

Úvod do analýzy dat

Měření a zpracování dat (MDS)

Ondřej Přibyl

Ústav aplikované matematiky
ČVUT v Praze, Fakulta dopravní



Ondřej Přibyl K611MDS

strana 1

Obsah prezentace

- Měřené veličiny
- Chyby měření
- Vizualizace dat
- Další aspekty analýzy dat
- Hlavní kroky při analýze dat

Ondřej Přibyl K611MDS

strana 2

Diskuze

- Jaký je rozdíl mezi:
 - DATY,
 - INFORMACÍ a
 - ZNALOSTMI?
- Uved'te na příkladech.

Ondřej Přibyl K611MDS

Data, informace a znalosti

Data

Jakékoli vyjádření (reprezentace) skutečnosti, schopné přenosu, interpretace či zpracování. Účelem dat je přenášet a dále zpracovávat odraz skutečnosti. Jsou to jakékoli zaznamenané poznatky či fakta.

Informace

Data, která mají smysl (význam). Jsou to sdělitelné (komunikovatelné) znalosti. Je to údaj, ke kterému si člověk přiřadí význam.

Znalost

To, co jednotlivec ví po osvojení dat a informací a po jejich začlenění do souvislostí.

Účelem znalostí je porozumět skutečnosti. Jako média pro uchování znalostí slouží lidská paměť, organizace, dokument nebo počítač.

Moudrost



zdroj: <http://knowledge management.ic.cz/informaceznalosti.doc>

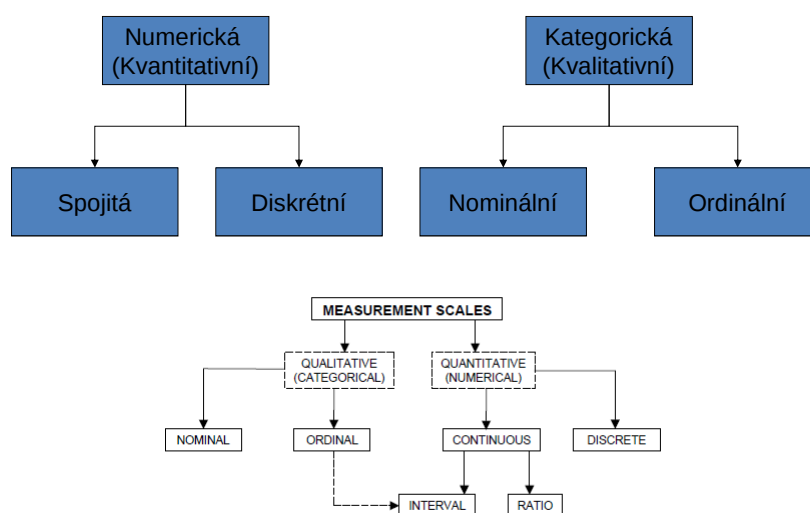
Ondřej Přibyl K611MDS

Diskuze

- Jaké znáte typy měřených dat?

Ondřej Příbyl K611MDS

Přehled typů měřených dat



Ondřej Příbyl K611MDS

Numerická data

Spojité data

- stat. znak, který může nabývat všech reálných hodnot v rámci konečného nebo nekonečného intervalu
- Příklady:
 - MTBF - doba do poruchy zařízení
 - Doba jízdy
 - Hmotnost vozidla

Diskrétní / Nespojité

- stat. znak který může nabývat v daném intervalu pouze izolovaných číselných hodnot
- zpravidla se jedná o přirozená čísla + 0, tedy $\{0, 1, 2, 3, \dots, n\}$
- Příklady:
 - Počet cest automobilem za týden
 - Počet dopravních nehod

Ondřej Přibyl K611MDS

Kategorická data

Nominální

- Nabývají konečného a nízkého počtu diskretních hodnot
- nelze nad nimi vytvořit uspořádání.
- Příklady:
 - Druhy dopravních prostředků
 - Barvy vozidel

Ordinální

- Od nominálních proměnných se liší v tom, že nad nimi lze vytvořit uspořádání.
- Příklady:
 - malý, střední, veliký
 - nikdy < občas < často < vždy
- **Binární** (speciální případ)
 - Nabývají hodnot 0 a 1

Marital status

1. Never married ☐
 2. Divorced ☐
 3. Widowed ☐

4. Married/Cohabiting ☐
 5. Separated ☐

Employee's performance

1. Excellent ☐
 2. Good ☐
 3. Average ☐

4. Poor ☐
 5. Very poor ☐

Ondřej Přibyl K611MDS

Příklady z dopravy (zatřídění a veličiny)

- Uvedte jednotky dané veličiny a klasifikujte ji dle typu
 - Intenzita dopravy
 - Obsazenost detektoru
 - Stupeň dopravy
 - Počet vozidel v domácnosti
 - Doba jízdy
 - Třídy vozidel
 - Hustota

Ondřej Přibyl K611MDS

Obsah prezentace

- Měřené veličiny
- **Chyby měření**
- Základní charakteristiky dat
- Vizualizace dat
- Další aspekty analýzy dat
- Hlavní kroky při analýze dat

Ondřej Přibyl K611MDS

strana 10

Kde vznikají chyby při měření dopravních dat?

Chyby...

Chyby zásadně ovlivňují měření, dělí se na náhodné (dynamická charakteristika) a systematické (statická Charakteristika)

- Chyba detektoru
 - Chyba měřicího zařízení
 - způsobena nedokonalostí měřicích přístrojů.
 - Chyba pozorovatele (chyby způsobené lidským faktorem)
 - do této kategorie patří nesprávná volba metody měření,
 - chybné zapojení přístrojů do obvodu,
 - nevhodná volba měřicího rozsahu,
 - chybné čtení údajů, atp.

Ondřej Příbyl K611MDS

Kde vznikají chyby při měření dopravních dat?

Chyby...

