

Přednáška 6

Pojmy z oblasti statistického testování

Typy testů

Normalita dat a její význam pro testování

Parametrické vs. neparametrické testy

- **Parametrické testy**

- Mají předpoklady o rozložení vstupujících dat (např. normální rozložení)
- Při stejném N a dodržení předpokladů mají vyšší sílu testu než testy neparametrické
- **Pokud nejsou dodrženy předpoklady parametrických testů, potom jejich síla testu prudce klesá a výsledek testu může být zcela chybný a nesmyslný**

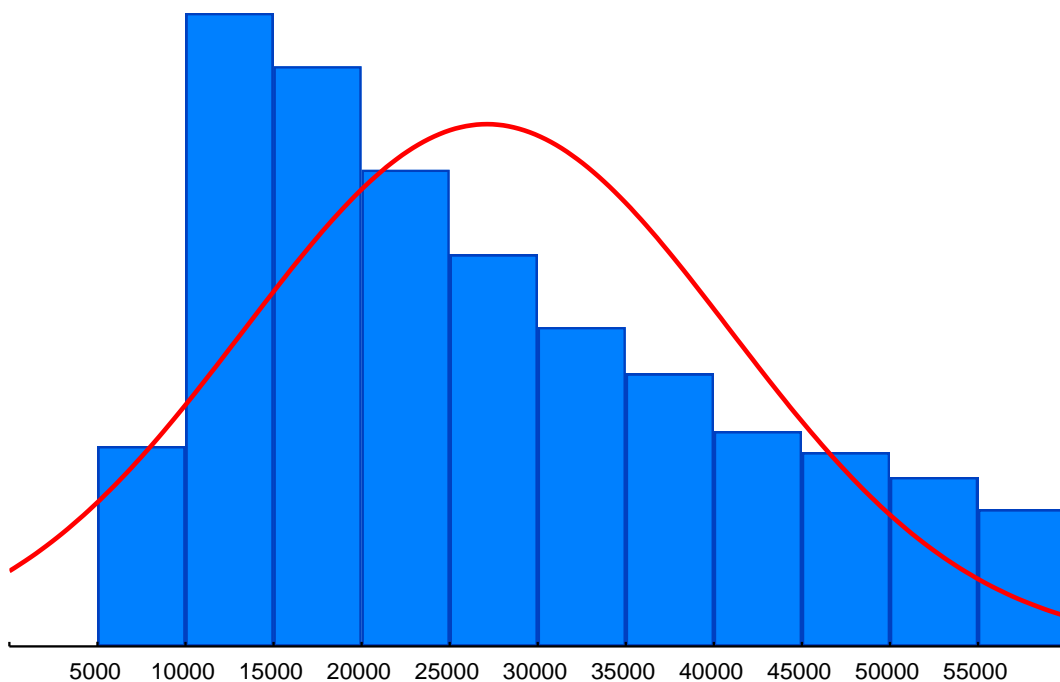
- **Neparametrické testy**

- Nemají předpoklady o rozložení vstupujících dat, lze je tedy použít i při asymetrickém rozložení, odlehlých hodnotách, či nedetekovatelném rozložení
- Snížená síla těchto testů je způsobena redukcí informační hodnoty původních dat, kdy neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí

Problémy parametrických a neparametrických testů

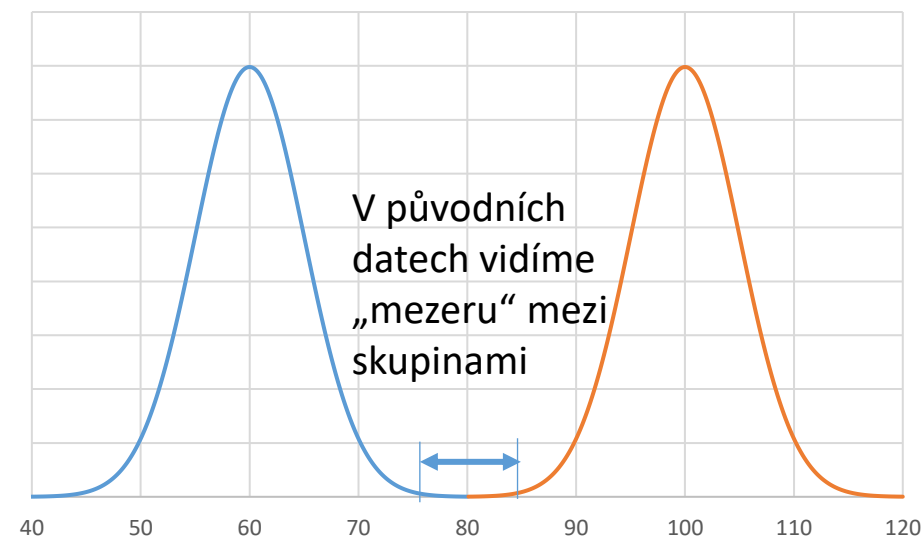
Parametrické testy

- Reálná data neodpovídají modelovému rozdělení



Neparametrické testy

- Díky převodu dat na pořadí ztrácíme část informace



Po převodu na pořadí o tuto informaci přicházíme.

