MUNI
MED

BIOSTATISTIKA

Typy proměnných

- **Kvalitativní (kategoriální) proměnná**
  lze ji řadit do kategorií, ale nelze ji kvantifikovat
  Příklad: pohlaví, HIV status, barva vlasů …

- **Kvantitativní (numerická) proměnná**
  můžeme ji přiřadit číselnou hodnotu
  Příklad: výška, hmotnost, teplota, počet hospitalizací …

Popis a vizualizace kvalitativních proměnných

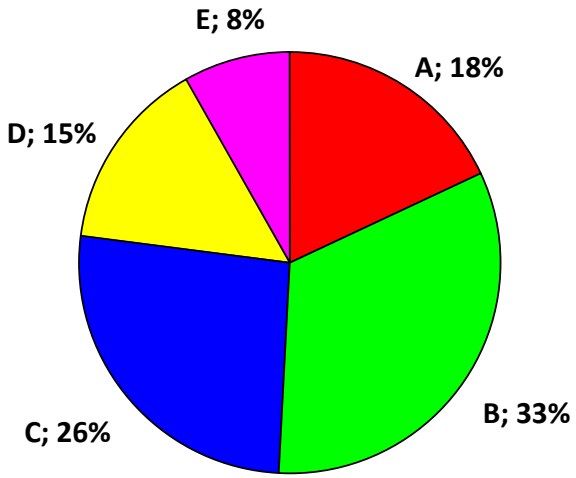
- **Popis kvalitativních dat:** četnost jednotlivých kategorií
- **Vizualizace kvalitativních dat:** koláčový nebo sloupcový graf

Příklad: Známka z biostatistiky (podzim 2014)

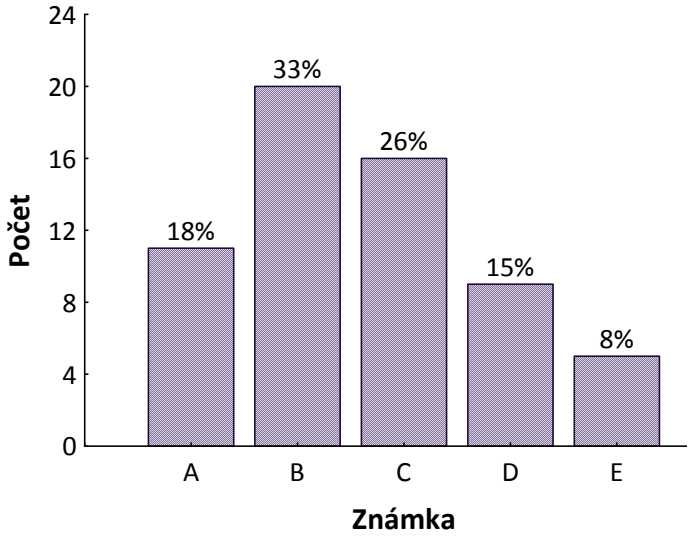
Frekvenční tabulka

Známka	n	%
A	11	18,0
B	20	32,8
C	16	26,2
D	9	14,8
E	5	8,2
F	0	0,0
Celkem	61	100,0

Koláčový graf



Sloupcový graf



MUNI MED

Popis kvantitativních dat

- **Popis kvantitativních dat:** charakteristika středu (průměr, medián aj.), charakteristika variability (rozptyl, rozsah hodnot, interkvartilové rozpětí aj.)

Příklad: Popis výšky pacientů (cm)

Popisné statistiky

Charakteristika	
N	61
Průměr (cm)	161,5
Medián (cm)	161,0
Sm. odchylka (cm)	4,7
Rozptyl (cm^2)	22,2
min-max (cm)	144 – 169
dolní-horní kvartil (cm)	158 - 164

Průměr a medián se téměř shodují. Co nám to říká?