Chyby zásadně ovlivňují měření, dělí se na náhodné (dynamická charakteristika) a systematické (statická Charakteristika)

- Chyba přenosu
 - Chyba způsobená výpadkem v přenosové cestě
- Chyba metody
 - jejich příčinou jsou různá zjednodušení vztahů pro výpočet měřené veličiny, zjednodušení zapojení, vliv spotřeby měřicího přístroje na jeho údaj, atd.
 - Tyto chyby je obvykle možno vypočítat a výsledek měření podle nich korigovat.

Ondřej Příbyl K611MDS

Úvod do problematiky

- Každé měření v sobě obsahuje **nejistotu správnosti** výsledku.
- **Systematické chyby**
 - jsou statického rázu, zkreslují výsledek stejným, kontrolovatelným způsobem bez ohledu na počet provedených měření.
 - zdroji těchto chyb je omezená přesnost přístrojů, použitá metoda měření a osobní chyby.
 - do chyb způsobených omezenou přesností spadají např. aditivní a multiplikativní chyby.
- Náhodné chyby

Ondřej Přibyl K611MDS

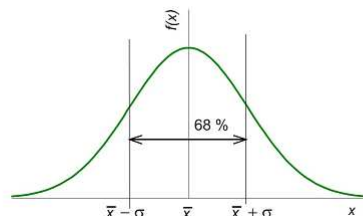
Úvod do problematiky

- Každé měření v sobě obsahuje **nejistotu správnosti** výsledku.
- Systematické chyby
- **Náhodné chyby**
 - jsou chyby, které se vyskytují zcela nepravidelně. Jejich výskyt je náhodný.
 - Náhodné chyby jsou způsobeny nekontrolovatelnými vlivy.
 - Tyto chyby nelze odstranit!
 - Zjistit je můžeme až při opakovaném měření.

Ondřej Přibyl K611MDS

Náhodné rozdělení chyb

- Normální Gaussovo rozdělení, střední hodnota odpovídá nejpravděpodobnější hodnotě opakovaného měření.

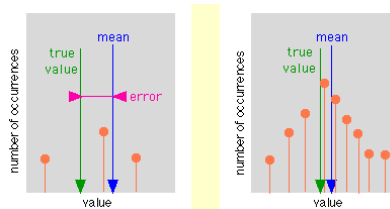


hustota pravděpodobnosti

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

- Výsledky platí pro velké množství měření ($n \rightarrow \infty$).

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$



Ondřej Přibyl K611MDS

„Přesnost“ versus „správnost“

Přesnost (precision)

- rozmezí statistické nejistoty výsledků
- Souvisí s náhodnými chybami
- Odpovídá reprodukovatelnosti měření.
- Vyjadřuje se jako rozptyl naměřených výsledků kolem průměru z n naměřených hodnot.
- Lze odhadnout statisticky

Správnost (accuracy)

- udává průměrnou odlehlost (vzdálenost) výsledků měření od skutečné hodnoty.
- Souvisí se systematickými chybami.
- Odpovídá odchýlení měření od teoretické hodnoty.
- Nelze ji odhadnout, je nutno ji stanovit s využitím standardů nebo měřením na více přístrojích.

Ondřej Přibyl K611MDS

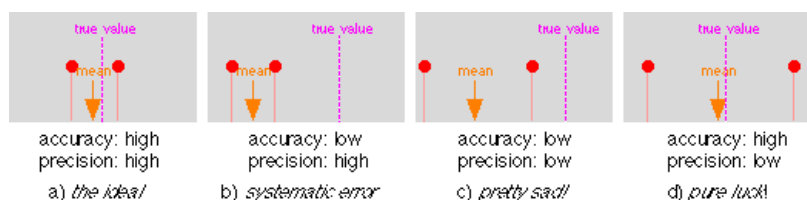
„Přesnost“ versus „správnost“

Přesnost (precision)

- rozmezí statistické nejistoty výsledků
- Přesnost přístroje lze odhadnout na základě statistické analýzy.

Správnost (accuracy)

- udává průměrnou odlehlost výsledků měření od skutečné hodnoty.
- Nelze ji odhadnout, je nutno ji stanovit s využitím standardů nebo měřením na více přístrojích.

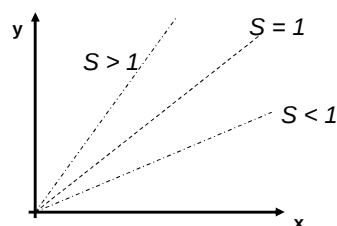


Ondřej Přibyl K611MDS

Citlivost (sensitivity) měřicího přístroje

- schopnost reagovat za stanovených podmínek na požadovanou změnu hodnoty měřené vstupní veličiny.
- Vyjadřuje se jako podíl změny přístrojového údaje (výstupní veličiny) k požadované změně měřené (vstupní) veličiny, která změnu údaje vyvolává.
- Na přístrojích s ručkovým ukazatelem je to velikost dílku stupnice, který odpovídá velikosti změny měřené veličiny,
- u digitálních přístrojů je to počet desetinných míst, s jakým je hodnota měřené veličiny udávána.

$$S = \Delta y / \Delta x$$



Ondřej Přibyl K611MDS

Diskuze

- „Při měření intenzity dopravy byla naměřena chyba 5 vozidel,,
– Je to hodně nebo málo?

