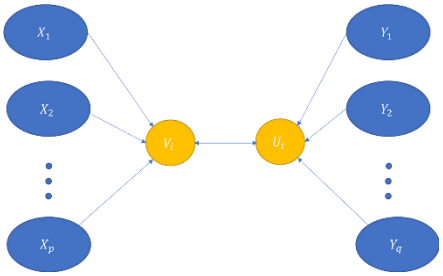


Kanonická korelační analýza výživových hodnot mléka

Celosvětově se snižující konzumace masa a nárůst vegetariánství s sebou nese i zvýšenou poptávku po náhradách mléčných výrobků. Někdy volíme místo klasického kravského mléka například jen bezlaktózovou variantu, jelikož narůstá i počet laktózových intolerancí v populaci. Vybrat si mléko, nejvhodnější pro svoje zdraví není zas tak komplikované při současné nabídce mlék.

V následujícím textu prozkoumáme, zda je možné se orientovat v nabídce mléka pouze na základě jednoho faktoru. Co je nevýhodnější zkoumat? Obsah bílkovin, množství Vápníku, nebo například stačí sledovat pouze energetickou hodnotu mléka?

Pro analýzu kanonických korelací využijeme soubor čítající 13 druhů mlék. Z toho je 6 mlék živočišných a 7 rostlinných. Do analýzy byla vybrána pouze mléka neslazená, která se prodávají již v tekuté formě jako nápoje (byla vynechána sušená mléka). Výživové hodnoty jednotlivých potravin jsou dostupné jak na obalech produktů, tak například na stránkách kaloricketabulky.cz, nebo mlekarenskelisty.cz, apod.

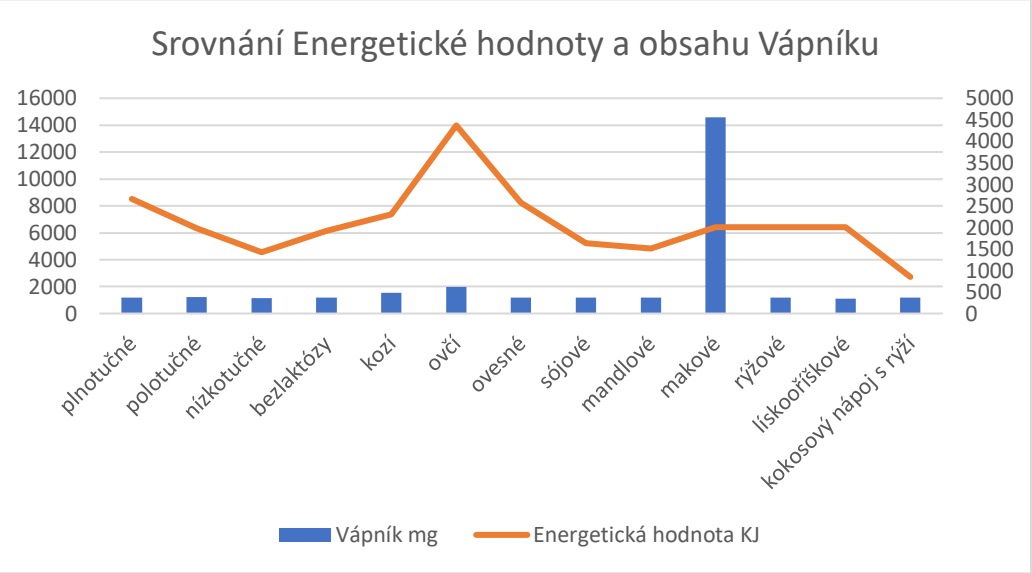
V krátkosti si můžeme představit obsah makroživin jednotlivých mlék z analýzy pomocí grafu:

Graf 1:

V grafu vidíme, že co do obsahu bílkovin a tuku vede ovčí mléko. Nejvyšší množství sacharidů má mléko rýžové.

Další důležitou složkou, která nás u mléka zajímá je Vápník. Nejbohatším přirozeným zdrojem Vápníku je mléko makové. Zatímco do mandlového mléka se Vápník uměle dodává. Jistě zajímavá informace z hlediska výživy je pak energetická hodnota mléka. Pro analýzu byl zvolen ukazatel KJ. Následující graf zobrazuje porovnání množství Vápníku a energetické hodnoty jednotlivých mlék.

