

Nové možnosti rozvoje vzdělávání na Technické univerzitě v Liberci

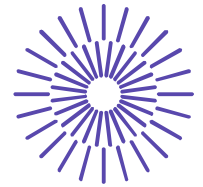
Specifický cíl A3: Tvorba nových profesně zaměřených studijních programů

NPO_TUL_MSMT-16598/2022



Téma 2: Příklad 2 – charakteristiky kardinální proměnné

Ing. Vladimíra Hovorková Valentová, Ph.D.



Zadání příkladu:

V distribučním centru jsou vedeny záznamy o počtu položek v odeslaných zásilkách. Data byla roztržiděna do tabulky prostého rozdělení četností. Charakterizujte pomocí vhodných charakteristik úroveň, variabilitu a šikmost dané proměnné. Všechny výsledky interpretujte.

Počet položek v zásilce (x_i)	Počet zásilek (n_i)	Podíl zásilek (p_i)	Kumulativní četnosti absolutní	Kumulativní četnosti relativní
1	2	0,050	2	0,050
2	7	0,175	9	0,225
3	5	0,125	14	0,350
4	12	0,300	26	0,650
5	6	0,150	32	0,800
6	3	0,075	35	0,875
7	1	0,025	36	0,900
8	4	0,100	40	1,000
Celkem	40	1,000	x	x

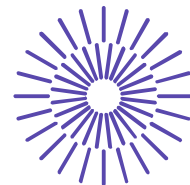
Řešení příkladu:

Sledovanou proměnnou je počet položek v zásilce. Tato proměnná je vyjádřena číselně a její hodnoty je možné hierarchicky seřadit a určit i rozdíly v jednotlivých hodnotách v kusech. Jde tedy o číselnou kardinální proměnnou. Úroveň hodnot kardinální proměnné lze charakterizovat pomocí modu, aritmetického průměru, mediánu či jiných kvantilů. Uvedme si tedy nejčastěji používané charakteristiky úrovně.

$$\hat{x} = 4$$

Nejčastěji se jsou do zásilek vloženy 4 položky.

$$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2} = \frac{x_{20.} + x_{21.}}{2} = \frac{4 + 4}{2} = 4$$



Jednotky, které se nacházejí na pořadovém místě 20 a 21 určíme nejlépe pomocí kumulativních absolutních četností. Hodnota mediánu nám říká, že 50 % zásilek obsahuje 4 položky, anebo méně (anebo více).

Pro výpočet aritmetického průměru zde volíme tvar vážený, protože data jsou uspořádána v tabulce rozdělení četností. Pomocný výpočet je uveden v tabulce níže.

Počet položek v zásilce (x_i)	Počet zásilek (n_i)	$x_i n_i$
1	2	2
2	7	14
3	5	15
4	12	48
5	6	30
6	3	18
7	1	7
8	4	32
Celkem	40	166

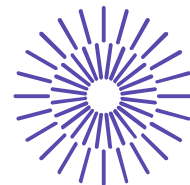
$$\bar{x} = \frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} = \frac{166}{40} = 4,15$$

Průměrně zásilka obsahuje 4,15 položky.

Variabilitu kardinální proměnné můžeme charakterizovat např. pomocí rozptylu, směrodatné odchylky či variačního koeficientu.

Pro výpočet rozptylu volíme rovněž vážený tvar a pro ukázkou ručního výpočtu tvar výpočtový:

$$s_x^2 = \frac{\sum_{i=1}^k x_i^2 n_i}{\sum_{i=1}^k n_i} - \left(\frac{\sum_{i=1}^k x_i n_i}{\sum_{i=1}^k n_i} \right)^2 = \frac{830}{40} - 4,15^2 = 3,5275$$



Pomocný výpočet k výpočtu rozptylu najdeme v tabulce níže. Rozptyl počtu položek v zásilkách je 3,5275.

Počet položek v zásilce (x_i)	Počet zásilek (n_i)	$x_i n_i$	$x_i^2 n_i$
1	2	2	2
2	7	14	28
3	5	15	45
4	12	48	192
5	6	30	150
6	3	18	108
7	1	7	49
8	4	32	256
Celkem	40	166	830

$$s_x = \sqrt{s_x^2} = \sqrt{3,5275} = 1,878$$

Počty položek v zásilkách se v průměru odlišují od průměrného počtu o 1,878 položky.

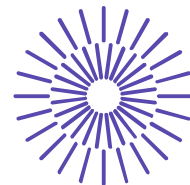
$$V_x = \frac{s_x}{\bar{x}} = \frac{1,878}{4,15} = 0,453$$

Variabilita počtu položek v zásilkách je spíše nižší. Směrodatná odchylka se podílí na aritmetickém průměru 45,3 %.