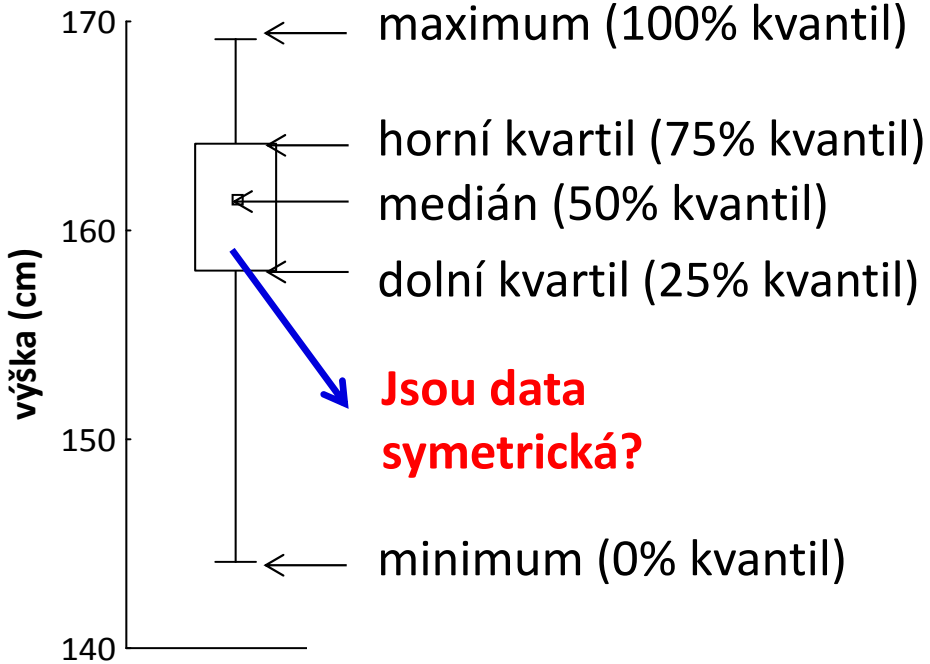
MUNI MED

Vizualizace kvantitativních dat

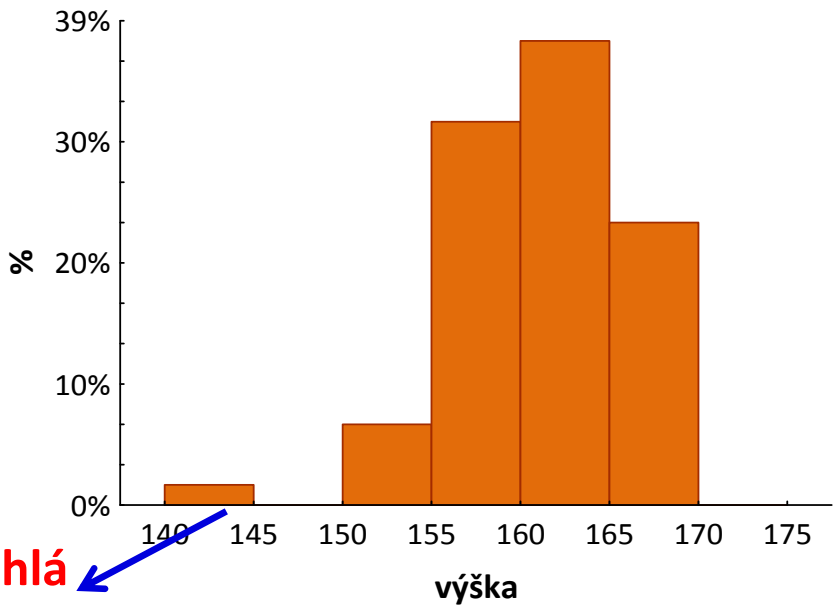
- **Vizualizace kvantitativních dat:** nejčastěji pomocí krabicového grafu nebo histogramu

Příklad: Popis výšky pacientů (cm)

Krabicový graf



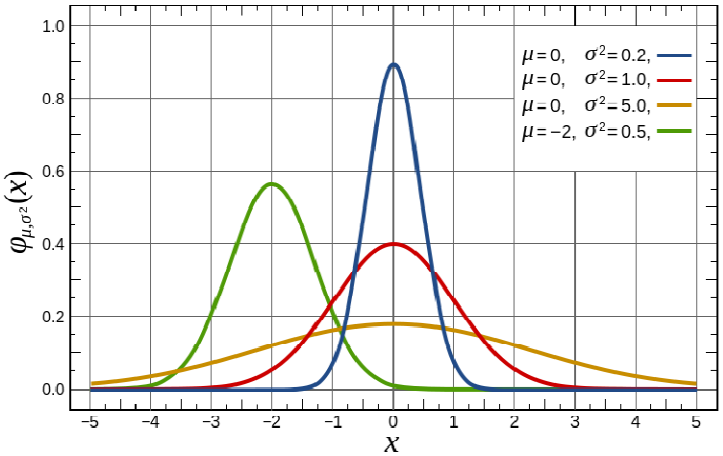
Histogram



MUNI MED

Normální rozdělení

- Nejklasičtějším modelovým rozdělením, od něhož je odvozena celá řada statistických analýz je tzv. **normální rozdělení**, známé též jako **Gaussova křivka**.
- Popisuje rozdělení pravděpodobnosti spojité náhodné veličiny, např. výška v populaci, chyba měření ...
- Je kompletně popsáno dvěma parametry:
  μ – střední hodnota
  σ^2 – rozptyl
  Označení: **N(μ, σ^2)**

