

7. SEMINÁŘ

DESKRIPTIVNÍ STATISTIKA

Oblasti využití statistiky v medicíně

- **Zvládání variability**
 - Variabilita: biologická, podmínek, měřících přístrojů - hodnocení variability, variabilita náhodná x nenáhodná
- **Diagnostika nemocí a identifikace zdravotních problémů společnosti**
 - Pravděpodobnostní závěry na základě mnoha údajů z předchozích obdobných případů (popis příznaků nemoci x počátek thalidomidové aféry)
- **Prognóza léčby a odhad přínosu zdravotnických programů**
 - Pravděpodobnostní odhad dalšího průběhu léčby (vychází z minulých zkušeností podobnými případy)
 - Také o aplikacích populačních zdr. opatřeních se vedou záznamy, které umožňují odhadovat úspěšnost příštích opatření

Oblasti využití statistiky v medicíně

- **Výběr vhodného medicínského postupu**
 - Dřívější zkušenosti + klinické zkoušky + další důležité aspekty dané metody (ekon. náklady, riziko pro společnost)
- **Řízení systému péče o zdraví**
 - Využívání soustavy rutinních statistik doplňovaných o výběrová šetření
 - velikost a struktura populace,
 - informace o populačních procesech – rození, umírání, migrace,
 - zdravotní stav populace,
 - životní prostředí,
 - životní styl,
 - zdravotnický systém

Počátky - popisná statistika

- **Statistika jako popis státu:**
 - **Popis a soupis** zemědělského, hospodářského a politického stavu země a obyvatelstva
 - **Politická aritmetika** – zachycení vývoje obyvatelstva
 - **Vyčerpávající šetření** – zachycení veškerého obyvatelstva pomocí sčítání lidu a vedení podrobných záznamů o demografických, geografických a hospodářských jevech

Moderní (induktivní) statistika

- 30. léta 20. století – rozvoj **teorie pravděpodobnosti** a revoluce ve statistice
- **Výběrová šetření** – nové možnosti:
 - nezjišťuje se každý jednotlivý detail – rozvoj metod umožňujících tvořit závěry o celku na základě výběrových šetření.
 - hlubší analýza výběrového souboru,
 - zkoumání mnoha dosud nezkoumaných jevů.

Statistika – základní pojmy

Statistika jako vědní obor

- Jejím předmětem jsou **hromadné jevy**
 - Vlastnosti, znaky a události, které se vyskytují ve velkém množství.
- Zabývá se **sběrem, popisem a analýzou dat.**
- **Data**
 - zjištěné (naměřené) hodnoty určitých vlastností
 - hodnoty jednotlivých vlastností se vyznačují **variabilitou**
- **Variabilita dat**
 - Důsledek působení velkého množství drobných **NÁHODNÝCH** vlivů, z nichž každý výslednou hodnotu sledované vlastnosti ovlivňuje jen nepatrně.

Náhoda ve statistice

- **Přirozený jev**, který lze zkoumat exaktními metodami **teorie pravděpodobnosti**.
- Má svoje **zákonitosti**, jsou-li sledované vlastnosti určovány pouze náhodnými vlivy, podléhají zákonitostem **náhody**.
- Pokud zjištěné údaje neodpovídají těmto zákonitostem, nezpůsobuje rozdíly v hodnotách vlastnosti **náhoda**, ale **systematické působení** nějakého faktoru.

Induktivní a deduktivní úvaha

Aplikace statistických metod se váže ke dvěma typům uvažování:

Deduktivní úvaha: využívání obecných znalostí k rozhodování v jednotlivých případech

- Obecný popis nemoci – stanovení diagnózy u konkrétního pacienta

Induktivní úvaha: zobecnění poznatků z jednotlivých případů na všechny možné případy

- Sledovaní pacienti na lék reagovali příznivě - zobecnění - reakce všech pacientů budou příznivé
- Každý závěr (rozhodnutí) je provázen určitou **nejistotou**, statistika umí tuto nejistotu vyčíslit.

Základní a výběrový soubor

- **Základní soubor** – soubor jednotek, jejichž vlastnosti chceme poznat
- **Výběrový soubor** – ta část souboru, u které skutečně probíhá statistické šetření

Výběrový soubor

- Vypovídá jen o tom základním souboru, ze kterého byl odvozen.
- **Reprezentativnost** výběrového souboru (dobře reprezentuje všechny známé i neznámé charakteristiky základního souboru).
- **Náhodný výběr** – je získán postupem, kdy každý prvek základního souboru má na začátku výběru stejnou naději být vybrán.

Metody náhodného výběru

- 1. Prostý náhodný výběr** – losováním, pomocí tabulek (generátoru) náhodných čísel
- 2. Náhodný výběr mechanický (systematický)** – vytvoříme seznam jednotek, ze kterého vybereme např. každou stou osobu, přičemž první osobu vybereme metodou prostého náhodného výběru.
- 3. Náhodný výběr oblastní (stratifikovaný)** – rozdělení do oblastí (strat) – např. rozdělíme soubor na muže a ženy a vybíráme prostým NV takový počet mužů a žen, aby byl zachován poměr mužů a žen v základním souboru.

