Skupiny států podle počtu medailí z olympijských her

Klastr analýza

Úvod

Od roku 1896 jsou novodobé olympijské hry společenským fenoménem, který zasahuje do života sportovců, architektů, politiků, služeb a mnoha dalších. Tato mimořádná událost je silně motivující nejen ke sportovním výkonům, ale i k vědeckým analýzám různých aspektů těchto her (Bačík, 2021). V následujícím textu si ukážeme pohled na počet medailí získanými sportovci jednotlivých zemí. Nebudeme provádět výčet všech zúčastněných států a jejich počet zlatých, stříbrných a bronzových medailí, ale budeme je dělit do skupin podle úspěšnosti. K tomu využijeme metodu zvanou klastr analýza, která nám toto umožní (Ambrožová, 2016).

Data

Data získaná z portálu Kaggle (2022) obsahují názvy všech států, které se kdy zúčastnily novodobých olympijských her, jejich počet účastí na letních i zimních hrách a počty medailí (viz data přiložená ke zprávě). Za dobu trvání se zúčastnilo 156 zemí světa, o nichž máme informace o počtu účastí, získaných zlatých, stříbrných i bronzových medailí zvlášť za LOH i ZOH i celkový počet medailí[1].

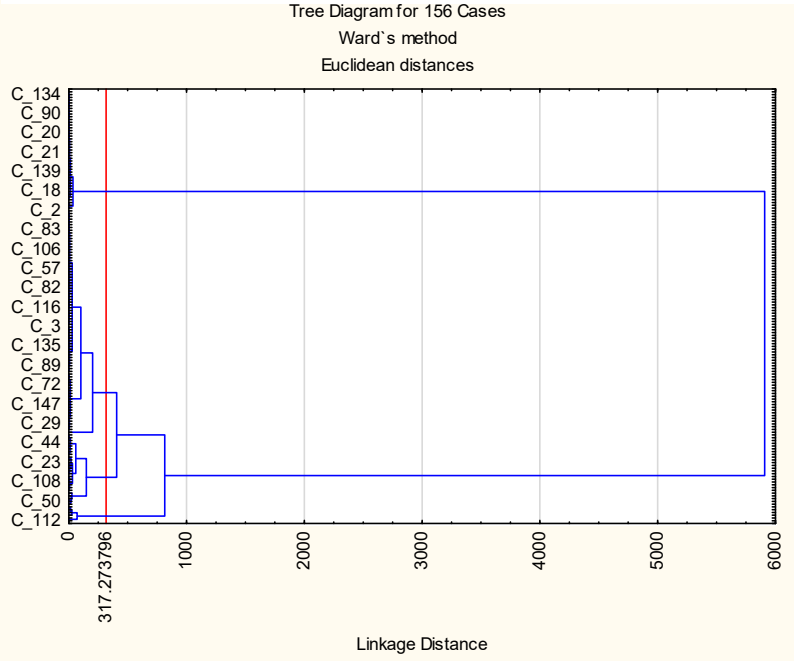
Cílem naší analýzy je vytvořit skupiny zemí podle výkonu, který podávají na OH a aby toto zařazení bylo co nejférovější. Jelikož je celkový počet medailí zkreslen počtem účastí, doplníme data o průměrný počet medailí získaných na jednu účast (jak pro zimní, tak pro letní OH), které dopočítáme ze zdrojových dat. Průměrováním vznikla v tabulce prázdná místa, jelikož docházelo k násobení nulou (v případě, že se stát zúčastnil pouze LOH a ne LOH). Protože program Statistika, ve kterém budou data analyzována, tyto prázdné buňky vynechává, byla data ještě dále upravena, a to funkcí *=iif(V19="", 0, V19+100)* pro zimní OH a obdobně i pro letní. Díky této úpravě budou oddělena data zúčastněných neúspěšných od nezúčastněných. Bohužel již nemáme data, která by nám sdělila informace o množství sportovců vyslaných za danou zemi.

