

Mlejnek, Ondřej

Statistické zpracování získaných dat

In: Mlejnek, Ondřej. *Paleolit východních svahů Dražanské vrchoviny*. Měřínský, Zdeněk (editor); Klápště, Jan (editor). 1. vyd. Brno: Masarykova univerzita, 2015, pp. 135-139

ISBN 978-80-210-7818-5

Stable URL (handle): <https://hdl.handle.net/11222.digilib/133587>

Access Date: 16. 02. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

6. STATISTICKÉ ZPRACOVÁNÍ ZÍSKANÝCH DAT

Hlavním cílem statistického zpracování dat získaných při surovinové a technologicko-typologické analýze souborů štípané industrie ze studované oblasti bylo ověřit datování analyzovaných souborů do jednotlivých technokomplexů. Pro tyto účely byla sestavena tabulka, do které byly zařazeny všechny statisticky hodnotitelné soubory z oblasti (kritériem bylo více než čtyřicet nástrojů v souboru) a ve které byly uvedeny indexy, jež jsou rozhodující pro přiřazení štípané industrie k jednotlivým technokomplexům. Indexy představovaly zastoupení daného jevu v souboru v procentech, a proto mohly nabývat hodnot od nuly po sto procent. Konkrétně šlo o zastoupení vybraných druhů kamenných surovin mezi silicity (rohovec typu Stránská skála, rohovec typu Troubky-Zdislavice, silicity z glacienních sedimentů), o podíl levalloiských produktů v debitáži (hodnoceny celé kusy a proximální zlomky), podíl plošně retušovaných hrotů a bifasů mezi nástroji, podíl aurignackých typů škrabadel a rydel mezi nástroji, podíl škrabadel, rydel, retušovaných čepelí, odštěpovačů a drásadel mezi nástroji, podíl čepelí v debitáži a zastoupení čepelí mezi čepelovitými polotovary (čepelími a čepelkami). Z hlediska definovaného cíle se jako nejvhodnější statistická metoda zdála být analýza shluků (*cluster analysis*), která umožňuje analyzovat jednotlivé případy (soubory štípané industrie) charakterizované více znaky (proměnnými, indexy) na základě podobnosti a odlišnosti (Meloun *et al.* 2005, 268–339). Grafickým výstupem analýzy shluků je mimo jiné tzv. dendrogram, v němž jsou jednotlivé případy uspořádány do shluků podle podobnosti tak, že nejpodobnější případy jsou si nejbliž a nejdlišnější zase nejdál. Předpoklad byl takový, že v případě použití indexů charakterizujících jednotlivé soubory by se měly lokality v dendrogramu seskupovat do shluků podle příslušnosti k různým technokomplexům zastoupeným v oblasti. Podobný postup použil také Z. Weber při analýze moravských souborů štípané industrie z počátku mladého paleolitu datovaných do počátku mladého paleolitu v práci K. Valocha věnované lokalitě Vedrovice V (Valoch *et al.* 1993, 79–93).

Pro statistické zpracování dat byl použit program Statistica Cz 10 od firmy StatSoft. Statistica je komplexní systém obsahující prostředky pro správu dat, jejich analýzu, vizualizaci a vývoj uživatelských aplikací. Poskytuje široký výběr základních i pokročilých technik speciálně vyvinutých pro podnikání, vytěžování dat, vědu a inženýrské aplikace (viz <http://www.statsoft.com/Products/STATISTICA/Product-Index/>). Držitelem licence je Masarykova univerzita (<http://www.muni.cz/ics/services/software/statistica>). V programu Statistica je možné na spravovaná data aplikovat jak základní

statistické metody, tak i analýzy vícerozměrných dat, kam spadá i analýza shluků.