Jednovýběrové a dvouvýběrové testy

- **Jednovýběrové testy (one-sample)**

- Srovnávají jeden vzorek (one sample, jednovýběrové testy) s referenční hodnotou (popřípadě se statistickým parametrem cílové populace)
- V testu je tedy srovnáváno rozložení hodnot (vzorek) s jediným číslem (referenční hodnota, hodnota cílové populace)
- Otázka položená v testu může být vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek

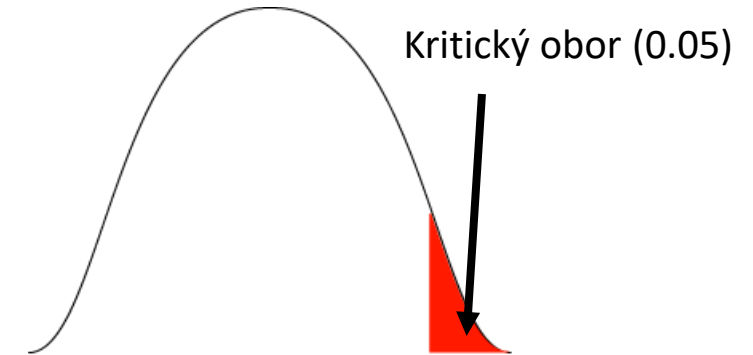
- **Dvouvýběrové testy (two-sample)**

- Srovnávají navzájem dva vzorky (two sample, dvouvýběrové testy)
- V testu jsou srovnávány dvě rozložení hodnot
- Otázka položená v testu může být opět vztažena k průměru, rozptylu, podílu hodnot i dalším statistickým parametrům popisujícím vzorek
- Kromě testů pro dvě skupiny hodnot existují samozřejmě i testy pro více skupin dat

Jednostranné a oboustranné hypotézy

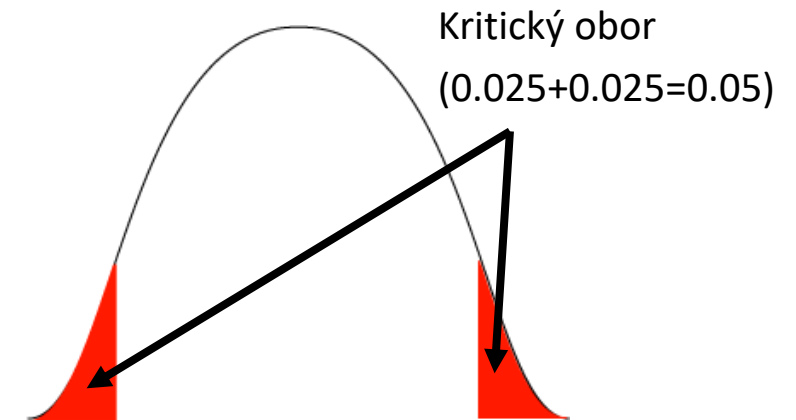
- **Jednostranné testy (one-tailed)**

- Hypotéza testu je postavena asymetricky, tedy ptáme se na větší než/ menší než
- Test může mít pouze dvojí výstup – jedna z hodnot je větší (menší) než druhá a všechny ostatní případy
- Pouze v případě jasné hypotézy – jinak může být napadnuto za účelovost (při vhodném výběru směru testování snažší potvrzení významnosti)



- **Oboustranné testy (two-tailed)**

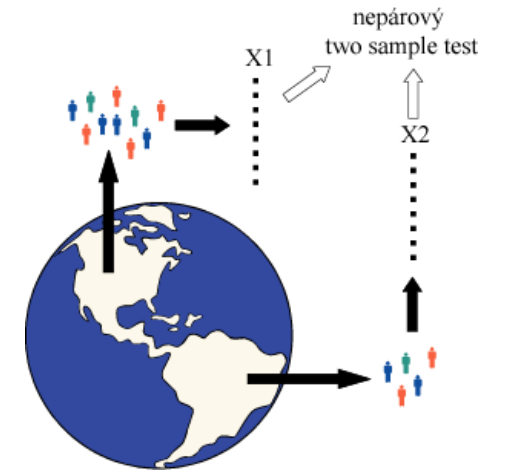
- Hypotéza testu se ptá na otázku rovná se/nerovná se
- Test může mít trojí výstup – menší - rovná se – větší než
- Situace nerovná se je tedy souhrnem dvou možných výstupů testu (menší+větší)
- Významově neutrální



