

The left side of the slide features a decorative design. It includes a vertical bar with a fine grid pattern, a solid green vertical bar, and several thin vertical lines. To the right of these bars are five green circles of varying sizes, arranged in a cluster. The text 'ANALÝZA DAT' is positioned to the right of the circles.

ANALÝZA DAT

ANALÝZA DAT

- rozbor dat a jejich syntézu
- Cíl analýzy: redukce, organizace, syntéza a sumarizace informací, aby bylo možné přisoudit získaným výsledkům význam
- V kvantitativním výzkumu slouží analýza dat k vyvrácení nebo potvrzení výzkumníkových hypotéz a to především pomocí **statické analýzy**



STATISTICKÁ ANALÝZA

Rozdělení:

- popisnou (deskriptivní) - zobrazuje a sumarizuje data, mezi charakteristiky deskriptivní statistiky patří např. střední hodnota nebo rozdělení četností
- inferenční (induktivní) - zobecňuje výsledky z výběrového vzorku na populaci
- **Proměnná** se v rámci statistické analýzy označuje jako **statistický znak** = jinak řečeno statistické znaky jsou popisované vlastnosti jednotek statistického souboru



STATISTICKÁ ANALÝZA

Dělení analýzy dle počtu analyzovaných znaků:

- jednorozměrnou – sleduje se jeden znak
- dvourozměrnou – sledují se dva znaky
- vícerozměrnou – sleduje se více znaků

Dělení statistických znaků:

- **kvantitativní** – vyjádřen číselně (např. věk, výška, hmotnost apod.)
 - spojitý – libovolná reálná čísla (hmotnost, výška, skóre apod.)
 - diskrétní (nespojitý) – veličiny (číselné hodnoty; např. počet dětí, počet let praxe apod.)
- **kvalitativní** – vyjádřen nenumerycky (např. pohlaví, zkušenost s prodělaným onemocněním, zaměstnání apod.)
 - alternativní – jsou možné pouze dvě varianty (např. muž nebo žena)
 - množný – je možné více variant (např. léčba: medikamentózní, chirurgická, režimová opatření)

PŘÍKLADY STATISTICKÝCH METOD

1) **Statistické třídění a organizace dat** - umožňuje zpřehlednění získaných dat a jejich uspořádání pomocí např.:

- třídění slovních znaků podle určitých kritérií (podle abecedy, podle četnosti výskytu apod.)
- třídění číselných znaků - převod dat např. do tabulek (tabulky neboli matice dat, se nejčastěji vytváří pomocí počítačových programů, jako je např. MS Excel; základní podoba matice dat se skládá ze sloupců – pro jednotlivé statistické znaky a řádků – pro jednotlivé případy neboli statistické jednotky)



TŘÍDĚNÍ A ORGANIZACE DAT

➤ **tabulky rozdělení četností**

- číselný znak má určitou hodnotu (obecně k), ke každé hodnotě je přiřazen počet výskytu (četnost)
- **četnost** lze vyjádřit jako **absolutní** (počtem výskytu) či **relativní** (často v %)

Vzorec pro výpočet **relativní četnosti**:

$$f_i = n_i / n$$



PŘÍKLAD JEDNOROZMĚRNÉHO ROZDĚLENÍ ČETNOSTÍ (FREKVENČNÍ ROZDĚLENÍ)

Četnost naměřených hodnot TK			
Pacient	Hodnoty TK (mm Hg)	Absolutní četnost (n_i)	Relativní četnost (f_i) vyjádřená v %
1.	> 180/100	18	10,1
2.	> 160/100 - 180/100	21	11,7
3.	> 110/70 - 120/75	24	13,4
4.	> 145/90 - 160/100	29	16,2
5.	<110/70	37	20,7
6.	> 120/75 – 145/90	50	27,9
Součet (n)	x	179	100,0



TŘÍDĚNÍ A ORGANIZACE DAT

➤ **grafické znázornění** – pomocí grafů a diagramů
lze znázornit:

- rozdělení četností

 - sloupcový graf/diagram

 - pruhový

 - výsečový graf

- znázornění vývoje

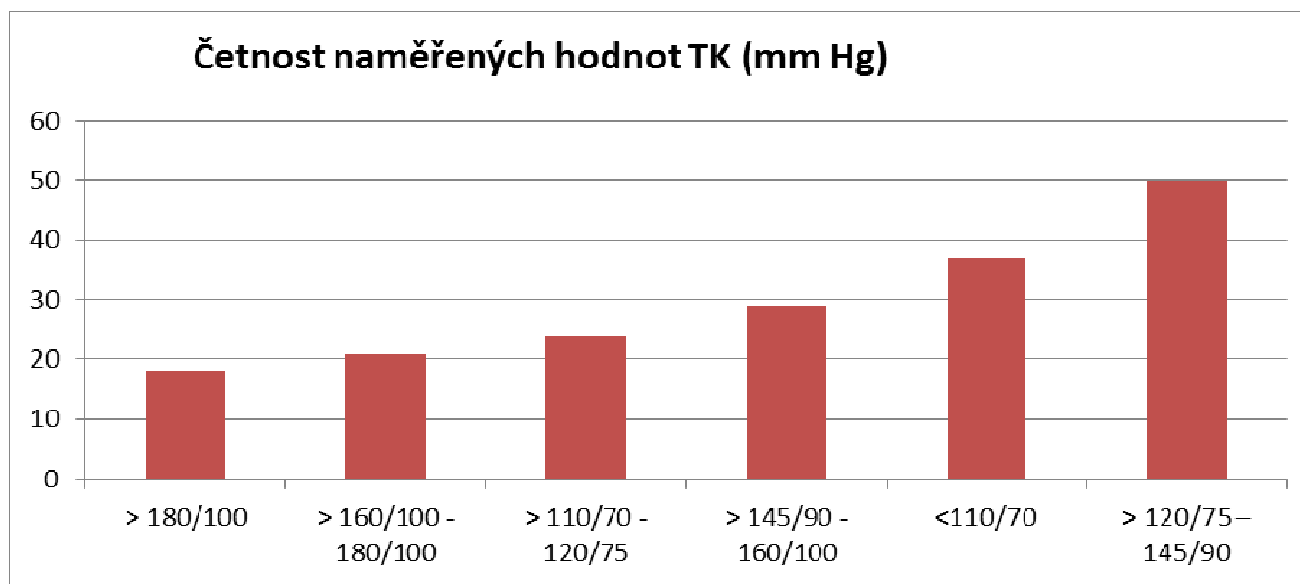
 - spojnicový graf (viz obr. 4 Příklad spojnicového grafu)