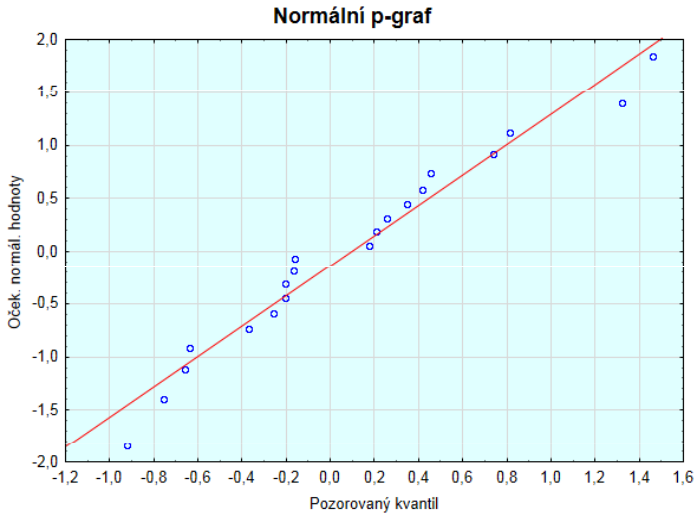


NORMALITA je klíčovým předpokladem řady statistických metod

MUNI MED

Vizuální ověření normality

- Pro hodnocení tvaru rozložení lze využít
  histogram nebo **normálně-pravděpodobnostný graf**



Pocházejí-li data z normálního rozložení, pak bude proložená křivka souhlasit s histogramem

Pocházejí-li data z normálního rozložení, pak body budou ležet okolo přímky

MUNI MED

Shapirův-Wilkův test normality

- Testy normality testují
  H_0: není rozdíl mezi zpracovávaným rozložením a normálním rozložením.

Shapirův-Wilkův test
Jde o neparametrický test použitelný i při velmi malých n (10) s dobrou sílou testu. Je zaměřen na testování symetrie.

Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.

Statistické testování – princip

Všechny statistické testy testují tzv. nulovou hypotézu. Proti ní stojí tzv. alternativní hypotéza.

- Nulová hypotéza H0 — H_0: sledovaný efekt je nulový
- Alternativní hypotéza HA — H_A: sledovaný efekt není nulový

Statistické testování odpovídá na otázku, zda je pozorovaný rozdíl náhodný či nikoliv.

- Testování nulové hypotézy probíhá většinou výpočtem tzv. testové statistiky a k ní je pak určena tzv. **p-hodnota**.