Nepárový vs. párový design

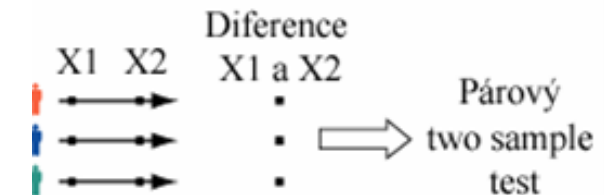
• Nepárový design

- Skupiny srovnávaných dat jsou na sobě zcela nezávislé (též nezávislý, independent design), např. lidé z různých zemí, nezávislé skupiny pacientů s odlišnou léčbou atd.
- Při výpočtu je nezbytné brát v úvahu charakteristiky obou skupin dat



• Párový design

- Mezi objekty v srovnávaných skupinách existuje vazba, daná např. člověkem před a po operaci, reakce stejného kmene krys atd.
- Vazba může být buď přímo dána nebo pouze předpokládána (v tom případě je nutné ji ověřit)
- Test je v podstatě prováděn na diferencích skupin, nikoliv na jejich původních datech



Důležité poznámky k testování hypotéz

- **Nezamítnutí nulové hypotézy neznamená automaticky její přijetí!** Může se jednat o situaci, kdy pro zamítnutí nulové hypotézy nemáme dostatečné množství informace.
- **Dosažená hladina významnosti testu** (ať už 5 %, 1 % nebo 10 %) **nesmí být slepě brána jako hranice pro existenci / neexistenci testovaného efektu.**
- **Malá p-hodnota nemusí znamenat velký efekt.** Hodnota testové statistiky a p-hodnota mohou být ovlivněny velkou velikostí vzorku a malou variabilitou pozorovaných dat.
- **Na výsledky testování musí být nahlíženo kriticky** – jedná se o závěr založený „pouze“ na jednom výběrovém souboru.
- **Statistická významnost** indikuje, že pozorovaný rozdíl není náhodný, ale nemusí znamenat, že je významný i ve skutečnosti. Důležitá je i **praktická (klinická) významnost.**

Statistické testy a normalita

- Normalita dat je jedním z předpokladů tzv. parametrických testů (testů založených na předpokladu nějakého rozložení) – např. t-testy
- Obecně lze říci, že každá statistická metoda, v jejímž algoritmu je obsažen výpočet průměru nebo směrodatné odchylky má předpoklad normálního rozložení
- Pokud data nejsou normální, neodpovídají ani modelovému rozložení, které je použito pro výpočet (t-rozložení) a test tak může lhát
- Řešením je tedy:
 - Transformace dat za účelem dosažení normality jejich rozložení
 - Neparametrické testy – tyto testy nemají předpoklady o rozložení dat (nebo jen minimální)

Typ srovnání	Parametrický test	Neparametrický test
2 skupiny dat nepárově:	Nepárový t-test	Mannův-Whitneyho test
2 skupiny dat párově:	Párový t-test	Wilcoxonův test, znaménkový test
Více skupin nepárově:	ANOVA (analýza rozptylu)	Kruskalův- Wallisův test
Korelace:	Pearsonův koeficient	Spearmanův koeficient

Testy normality

- Testy normality pracují s nulovou hypotézou, že není rozdíl mezi zpracovávaným rozložením a normálním rozložením. Vždy je ovšem dobré prohlédnout si i histogram, protože některé odchylky od normality, např. bimodalitu některé testy neodhalí.

Chí-kvadrát test dobré shody

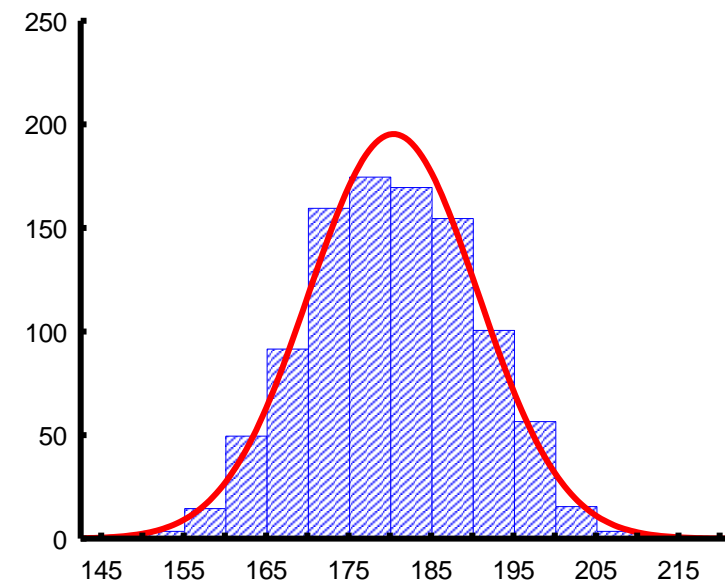
- V testu dobré shody jsou data rozdělena do kategorií (obdobně jako při tvorbě histogramu), tyto intervaly jsou normalizovány (převedeny na normální rozložení) a podle obecných vzorců normálního rozložení jsou k nim dopočítány očekávané hodnoty v intervalech, pokud by rozložení bylo normální. Pozorované normalizované četnosti jsou poté srovnány s očekávanými četnostmi pomocí χ^2 testu dobré shody. Test dává dobré výsledky, ale je náročný na n , tedy množství dat, aby bylo možné vytvořit dostatečný počet tříd hodnot.