Šikmost rozdělení počtu položek v zásilce můžeme posoudit pomocí jednoduché míry šikmosti:

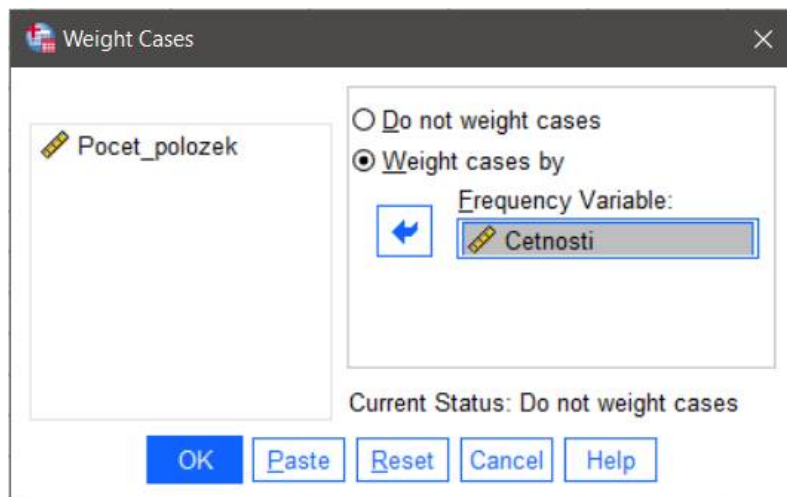
$$\alpha' = \frac{n' - n''}{n} = \frac{26 - 14}{40} = 0,3$$

Počet hodnot menších než aritmetický průměr (4,15) je 26 (vyčteme z tabulky rozdělení četností), počet hodnot větších než aritmetický průměr je 14. Výsledná hodnota nám říká, že rozdělení počtu položek v zásilce je kladně zešikmené. To znamená, že zde najdeme více zásilek s podprůměrným počtem položek.

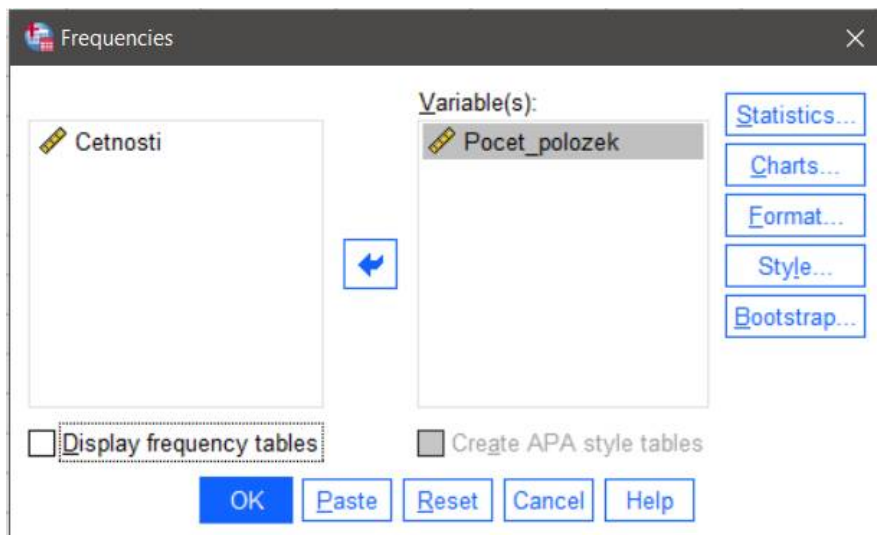
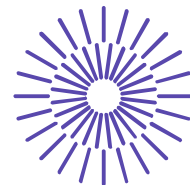


Pokud budeme chtít dané charakteristiky spočítat pomocí programu SPSS, zadáme nejprve data do datového listu. Jak tedy zadat tabulku rozdělení četností?