Etapy statistického šetření

- 1) Plán šetření (cíl, studium literatury, statistická jednotka, základní soubor, sledované znaky, způsob a přesnost měření, forma záznamu, způsob a rozsah výběru, statistické zpracování, pracovní a testované hypotézy, přínos a náklady výzkumu, pilotní studie)
- 2) Sběr dat (dodržování pravidel těmi, kdo sběr dat provádějí)
- 3) Popis a technické zpracování (deskriptivní statistika)
- 4) Rozbory a závěry (induktivní statistika)

Dvě základní oblasti statistiky

- **Popisná statistika**
- **Induktivní statistika**

Deskriptivní statistika - popis dat

Deskriptivní statistika

1. Statistické třídění
2. Prezentace (vizualizace) dat
3. Statistické charakteristiky

Statistické třídění

Třídění

- zpřehlednění souboru dat
- popis struktury souboru
- rozložení četností

Způsob třídění závisí na typu veličiny.

Výsledky třídění uvádíme v tabulkách – tzv. tabulky rozdělení četností.

Třídění: typy znaků (veličin)

50ti-letý muž, měří 170 cm, váží 90 kg, vitální kapacitu plic má 4,62 l, prodělal zánětlivé plicní onemocnění, má středoškolské vzdělání a je nekuřák.

Charakteristika → znak (veličina, proměnná) → obor hodnot znaku

Typy znaků:

Kvalitativní (kategoriální)

- Nominální

- Alternativní (test+, test-)
- Množné

- Pořadové (ordinální)

Kvantitativní (intervalové)

- Spojité

- Diskrétní

Třídění kvalitativních veličin

- Kategorie třídění jsou předem dány.
- Jde o **výčet všech hodnot**, kterých může sledovaný znak nabývat (např. znak pohlaví - hodnoty znaku: muž, žena; vzdělání – hodnoty znaku: ZŠ, SŠ, VŠ).

Třídění: jednostupňové a vícestupňové

- Třídění podle **jednoho znaku**.
- Třídění podle **dvou a více znaků** současně.

Jednostup. třídění

	CELKEM
Nekuřák	120
Slabý kuřák	60
Silný kuřák	20
CELKEM	200

Dvoustupňové třídění

	ZŠ	SŠ	VŠ	CELKEM
Nekuřák	20	40	60	120
Slabý kuřák	35	10	15	60
Silný kuřák	12	7	1	20
CELKEM	67	57	76	200

Kvalitativní znaky

Tab. 1.: Rozložení souboru 200 mužů podle kuřáckých zvyklostí a vzdělání (absolutní počty a procenta)

	ZŠ		SŠ		VŠ		CELKEM	
Nekuřák	20	<i>16,7</i>	40	<i>33,3</i>	60	<i>50,0</i>	120	<i>100,0</i>
	<i>29,9</i>		<i>70,1</i>		<i>78,9</i>		<i>60,0</i>	
Slabý kuřák	35	<i>58,3</i>	10	<i>16,7</i>	15	<i>25,0</i>	60	<i>100,0</i>
	<i>52,2</i>		<i>17,5</i>		<i>19,7</i>		<i>30,0</i>	
Silný kuřák	12	<i>60,0</i>	7	<i>35,0</i>	1	<i>5</i>	20	<i>100,0</i>
	<i>17,9</i>		<i>12,3</i>		<i>1,4</i>		<i>10,0</i>	
CELKEM	67	<i>33,5</i>	57	<i>28,5</i>	76	<i>38,0</i>	200	<i>100,0</i>
	<i>100,0</i>		<i>100,0</i>		<i>100,0</i>		<i>100,0</i>	

Třídění kvantitativních veličin

- Třídy vytváříme teprve na základě získaných dat
- Dochází k **redukci dat** ve prospěch přehlednosti
- **Vytváření intervalů:**
 - počet intervalů
 - délka intervalů
 - hranice intervalů
- **Musíme brát v úvahu:**
 - počet dat (velikost souboru)
 - přesnost měření
 - cíl třídění

Prezentace dat v tabulkách

- Výsledky třídění uvádíme v tabulkách – tzv. **tabulky rozdělení četností**.
- **Četnosti:**
 - absolutní
 - relativní
 - kumulativní absolutní
 - kumulativní relativní

Kvantitativní znaky

Tab. 1.: Rozložení vitální kapacity plic u 200 mužů ve věku 40-50 let (v litrech)

interval	střed	četnost		kumulativní četnost	
		absolut.	relat. %	absolut.	relat. %
3,00 – 3,39	3,20	6	3,0	6	3,0
3,40 – 3,79	3,60	9	4,5	15	7,5
3,80 – 4,19	4,00	16	8,0	31	15,5
4,20 – 4,59	4,40	36	18,0	67	33,5
4,60 – 4,99	4,80	52	26,0	119	59,5
5,00 – 5,39	5,20	44	22,0	163	81,5
5,40 – 5,79	5,60	22	11,0	185	92,5
5,80 – 6,19	6,00	11	5,5	196	98,0
6,20 – 6,59	6,40	4	2,0	200	100,0
celkem		200	100,0	200	100,0