Ondřej Přibyl K611MDS

Chyby měření

- Absolutní chyba měření
- y_N ... naměřená hodnota
- y_S ... správná hodnota

$$\Delta_y = y_N - y_S$$

- Relativní chyba měření

$$\delta_y = \frac{\Delta_y}{y_S}$$

- Relativní chyba senzoru
- Poměr maximální absolutní chyby měření vůči rozsahu hodnot měřené veličiny

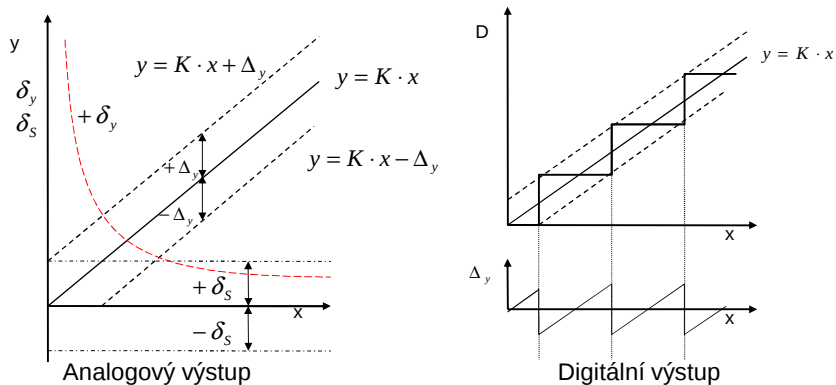
$$\delta_s = \frac{(\Delta_y)_{\max}}{y_{\max} - y_{\min}}$$

Ondřej Přibyl K611MDS

Chyby měření

Aditivní chyba měření

- Způsobena posunem jmenovité lineární charakteristiky
- Absolutní chyba měření je konstantní
- Relativní chyba měření závisí hyperbolicky na x

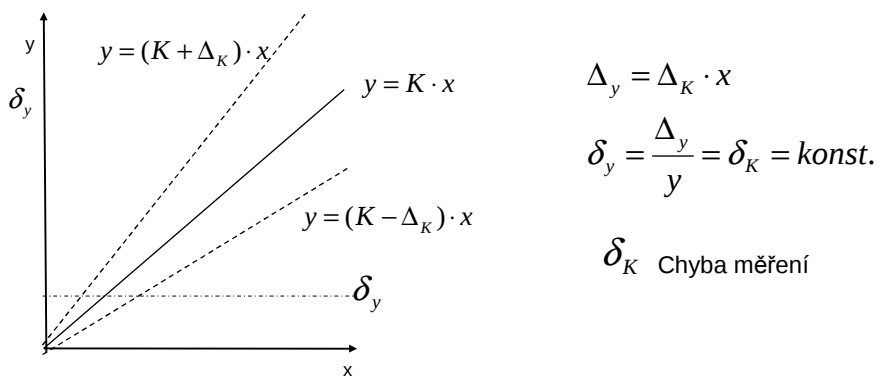


Ondřej Příbyl K611MDS

Chyby měření

Multiplikativní chyba měření

- Je ekvivalentní změně citlivosti senzoru
- Absolutní chyba je závislá na hodnotě měřené veličiny
- Relativní chyba měření je konstantní



$$\Delta_y = \Delta_K \cdot x$$

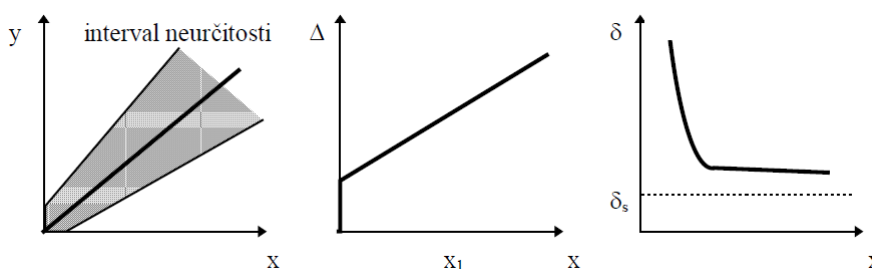
$$\delta_y = \frac{\Delta_y}{y} = \delta_K = \text{konst.}$$

δ_K Chyba měření

Ondřej Příbyl K611MDS

Chyby měření

- Kombinovaná chyba měření
- Kombinace aditivní a multiplikativní chyby



Ondřej Příbyl K611MDS

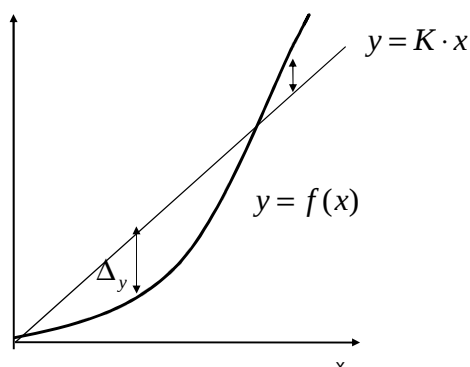
Chyby měření

Chyba linearity

- Dána odchylkou od ideální lineární charakteristiky
- je udávána vztahem:

$$\delta_L = \left(\frac{y_n - y_L}{y_{\max} - y_{\min}} \right)_{\max}$$

- kde y_L je definována ideální funkcí $y = y_0 + Kx$,
- parametr K lze odhadnout pomocí lineární regrese.

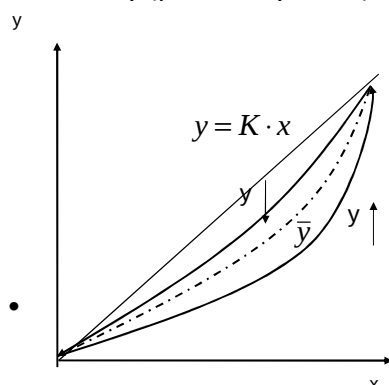


Ondřej Příbyl K611MDS

Chyby měření

Chyba hystereze

- Vyjadřuje závislost měření na předchozích stavech měřené veličiny (paměťový efekt)



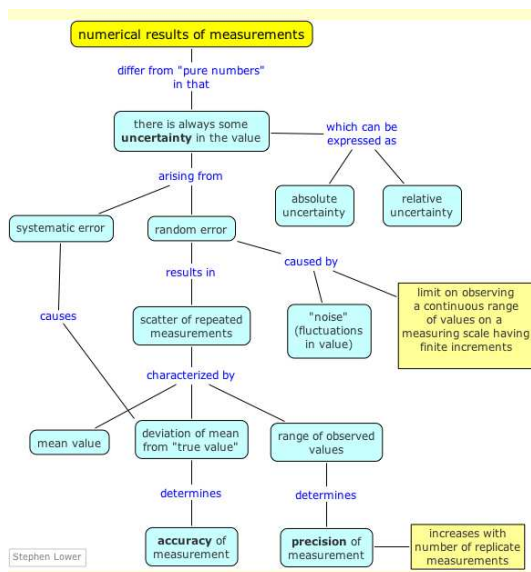
$$\delta_s = \left(\frac{y \downarrow - y \uparrow}{y_{\max}} \right)_{\max} = \left(\frac{\Delta_{yH}}{y_{\max}} \right)_{\max}$$

$$\delta_s = \left(\frac{y - \bar{y}}{y_{\max}} \right)_{\max}$$

kde $\bar{y}$ je střední hodnota
vzestupné a klesající závislosti y .