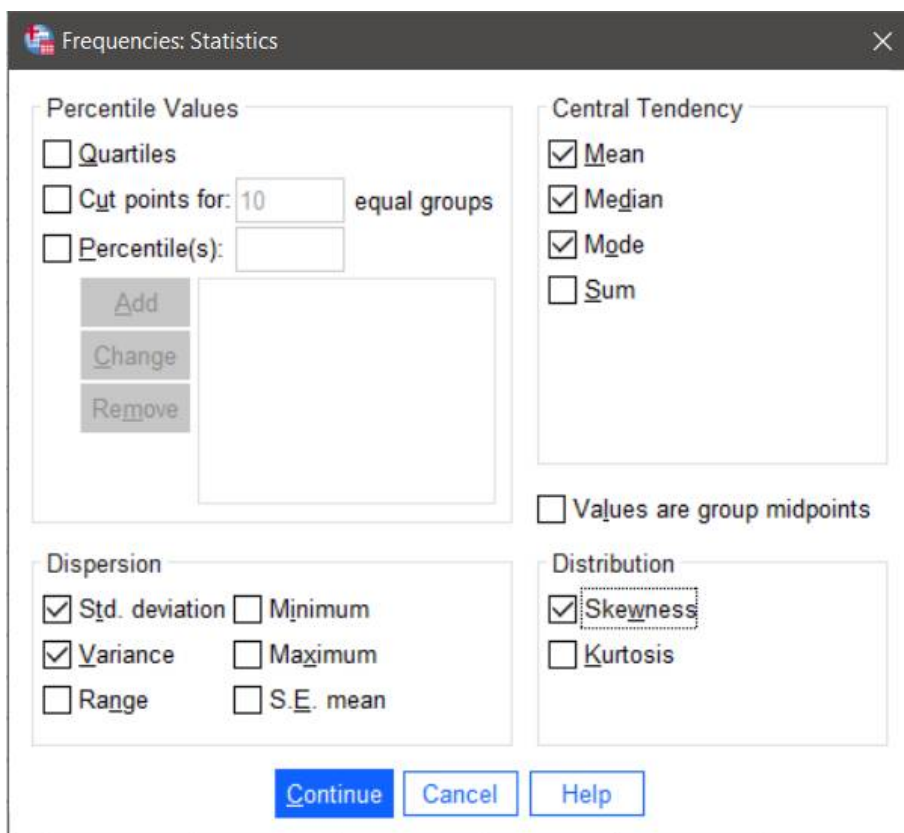
Nejprve zadáme do dvou sloupců jako dvě samostatné proměnné první dva sloupce tabulky rozdělení četností, tj. obměny proměnné (x_i) a absolutní četnosti výskytu těchto obměn (n_i). Pro spojení těchto dvou samostatných proměnných do tabulky volíme posloupnost procedur **Data – Weight Cases** a vstupní panel vyplníme následovně:

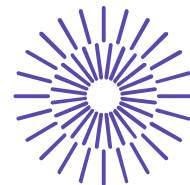


Po potvrzení klávesou OK bude počet položek „vážen“ četnostmi. Nyní již lze přejít k výpočtu popisných charakteristik. K tomu je potřeba zvolit **Analyze – Descriptive Statistics – Frequencies**. Vstupní panel vyplníme takto:



Dále zmáčkneme tlačítko *Statistics*, kde vybereme charakteristiky, které chceme vypočítat:





Mean – aritmetický průměr.

Median – medián (50% kvantil).

Mode – modus.

Std. Deviation – směrodatná odchylka VÝBĚROVÁ. SPSS nepočítá klasickou směrodatnou odchylku.

Variance – rozptyl VÝBĚROVÝ. SPSS nepočítá hodnotu klasického rozptylu.

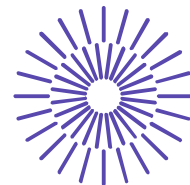
Skewness – momentová míra šikmosti; neshoduje se s jednoduchou mírou šikmosti, kterou jsme počítali výše, ale orientačně ji lze použít.

SPSS nám poskytne tento výstup:

Statistics		
Pocet_polozek		
N	Valid	40
	Missing	0
Mean		4,15
Median		4,00
Mode		4
Std. Deviation		1,902
Variance		3,618
Skewness		,527
Std. Error of Skewness		,374

Jak vidíme, hodnoty směrodatné odchylky a rozptylu se s našimi výpočty neshodují, protože my jsme počítali hodnoty klasického rozptylu a klasické směrodatné odchylky. Jednoduchou úpravou ale docílíme stejných hodnot, neboť platí:

$$s_x^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{n - 1} \cdot \frac{n - 1}{n}$$



Tedy:

$$s_x^2 = s_x'^2 \cdot \frac{n-1}{n}$$

$$s_x^2 = 3,618 \cdot \frac{39}{40} = 3,5276$$

$$s_x = \sqrt{3,5276} = 1,878$$

Drobné odchylky mohou vzniknout jako chyba ze zaokrouhlování.

Variační koeficient ani jednoduchá míra šikmosti není v programu SPSS k dispozici, je potřeba spočítat je ručně nebo např. pomocí MS Excel.