Třídění kvantitativních veličin

- Stejně dlouhé intervaly
- Nestejně dlouhé intervaly

Věk	Abs. četnost
1	18
2	43
3	50
4	60
5	36
6	25
7	22
8	21
9	6
10	5
11-15	14
16-20	3
Celkem	303

**Prezentace (vizualizace)
dat**

Prezentace dat

Prezentace dat v grafech

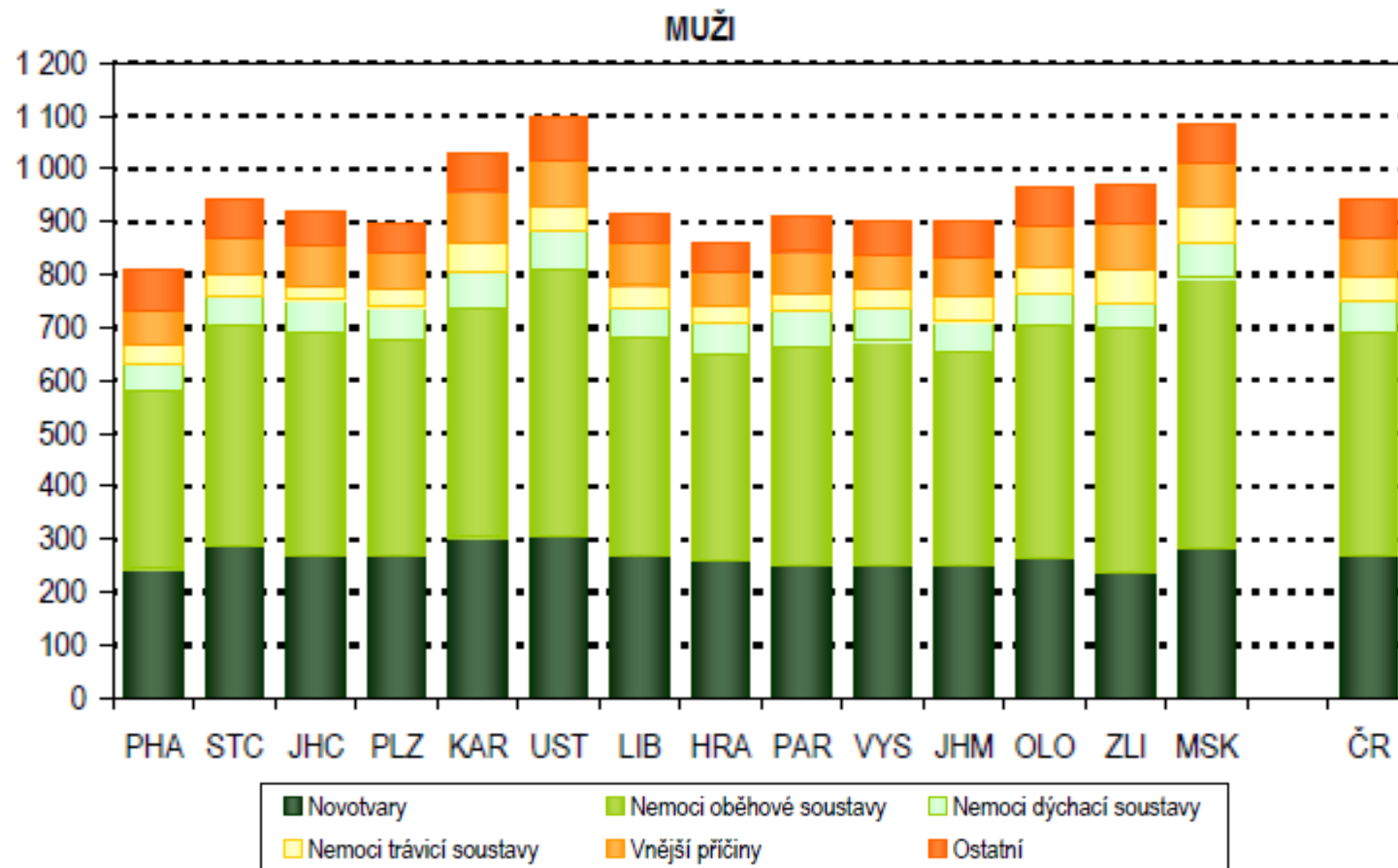
1. **Četnost** jednotlivých kategorií (tříd, intervalů)
2. **Tvar rozložení četností**
 - Symetrické x asymetrické
 - Jednovrcholové x dvouvrcholové
 - Výběr vhodných statistických ukazatelů
 - Výběr vhodného teoretického rozložení četností při odhadu parametrů a testování hypotéz

Prezentace dat v grafech

- **Kvalitativní veličiny**
 - Sloupcový graf (sloupce oddělené mezerou)
 - Výsečový graf (struktura)
 - Kartogram (regionální srovnání)