Graf 2:



Zde je krásně patrná převaha makového mléka, co se obsahu Vápníku týče, oproti ostatním mlékům. Ovčí mléko a kravské plnotučné mléko vyčnívá v energetické hodnotě.

Nyní přejdeme k samotné Kanonické korelaci. Tato analýza patří mezi metody vícerozměrné analýzy dat. Jejím cílem je nalézt vztahy mezi dvěma skupinami proměnných. Můžeme se na ni dívat jako na rozšíření vícenásobné lineární regrese, při níž se hledá vztah jedné (tzv. vysvětlované) proměnné se skupinou (tzv. vysvětlujících) proměnných. Zatímco principem kanonické korelační analýzy je hledání lineární kombinace jedné **skupiny p** proměnných, která nejlépe koreluje s lineární kombinací druhé **skupiny q** proměnných. Příkladem může být hledání vztahů mezi parametry počasí (průměrné denní srážky, vlhkost, počet hodin slunečního záření) s výnosem plodin (výška rostlin, hmotnost po usušení, počet listů). Či například jak souvisí obsah živin potravin s energetickou hodnotou a PHE. Což je přesně náš případ. Metoda nevyžaduje

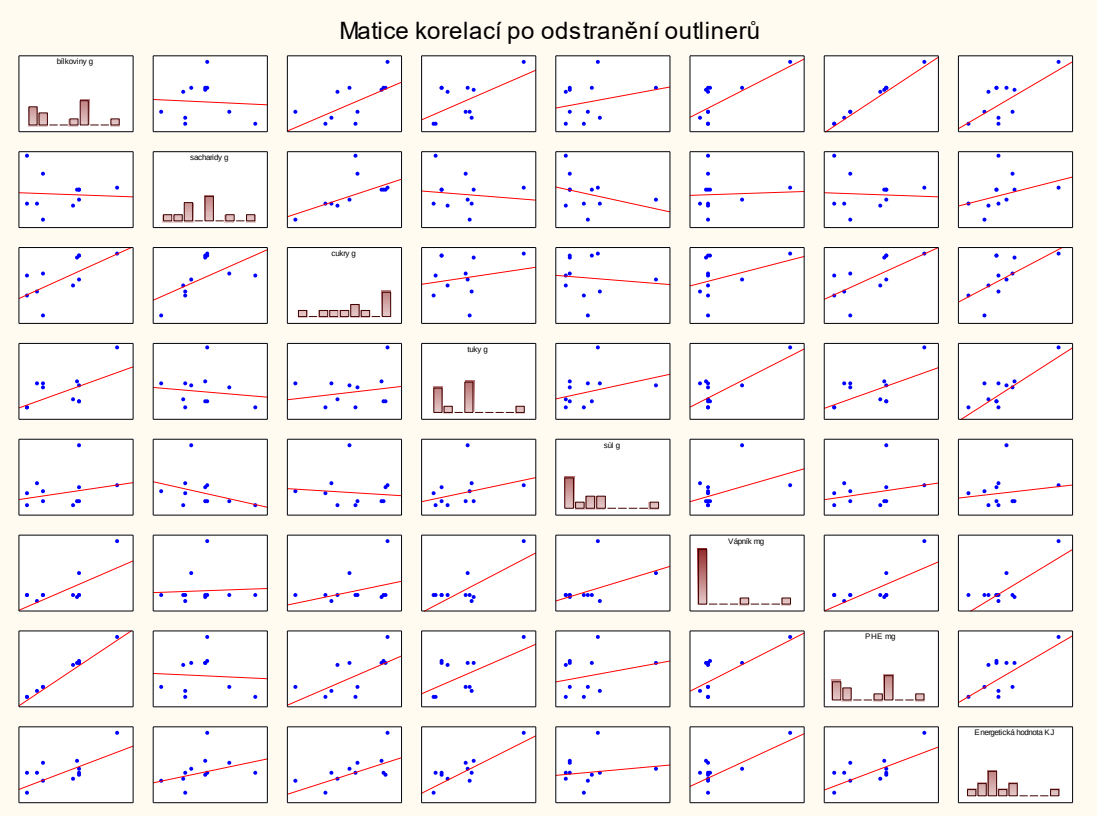
normální rozdělení proměnných. Naopak vyžaduje alespoň 10 subjektů a měření nesmí obsahovat nulové hodnoty. Proměnné by měly být lineární (Koriťáková, 2020).[1]

