoucím výzkumu by bylo vhodné rozšířit klasifikaci o další proměnné (věk, způsob ukončení hospitalizace, komorbidita).

Použité zdroje:

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Národní katalog otevřených dat. (n.d.). *Zdravotnická statistika: Psychiatrická hospitalizace*. <https://data.gov.cz>

Oldham, M. A., Chahal, K., & Lee, H. B. (2019). A systematic review of proactive psychiatric consultation on hospital length of stay. *General Hospital Psychiatry*, 61, 1–9. <https://doi.org/10.1016/j.genhosppsych.2019.08.001>

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>

Romney, D. M. (1984). Factors influencing length of stay in psychiatric day care. *Canadian Journal of Psychiatry / Revue canadienne de psychiatrie*, 29(4), 312–314. <https://doi.org/10.1177/070674378402900406>

Statgraphics Technologies, Inc. (2023). *Statgraphics Centurion 19* [Software]. <https://www.statgraphics.com>

Světová zdravotnická organizace. (2019). *International statistical classification of diseases and related health problems: 10th revision (ICD-10)*. <https://icd.who.int/>