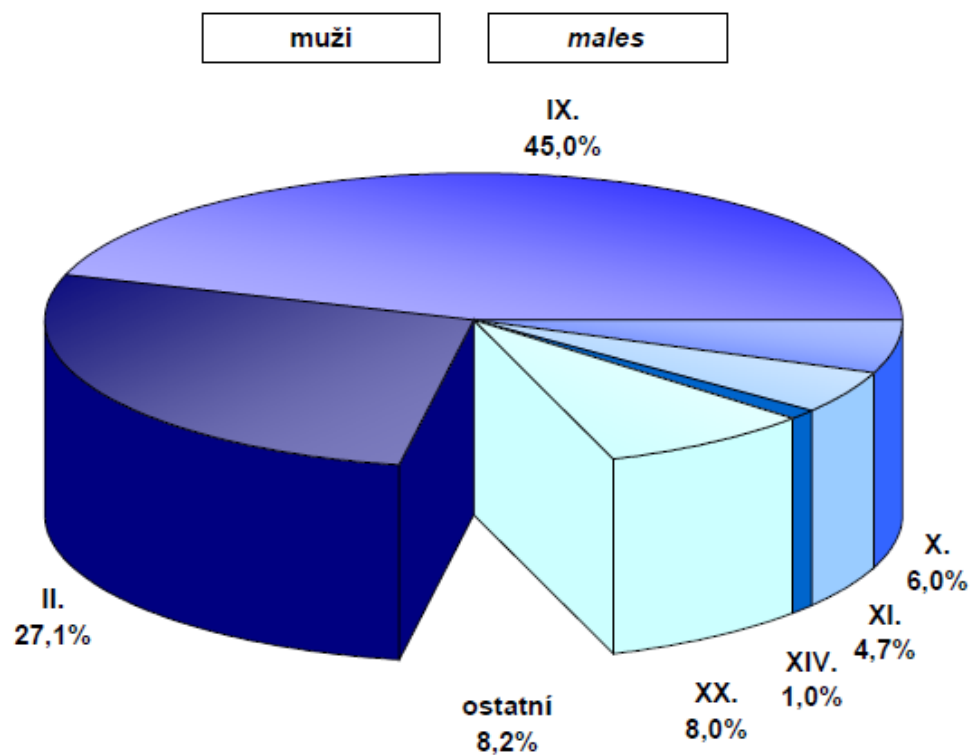
Sloupcový graf

2. Standardizovaná úmrtnost podle příčin smrti a kraje bydliště (na 100 000 osob)



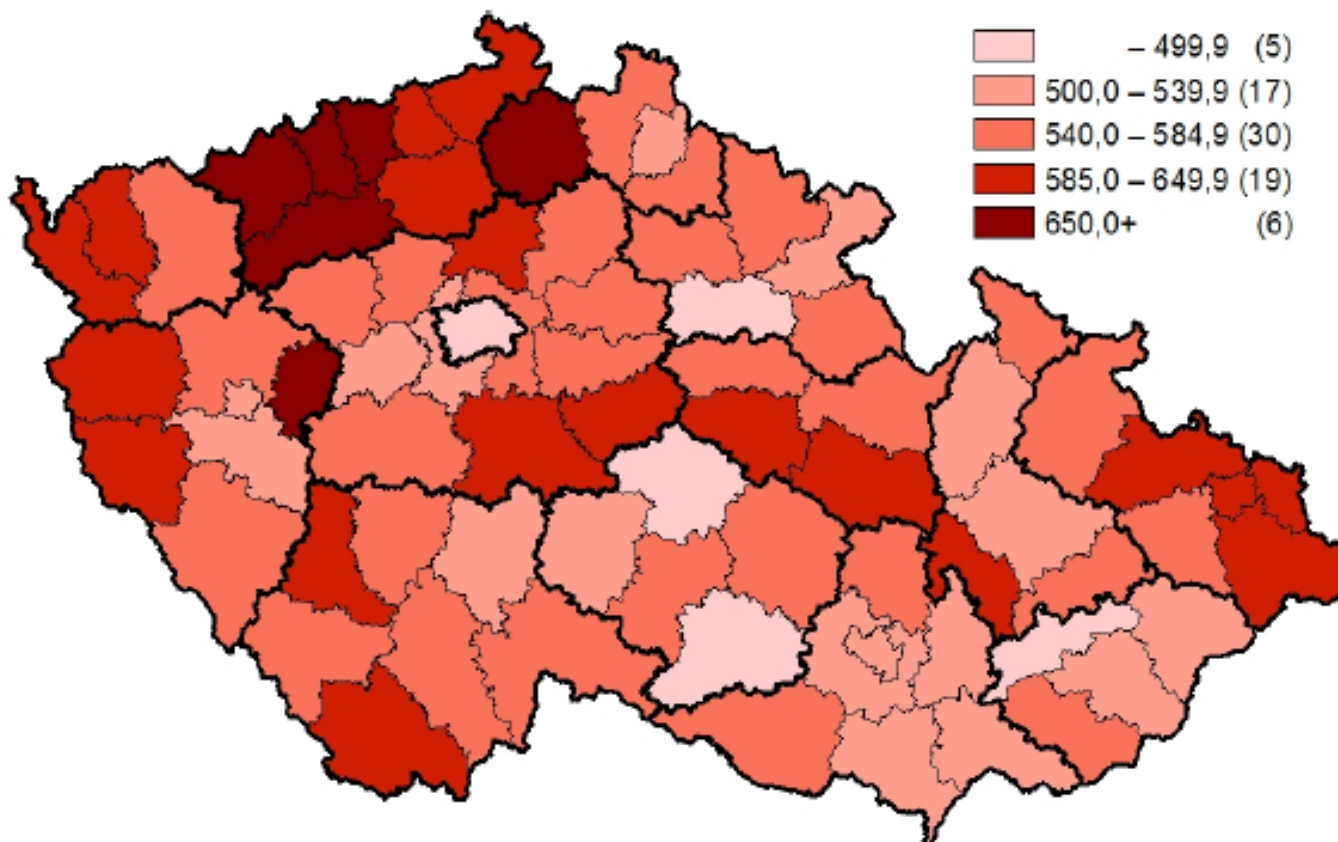
Výsečový graf

Standardizovaná úmrtnost podle příčin smrti (MKN-10)