Zpočátku byla pomocí programu Statistica z analyzovaných indexů sestavena korelační matice (příloha 6). Korelační analýza studuje vzájemnou závislost všech dvojic proměnných. Míru závislosti určuje index, který může nabývat hodnot od -1 do $+1$. Hodnoty blízké se -1 znamenají silnou zápornou závislost (nízká hodnota jedné proměnné odpovídá ve většině případů vysoké hodnotě jiné proměnné), naopak hodnoty blízké se $+1$ znamenají silnou kladnou závislost (vysoká hodnota jedné proměnné odpovídá ve většině případů vysoké hodnotě jiné proměnné). Výsledky korelační analýzy je možné využít pro studium vzájemné závislosti jednotlivých proměnných a také pro určení proměnných, které je vhodné z dalších statistických analýz vynechat, protože výrazně korelují s jinou proměnnou a výsledky dalších statistických analýz by touto skutečností byly ovlivněny.

Při pohledu na výslednou korelační matici (příloha 6) lze za statisticky významné považovat zejména zápornou závislost mezi podílem SGS mezi silicity a podílem plošně retušovaných hrotů a bifasů mezi nástroji, kladnou závislost mezi podílem SGS mezi silicity a zastoupením rydel mezi nástroji, z toho vyplývající zápornou závislost mezi podílem plošně retušovaných hrotů a bifasů mezi nástroji a zastoupením rydel mezi nástroji a kladnou závislost mezi podílem rydel mezi nástroji a podílem aurignackých rydel mezi nástroji. Další závislosti jsou již statisticky méně významné. Je tedy zřejmé, že plošně retušované hroty a bifasy se vyskytují zejména v souborech s nižším podílem SGS v silicitové části industrie, a naopak podíl rydel je vyšší v souborech s vyšším zastoupením SGS. Podíl aurignackých rydel mezi nástroji je vyšší v souborech s vyšším zastoupením rydel obecně.

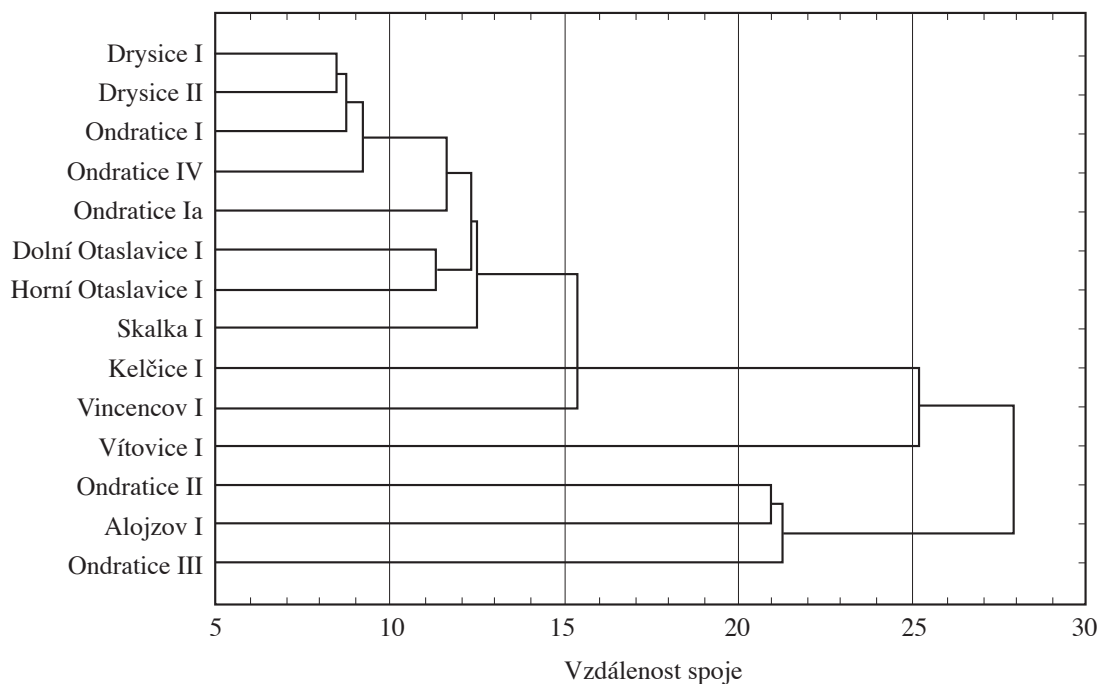
V další fázi bylo nutné vybrat proměnné (indexy), které by se nejlépe hodily pro analýzu shluků s cílem rozdělení jednotlivých případů (souborů štípané industrie) do shluků podle technokomplexů. Tyto technokomplexy jsou definovány zejména na základě technologicko-typologických charakteristik, z tohoto důvodu byly nakonec z analýzy shluků vypuštěny proměnné týkající se zastoupení jednotlivých surovin (indexy 2–4). V případě některých souborů byly využity údaje získané z literatury, kde ovšem nebylo možné některé indexy zjistit. Program Statistica dokáže pracovat i s neúplnou vstupní tabulkou. Chybějící údaje jsou v takovém případě nahrazeny průměrnými hodnotami, což však výsledek poněkud zkreslí. Proto byly ze shlukové analýzy vyloučeny také proměnné udávající podíl čepelí, čepelí a jader (indexy 13–15). Jako vhodné proměnné pro analýzu

