

ZÁVISLOST STŘEDNÍ HODNOTY A INTERVALU SPOLEHLIVOSTI SOUBORU ŘEPKY NA JEHO VELIKOSTI A ZPŮSOBU STATISTICKÉHO ZPRACOVÁNÍ

Dependence of the mean and confidence interval oilseed file on its size and the method of statistical processing

Josef PECEN, Iva ČERNÁ, Petra ZABLOUDILOVÁ

Česká zemědělská univerzita v Praze

Summary: The paper deals with the use of selected statistical methods for processing different sized sets of measured values rape seed weight in the range of 8-48 seeds of two varieties of oilseed rape, EXAGONE and JESPER. The results are mean values and confidence intervals for all processed files depending on their size and mixing ratio of the two varieties of rapeseed. When deciding on the selection of methods of statistical processing of the file is used as a histogram file.

Key words: mean of file, confidence interval, nonparametric statistical methods, file

Souhrn: Příspěvek se zabývá použitím vybraných statistických metod pro zpracování různě velkých souborů naměřených hodnot hmotnosti semen řepky v rozsahu 8 až 48 semen pro dvě odrůdy řepky olejné, EXAGONE a JESPER. Výsledkem jsou střední hodnoty a intervaly spolehlivosti všech zpracovávaných souborů v závislosti na jejich velikosti a poměru smísení obou odrůd řepky. Při rozhodování o výběru metody statistického zpracování souboru je využito i histogramu souboru.

Klíčová slova: střední hodnota souboru, interval spolehlivosti, neparametrické statistické metody, soubor

Úvod

Vyjádřit řadu naměřených dat jediným údajem (střední hodnotou) je snaha velmi stará a nejspolehlivějším nástrojem je vhodný způsob zpracování naměřených dat. Postupně byly vyvinuty stále sofistikovanější metody, které více respektují jak počet takto získaných dat, tak i jejich rozdělení, četnosti včetně vlastností toho, co bylo opakovaně měřeno. Většina používaných metod zpracování poskytuje obecnější poznatky, které se uplatňují i jinde. Například při zpracování obrazu (obrazové analýze) hraje důležitou roli jak histogram souboru, tak i jeho střední hodnota (Hlaváč a Šonka 1992). Tyto veličiny se pro některé materiály nedají získat přímo nebo jen obtížně a tak nezbývá než se pokusit tyto informace získat jinak.

Například při analýze obrazu textury brikety z biomasy nebo digestátu jsou základními informacemi počet a velikost částic. Různě velké soubory dat o těchto částicích se zpracovávají obdobně jako například soubory malých dat. Proto byla zvolena semena řepky, jako snadno dostupný materiál a v některých ohledech podobné chování malých částic biomasy. U materiálu biomasy a tedy i digestátu můžeme velikost částic kontrolovat třeba síťovou analýzou a při použití semen řepky jde velikost semen určit zjištěním hmotnosti každého semena řepky. Poznatky získané snadněji pomocí souboru semen řepky jde potom částečně využít například pro analýzu obrazu i jinde.

Materiál a metody

Tabulka 1: Počty semen odrůd řepky EXAGONE (A) a JESPER (B) pro jednotlivé míšící poměry.

A EXAGONE	B JESPER	A + B	A : B
A = 48	0	48	
A1 = 8	B1 = 40	48	1 : 5
A2 = 16	B2 = 32	48	1 : 2
A3 = 24	B3 = 24	48	1 : 1
A4 = 32	B4 = 16	48	2 : 1
A5 = 40	B5 = 8	48	5 : 1
0	B = 48	48	

Pozn.: Semena obou odrůd byla vždy vybrána náhodně na základě generátoru náhodných čísel.