- znázornění umístění, rozptylu

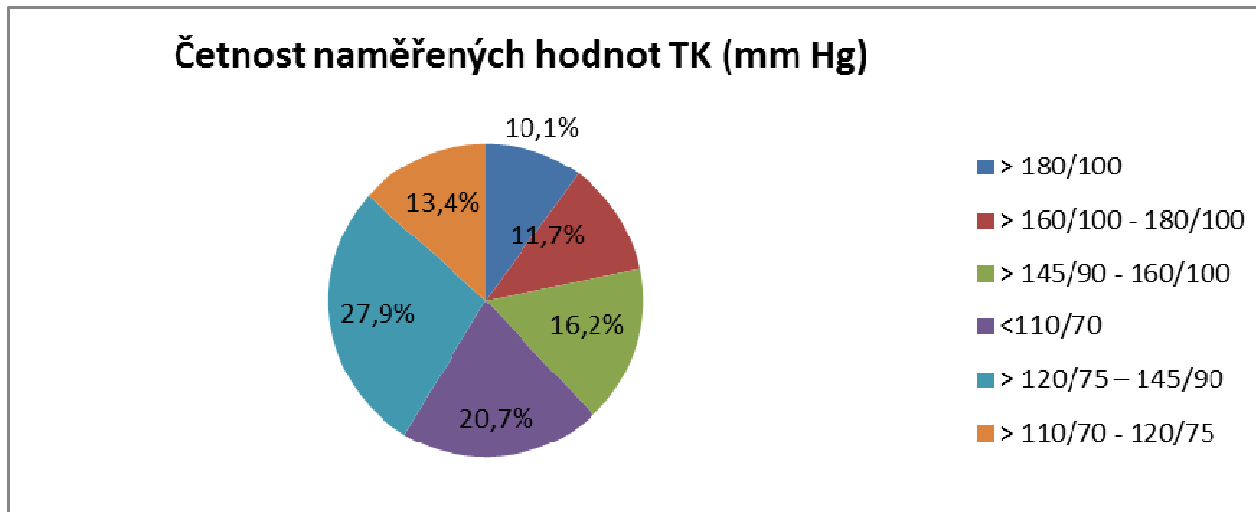
 - XY bodový



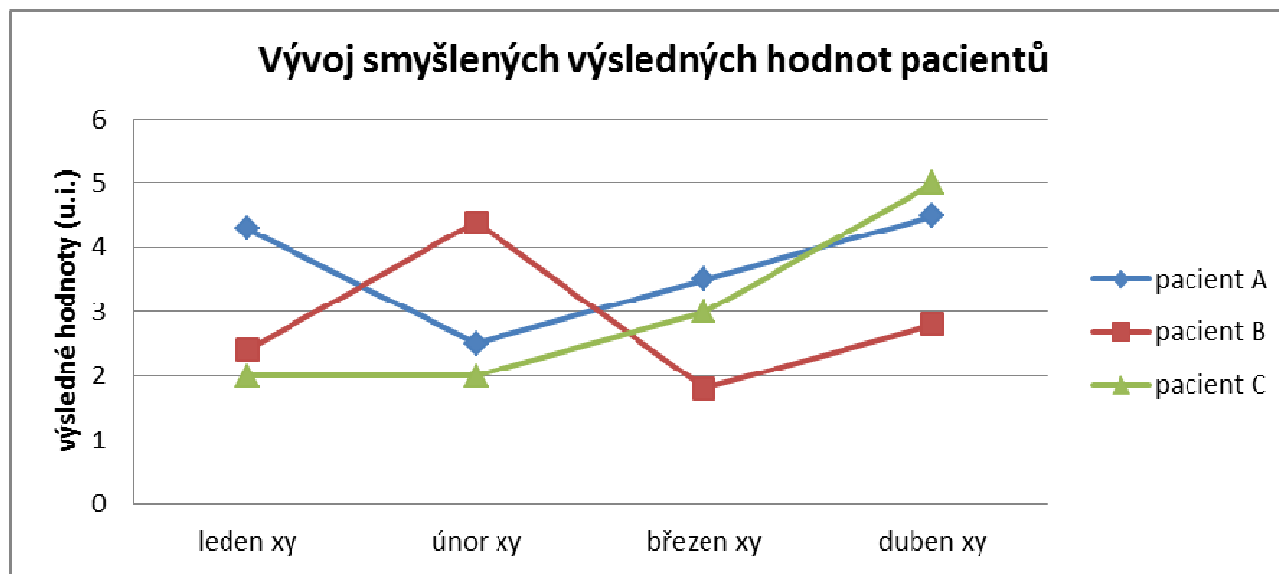
PŘÍKLAD SLOUPCOVÉHO DIAGRAMU



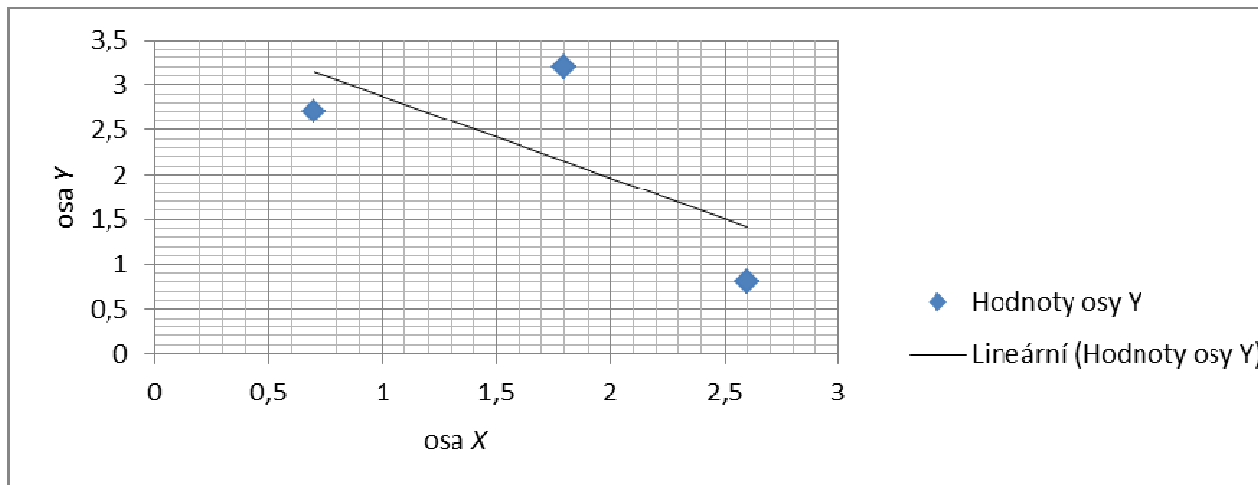
PŘÍKLAD VÝSEČOVÉHO GRAFU



PŘÍKLAD SPOJNICOVÉHO GRAFU



PŘÍKLAD XY BODOVÉHO GRAFU



TŘÍDĚNÍ A ORGANIZACE DAT

- **míry polohy** patří mezi základní popisné charakteristiky a jedná se o:
 - maximum – nejvyšší hodnota
 - minimum – nejnižší hodnota
 - modus – hodnota, která se nejčastěji vyskytuje
 - medián – hodnota nacházející se uprostřed pořadí
 - aritmetický průměr – součet pozorovaných hodnot dělený jejich počtem
- **míry rozptýlenosti** vypovídají o vyrovnanosti jednotek souboru, jak jsou hodnoty souboru rozptýleny a také jak je jednotlivé hodnoty znaku od sebe liší.
 - variační řada – seřazení dat podle velikosti
 - variační rozpětí (R) – jde o rozdíl maximální a minimální hodnoty sledovaného znaku; nevýhodou variačního rozpětí je, že výsledek může být zkreslen, jsou-li v daném souboru odlehlé hodnoty

Vzorec pro výpočet variačního rozpětí: $R = x_{max} - x_{min}$



TŘÍDĚNÍ A ORGANIZACE DAT