[1] *Data a další informace o této zprávě jsou dostupné na adrese https://dostal.vyzkum-psychologie.cz/stat4?i=227*

Klastr analýza průměrného počtu získaných medailí na jednu účast

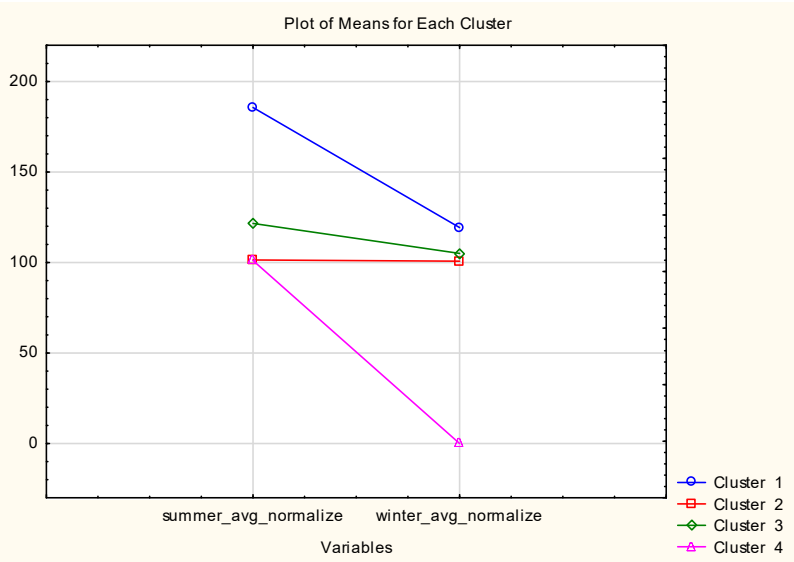
Jako první krok analýzy jsme zvolili klastr analýzu zvanou **spojování a hierarchické shlukování** a za pomoci Wardovy metody slučování jsme získali povědomí o možném počtu klastrů, s nimiž budeme dále pracovat (Obrázek 1).

Obrázek 1: Stromový graf s dělením na klastry



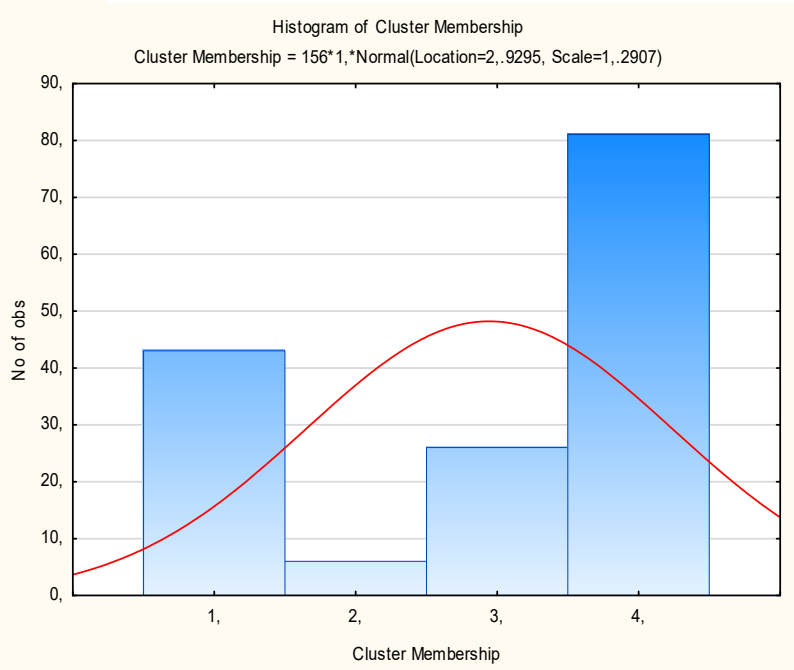
Na základě euklidovských vzdáleností viditelných na stromovém diagramu jsme stanovili 4 klastry jako vhodný počet pro rozdělení zemí do skupin dle úspěšnosti. Jelikož už známe počet klastrů, které chceme dále analyzovat můžeme přistoupit k **metodě K-průměrů**. Graf průměrů jednotlivých klastrů zobrazuje Obrázek 2.

Obrázek 2: Graf průměrů pro každý klastr



Z grafu lze vyčíst, že všechny státy, které se kdy zúčastnily se alespoň jednou zúčastnily v letních olympijských hrách. Díky rozdělení do čtyř klastrů je vidět, že existuje skupina, která je výrazně úspěšnější než ostatní. Také existuje skupina zemí, které si vedou dobře a získávají medaile, ale ne tak velký počet. Zároveň existuje skupina zemí, které se zúčastňují, ale nezískávají velký počet medailí, a nakonec poslední skupina, která se účastní pouze letních her, nikoli zimních. Velikosti těchto skupin jsou znázorněny v histogramu klastrů (Obrázek 3).