Ondřej Příbyl K611MDS

Přehled



Stephen Lower

Ondřej Příbyl K611MDS

Obsah prezentace

- Měřené veličiny
- Chyby měření
- **Základní charakteristiky dat**
- Vizualizace dat
- Další aspekty analýzy dat
- Hlavní kroky při analýze dat

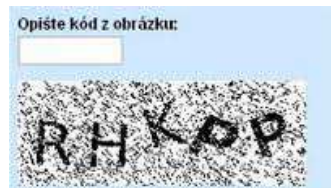
Ondřej Přibyl K611MDS

Co je to Průzkumová analýza dat?

- První krok při analýze nových dat
- Kombinace grafických, semigrafických a číselných tabulkových postupů, které podají základní informace o vlastnostech souboru

Cíle

- získat přehled o datech, jejich kvalitě a vlastnostech
- vybrat vhodný nástroj pro předzpracování dat
- využít lidských schopností dříve než je vybrán automatický nástroj (Lidé jsou schopni rozpoznat charakteristiky dat, které nemohou být rozpoznány (nebo jen velmi obtížně) automatickými systémy)



Ondřej Přibyl K611MDS

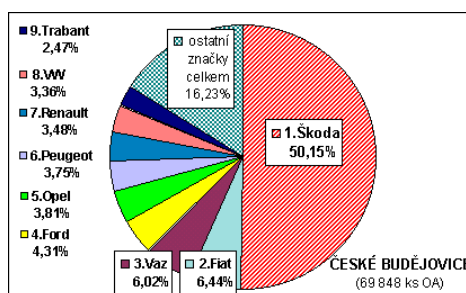
Frekvence atributu a rozsah hodnot

- **Frekvence atributu**

- Procentuální vyjádření četnosti výskytu dané hodnoty v datech
- Na příklad v ČR je frekvence výskytu vozidel Škoda 50,15%

- **Rozsah hodnot**

- Rozdíl mezi maximální a minimální hodnotou daného atributu



Ondřej Přibyl K611MDS

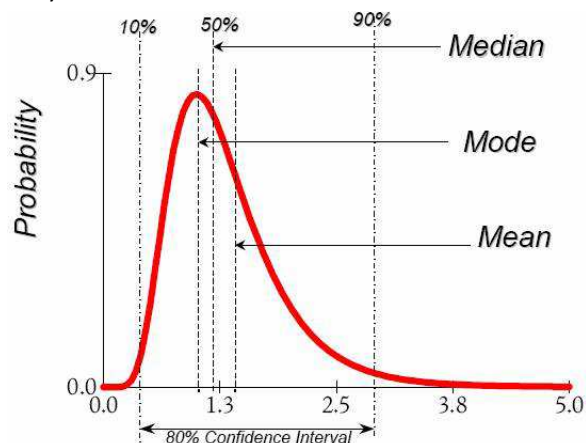
Modus, Medián a Aritmetický průměr atributu

- Aritmetický průměr (mean)
 - statistická veličina, která v vyjadřuje typickou hodnotu
 - součet všech hodnot vydělený jejich počtem.
- Modus atributu (mode)
 - **nejčastější** hodnota v daném statistickém souboru
 - je to hodnota znaku s největší relativní četností.
 - určení předpokládá rozřídění souboru podle **obměn** znaku.
- Medián (median)
 - hodnota, jež dělí řadu podle velikosti seřazených výsledků na dvě stejně početné poloviny.
 - Platí, že 50 % hodnot je menších nebo rovných a 50 % hodnot je větších nebo rovných mediánu.

Ondřej Přibyl K611MDS

Modus, Medián a Aritmetický průměr atributu

- Aritmetický průměr (mean)
- Modus atributu (mode)
- Medián (median)



Ondřej Přibyl

K611MDS

strana 31

Příklad

- Nalezni
 - Modus,
 - Medián
 - Aritmetický průměr (mean)
 - Rozsah hodnot a
 - Frekvenci výskytu hodnoty 13
- pro následující hodnoty: 13, 18, 13, 14, 13, 16, 14, 21, 13

Ondřej Přibyl

K611MDS

Příklad

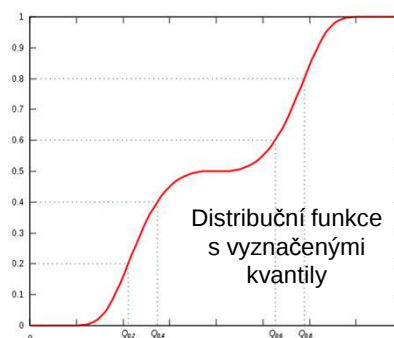
Řešení:

- Aritmetický průměr:
– $(13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13) \div 9 = 15$
- Medián (uspořádej vzestupně): 13, 13, 13, 13, 14, 14, 16, 18, 21
– Z devíti čísel je střední $(9 + 1) \div 2 = 5$ té číslo : 14
- Modus: nejčastěji se opakující číslo: 13
- Rozsah hodnot: $21 - 13 = 8$
- Frekvence výskytu hodnoty 13: $4 / 9 * 100 = 44,4\%$

Ondřej Přibyl K611MDS

Kvantily

- dělí soubor seřazených hodnot na několik stejně velkých částí.
- **Medián** - kvantil $Q_{0,5}$.
– Kvantil rozděluje statistický soubor na dvě stejně početné
- **Kvartil** (rozdělují statistický soubor na čtvrtiny.)
– 25 % prvků má hodnoty menší než dolní kvartil $Q_{0,25}$ a
– 75 % prvků hodnoty menší než horní kvartil $Q_{0,75}$
- **Percentil**
– Percentil dělí statistický soubor na setiny. Jako k -tý percentil označujeme $Q_k / 100$.