Jen pro dovysvětlení, PHE (Fenylalanin) je esenciální aminokyselinou, kterou si naše tělo nedokáže samo vyrobit. Obsahují ji bílkovinové produkty. Nalezneme ji proto i v čokoládě, sýrech, kvasnicích, ořiščích, ale též v mladém ječmeni.

Pro statistickou analýzu byl zvolen program Statistica. Postup zadání Kanonické korelace: Statistica - > Mult/Exploratory -> Canonical ->Variables ALL -> OK.

V tomto bodě se zastavíme a zkontrolujeme na záložce Descriptives položku Matrix_Plot of Correlations. Zobrazí se nám matice bodových grafů. Zkontrolujeme, zda se nevyskytují v souboru outlinery a zda jsou korelace lineární. V této analýze vyšlo, že makové mléko díky svému obsahu Vápníku je outliner. Proto je třeba ho ze souboru vyřadit. Nízkotučné mléko je též outliner v tucích. Proto ho též z analýzy vyřadíme. Pracujeme dále jen s 11 mléky.

Graf 3: výsledná matice korelací jednotlivých živin po odstranění outlinerů ze souboru.



[1] *Data a další informace o této zprávě jsou dostupné na adrese https://dostal.vyzkum-psychologie.cz/stat4?i=346*

Dále pokračujeme v programu Statistica takto: Variables for Canonical Analyses -> první/levý sloupec korelovaných proměnných, druhý/pravý sloupec korelovaných proměnných (zde vybereme PHE a KJ).

Nejprve se seznámíme s výsledky kanonické korelace všech živin s PHE a energetickou hodnotou (KJ). Jistě nás nepřekvapí výsledky: Kanonické R = 0,99994, chí^2 = 81,83, p=0. Korelace PHE a KJ = 0,54. Tedy poměrně silná korelace.

Tabulka 1: Souhrn Kanonické analýzy: Kanonické R = 0,99994; Chi²=81,825; p=0,0000.

	Levá strana	Pravá strana
Počet proměnných	6	2
Extrahovaný rozptyl	54,35 %	100 %
Celková redundance	54,3 %	99,92 %
Proměnná 1	bílkoviny g	PHE mg
Proměnná 2	sacharidy g	Energetická hodnota KJ
Proměnná 3	cukry g	
Proměnná 4	tuky g	
Proměnná 5	sůl g	
Proměnná 6	Vápník mg	

Další z provedených testů analýzy je Chí^2 Test. V testu se zobrazují výsledky Lambdy. V našem případě se Lambda blíží nule.

Tabulka 2: Chí^2 Test s postupně odstraněnými kořeny (Roots)

	Canonical R	Canonicl R-sqr.	Chi-sqr.	df	p	Lambda
Root 0	0,999936	0,999873	81,82	12	0,00000	0,0000
Root 1	0,998643	0,997288	32,51	5	0,000005	0,0027

Další částí analýzy je korelace položek levého sloupce a korelace levého a pravého sloupce. Korelace levého sloupce je vzájemnou korelací jednotlivých živin. Nejvyšší korelaci nalezneme mezi tuky a Vápníkem (0,78), dále bílkovinami a Vápníkem (0,71) a bílkovinami a cukry (0,66).

Tabulka 3: Korelace položek levé strany

	bílkoviny g	sacharidy g	cukry g	tuky g	sùl g	Vápník mg
bílkoviny g	1,00	-0,06	0,66	0,60	0,25	0,71
sacharidy g	-0,06	1,00	0,58	-0,12	-0,33	0,05
cukry g	0,66	0,58	1,00	0,20	-0,11	0,35
tuky g	0,60	-0,12	0,20	1,00	0,32	0,78
sùl g	0,25	-0,33	-0,11	0,32	1,00	0,44
Vápník mg	0,71	0,05	0,35	0,78	0,44	1,00

Korelace položek levé a pravé strany vychází po zaokrouhlení jako dokonalá korelace PHE a bílkovin. To není překvapení, jelikož PHE se vyskytuje pouze v bílkovinách. Energetická hodnota silně koreluje s tuky, Vápníkem, bílkovinami. Slabá korelace je pouze se solí.