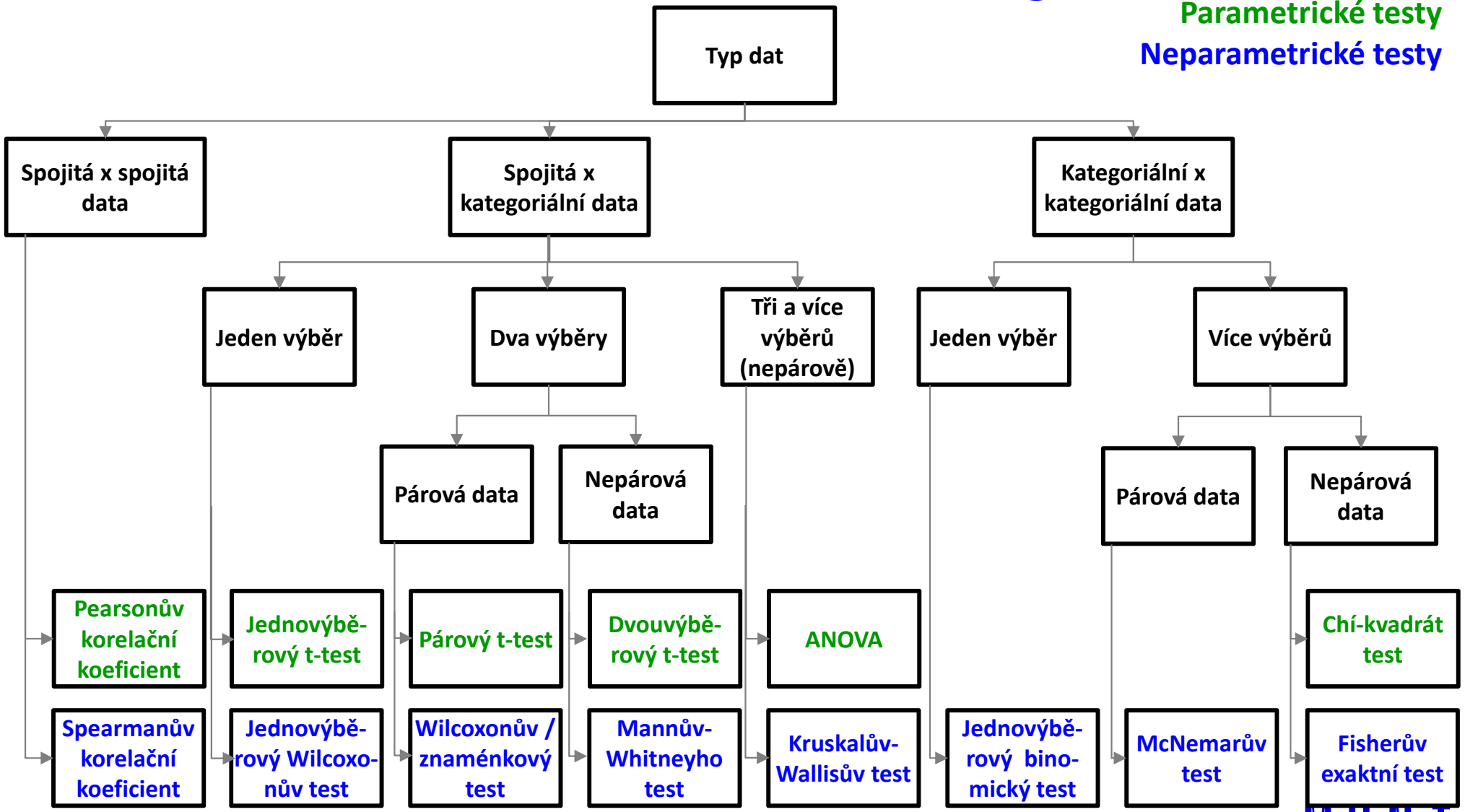
Způsoby testování: P-hodnota

- Významnost hypotézy hodnotíme dle získané **p-hodnoty**, která vyjadřuje pravděpodobnost, s jakou číselné realizace výběru podporují H_0, je-li pravdivá.
- P-hodnotu porovnáme s hladinou významnosti **α** (stanovujeme ji na 0,05).
- P-hodnotu získáme při testování hypotéz ve statistickém softwaru.

Je-li **p ≤ α**, pak H_0 zamítáme na hladině významnosti α a přijímáme H_A.
Je-li **p > α**, pak H_0 nezamítáme na hladině významnosti α.

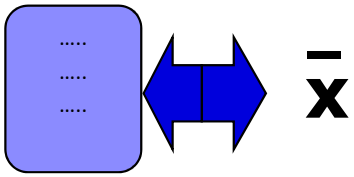
Základní statistické testy

Parametrické testy
Neparametrické testy



MUNI MED

Jednovýběrový test



1. **Stanovení nulové a alternativní hypotézy:**

   H_0: Průměr výběru je rovný referenční hodnotě.

   H_A: Průměr výběru není rovný referenční hodnotě.

2. Ověření normality rozdělení hodnot výběru

   (vizuálně i statistickým testem: Shapiro-Wilkův test).

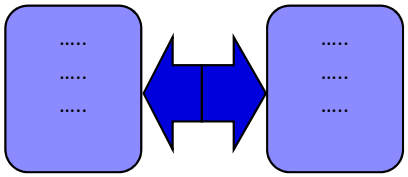
   Předpoklad splněn => **jednovýběrový t-test**

   Předpoklad nesplněn => **Wilcoxonův test, znaménkový test**

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**. Když je vypočítaná p-hodnota menší než zvolená hladina významnosti $\alpha = 0,05$, zamítáme nulovou hypotézu.

MUNI MED

Párový test



1. **Stanovení nulové a alternativní hypotézy:**

   $\mathbf{H_0}$**:** Průměry před a po léčbě se neliší.

   $\mathbf{H_A}$**:** Průměry před a po léčbě se liší.

2. Spočítání diference hodnot a prohlédnutí jejich průběhu.

   Ověření normality rozdělení diferencí

   (vizuálně i statistickým testem: Shapiro-Wilkův test).

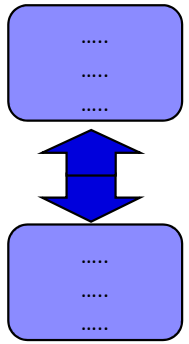
   Předpoklad splněn => **párový t-test**

   Předpoklad nesplněn => **párový Wilcoxonův/znaménkový t.**

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**. Když je vypočítaná p-hodnota menší než zvolená hladina významnosti $\alpha = 0{,}05$, zamítáme nulovou hypotézu.

MUNI MED

Dvouvýběrový test



1. **Stanovení nulové a alternativní hypotézy:**

   H_0: Průměry obou skupin jsou shodné.

   H_A: Průměry obou skupin nejsou shodné.

2. Prohlédnutí průběhu dat, určení průměru, mediánu

   Ověření normality dat (vizuálně i Shapiro-Wilkovým testem)

   Ověření homogenity rozptylů (F-testem)

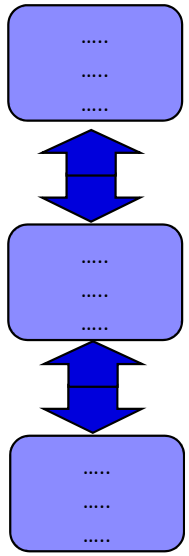
   Předpoklady splněny => **nepárový dvouvýběrový t-test**

   Předpoklady nesplněny => **Mannův-Whitneyův U test**

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**. Když je vypočítaná p-hodnota menší než zvolená hladina významnosti $\alpha = 0,05$, zamítáme nulovou hypotézu.