Pro popsání experimentu byly zcela náhodně vybrány dvě odrůdy řepky olejné, EXAGONE a JESPER ze sklizně v roce 2012 na jednom stanovišti. Konečný soubor 50 semen od každé odrůdy byl získán postupným dělením na půlku z původní hmotnosti vzorku asi 100 g na konečnou hmotnost 2 g. Z tohoto množství

bylo poté podle generátoru náhodných čísel vybráno od každé odrůdy 50 semen, která byla jednotlivě zvážena (na celé mg). Z těchto dvou základních souborů byly opět na základě generátoru náhodných čísel vybrány postupně soubory: A, A1, A2, A3, A4, A5 a B, B1, B2, B3, B4, B5 s počty semen, které jsou uvedeny v tabulce 1.

Tyto soubory byly postupně kombinovány podle tabulky 1 a zpracovávány různými statistickými metodami, takže byly získány pro různé kombinace a velikosti souborů jejich různé střední hodnoty a intervaly spolehlivosti, v závislosti i na metodě zpracování naměřených souborů. Výsledky jsou v přehledné formě uvedeny v tabulce 2 včetně histogramů všech použitých souborů na obrázku 1. Jak střední hodnoty, tak i dolní a horní meze intervalu spolehlivosti střední hodnoty jsou v tabulce uvedeny přímo v mg hmotnosti semen. Pro lepší názornost byly použity pro zpracování souborů parametrické i neparametrické statistické metody, aby mohl vyniknout rozdíl mezi nimi. V uvedené tabulce mají uvedené zkratky následující obsah:

- AP aritmetický průměr
- GP geometrický průměr
- HP harmonický průměr
- M..... medián
- HM..... Hornova metoda
- LT lambda transformace
- SH střední hodnota souboru
- IS..... interval spolehlivosti střední hodnoty
- šířka inter. ... šířka intervalu spolehlivosti

Část tabulky označená jako „Charakteristika souborů“ obsahuje ve sloupcích počet semen řepky stejné hmotnosti podle jednotlivých souborů. Údaje uvedené v tab. 2 jsou na druhém desetinném místě zaokrouhlené. Vlastnosti většiny uvedených metod najde čtenář i ve statistických příručkách nebo stručný výtah jejich vlastností (Diggle a Chetwynd 2011, Krámer 2005).

Výsledky a diskuse

Pohled na histogramy všech použitých souborů na obrázku 1 dovoluje si udělat představu o rozdělení četností v jednotlivých souborech a jejich odchylkách od normálního rozložení.

Minimální velikost pro soubory s normálním rozložením bývá udávána počtem 20 (někdy 30) naměřených hodnot (Hawkins, 2005), obecně záleží i na tom, co se měří. Z tabulky 2 dále plyne, že střední hodnoty souboru kolem 20 naměřených hodnot již nemají extrémní odchylky ve velikosti středních hodnot, s šířkou intervalu spolehlivosti je to horší.

Nicméně z tabulky 2. je vidět, že nejčastěji uplatňovaný aritmetický průměr nevyjadřuje pro malé soubory nejlépe střední hodnotu. Z tabulky rovněž plyne, že neparametrické metody (u nich nezáleží na rozdělení četnosti) obecně poskytují lepší výsledky, zejména pro malé soubory. To je poznatek, který je potvrzen i autorem Hawkins

(2005) a zejména autory Dempír a Dohnal (2005). Podrobné teoretické zdůvodnění je možné nalézt v publikaci autorů Diggle a Chetwynd (2011). Vzhledem k tomu, že většina naměřených souborů nedisponuje normálním rozložením dat, je použití neparametrických metod zpracování takovýchto souborů naprosto oprávněné. Navíc jsou většinou naměřené soubory dat malých velikostí. Z použitých neparametrických metod v uvedeném příkladu má nejlepší výsledky (pro malé i větší soubory) použití metody Lambda transformace. Ta je sice pracnější, ale dosažené výsledky jsou objektivnější. Těchto neparametrických metod je celá řada, ale jejich praktické využití při zpracování podobných dat není tolik rozšířené. Tomu trochu brání i nutnost použití počítačů, protože při použití kalkulačky jsou tyto úlohy časově náročnější. Je zajímavé, že i jednoduchá metoda GP poskytuje docela dobré výsledky, jak je vidět z tabulky, stejně tak i metoda HP.