Kolmogorovův - Smirnovův test

- Tento test je často používán, dokáže dobře najít odlehlé hodnoty, ale počítá spíše se symetrií hodnot než přímo s normalitou. Jde o neparametrický test pro srovnání rozdílu dvou rozložením. Je založen na zjištění rozdílu mezi reálným kumulativním rozložením (vzorek) a teoretickým kumulativním rozložením. Měl by být počítán pouze v případě, že známe průměr a směrodatnou odchylku hypotetického rozložení, pokud tyto hodnoty neznáme, měla by být použita jeho modifikace – Lilieforsův test.

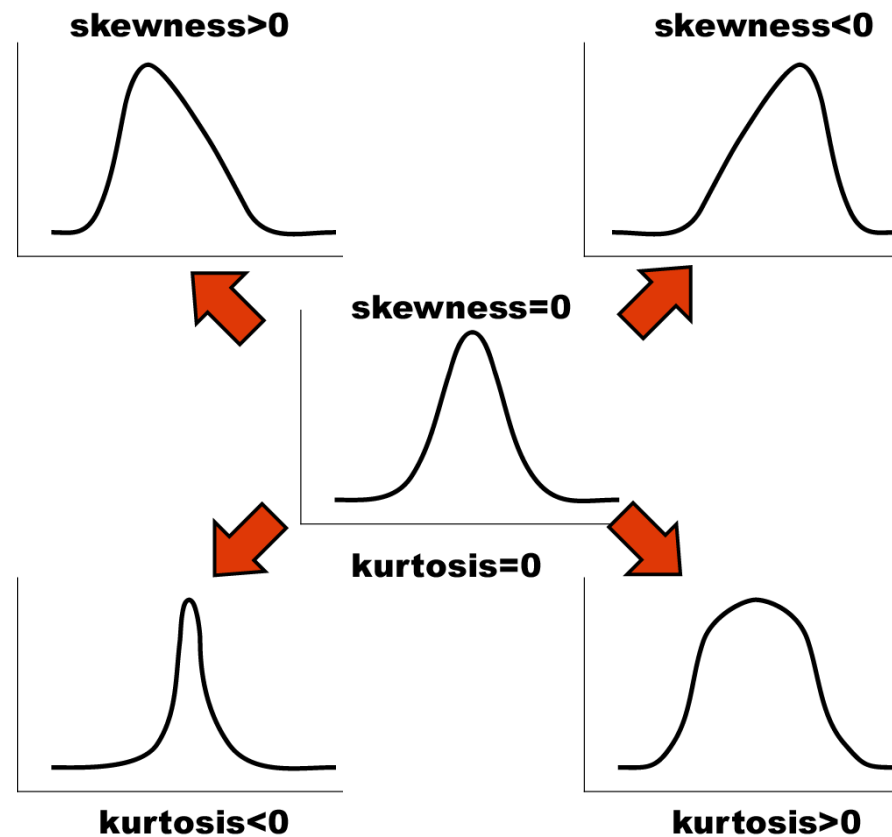
Shapiro-Wilkův test

- Jde o neparametrický test použitelný i při velmi malých n (10) s dobrou silou testu, zvláště ve srovnání s alternativními typy testů, je zaměřen na testování symetrie.