Obrázek 3: Histogram klastrů



1. klastr obsahuje všechny země, které se neúčastní zimních her, tam patří například Tanzanie, Tunis, Uganda, Kuba a jiní. Další klastry jsou seřazeny dle úspěšnosti, tedy od těch, co získali největší počet medailí (kam patří USA, Východní Německo, Rusko, Sovětský svaz, ROC a Unified Team) přes ty, kteří získali velký počet medailí, kam patříme my, Velká Británie, Austrálie, Rakousko a jiní až po ty, kteří získali malý počet medailí, jako jsou Egypt, Řecko, Dánsko a další. Více údajů lze vyčíst z excelovské tabulky přiložené k této zprávě (po seřazení dat dle klastru, nyní jsou řazeny dle abecedy), ale pro zajímavost zmíním, jak si vede Česká republika.

Česká republika, jako samostatný stát se aktivně zapojila zatím pouze 7x do letních a 8x do zimních olympijských her. Takto malý počet účastí je způsoben datem vzniku naší republiky. Úspěšnost naší země naleznete v Tabulce č 1. Pokud bychom chtěli získat celkový

počet medailí, museli bychom sečíst výsledky České republiky a Československa. Jestli bychom ale chtěli být přesný a započítávat pouze výkon českých sportovců, chybí nám potřebná data.

Tabulka 1: Výsledky českých sportovců

Cluster Membership	countries	s_participations	w_participations	total_gold	total_silver	total_bronze	total_total	summer_avg	winter_avg
3	Czech Republic	7	8	29	32	40	101	9,.57	4,.25
3	Czechoslovakia	16	16	51	57	60	168	8,.94	1,.56

Euklidovskou vzdálenost jednotlivých klastrů od sebe navzájem je možné vyčíst z Tabulky č. 2. Kvůli přidání čísla 100 ke všem státům, které se zúčastnily zimních olympijských her, je největší vzdálenost 103,3965, která je ale vytvořena uměle při úpravě dat. Bez úpravy, by to byla velikost 3,3965, tedy malá. Úspěšnost 1. a 4. klastru je tedy velmi podobná, ale liší se ne/účastí na ZOH.

Tabulka 2: Euklidovské vzdálenosti jednotlivých klastrů

Cluster Number	Euclidean Distances between Clusters (Spreadsheet15) Distances below diagonal Squared distances above diagonal			
	No. 1	No. 2	No. 3	No. 4
No. 1	0.0000	3719.920	2153.776	10690.84
No. 2	60.9912	0.000	212.663	5076.98
No. 3	46.4088	14.583	0.000	5726.08
No. 4	103.3965	71.253	75.671	0.00

Závěr

Úspěšnost států se v olympijských hrách velmi liší a je možné je díky využití klastr analýzy rozdělit do čtyř skupin. Tato data ale nejsou stoprocentně ideální pro hluboké prozkoumání úspěšnosti. Pokud bychom měli více proměnných, mohli bychom tento model rozšířit o další rozměry.

Předmětem dalšího zkoumání by v dalších pracích mohla být souvislost mezi úspěšností a počtem vyslaných sportovců danou zemí, nebo také jaký má vliv velikost státu. Pokud bychom předpokládali, že čím větší počet lidí v zemi, tím větší je pravděpodobnost úspěchu, byla by to také zajímavá informace a model by se dal takto velmi zajímavě rozšířit. Další zajímavou proměnnou by byla informace o počtu disciplín, ve kterých státy bojují o medaile. Bohužel to jsou informace, které naše data neobsahují a nelze je tedy uvést v této analýze.

Literatura

Ambrožová, K. (2016). *Shluková analýza: Cluster analysis*. https://is.muni.cz/el/1431/jaro2016/Z2069/um/54271982/56414665/Cv9_prezentace_pondeli.pdf

Bačík, V. (2021). *Olympic medalists of the modern summer Olympic games 1896–2016*. Journal of Maps, 17(1), 145-153. https://doi.org/10.1080/17445647.2021.1996475

Kaggle. (2022). *Countries Olympics Medals since 1896: Countries performance in both Olympics combined since the start*. Retrieved March 9, 2024, from https://www.kaggle.com/datasets/surajjha101/countries-olympics-medals-since-1896/data