Ondřej Přibyl K611MDS

Obsah prezentace

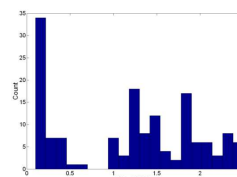
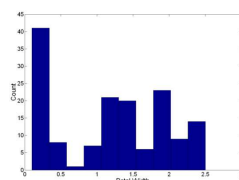
- Měřené veličiny
- Chyby měření
- Základní charakteristiky dat
- **Vizualizace dat**
- Další aspekty analýzy dat
- Hlavní kroky při analýze dat

Vizualizace

- převedení dat do vizuální či tabulkové podoby pro potřeby analýzy dat
- velmi silným nástrojem pro **průzkumovou analýzu** dat.
 - Lidé mají velkou schopnost analyzovat velké množství dat prezentované vizuálně
 - Je možné identifikovat obecné trendy a struktury
 - Je možné identifikovat obecné outliery
- Techniky:
 - Histogram
 - Box plot
 - Korelační diagram

Vizualizační techniky: Histogram

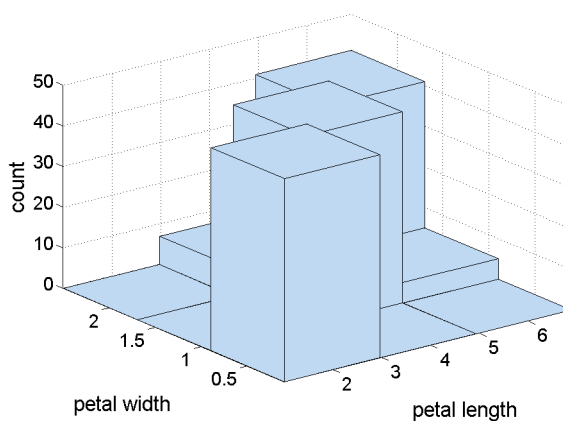
- Histogram
 - Rozdělí hodnoty do intervalů a zobrazí jejich četnosti
 - Výška sloupce udává počet objektů v daném intervalu
 - říká zda je soubor homogenní, nebo zda se rozpadá do dílčích menších podsouborů
 - jen jedna nejčetnější hodnota (homogenní soubor)
 - více hodnot s většími četnostmi
 - někdy lze zjistit přítomnost extrémních výchylek v datech



Ondřej Přibyl K611MDS

Dvoudimensionální Histogram

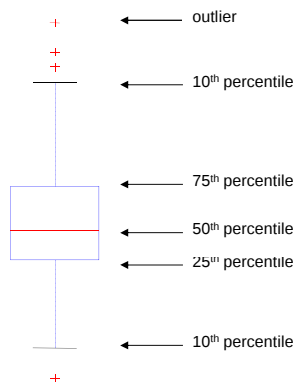
- Zobrazuje společné rozdělení dvou atributů



Ondřej Přibyl K611MDS

Vizualizační techniky: Box Plots

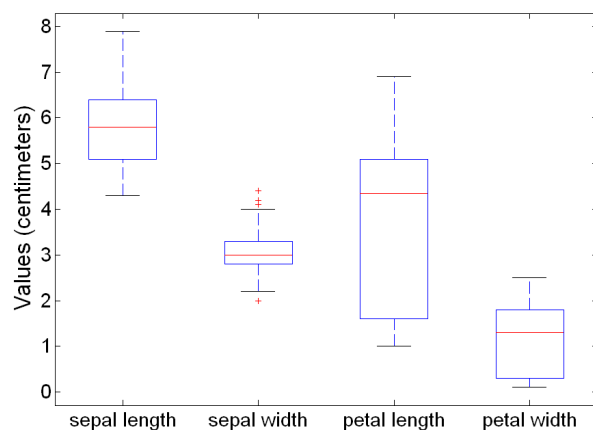
- Box Plots
 - grafické zobrazení tzv. 5ti číselného souhrnu
 - Autor J. Tukey



Ondřej Přibyl K611MDS

Příklad Box Plots

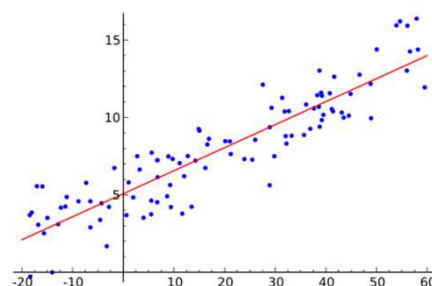
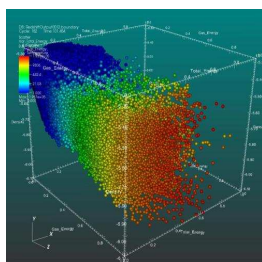
- Box plots se využívají k porovnání atributů