Standardized mortality rate by cause of death (ICD-10)



Kartogram

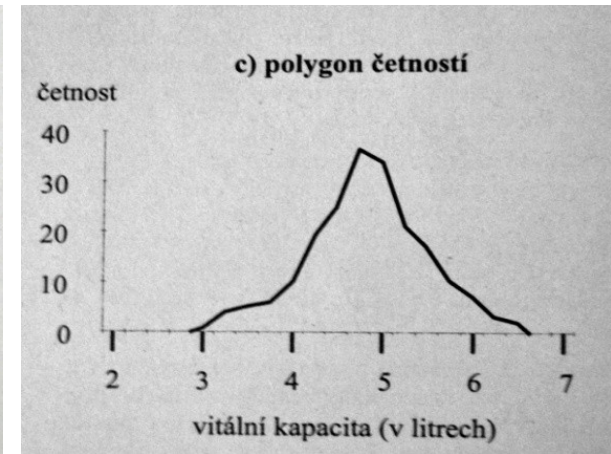
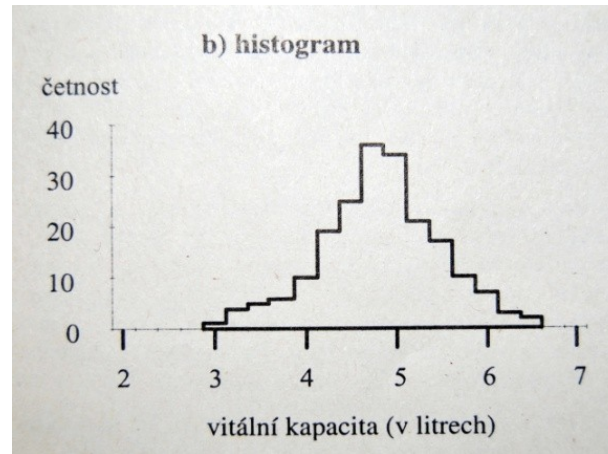
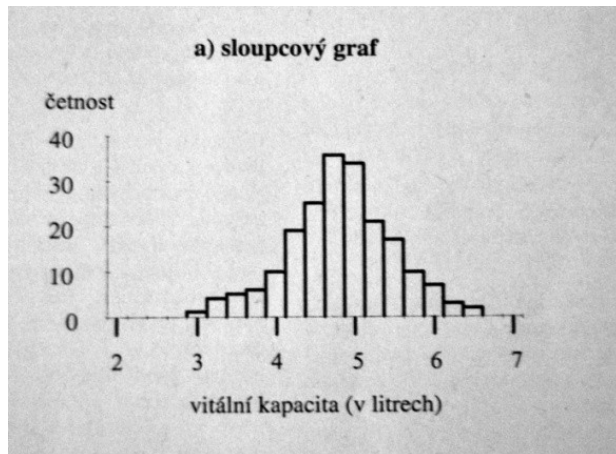
7. Standardizovaná úmrtnost žen (na 100 000 osob)