MUNI MED

Test pro více nezávislých výběrů



1. **Stanovení nulové a alternativní hypotézy:**

   H_0: Střední hodnoty všech skupin jsou shodné.

   H_A: Aspoň jedna dvojice středních hodnot se liší.

2. Prohlédnutí průběhu dat, určení průměru, mediánu

   Ověření normality dat (vizuálně i Shapiro-Wilkovým testem)

   Ověření homogenity rozptylů (Levenův test)

   Předpoklady splněny => **ANOVA**

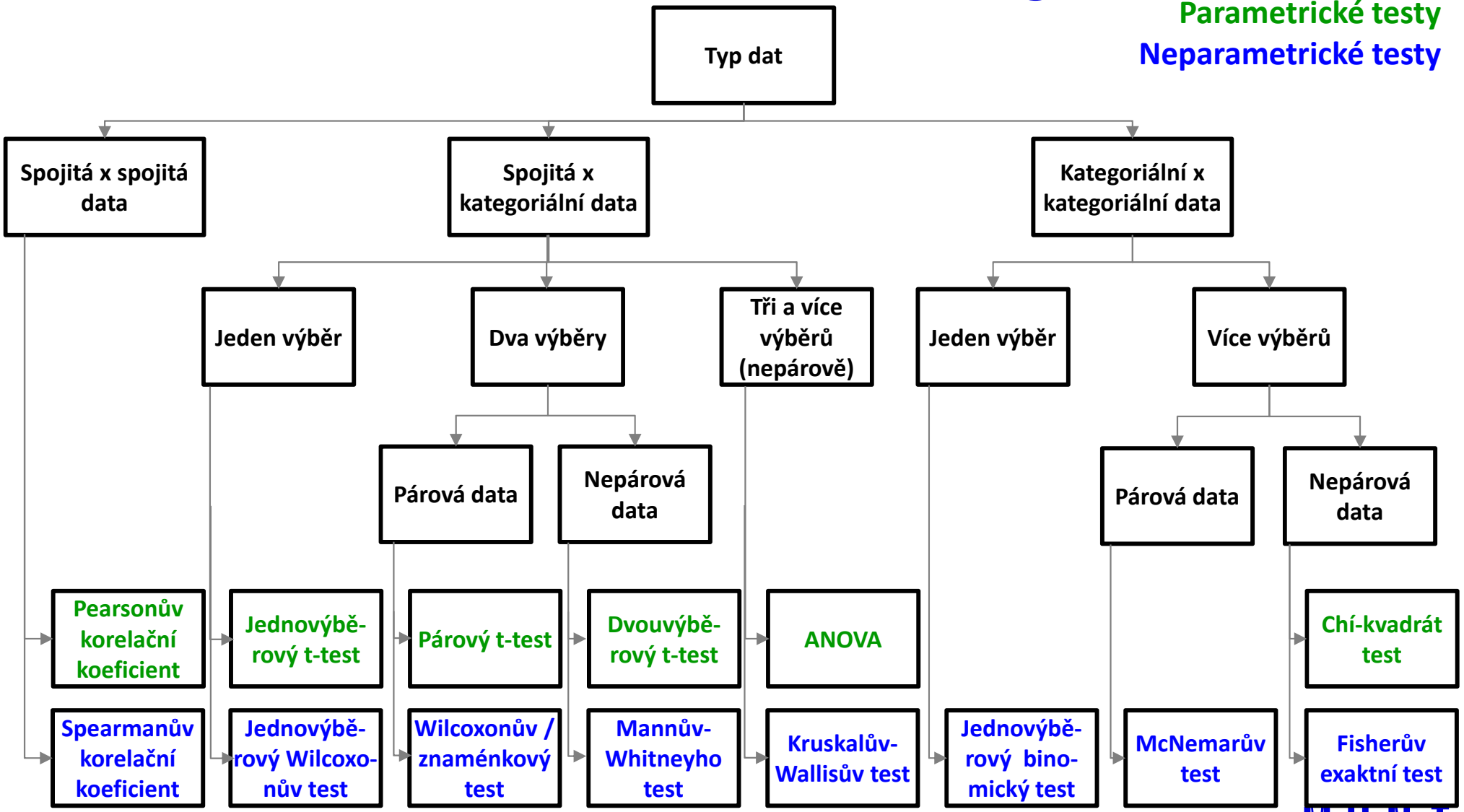
   Předpoklady nesplněny => **Kruskalův-Wallisův test**

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**.