Analyzované soubory	Sledované indexy							
	1	2	3	4	5	6	7	8
Dryšice I	6,73	7,88	3,62	27,37	2,83	3,54	0,24	13,68
Dryšice III	8,18	17,49	3,08	21,65	3,06	4,37	2,62	16,59
Ondratice I	8,01	14,98	12,24	26,41	5,44	6,67	1,11	17,22
Ondratice II	0,00	0,00	6,24	68,44	0,00	0,00	18,52	11,11
Ondratice IV	10,22	3,09	6,56	35,14	3,17	6,67	0,79	15,07
Ondratice VIII	1,67	0,46	6,43	72,28	0,00	3,06	3,93	9,61
Vítovice I	0,00	2,85	1,42	32,48	3,64	3,64	1,82	16,36
Dolní Otaslavice I	1,49	0,00	8,56	57,41	1,18	4,12	4,71	22,94
Horní Otaslavice I	2,03	0,22	12,69	40,65	1,44	6,47	0,72	16,55
Kelčice I					5,59	8,94	3,57	25,70
Vincencov I	0,00				11,04	3,07	1,84	12,88
Alojzov I	0,00					1,80	15,00	7,51
Skalka I	0,00	5,09	9,45	50,91	2,33	11,63	0,00	23,26
Ondratice Ia					9,29	1,33	0,00	15,49

Analyzované soubory	Sledované indexy							
	9	10	11	12	13	14	15	
Dryšice I	11,79	16,51	9,91	15,57	26,92	3,21	4,95	
Dryšice III	17,03	14,85	13,10	11,79	30,78	0,96	8,81	
Ondratice I	12,78	16,11	5,00	11,11	24,59	2,46	5,02	
Ondratice II	66,67	5,56	9,26	3,70	39,62	28,11	4,61	
Ondratice IV	11,11	14,29	10,32	23,02	48,98	0,00	12,57	
Ondratice VIII	50,22	14,85	6,99	6,11	40,75	6,04	3,74	
Vítovice I	5,45	25,45	30,91	9,09	26,98	4,71	3,11	
Dolní Otaslavice I	22,94	21,18	7,65	11,18	53,51	20,09	8,98	
Horní Otaslavice I	25,18	14,39	7,91	15,11	47,67	3,62	11,76	
Kelčice I	15,64	6,15	7,26	25,14	24,95			
Vincencov I	17,18	12,27	12,88	25,15	34,93			
Alojzov I	57,41			5,80				
Skalka I	25,58	6,98	11,63	18,60	34,91	13,83	2,15	
Ondratice Ia	3,98	19,47	6,64	15,93				