Prezentace dat v grafech

- **Kvantitativní veličiny**
 - Sloupcový graf
 - Histogram
 - Polygon četností

Prezentace kvantitat. dat



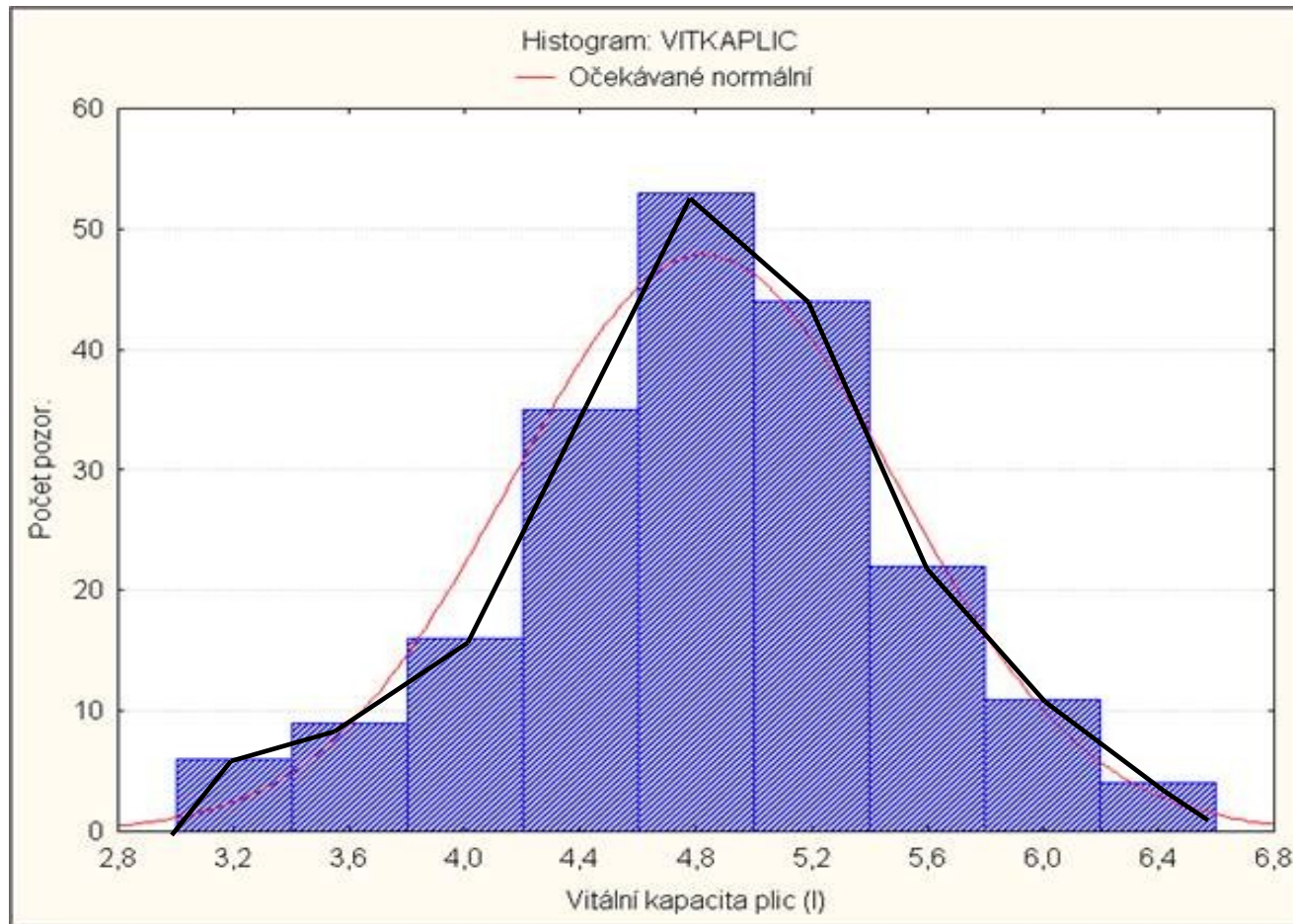
osa **X** : naměřené hodnoty sledované veličiny (VKP v l)

osa **Y** : četnost intervalů (abs. nebo v %)

Tvar rozložení četností:

- Symetrické x asymetrické
- Jednovrcholové x vícevrcholové
- Podoba s teoretickými modely rozložení četností

Prezentace kvantitativních dat



Statistické ukazatele

Statistické ukazatele

- a) **Relativní ukazatele** - (viz rutinní statistiky: ukazatele frekvence, ukazatele struktury, indexy)
- b) **Střední hodnoty (ukazatele polohy)**
- c) **Ukazatele variability**

VOLBA VHODNÝCH UKAZATELŮ POLOHY
A VARIABILITY ZÁVISÍ **NA TYPU** SLEDOVANÉHO
ZNAKU (nominální x ordinální x intervalový) **A NA TVARU**
ROZLOŽENÍ ČETNOSTÍ (symetrické x asymetrické).