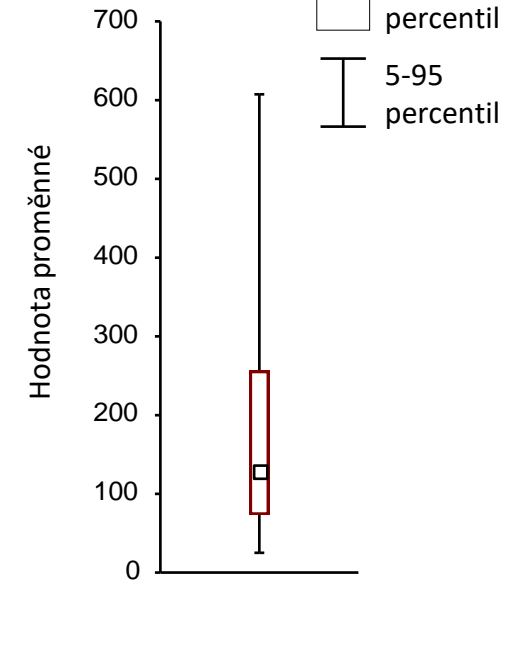
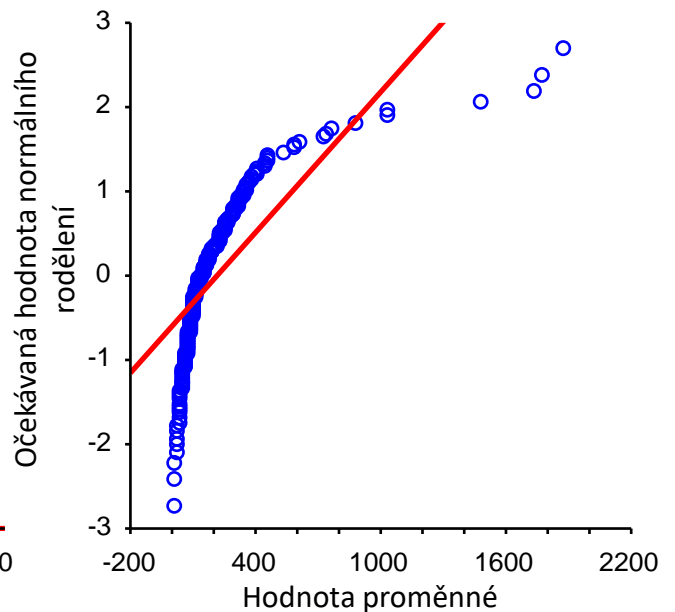
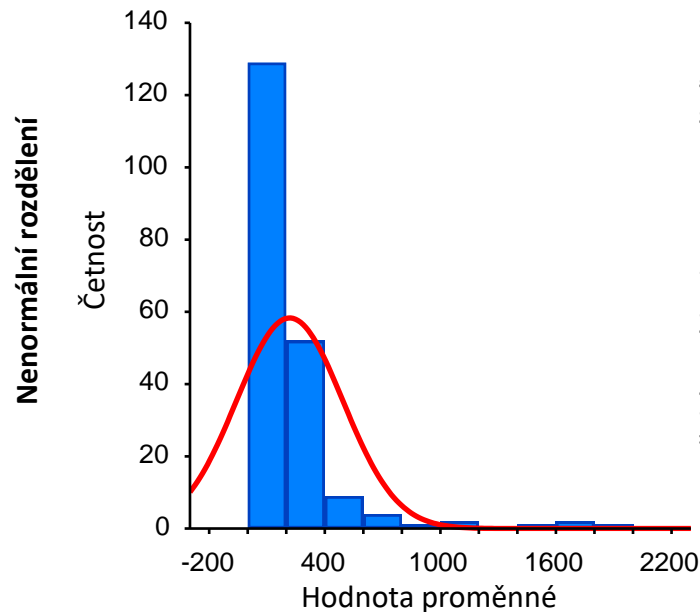
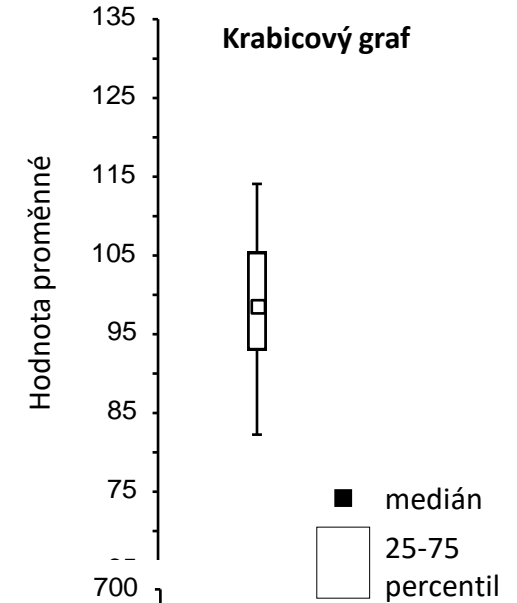
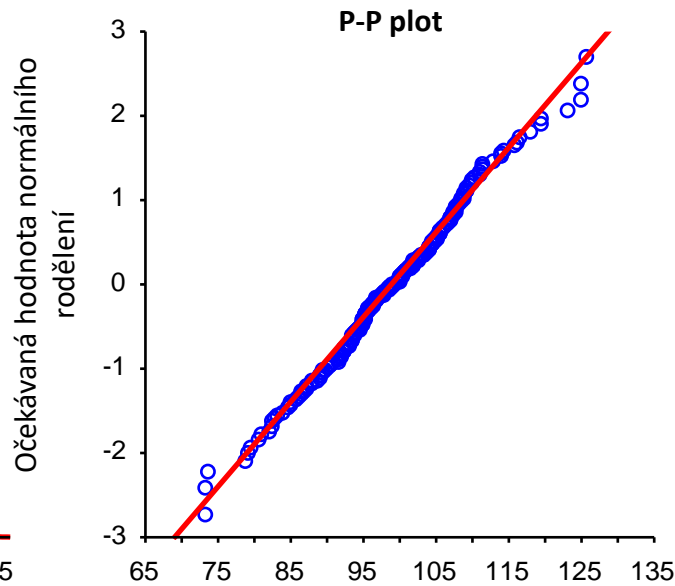
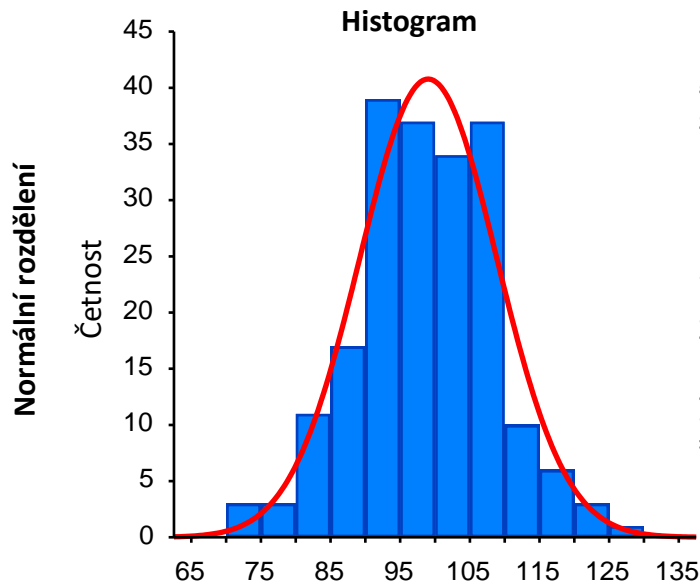