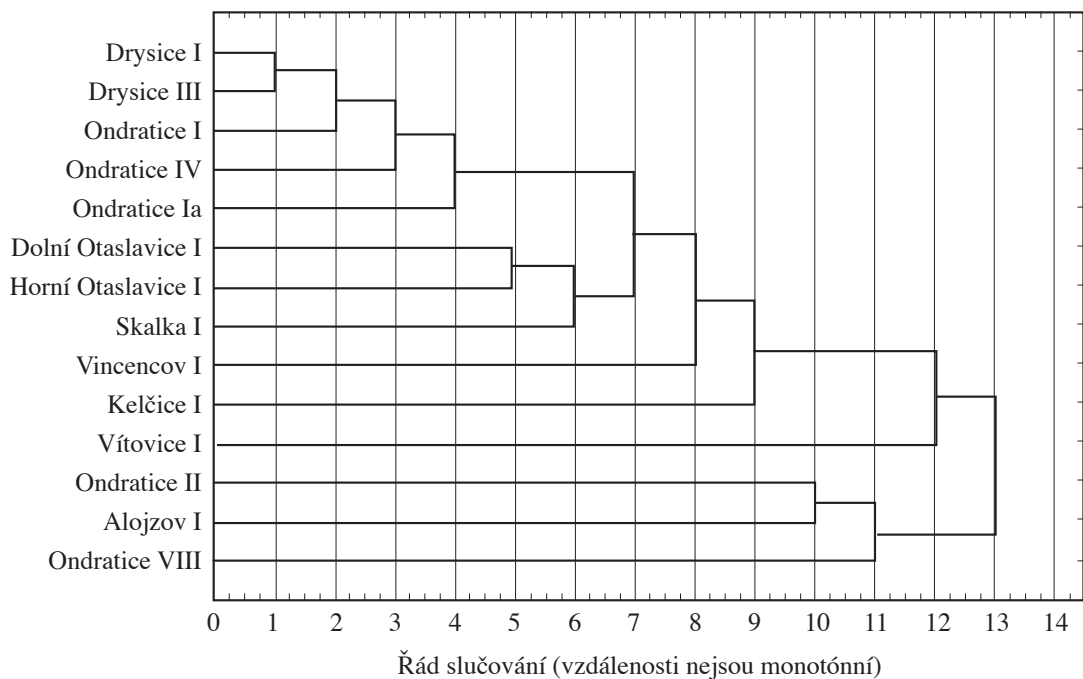
Tab. 6: Vstupní tabulka vybraných indexů použitá pro statistické analýzy. Indexy – 1: podíl levalloiských produktů v debitáži (zkoumány celé kusy a proximální zlomky), 2: podíl rohovce typu Stránská skála mezi silicity, 3: podíl rohovce typu Troubky-Zdislavice mezi silicity, 4: podíl silicity z glacienních sedimentů (SGS+SGS?) mezi silicity, 5: podíl plošně retušovaných hrotů a bifasů mezi nástroji, 6: podíl aurignackých typů škrabadel mezi nástroji, 7: podíl aurignackých typů rydel mezi nástroji, 8: podíl škrabadel mezi nástroji, 9: podíl rydel mezi nástroji, 10: podíl retušovaných čepelí mezi nástroji, 11: podíl odštěpovačů mezi nástroji, 12: podíl drásadel mezi nástroji, 13: podíl čepelí v debitáži, 14: podíl čepelí mezi čepelovitými polotovary (čepelími a čepelkami), 15: podíl jader. Uvedené indexy byly u lokalit Dryšice I, Dryšice III, Ondratice I, Ondratice II, Ondratice IV, Ondratice VIII, Vítovice I, Dolní Otaslavice I, Horní Otaslavice I a Skalka I vypočítány na základě analýz autora této práce, v případě ostatních lokalit na základě údajů uvedených v literatuře (Kelčice I: Svoboda – Přichystal 1990; Vincencov I: Svoboda – Přichystal 1987; Alojzov I: Oliva 1987, 112–113; Ondratice Ia: Oliva 2004). V případě prázdných polí v tabulce nebylo možné tyto údaje z literatury zjistit.