Ondřej Přibyl K611MDS

Vizualizační techniky: Korelační diagram

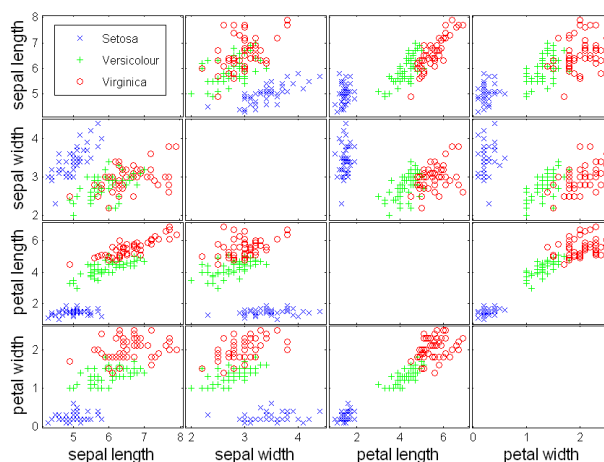
- též bodový graf (anglicky Scatter plot)
- matematické schéma užívající kartézských souřadnic pro zobrazení souboru dat o dvou (či tří) proměnných (na osy).
- Takto je možné jednoduše zjistit vzájemný vztah (korelaci) mezi oběma proměnnými



Ondřej Přibyl K611MDS

Pole korelačních diagramů

- Vícerozměrné zobrazení je nepřehledné
- Zobrazuje vzájemné vztahy více proměnných



Ondřej Přibyl K611MDS

Obsah prezentace

- Měřené veličiny
- Chyby měření
- Základní charakteristiky dat
- Vizualizace dat
- **Další aspekty analýzy dat**
- Hlavní kroky při analýze dat

Co jsou data?

- Kolekce datových objektů a jejich atributů
- Atribut
 - vlastnost či charakteristika objektu
 - Příklad: barva vozidla, objem motoru, a další
- Datový objekt (rekord (DB), instance, vzorek, entita, ...) je popsán kolekcí atributů

Objekty

Atributy

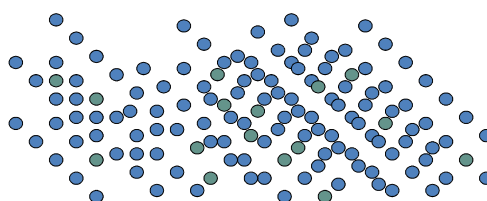
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Populace versus náhodný výběr

- Základní soubor (populace) - Všechny jednotky
 - např. všichni řidiči v ČR
 - Označují se písmeny řecké abecedy (μ , σ , ...)
- Výběrový soubor - vybrané jednotky, náhodný výběr
 - např. všichni řidiči, kteří v konkrétním dni jeli autem a stali se účastníky dotazníku
 - Označují se písmeny latinské abecedy (, s, ...)



Zdroj <http://www.nedarc.org/>



Ondřej Přibyl K611MDS

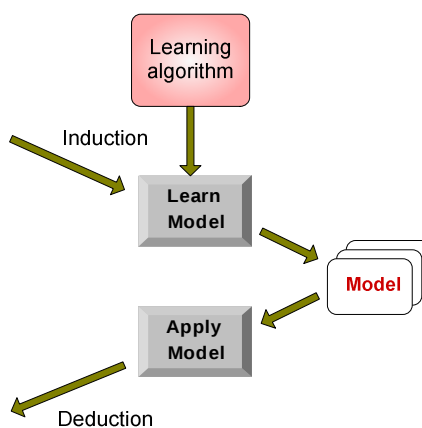
Aplikace modelu

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

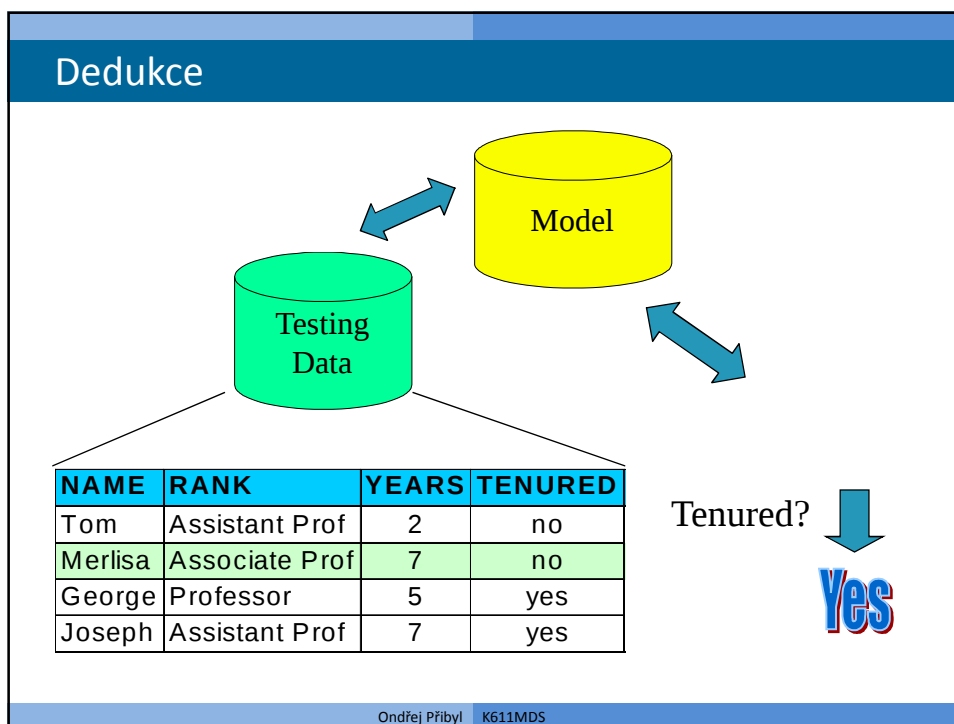
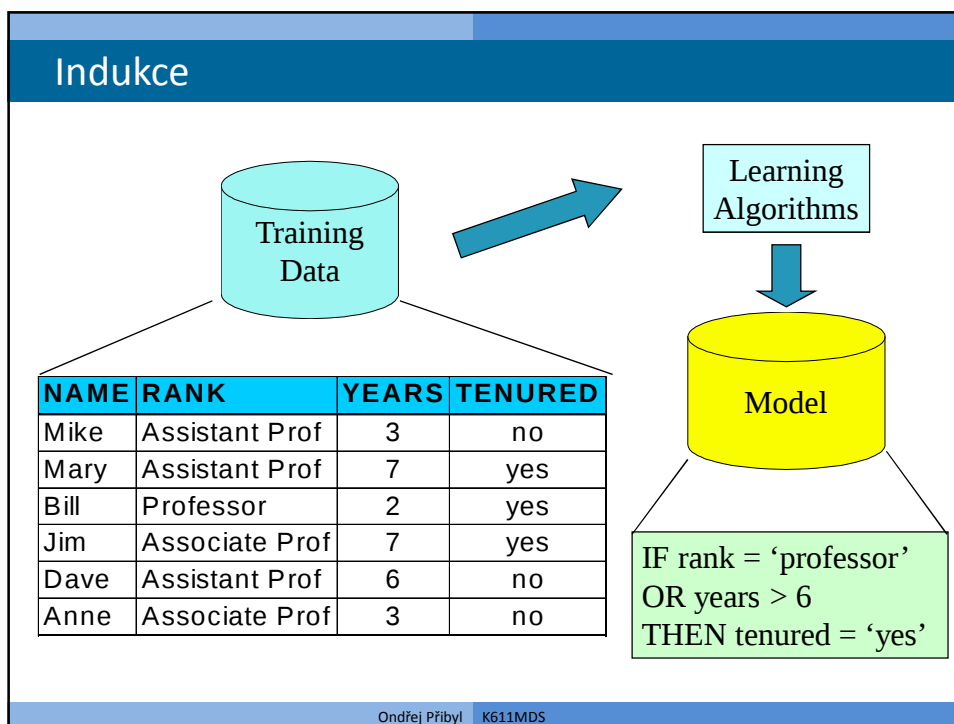
Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set