   Když je $p < \alpha$, zamítáme nulovou hypotézu. Dalším, tzv. **post hoc testem** hledáme dvojici s odlišnou střední hodnotou.

Základní statistické testy

Parametrické testy
Neparametrické testy



MUNI MED

Typy proměnných

- **Kvalitativní (kategoriální) proměnná**
  lze ji řadit do kategorií, ale nelze ji kvantifikovat
  Příklad: pohlaví, HIV status, barva vlasů …

- **Kvantitativní (numerická) proměnná**
  můžeme ji přiřadit číselnou hodnotu
  Příklad: výška, hmotnost, teplota, počet hospitalizací …

Popis a vizualizace kvalitativních proměnných

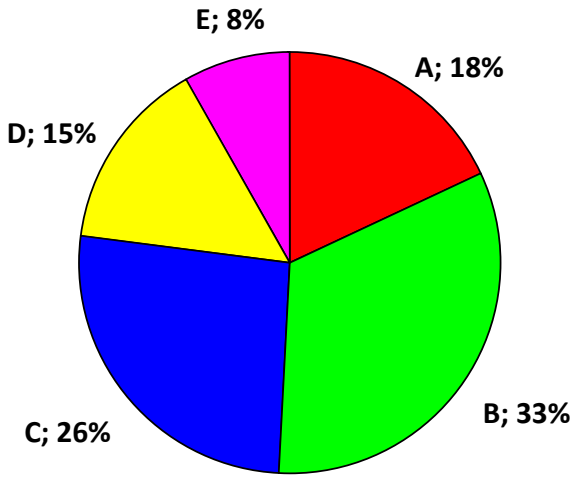
- **Popis kvalitativních dat:** četnost jednotlivých kategorií
- **Vizualizace kvalitativních dat:** koláčový nebo sloupcový graf

Příklad: Známka z biostatistiky (podzim 2014)

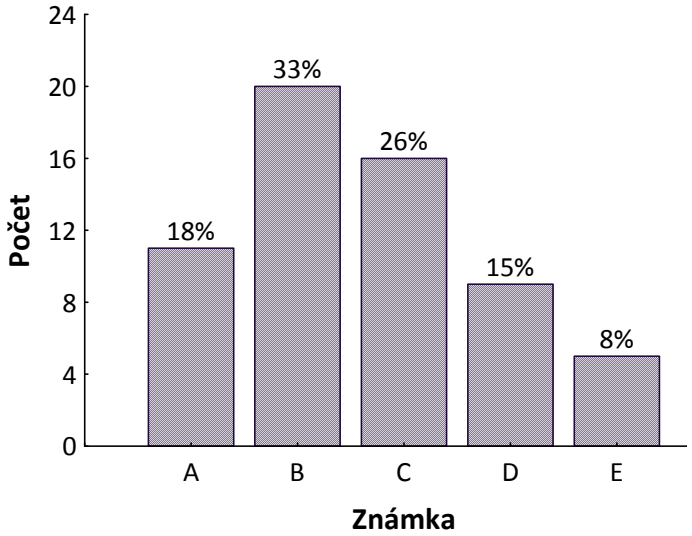
Frekvenční tabulka

Známka	n	%
A	11	18,0
B	20	32,8
C	16	26,2
D	9	14,8
E	5	8,2
F	0	0,0
Celkem	61	100,0

Koláčový graf



Sloupcový graf



MUNI MED

Kontingenční tabulka – vztah kategoriálních proměnných

- Řádky (r) hodnotami (kategoriemi) první proměnné, sloupce (c) hodnotami druhé proměnné.
- V buňkách tabulky jsou uvedeny počty případů s hodnotou první proměnné odpovídající příslušnému řádku a druhé proměnné s hodnotou odpovídající příslušnému sloupci.

	Nemocný	Zdravý	Celkem
Muž	45	11	56
Žena	25	6	31
Celkem	70	17	87

Analýza kontingenčních tabulek

- Analýza kontingenčních tabulek umožňuje analyzovat **vazbu mezi dvěma kategoriálními proměnnými** (pomocí chí-kvadrát testu, tj. srovnáním pozorovaných a očekávaných četností)
- Umožňuje testovat:

Hypotézu o nezávislosti:

H_0: Proměnné jsou nezávislé; H_A: Proměnné jsou závislé.

Hypotézu o shodě struktury:

H_0: Procentuální zastoupení kategorií proměnné je stejné ve srovnávaných výběrech; H_A: … není stejné

Hypotézu o symetrii:

H_0: (pokus nemá vliv na výskyt daného znaku)

H_A: $n_{ij} \neq n_{ji}$ (pokus má vliv na výskyt daného znaku)

Testování nezávislosti dvou kategoriálních proměnných

1. **Stanovení nulové a alternativní hypotézy:**

   H_0: Dvě kategoriální proměnné jsou nezávislé.