Str. diagram pro 14 případů
Jednoduché spojení
Euklid. vzdálenost



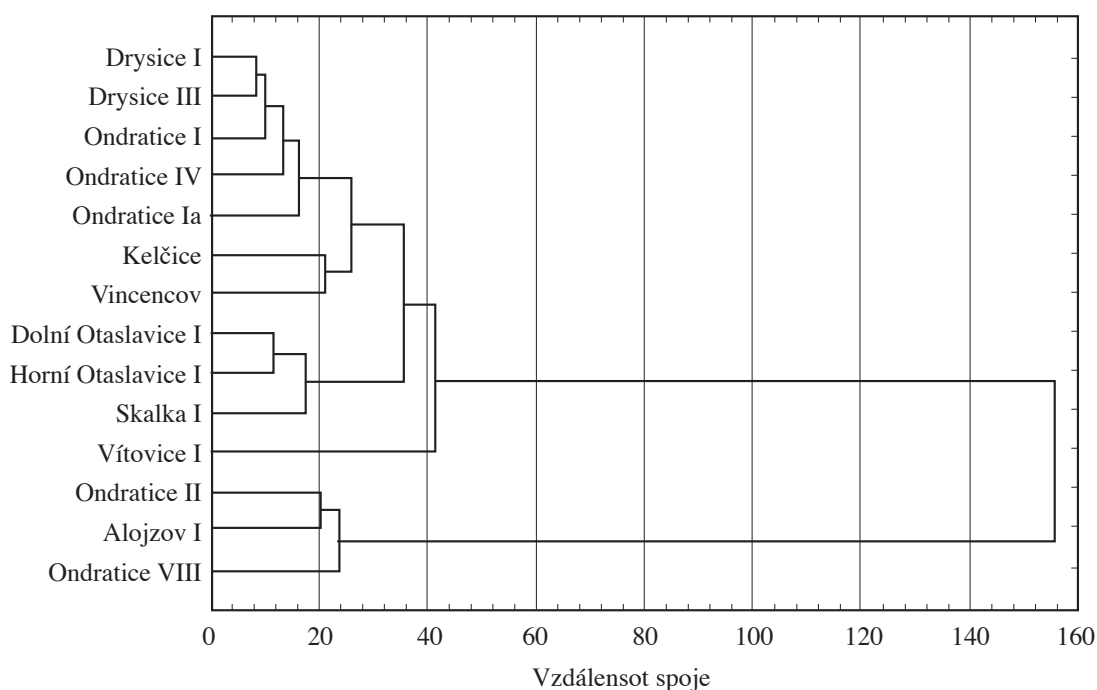
Obr. 125: Výsledky analýzy shluků aplikované na statisticky významné soubory ve studované oblasti. Jako proměnné byly použity indexy 1 a 5–12 z tabulky 6. Metoda spojování: jednoduché spojení.

Str. diagram pro 14 případů
Vážený centriod skupin dvojic (medián)
Euklid. vzdálenost



Obr. 126: Výsledky analýzy shluků aplikované na statisticky významné soubory ve studované oblasti. Jako proměnné byly použity indexy 1 a 5–12 z tabulky 6. Metoda spojování: metoda váženého centroidu skupin dvojic (mediánová metoda).

Str. diagram pro 14 případů
Wardova metoda
Euklid. vzdálenost



Obr. 127: Výsledky analýzy shluků aplikované na statisticky významné soubory ve studované oblasti. Jako proměnné byly použity indexy 1 a 5–12 z tabulky 6. Metoda spojování: Wardova metoda.

shluků byly tedy vybrány indexy 1 (podíl levalloiských produktů v debitáži), 5 (podíl listovitých hrotů a bifasů mezi nástroji), 6 (zastoupení aurignackých škrabadel), 7 (zastoupení aurignackých rydel), 8 (podíl škrabadel), 9 (podíl rydel), 10 (zastoupení retušovaných čepelí), 11 (podíl odštěpovačů) a 12 (zastoupení drásadel).

Při analýze shluků byla v programu Statistica zvolena možnost spojování (hierarchického shlukování), jehož výsledkem je dendrogram. Jako případy byly zvoleny soubory z jednotlivých lokalit a jako proměnné vybrané indexy (1, 5–12). Pro shlukování byly určeny případy (soubory z jednotlivých lokalit) na základě podobnosti a odlišnosti vybraných proměnných (indexů). Jako míra vzdálenosti byla zadána euklidovská vzdálenost, zvaná také geometrická metrika (viz Meloun *et al.* 2005, 269, obr. 9.1a), která představuje standardní typ vzdálenosti. Program Statistika umožňuje také přidání vah pro vybrané proměnné. V případě využití této funkce by hodnoty vybraných proměnných měly při analýze větší váhu. Této funkce nebylo využito, protože by šlo o subjektivní výběr, který by ovlivnil výsledky analýzy. Využita byla naopak funkce, která umožňuje nahradit chybějící proměnné průměrnými hodnotami. Jako pravidlo spojování byla použita metoda jednoduchého spojení, metoda váženého centroidu skupin dvojic (mediánová metoda) a Wardova metoda (ke způsobům shlukování viz Meloun *et al.* 2005, 275–277; Valoch *et al.* 1993, 80–81).