Ukazatele polohy

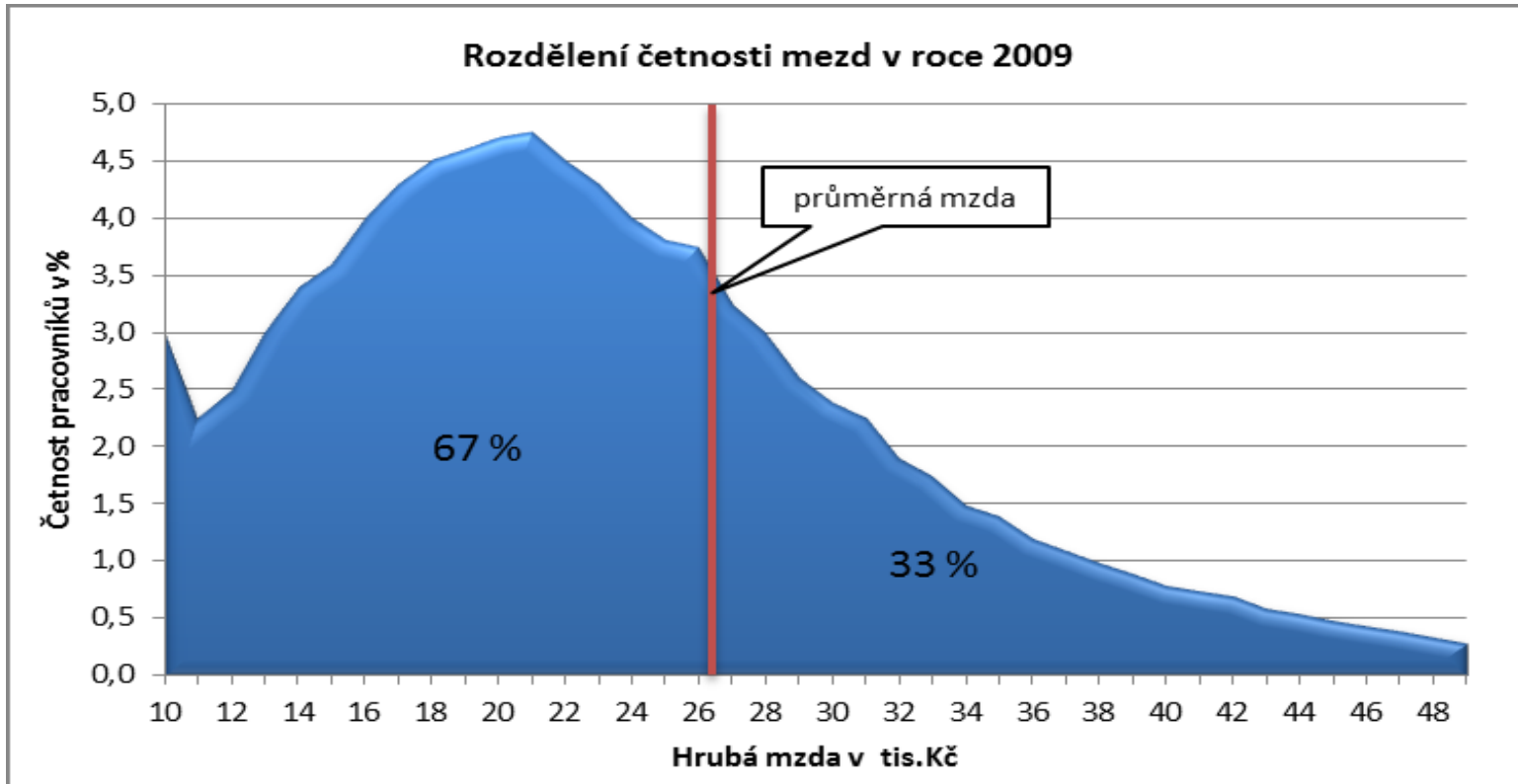
Ukazatele polohy

- **Aritmetický průměr (m):**
 - sečteme pozorované hodnoty a vydělíme je počtem sledovaných jednotek
- **Medián (m_e):**
 - hodnota, která je právě uprostřed všech pozorování, která jsme seřadili podle velikosti
- **Modus (m_o):**
 - třída (kategorie) s nejvyšší četností
- **Kvantil (percentil, decil, kvartil)**
 - pořadový ukazatel, obměna mediánu

Ukazatele polohy

- **Typ znaku:**
 - nominální: modus
 - ordinální: modus, medián, percentil (kvantil)
 - intervalové: modus, medián, percentil (kvantil), průměr
- **POZOR NA INTERPRETACI ARITMETICKÉHO PRŮMĚRU U ASYMETRICKÝCH ROZLOŽENÍ.**
- **ARITMETICKÝ PRŮMĚR JE CITLIVÝ NA VYCHÝLENÉ HODNOTY.**
- **VHODNĚJŠÍM UKAZATELEM POLOHY U ASYMETRICKÝCH ROZLOŽENÍ MŮŽE BÝT MEDIÁN.**

Ukazatele polohy



$$m = 26\,700 \quad m_0 = 20\,000 \quad m_e = 22\,000$$

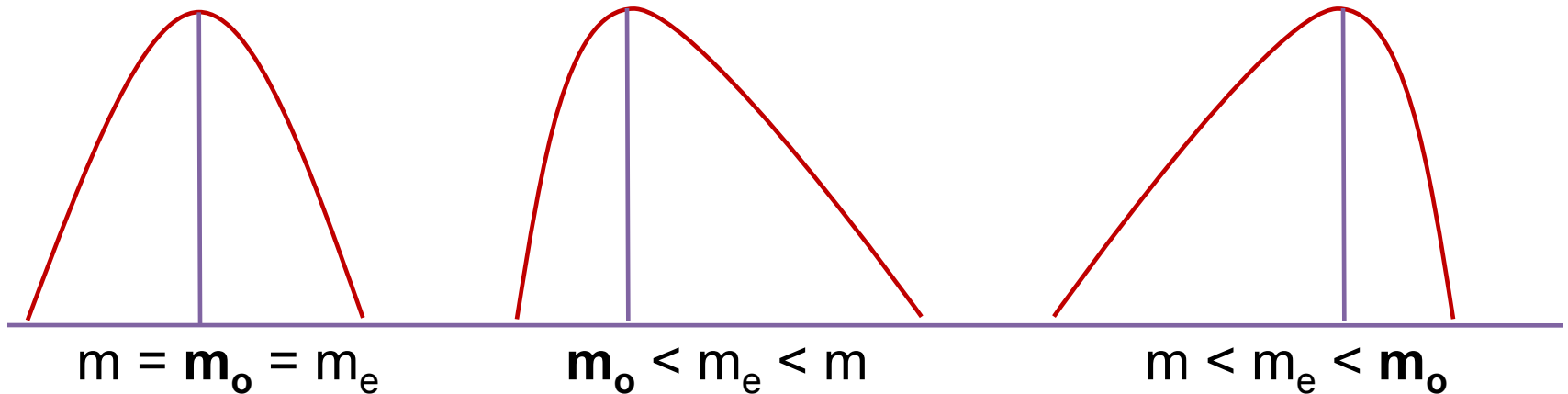
Ukazatele polohy

- Ukazatele polohy u symetrického a asymetrického rozložení

symetrické

pravostr. asym.