Závěr

- Při rozhodování o použití vhodné metody na zpracování souboru naměřených dat je třeba brát v úvahu nejen velikost souboru, ale i jeho histogram.
- Při zpracování souboru s počtem naměřených dat dvacet a méně raději používat některou neparametrickou statistickou metodu
- Vždy určit Interval spolehlivosti, jako základní charakteristiku souboru.
- Nezpracovávat soubor s méně než pěti naměřenými hodnotami.

Použitá literatura

- [1] Diggle, P., J., Chetwynd, A., G., Statistics and Scientific Metod. Oxford Universty Press, New York, USA, 2011, 172 pp. ISBN: 978-0-19-954318-2.
- [2] Krámer, W.: Statistika do vesty. BARONET, Praha 2005. 124 s. ISBN 80-7214-848-6.
- [3] Hawkins, D., M., Biomeasurement: Understanding, Analysing, and Communicating Data in the Biosciences. Oxford Universty Press, New York, USA, 2005, 300 pp. ISBN-13:978-0199265152.
- [4] Dempír, J., Dohnal, L. Některé robustní postupy určení střední hodnoty a rozptýlení souboru výsledků a jejich použití. Klin. Biochem. Metab., 13(34); 2005, No. 3, p. 139-144.
- [5] Hlaváč, V., Šonka, M.: Počítačové vidění. Grada a.s. 1992. 272 s. ISBN 80 85424 67 3.