Tabulka 4: Korelace položek levé strana a pravé strany

	bílkoviny g	sacharidy g	cukry g	tuky g	sùl g	Vápník mg
PHE mg	1,00	-0,06	0,66	0,59	0,25	0,70
Energetická hodnota KJ	0,71	0,35	0,62	0,85	0,15	0,80

To, co nás v kanonické korelaci velmi zajímá, jsou jednotlivé váhy = kanonické koeficienty.

V kanonické korelační analýze jsou kanonické koeficienty **a** a **b** ve vztazích:

$$U1 = a_1 y_1 + a_2 y_2 + ... + a_p y_p$$

$$V1 = b_1 x_1 + b_2 x_2 + ... + b_p x_p$$

Koeficienty jsou hledány tak, aby maximalizovaly korelaci mezi proměnnými $\mathbf{U_1}$ a $\mathbf{V_1}$. Po nalezení nejlepších možných odhadů **a** a **b** se $\mathbf{U_1}$ nazývá první kanonická proměnná závisle proměnných y a $\mathbf{V_1}$ první kanonická proměnná nezávisle proměnných x. Obě kanonické proměnné mají průměr roven nule. Korelace mezi U_1 a V_1 se nazývá **první kanonická korelace** a čtverec této korelace je nazýván **vlastní číslo** (Meloun & Militický, 2002).

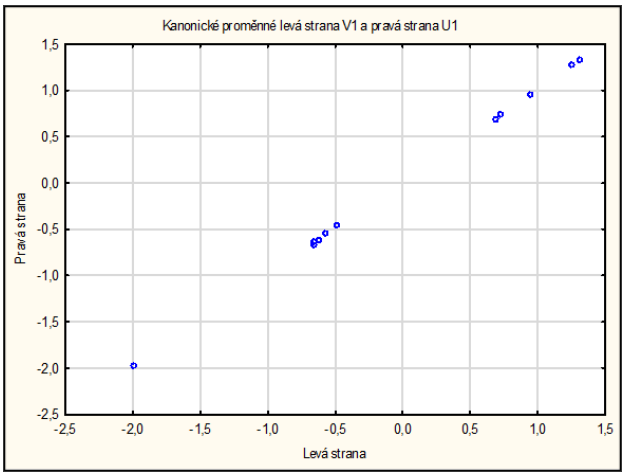
Nejprve se seznámíme s kanonickými koeficienty. Analýza spočítala dva modely. V jednom má větší váhu PHE (Root 1) a v druhém mají větší váhu KJ (Root 2). Zde je poměrně důležité si povšimnout, že váhy PHE jsou záporné. Zatímco KJ jsou kladné.

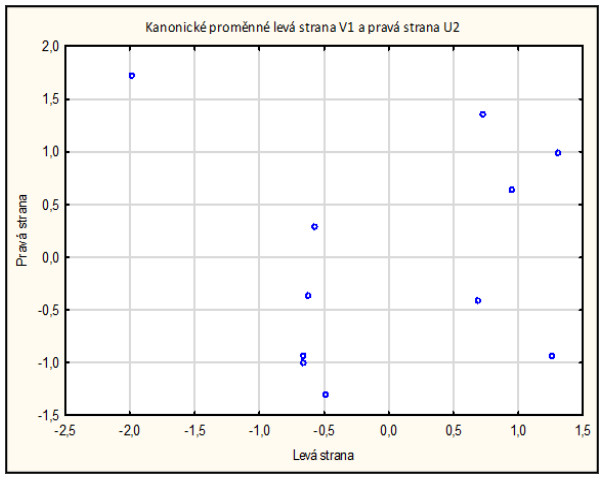
Tabulka 5 a 6: Kanonické váhy levé a pravé strany pro kořeny 1 a 2 (Root 1, 2).