Ondřej Přibyl K611MDS



Jak vyhodnotit výsledný model?

- Podle prediktivní schopnosti (ne rychlost algoritmu)
- Matice záměn (Confusion Matrix)
 - A je počet správných predikcí, že daná instance je negativní
 - B je počet nesprávných predikcí, že daná instance je negativní
 - C je počet nesprávných predikcí, že daná instance je pozitivní
 - D je počet správných predikcí, že daná instance je pozitivní

	PREDIKOVANÁ TŘÍDA		
SKUTEČNÁ TŘÍDA		Třída=Ano	Třída=Ne
	Třída=Ano	a	b
	Třída=Ne	c	d

a: TP (true positive)
 b: FN (false negative)
 c: FP (false positive)
 d: TN (true negative)

Ondřej Přibyl K611MDS

Metriky pro porovnání metod

- Přenost (accuracy)

$$\text{Accuracy} = \frac{a + d}{a + b + c + d} = \frac{TP + TN}{TP + TN + FP + FN}$$

	PREDIKOVANÁ TŘÍDA		
SKUTEČNÁ TŘÍDA		Třída=Ano	Třída=Ne
	Třída=Ano	a	b
	Třída=Ne	c	d

a: TP (true positive)
 b: FN (false negative)
 c: FP (false positive)
 d: TN (true negative)

Ondřej Přibyl K611MDS

Problémy s touto metrikou (přesnost)

- Máme problém s dvěma třídami
 - Počet vzorků pro třídu 0 = 9990
 - Počet vzorků pro třídu 1 = 10
- Pokud model predikuje vše do třídy 0, přesnost je
 - $9990/10000 = 99.9 \%$

Ondřej Přibyl K611MDS

Cost Matrix (cena)

	PREDIKOVANÁ TŘÍDA		
	C(i j)	Třída=Ano	Třída=Ne
SKUTEČNÁ TŘÍDA	Třída=Ano	C(Ano Ano)	C(Ne Ano)
	Třída=Ne	C(Ano Ne)	C(Ne Ne)

$C(i|j)$: Cena za špatnou klasifikaci vzorku z třídy j do třídy i

Ondřej Přibyl K611MDS

Výpočet ceny klasifikace

Cost Matrix	PREDICTED CLASS		
	C(i j)	+	-
	ACTUAL CLASS		
	+	-1	100
	-	1	0

Model M_1	PREDICTED CLASS		
		+	-
	ACTUAL CLASS		
	+	150	40
	-	60	250

Accuracy = 80%

$$(150+250)/(150+250+60+40)*100$$

Cost = 5890

$$150*(-1)+40*100+60*1+250*0$$

Model M_2	PREDICTED CLASS		
		+	-
	ACTUAL CLASS		
	+	250	45
	-	5	200

Accuracy = 90%

Cost = 4255

Ondřej Přibyl K611MDS

Cost-Sensitive Measures

- Precision is biased towards $C(\text{Yes}|\text{Yes})$ & $C(\text{Yes}|\text{No})$
- Recall is biased towards $C(\text{Yes}|\text{Yes})$ & $C(\text{No}|\text{Yes})$
- F-measure is biased towards all except $C(\text{No}|\text{No})$

$$\text{Precision (p)} = \frac{a}{a+c}$$

$$\text{Recall (r)} = \frac{a}{a+b}$$

$$\text{F-measure (F)} = \frac{2rp}{r+p} = \frac{2a}{2a+b+c}$$

$$\text{Weighted Accuracy} = \frac{w_1a + w_4d}{w_1a + w_2b + w_3c + w_4d}$$

Ondřej Přibyl K611MDS

Obsah prezentace

- Měřené veličiny
- Chyby měření
- Vizualizace dat
- Další aspekty analýzy dat
- Hlavní kroky při analýze dat

