

Základní deskriptivní statistika

Explorace dat

Vizualizace dat

s použitím programu Statistica 7

Kateřina Vlčková, Ph.D.

Centrum pedagogického výzkumu
PdF MU Brno

1

Deskriptivní statistika

- Slouží
 - Ke kontrole dat
 - Základnímu popisu dat
 - Předzpracování dat pro další analýzy
 - A k exploraci dat
 - Exploraci problematiky

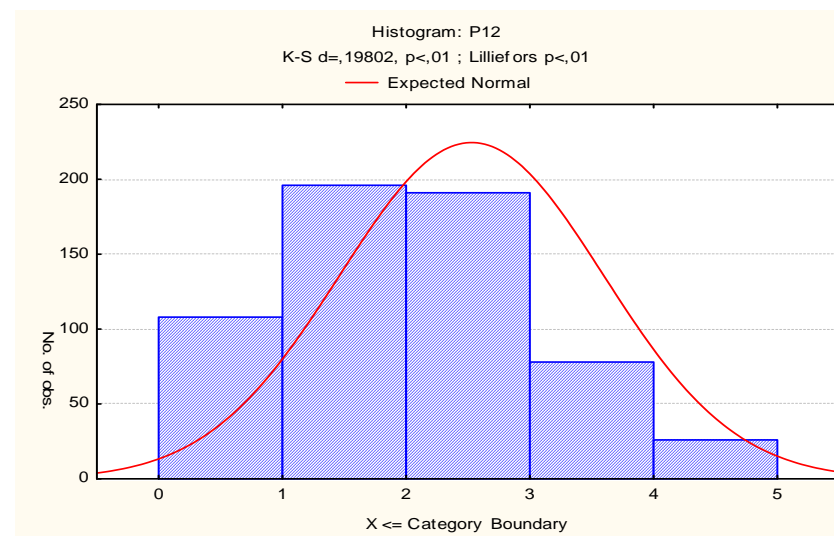
2

Četnosti

- **Frekvence** (frequencies)
 - Absolutní četnosti
 - Relativní četnosti
 - Kumulativní četnosti

Frequency table: P12 (database_strategie)						
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
1	108	108	18,03005	18,0301	17,82178	17,8218
2	196	304	32,72120	50,7513	32,34323	50,1650
3	191	495	31,88648	82,6377	31,51815	81,6832
4	78	573	13,02170	95,6594	12,87129	94,5545
5	26	599	4,34057	100,0000	4,29043	98,8449
Missing	7	606	1,16861		1,15512	100,0000

Histogram (vizualizace dat)



Deskriptivní statistika

P12 – vracím se k tomu, co jsem se učil dříve

Descriptive Statistics (database_strategie)												
Variab	Valid	Mean	Confiden	Confiden	Geometri	Harmon	Medial	Mode	Frequen	Sum	Minimul	Maximu
			-95,000	+95,000	Mean	Mean			of Mode			
P12	59	2,5292	2,4438	2,6146	2,2875	2,0348	2,0000	2,0000	19	1515,0	1,0000	5,0000

Descriptive Statistics (database_strategie)													
Variab	Lower	Upper	Percentil	Percentil	Range	Quartil	Varianc	Std.De	Standar	Skewnes	Std.Err.	Kurtosi	
	Quartil	Quartil	10,0000	90,0000		Range		Error	Error	Skewnes	Skewnes	Kurtosi	
P12	2,0000	3,0000	1,0000	4,0000	4,0000	1,0000	1,1325	1,0641	0,0434	0,3410	0,0998	-0,4478	0,1993

5

Deskriptivní statistika

- **N**: Počet platných případů N (valid N)
- **Suma**
- **Minimum, maximum**

6

Deskriptivní statistika použití měr centrální tendence

- **Průměr** (arithmetical average)
 - Součet všech údajů vydělený jejich počtem
 - Kdy se používá:
 - Min. intervalová data
 - Symetrické rozdělení
 - Chceme použít stat. testy

7

Deskriptivní statistika použití měr centrální tendence

- **Me: Medián**
 - Dělí řadu podle velikosti seřazených prvků do dvou stejných polovin
 - 0, 1, 2, 5, 8, 9, 10 Me = 5
 - Kdy se používá?
 - Min. ordinální měřítko
 - Chceme znát střed rozdělení
 - Data mohou obsahovat odlehlé hodnoty
 - Rozdělení dat je zešikmené

8

Deskriptivní statistika

použití měr centrální tendence

- **Mo: Modus, četnost modu**

- Nejčastější hodnota
- Zejména u kategoriálních dat
- Zobrazení pomocí histogramu
- Kdy se používá?
 - Jestliže rozdělení má více vrcholů
 - Chceme jen základní přehled
 - Slovem „průměrně“ se myslí nejčastější hodnota
 - Data min. v ordinálním měřítku

9

Deskriptivní statistika

Míry rozptýlenosti (measurement of variability)

- **Rozptyl (variance)**

- Čtverec vzdálenosti hodnot jednotlivých případů od průměru všech případů
- Průměrná kvadratická odchylka měření od aritmetického průměru

- **Variační rozpětí (range)**

- Rozdíl mezi nejnižší a nejvyšší hodnotou
- U nominálních dat logicky jen min a max
- Lepší: **interkvartilové rozpětí** – mezi 75tým a 25tým kvantilem

- **Směrodatná odchylka**

- Měří rozptýlenost kolem průměru

10

Deskriptivní statistika

Míry rozptýlenosti
založené na empirických kvantilech

- **Kvantil**

- Hodnota pod níž leží definovaná část údajů

- **Percentily 10, 90 (percentiles)**

- Procento případů, s hodnotami, které jsou nad a pod

- **Dolní a horní kvartil (quartiles)**

- Kvartily rozdělují vzorek do 4 skupin
- 25tý, 50tý, 75tý percentil

11

Deskriptivní statistika

Míry špičatosti a šikmosti

- **Šikmost (skewness)**

- Zešikmenost, nesymetrie dat

- **Špičatost (kurtosis)**

- Odchylka špičatosti od normálního rozdělení

12

Krabicový graf

- Popis pomocí pěti hodnot
- Zachytíme i odlehle hodnoty (outliers)