- **rozptyl** (variance) – vyjadřuje odchýlení hodnot určitého souboru od svého průměru; jde o průměr druhých mocnin odchylek od průměru
- **směrodatná odchylka** (SD) – je druhá odmocnina rozptylu; jde o odhad rozmístění hodnot
- **variační koeficient** – srovnává variabilitu dvou či více souborů nezávisle na jednotkách měření; jde o procentuálně vyjádřený poměr směrodatné odchylky k aritmetickému průměru, jinak řečeno vyjadřuje jaké % aritmetického průměru je tvořeno směrodatnou odchylkou (čím vyšší variační koeficient, tím „hůře“ aritmetický průměr zastupuje data souboru)



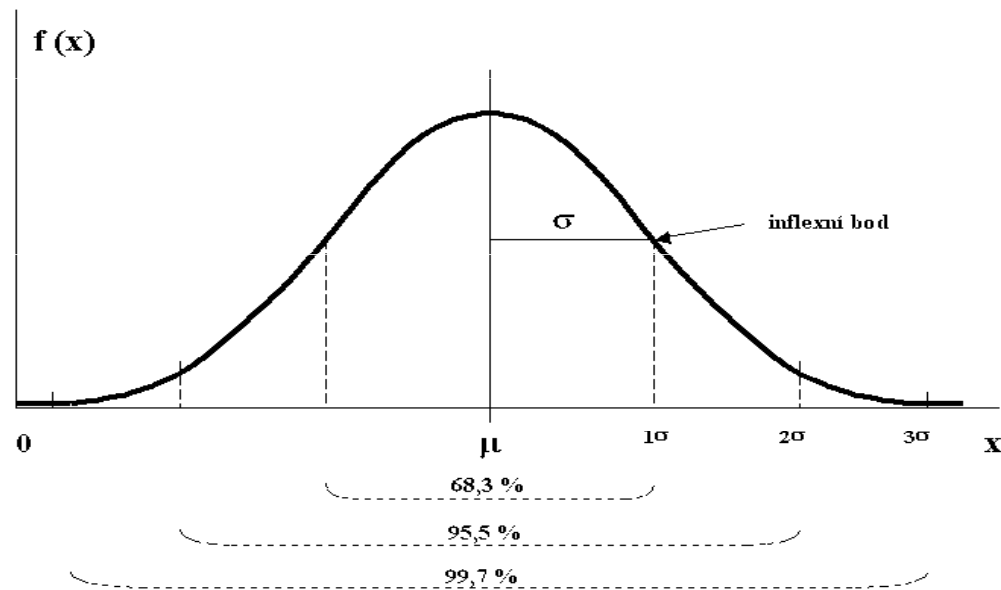
GAUSSOVA KŘIVKA

- Hustota normálního rozdělení
- Zvonová křivka

- Tvar Gaussovy křivky je dán:
 - střední hodnotou (μ) – určuje polohu maximální četnosti výskytu veličiny
- směrodatná odchylka (σ) – určuje tvar (šíři) křivky



GAUSSOVA KŘIVKA



Legenda: $f(x)$ - četnost výskytu sledované veličiny; X – charakteristická hodnota sledované veličiny; μ – střední hodnota, parametr polohy; σ – směrodatná odchylka charakteristické hodnoty sledované veličiny, parametr rozptýlení; inflexní bod – bod obratu křivky, je dán umístěním směrodatné odchylky