Šikmost a špičatost jako testy normality

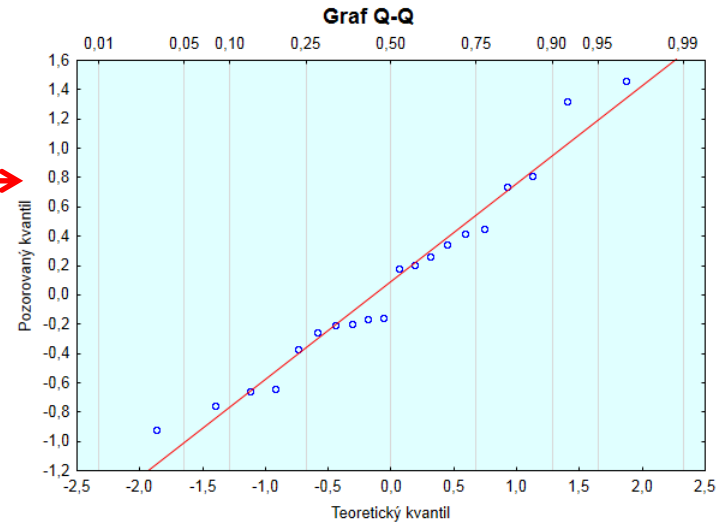
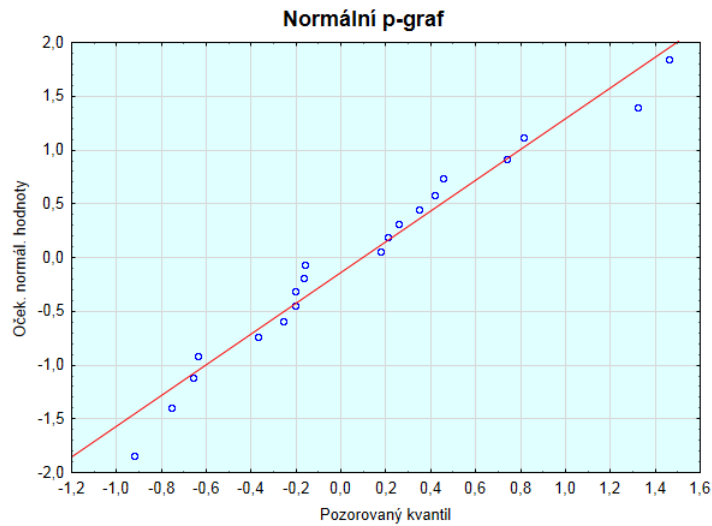
- Parametry normálního rozdělení, skewness a kurtosis mohou být využity pro testování normality, ale pouze pro velké vzorky (šikmost – 100, špičatost – 500).