   H_A: Dvě kategoriální proměnné jsou závislé.

2. Vypočítání pozorovaných a očekávaných četností

   Ověření podmínky dobré aproximace (týká se oček. četností)

   Předpoklad splněn => **Pearsonův chí-kvadrát test**

   Předpoklad nesplněn => **Fisherův exaktní test**

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**.

   Když je $p < \alpha$, zamítáme nulovou hypotézu.

Testování shody struktury dvou kategoriálních proměnných

1. **Stanovení nulové a alternativní hypotézy:**

   H_0: Pravděpodobnostní rozdělení kategoriální proměnné je stejné v různých populacích.

   H_A: Pravděpodobnostní rozdělení kategoriální proměnné není stejné v různých populacích.

2. Vypočítání pozorovaných a očekávaných četností

   Ověření podmínky dobré aproximace (týká se oček. četností)

   Předpoklad splněn => **Pearsonův chí-kvadrát test**

   Předpoklad nesplněn => **Fisherův exaktní test**

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**.

   Když je $p < \alpha$, zamítáme nulovou hypotézu.

MUNI MED

Testování symetrie – McNemarův test

- **Hypotéza o symetrii:** Opakovaně sledujeme binární proměnnou a zajímá nás, zda došlo ke změně jejího rozdělení.
  Příklad: *Výskyt bolesti před a po užití léku.*
- **H_0:** (pokus nemá vliv na výskyt daného znaku)

Četnost	Po: ANO	Po: NE	
Před: ANO	a	b	a + b
Před: NE	c	d	c + d
	a + c	b + d	N

Teoretická pravděpodobnost	Po: ANO	Po: NE	
Před: ANO	n_{11}	n_{12}	$n_{1.}$
Před: NE	n_{21}	n_{22}	$n_{2.}$
	$n_{.1}$	$n_{.2}$	

- Testová statistika: Pokud je větší než kritická hodnota rozdělení o jednom stupni volnosti (vhodné pro počty údajů b + c > 8), pak nulovou hypotézu zamítáme.

MUNI MED

Testování hypotézy o symetrii

1. **Stanovení nulové a alternativní hypotézy:**

   $\mathbf{H_0}$: Pokus nemá vliv na výskyt daného znaku.

   $\mathbf{H_A}$: $n_{ij} \neq n_{ji}$ Pokus má vliv na výskyt daného znaku.