GAUSSOVA KŘIVKA – NORMÁLNÍ ROZDĚLENÍ

- Gaussovo normální rozdělení (μ , σ):
 - v rozmezí hodnot $\mu \pm 1\sigma$ se vyskytuje 68,3 % všech jedinců populace
 - v rozmezí hodnot $\mu \pm 2\sigma$ se vyskytuje 95,5 % všech jedinců populace
 - v rozmezí hodnot $\mu \pm 3\sigma$ se vyskytuje 99,7 % všech jedinců populace
 - zbylé 0,3 % hodnot (na koncích osy x) jsou ze statistického hlediska považovány za odlehlé hodnoty a pro další analýzu jsou vyřazeny



PŘÍKLADY STATISTICKÝCH METOD

2) Analýza závislosti

- **korelační analýza** – zkoumá vztahy proměnných;
korelační analýza se nezabývá kauzalitou vztahů

- Pearsonův korelační koeficient (r) – určuje míru lineární závislosti mezi dvěma spojitými proměnnými X a Y

Vzorec pro výpočet Pearsonova korelačního koeficientu:

$$r = \frac{1}{n - 1} \sum \frac{x - \bar{x}}{s_x} * \frac{y - \bar{y}}{s_y}$$

Legenda: s_x – směrodatná odchylka veličiny x ; s_y - směrodatná odchylka veličiny y ; \bar{x} - průměr znaků x ; \bar{y} - průměr znaků y



PŘÍKLADY STATISTICKÝCH METOD

- Hodnocení korelačního koeficientu:

1 = přímá závislost; pozitivní korelace; vzrůstající hodnota X i Y ; hodnotu Y lze stanovit na základě znalosti hodnoty X

0 = naprostá nezávislost; neexistuje statisticky zjistitelná lineární závislost (fakticky mohou na sobě ale veličiny záviset); hodnotu Y nelze odhadnout na základě znalosti hodnot X

-1 = negativní korelace; nepřímá závislost; když hodnota X vzrůstá, tak hodnota Y klesá; hodnotu Y lze stanovit na základě znalosti hodnoty X (DISMAN, 2002)



PŘÍKLADY STATISTICKÝCH METOD

- Spearmanův koeficient pořadové korelace (ρ nebo r_s) – měří sílu vztahu mezi X a Y ; výhodou této analýzy je, že zachycuje lineární vztahy a je rezistentní vůči odlehlým hodnotám; pro výpočet se užívá Pearsonův koeficient, ale místo původních hodnot se používají jejich pořadová čísla
- **regresní analýza** – zkoumá vztahy mezi závisle proměnnou (Y) a nezávisle proměnnou (X)



PŘÍKLADY STATISTICKÝCH METOD

3) Analýza kategoriálních dat – kategoriální data lze zachycovat do dvou a více rozměrných tabulek četností či relativních četností; klasifikace kategorií vždy odpovídá daným proměnným (závisle či nezávisle proměnné);

- χ^2 (chí kvadrát) **test dobré shody** - používá se k zjištění, zda data odpovídají předpokládanému rozdělení nebo rozdílu mezi pozorovanými a očekávanými četnostmi;
- pro χ^2 je nutné mít stanovenou tzv. nulovou hypotézu (H_0), tedy testovanou hypotézu, která se zabývá vztahem znaků (př. znak 1 a 2 jsou nezávislé);
- test dobré shody testuje nulovou hypotézu (H_0) vůči alternativní (H_1)



PŘÍKLADY STATISTICKÝCH METOD

- **závislost kategoriálních proměnných** – provádí se pomocí statistické analýzy četnostních tabulek, které zpřehledňují vztahy více statistických znaků

- kontingenční tabulky – tabulky vzniklé tříděním podle proměnných (každá jednotka statistického souboru může být klasifikována podle dvou kritérií – A, které mají r úrovní a B, které mají s úrovní; tento typ kontingenční tabulky je označován jako $(r \times s)$, viz obr 4 Příklad kontingenční tabulky; rozeznáváme různé druhy kontingenčních tabulek, kromě obecných $r \times s$, dále např. čtyřpolní tabulku (typ 2×2)



PŘÍKLAD KONTINGENČNÍ TABULKY

Úrovně	B_1	B_2	...	B_s	Součty řádkové
A_1	n_{11}	n_{12}	...	n_{1s}	$n_{1.}$
A_2	n_{21}	n_{22}	...	n_{2s}	$n_{2.}$
.
.
.
A_r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r.}$
Součty sloupcové	$n_{.1}$	$n_{.2}$...	$n_{.s}$	n