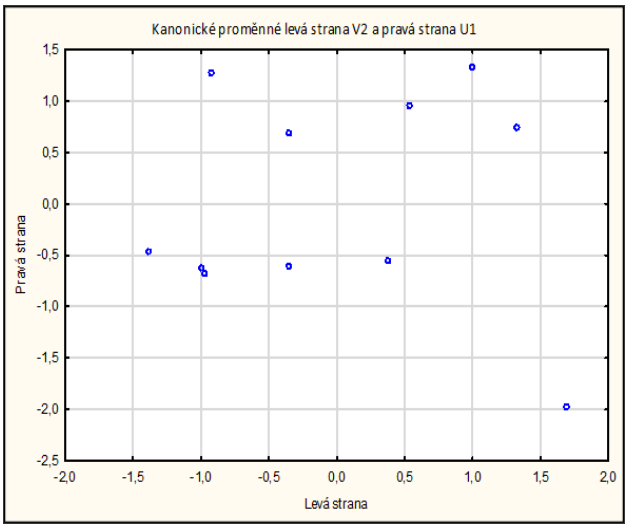
Kanonické váhy levá strana				Kanonické váhy pravá strana		
	Root 1	Root 2			Root 1	Root 2
bílkoviny g	-1,01	-0,60		**PHE mg**	-1,03	-0,97
sacharidy g	0,02	0,61		**Energetická hodnota KJ**	0,04	1,41
cukry g	-0,01	0,07				
tuky g	0,04	1,07				
sùl g	0,00	0,01				
Vápník mg	0,00	-0,03				

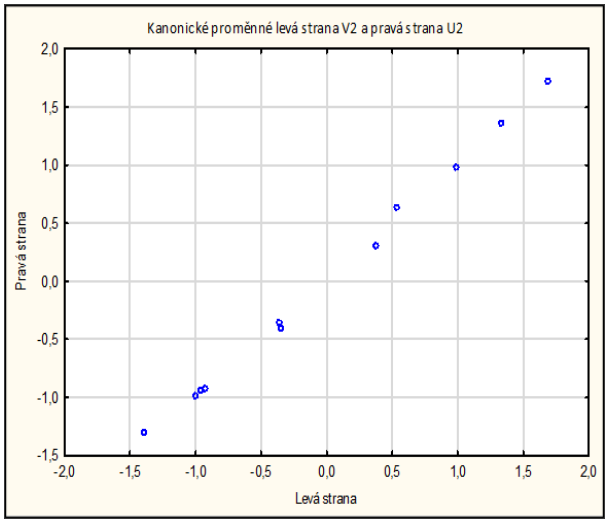
Jako poslední si představíme všechny kanonické korelace VU. Zleva doprava a shora dolů vidíme kanonické korelace V1U1, V1U2, V2U1, V2U2.

Grafy 4 až 7: Kanonické korelace VU









Tato cvičná analýza přinesla výsledky, které se daly očekávat. Silnější model je model druhý s primární větví KJ. Energetickou hodnotu mléka nejlépe vyjadřuje obsah tuků. PHE nejvíce koreluje s množstvím bílkovin. Žádná z potravin není výjimečná ve všech oblastech. I zde by se dala uplatnit poučka, že nejlepší strava je ta, která je pestrá, a proto je dobré zařazovat do stravy rozličná mléka.

Zdroje:

Example1: Canonical Correlation (1995-2022). Users Guide. TIBCO Statistica® 14.0.1. Dostupné 10. dubna 2024 z: https://docs.tibco.com/pub/dsc-stat/14.0.1/doc/html/UsersGuide/_shared/example1-canonical-correlation.htm

Koriťáková, E. (2. dubna 2020). CCA-Výukový text. Pokročilé metody analýzy dat v neurovědách. Informační systém Masarykovi Univerzity. Dostupné 10. dubna 2024 z: https://is.muni.cz/el/med/jaro2020/DSAN02/um/53004109/CCA_vyukovy_text.pdf

Meloun, M. & Militický, J. (2002). Kdy kanonická korelace a kdy vícerozměrná lineární regrese? Konference, Komorní Lhotka. Dostupné 9. dubna 2024 z: https://meloun.upce.cz/docs/publication/123.pdf