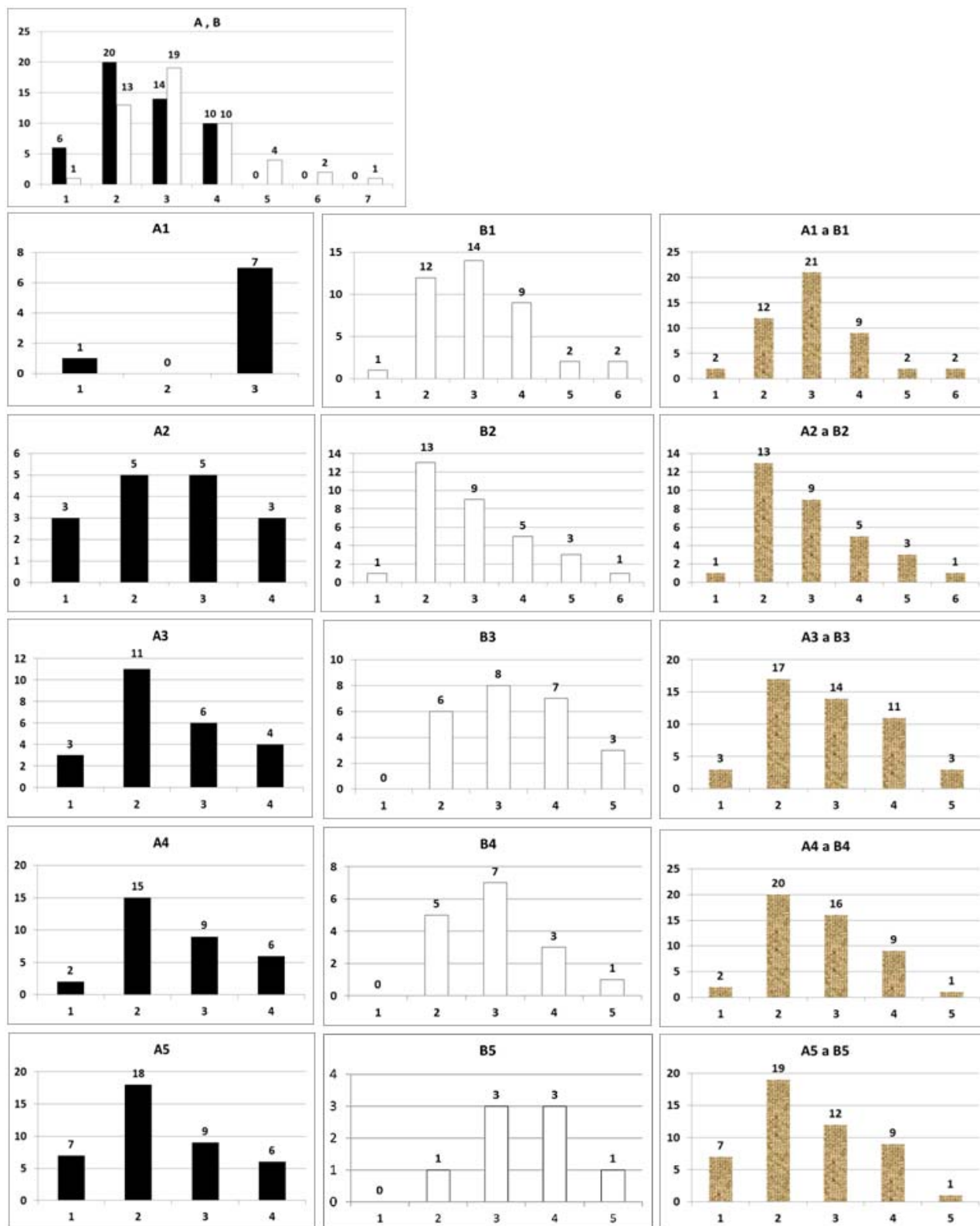
Kontaktní adresa

doc. Ing. Josef Pecen, CSc., ČZU v Praze, Fakulta tropického zemědělství, Kamýcká 129, 161 21 Praha 6 - Suchbátka e-mail: Pecen@ftz.czu.cz

Tato práce byla realizována za finanční podpory grantu IGA České zemědělské univerzity v Praze, reg. číslo 2 014 030.



Obrázek 1: Histogramy všech zpracovaných souborů obou odrůd řepky.



Pozn.: A,B jsou celé soubory obou odrůd řepky, každý v počtu 48 náhodně vybraných semen. Jejich podíly jsou označeny stejně jako v tabulce 1. Černé sloupce v histogramech jsou odrůda EXAGONE a bílé sloupce odrůda JESPER. Šrafované sloupce jsou vždy součet. Semena obou odrůd byla vždy vybrána náhodně na základě generátoru náhodných čísel.

Tabulka 2: Střední hodnoty a intervaly spolehlivosti (pro $\alpha = 0,05$) pro různé poměry smíchání obou odrůd řepky a pro vybrané metody.