Ondřej Příbyl K611MDS

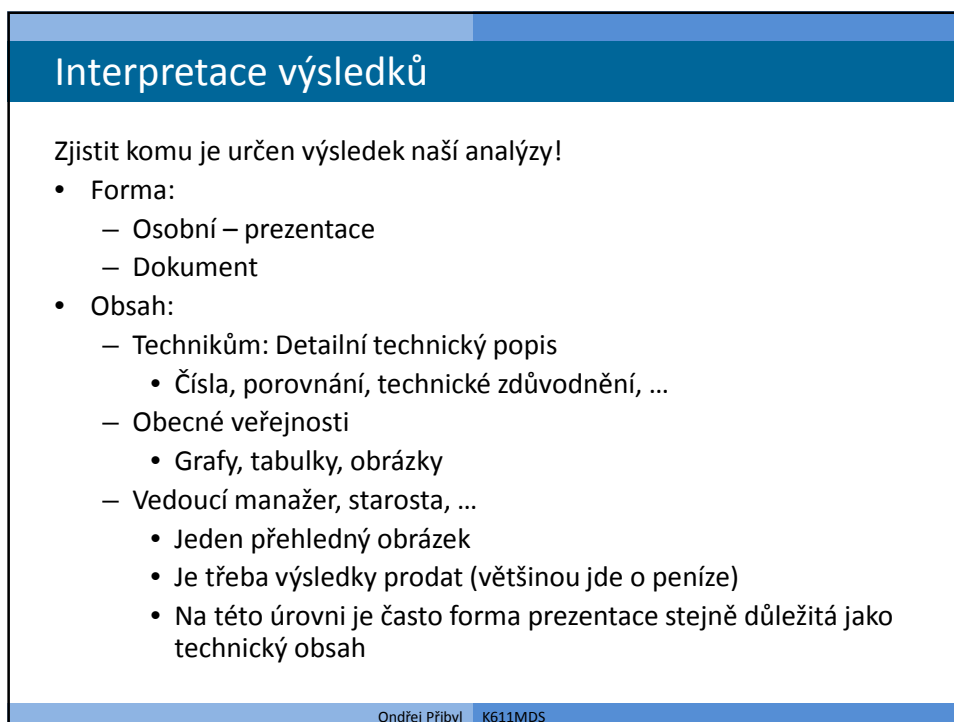
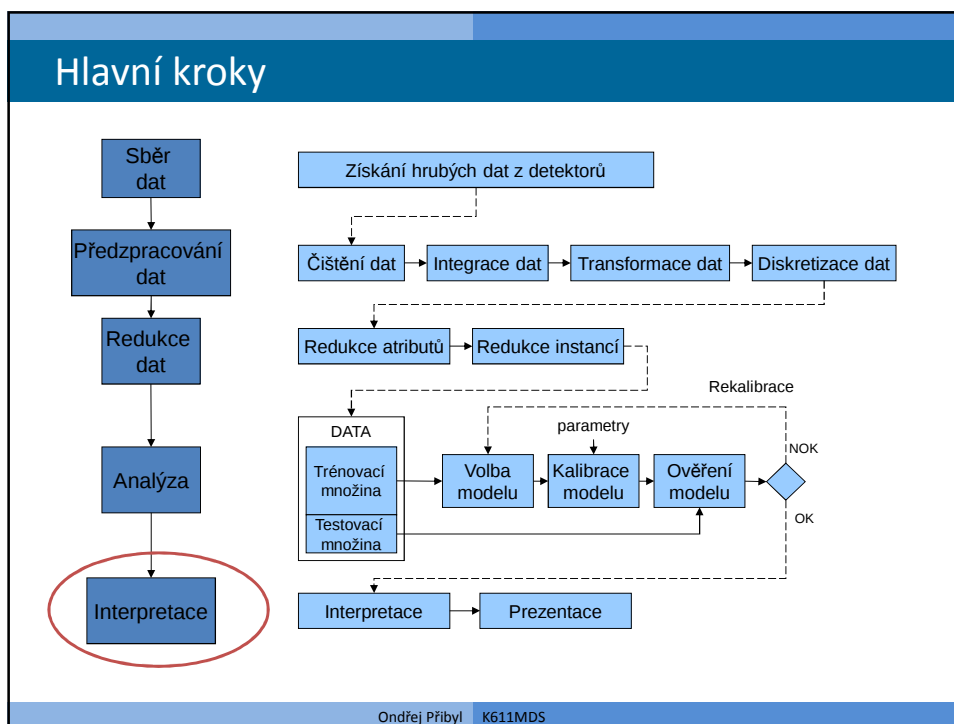
strana 55

Diskuze

- Co je třeba udělat s daty před aplikací vlastního matematického modelu?
 - Uvedte na příkladech



Ondřej Příbyl K611MDS



Co by v prezentaci výsledků nemělo chybět

- Název, datum, kontakty
- Stručný úvod (o co vlastně jde)
- Představení autorů, poděkování sponzorům
- Cíle
- Popis použité metody
 - Proč byla použita tato metoda a ne jiná
 - Základy
- Všechny předpoklady použité při zpracování
 - Je třeba zajistit reprodukovatelnost výsledků
- Výsledky
 - Graficky, přehledně (Je třeba je správně „prodat“)
- Závěr
 - Co se podařilo, ale i co se nepodařilo
 - Další kroky

Ondřej Přibyl K611MDS

Děkuji za pozornost

Ondřej Přibyl K611MDS

Témata ke zkoušce

- Data, informace a znalosti
 - Definice a vztah
- Typy měřených veličin
 - Klasifikace s příklady
- Chyby měření
 - Klasifikace
 - Vzorce
 - Příklady a výpočet
- Populace versus náhodný výběr
- Přehled kroků analýzy dat
- Zdroje:
 - Ondřej Šlapák „Data, informace, znalosti“, ELECTRONIC JOURNAL FOR PHILOSOPHY/2003, ISSN 1211-0442
 - <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>
 - <http://www.physics.unc.edu/~deardorf/uncertainty/definitions.html>
 - <http://www.csse.monash.edu.au/~smarkham/resources/scaling.htm>
 - <http://people.math.sfu.ca/~cschwarz/Stat-301/Handouts/node5.html>
 - Meloun M., Militký J.: Zpracování experimentálních dat, Plus Praha 1994