Jako forma znázornění výsledku analýzy byl vybrán horizontální graf hierarchického stromu.

Z porovnání grafů s výsledky shlukové analýzy pomocí tří různých metod shlukování vyplývá, že nejodlišnější shluk tvoří tři epiaurignacké soubory z lokalit Ondratice II, Ondratice VIII a Alojzov I, které jsou si vzájemně velmi podobné. Zbývající lokality tvoří druhou skupinu, ve které je možné rozlišit několik podskupin. Výraznou podskupinu tvoří ve všech použitých metodách spojování shluk pěti podobných souborů z lokalit Ondratice I/Želeč, Ondratice Ia, Ondratice IV, Drysice I a Drysice III, které je možné zařadit do tzv. ondratického typu (Mlejnek *et al.* 2012, 311) datovatelného na počátek mladého paleolitu. O další podskupině je možné mluvit v případě aurignackých lokalit na Prostějovsku (Dolní Otaslavice I, Horní Otaslavice I). Samostatné skupiny potom tvoří aurignacké soubory ze Skalky u Prostějova (ale ten je při použití mediánové a Wardovy metody spojování zařazen do shluku s otaslavickými lokalitami), z Kelčic I a z Vítovic I a také szeletský soubor z Vincencova I. Umístění těchto souborů je při použití různých metod spojování rozdílné, vždy se však nacházejí v prostoru mezi epiaurignackými industriemi na jedné straně a soubory ondratického typu na straně druhé.

Analýza shluků založená na porovnávání vybraných technologicko-typologických indexů statisticky hodnotitelných souborů ve studované oblasti tedy prokázala vzájemnou

podobnost souborů tzv. ondratického typu (szeletien levalloiské facie, bohunicien), stejně jako vzájemnou podobnost souborů řazených do epiaurignacienu. Vzájemná podobnost aurignackých souborů je již méně zřejmá. Szeletská lokalita Vincencov I tvoří samostatnou skupinu a z ostatních souborů je nejpodobnější industrii z lokality Kelčice I.

Na závěr je třeba uvést mnohá úskalí použití analýzy shluků k rozřídění analyzovaných souborů do jednotlivých skupin odpovídajících technokomplexům zastoupeným v oblasti. Ze statistického hlediska je třeba přiznat, že shluková analýza se častěji používá v případech, kdy počet analyzovaných objektů (případů) výrazně početně převyšuje počet proměnných (např. v marketingu), a čtrnáct analyzovaných lokalit se může zdát pro použití této metody jako nedostatečný počet. Podobným způsobem však byla tato metoda úspěšně aplikována již Z. Weberem na soubory z moravských lokalit datovaných do počátku mladého paleolitu (Valoch *et al.* 1993, 79–89). Dále je třeba upozornit na poměrně subjektivní způsob výběru proměnných

(indexů). Vybrané indexy byly ovšem zvoleny tak, aby co nejlépe odrážely rozdíly mezi technokomplexy zastoupenými v oblasti. Z archeologického hlediska je problematické zejména použití souborů pocházejících z povrchových lokalit, kde je možné předpokládat smíchání obsahu více kulturních vrstev přináležejících různým technokomplexům. V případě některých lokalit je to dokonce více než pravděpodobné. Dalším metodologickým problémem je poměrně malý počet nástrojů nalezených na některých analyzovaných lokalitách. V případě monokulturních lokalit jsou početnější soubory statisticky průkaznější než soubory chudší. Přes všechny výše uvedené metodologické problémy je zřejmé, že analýza shluků může být v kombinaci s klasickou statistickou metodou používanou pro mladopaleolitickou štípanou industrii (*de Sonneville-Bordes – Perrot 1953; Klíma 1956b*) vhodným nástrojem při přiřazování souborů štípané industrie k jednotlivým technokomplexům. Jde také o vhodnou metodu, jak patřičně znázornit podobnost či odlišnost různých industrií.