Charakteristika souborů									AP				GP				HP				M				HM				LT			
Označení souboru	Počet semen	Struktura souborů							SH	IS			SH	IS			SH	IS			SH	IS			SH	IS						
		1	2	3	4	5	6	7		min	max	šířka inter.		min	max	šířka inter.		min	max	šířka inter.		min	max	šířka inter.		min	max	šířka inter.				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
A	50	6	20	14	10				2,56	2,29	2,82	0,53	2,36	0	6,21	6,29	2,16	1,94	2,38	0,44	2	1,77	2,22	0,45	2,5	2,16	2,83	0,67	2,39	2,09	2,73	0,64
A1	8	1		7					2,75	2,16	3,34	1,18	3	1,89	3,62	1,73	2,4	1,84	2,96	1,11	3,0	2,45	3,55	1,11	2,5	1,94	3,06	1,13	2,64	1,95	3,53	1,58
B1	40	1	13	14	9	2	1		3,13	2,18	4,07	1,9	2,27	2,16	3,22	1,06	2,83	2,33	3,33	0,93	3	2,5	3,49	0,99	3,5	3,16	3,84	0,68	2,9	2,55	3,29	0,75
+	48	2	13	20	9	2	2		3,06	2,16	3,96	1,8	2,61	2,58	3,2	0,62	2,66	2,21	3,11	0,91	3	2,55	3,45	0,91	4	3,66	4,34	0,67	2,9	2,57	3,28	0,7
A2	16	3	5	5	3				2,35	1,94	2,77	0,83	2,27	2,57	4,75	2,18	2,02	1,77	2,27	0,5	2,0	1,75	2,25	0,5	2,5	2,06	2,93	0,5	2,26	1,59	3,14	1,55
B2	32	1	13	9	5	3	1		2,97	2,03	3,9	1,87	2,75	2,38	3,19	0,81	2,56	2	3,11	1,11	3	2,44	3,55	1,11	3	2,39	3,61	1,22	3,07	2,35	3,97	1,62
+	48	4	18	14	8	3	1		2,812	2,46	3,16	0,69	2,58	2,28	2,92	0,64	2,36	2,13	2,58	0,45	3	2,77	3,22	0,45	3	2,59	3,41	0,82	3,07	2,47	3,79	1,32
A3	24	3	11	6	4				2,46	2,07	2,85	0,78	2,28	1,91	2,71	0,78	2,08	1,86	2,4	0,61	2	1,68	2,32	0,64	2,5	2,18	2,81	0,63	2,26	1,7	2,96	1,26
B3	24	0	6	8	7	3			3,29	2,87	3,71	0,84	3,14	2,75	3,58	0,83	2,32	1,99	2,64	0,64	3	2,55	3,45	0,9	3,5	3,8	3,82	0,64	2,66	1,98	3,54	1,56
+	48	3	17	14	11	3			2,87	2,57	3,17	0,6	2,87	2,57	3,17	0,6	2,67	2,38	3	0,62	3	2,55	3,45	0,9	3	2,37	3,63	0,99	2,66	2,16	3,26	1,1
A4	32	2	15	9	6				2,59	2,28	2,91	0,63	2,44	2,4	2,48	0,08	2,26	1,98	2,53	0,55	2	1,72	2,28	0,56	2,5	2,2	2,8	0,6	2,46	2,13	2,86	0,71
B4	16	0	5	7	3	1			3	2,23	3,77	1,54	2,75	2,66	2,83	0,17	2,77	2,37	3,16	0,79	3	2,6	3,39	0,79	3	2,13	3,87	1,74	2,99	2,43	3,41	0,98
+	48	2	20	16	9	1			2,73	2,47	2,98	0,51	2,58	2,33	2,85	0,52	2,42	1,97	2,87	0,9	3,0	2,55	3,45	0,9	2,5	2,24	2,72	0,48	2,6	2,3	2,91	0,61
A5	40	7	18	9	6				2,35	2,05	2,65	0,6	2,15	1,83	2,52	0,69	1,95	1,7	2,2	0,5	2	1,75	2,24	0,49	2,5	2,21	2,78	0,57	2,18	1,86	2,55	0,69
B5	8	0	1	3	3	1			3,5	2,74	4,25	1,51	3,38	2,69	4,25	1,56	3,28	2,73	3,83	1,1	3,5	2,94	4,05	1,11	3,5	2,94	4,06	1,12	3,4	2,7	4,25	1,55
+	48	7	19	12	9	1			2,54	2,19	2,88	0,68	2,32	2	2,69	0,69	2,09	1,87	2,31	0,44	2	1,77	2,22	0,45	2,5	2,24	2,76	0,52	2,35	2,03	2,72	0,69
B	50	1	13	18	11	4	2	1	3,28	2,91	3,64	0,73	3,06	1,28	7,29	6,01	2,5	2,056	2,94	0,88	3	2,57	3,44	0,87	3	2,33	3,63	1,3	3,09	2,74	3,49	0,75

Pozn.: Čísla 1...7 (sloupce 3 až 9 tabulky) v tabulce pod nápisem „Struktura souborů“ jsou hmotnosti jednotlivých semen řepky v mg a ve sloupcích pod nimi jsou uvedeny jejich počty v každém souboru.