levostr. asym.



Ukazatele variability

Ukazatele variability

Proč nestačí ukazatele polohy k výstižnému popisu dat?

Př.

1. sk.: 3,08 4,42 5,05 5,67 6,59 **m = 4,96**

2. sk.: 4,86 4,90 4,91 5,03 5,11 **m = 4,96**

Obě skupiny mají stejný průměr, liší se ale kolísáním hodnot, tj. **VARIABILITOU**.

Ukazatele variability

Spolu se střední hodnotou by se měl vždy udávat příslušný ukazatel variability!

- **Rozpětí** (u malých souborů, kde $n \leq 10$)
- **Rozptyl - směrodatná odchylka (nejč.) - variační koeficient**
 - uvádějí se s aritmetickým průměrem (u symetrických rozdělení)
- **Kvantily** (percentily, decily, kvartily)
 - uvádějí se s modem či medián (asymetrický rozdělení)
 - lze je ale samozřejmě použít i s aritmetickým průměrem

Ukazatele variability

Rozpětí:

- max. - min. Pro $n \leq 10$

Rozptyl (s^2):

- Průměr čtverců odchylek aritmetického průměru od jednotlivých měření:

- $$s^2 = \frac{\sum(x_i - m)^2}{n-1}$$

1.sk.: 3,08 4,42 5,05 5,67 6,59 m = 4,96

$$\begin{array}{r} 4,96 - 3,08 = 1,88 \\ - 4,42 = 0,54 \\ - 5,05 = -0,09 \\ - 5,67 = -0,71 \\ - 6,59 = -1,63 \end{array} \quad \begin{array}{r} 3,53 \\ 0,29 \\ 0,01 \\ 0,50 \\ 2,66 \end{array}$$

$$s^2 = 6,99/4 = 1,75$$

- Udává se ve čtvercích jednotek sledovaného znaku, tj. zde v litrech²

Ukazatele variability

Směrodatná odchylka (s):

- Odmocněný rozptyl, $s = \sqrt{s^2}$
 - Ukazatel variability udávaný ve stejných jednotkách jako sledovaný znak
 - Vypovídá o tom, o kolik se většina hodnot sledovaného znaku odchyluje od průměru
 - $m \pm 1s$ interval, ve kterém leží 68% naměřených hodnot
 - $m \pm 2s$ interval, ve kterém leží 95% naměřených hodnot
 - $m \pm 3s$ interval, ve kterém leží 99% naměřených hodnot
- **Příklad:** vypočítejte, v jakém intervalu leží 68% hodnot VKP v našem souboru 200 mužů

Ukazatele variability

Variační koeficient (v.k.)

- Relativní ukazatel variability
- Udává, jaký podíl tvoří směrodatná odchylka z průměru.

Ukazatele variability

Variační koeficient (v.k.)

- Slouží ke srovnání variability 2 souborů, jejichž průměry se značně liší

Př.: VKP u mužů a u žen

M: $m = 4,80$ $s = 0,66$ $v.k. = 13,8\%$

Ž: $m = 3,90$ $s = 0,42$ $v.k. = 10,8\%$

- Slouží ke srovnání variability znaků uváděných v různých jednotkách

Př.: VKP (l), výška (cm) a hmotnost mužů (kg)

VKP: $m = 4,80$ $s = 0,66$ $v.k. = 13,8\%$

Výška: $m = 178$ $s = 4$ $v.k. = 2,2\%$

Hmotnost: $m = 82$ $s = 6$ $v.k. = 7,3\%$

Ukazatele variability

Kvantily – percentily, decily, kvartily

- Kvantily dělí soubor dat uspořádaných podle velikosti na části obsahující stejný podíl z celkového počtu jednotek
- Variabilita se určuje pomocí intervalu, ve kterém se pohybuje nejčastěji 80% ($P_{10} - P_{90}$) nebo 50% ($P_{25} - P_{75}$) pozorování.
- Postup výpočtu:
 1. Určíme hodnotu pozorování, které představuje 10. percentil = dolní hranice intervalu
 2. Určíme hodnotu pozorování, které představuje 90. percentil = horní hranice intervalu
- **Vhodné ukazatele variability pro asymetrická rozložení**

Příklad

Porodní délka 5 novorozenců v cm:

49, 50, 50, 51, 53

Vypočítejte:

- Aritmetický průměr
- Rozptyl
- Směrodatnou odchylku
- Variační koeficient