2. Vypočítání pozorovaných četností

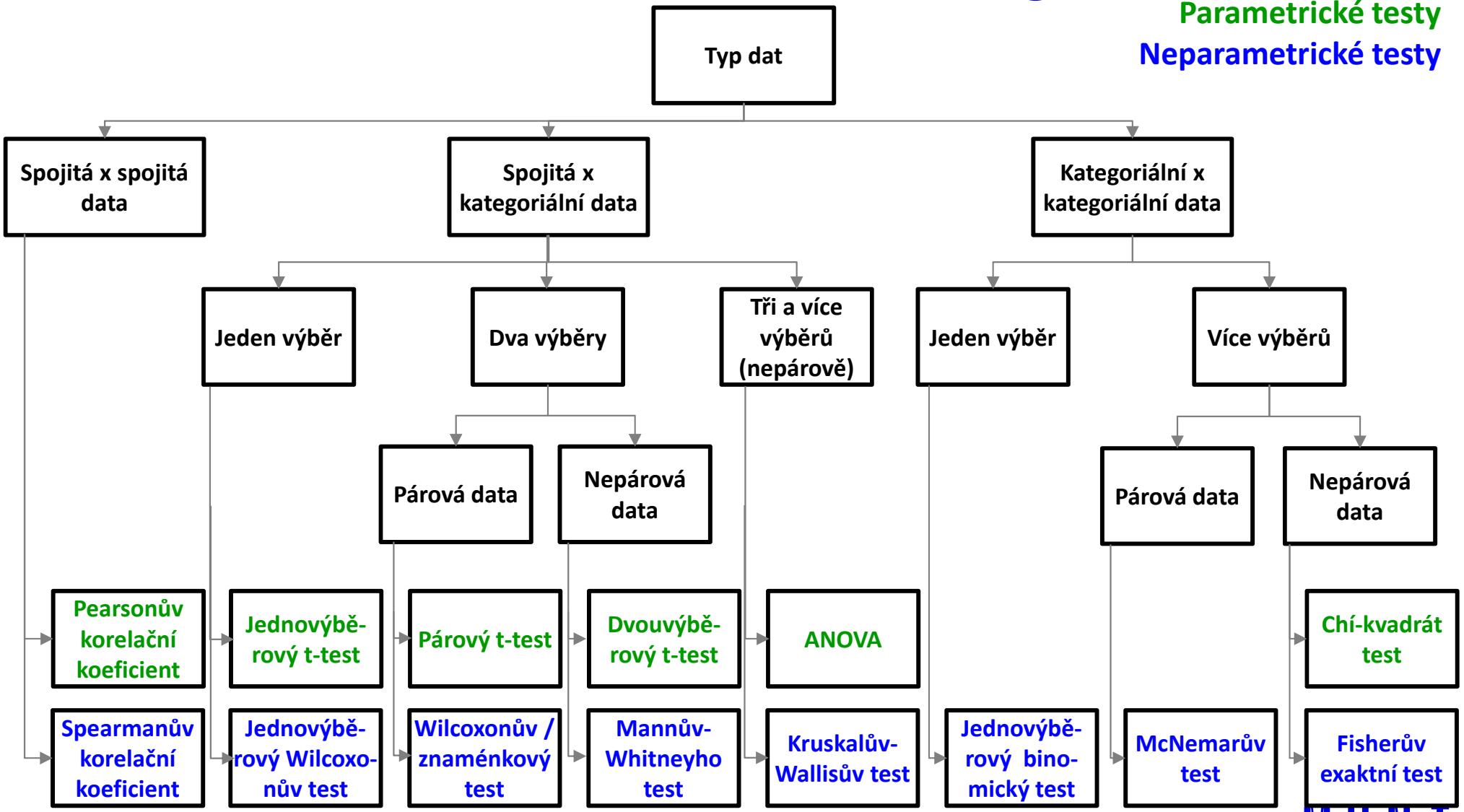
   McNemarův test

3. **Vypočítaní** hodnoty testové statistiky a **p-hodnoty**.

   Když je p < α, zamítáme nulovou hypotézu.

Základní statistické testy

Parametrické testy
Neparametrické testy

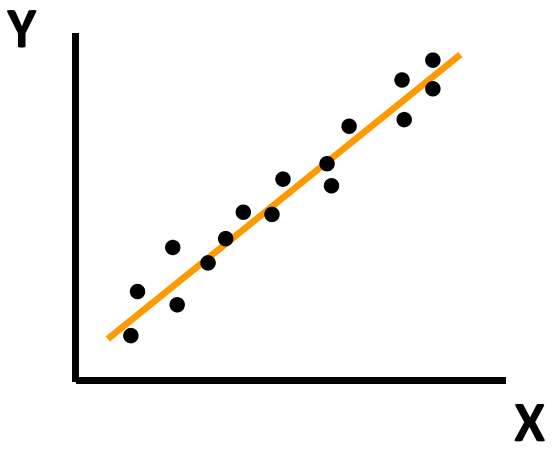


MUNI MED

Korelace

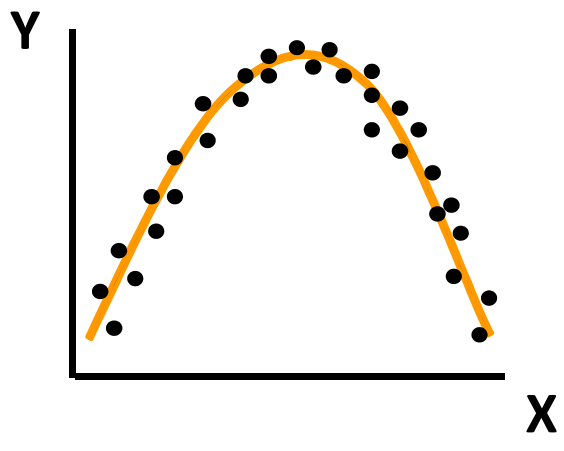
– **Korelační analýza** je využívána pro vyhodnocení míry **vztahu** dvou spojitých proměnných. Obdobně jako jiné statistické metody, i korelace mohou být **parametrické** nebo **neparametrické**.

Kladná korelace



Záporná korelace

Bez korelace



MUNI MED

Korelační koeficienty

- **Korelační koeficient** (*r*) – kvantifikuje míru vztahu mezi dvěma spojitými proměnnými *X* a *Y*.
- **Pearsonův korelační koeficient** je parametrický; hodnotí míru lineární závislosti mezi dvěma spojitými proměnnými.
  Předpoklad: proměnné pocházejí z tzv. dvourozměrného **normálního rozdělení** (pro každou hodnotu *X* má proměnná *Y* normální rozdělení a pro každou hodnotu *Y* má proměnná *X* normální rozdělení)
- **Spearmanův korelační koeficient** je neparametrický; hodnotí míru závislosti pořadí hodnot dvou spojitých proměnných.
- Hodnota ***r*** je **kladná**, když vyšší hodnoty *X* souvisí s vyššími hodnotami *Y*. Naopak hodnota ***r*** je **záporná**, když nižší hodnoty *X* souvisí s vyššími hodnotami *Y*.

MUNI MED