Vizuální hodnocení normality I

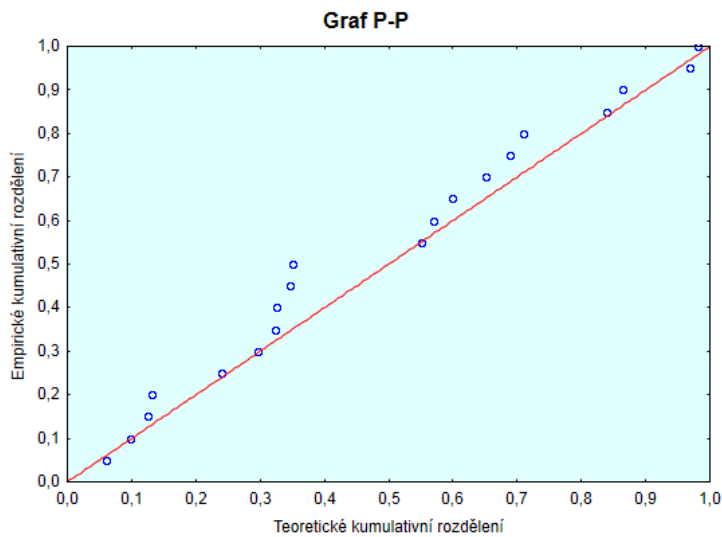


Vizuální hodnocení normality II



???

- Pouze výměna os
- Znázorněn pozorovaný a teoretický kvantil



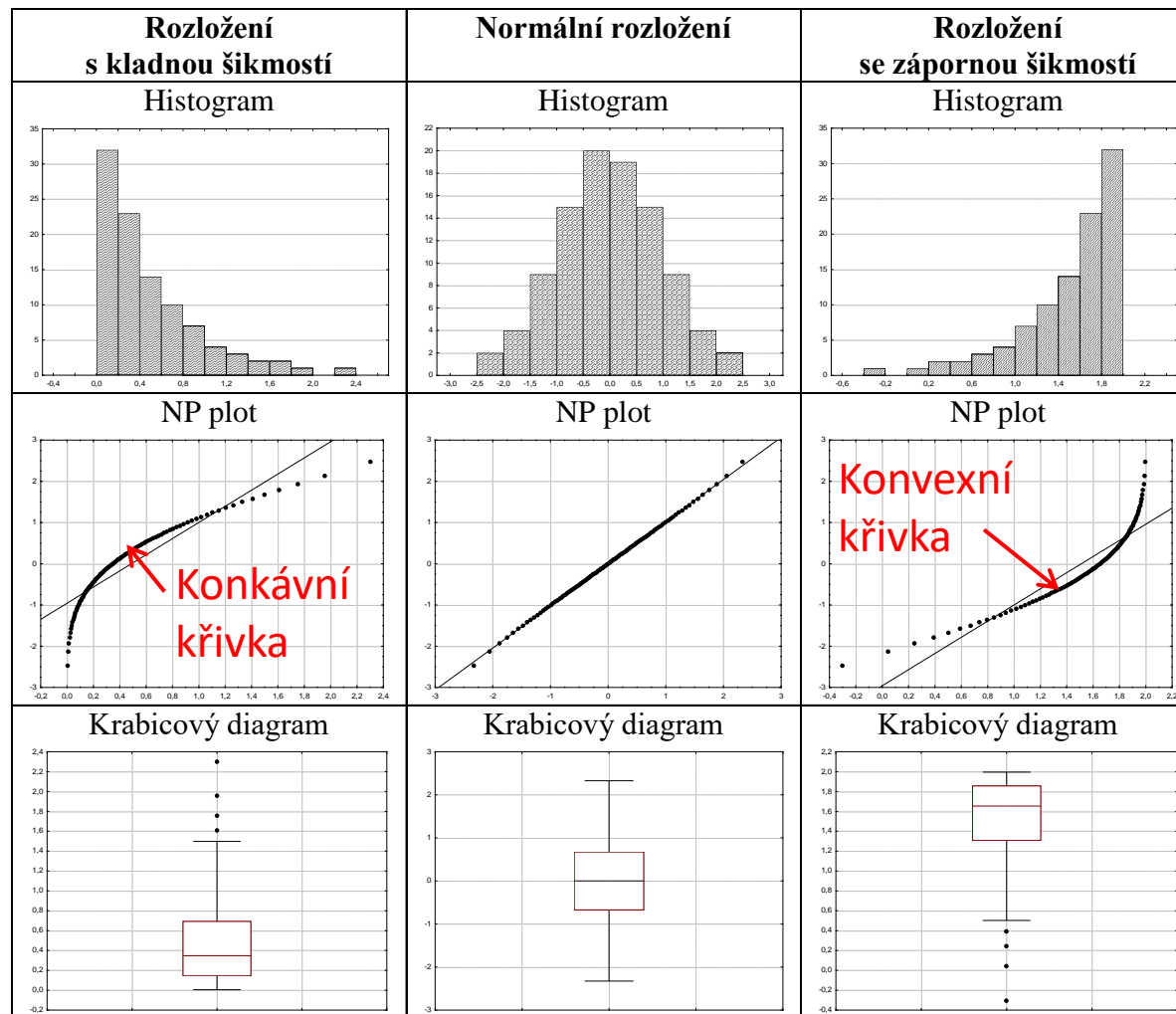
- Vykresleno kumulativní rozdělení

PAMATUJ:

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



Vizuální hodnocení normality III



Výukové materiály: Výpočetní statistika,
RNDr. Marie Budíková, Dr., 2011

Parametrické jednovýběrové statistické testy

Jednovýběrový t-test

Jednovýběrový test rozptylu

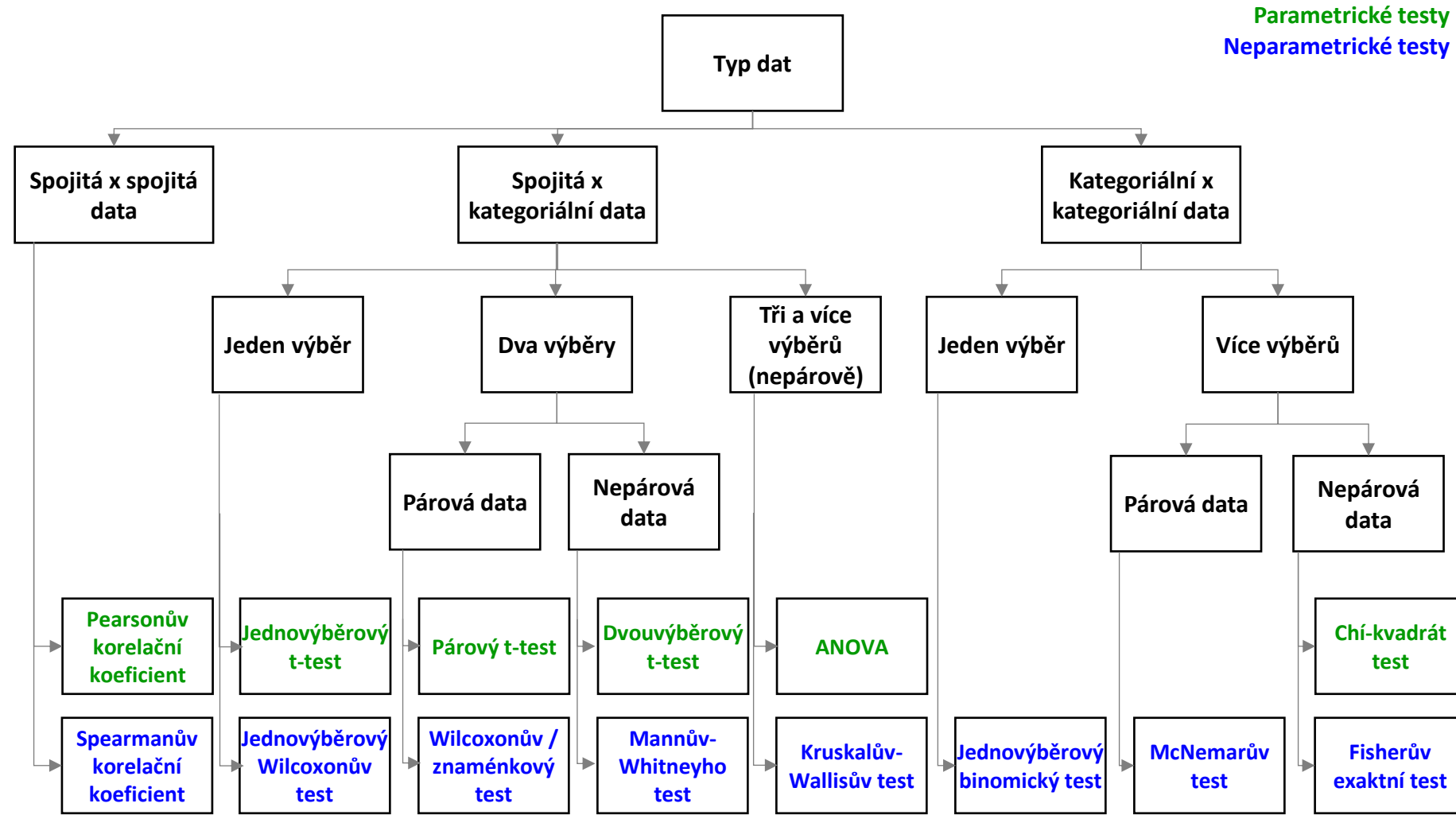
Anotace

- Jednovýběrové statistické testy srovnávají některou popisnou statistiku vzorku (průměr, směrodatnou odchylku) s jediným číslem, jehož význam je ze statistické hlediska hodnota cílové populace
- Z hlediska statistické teorie jde o ověření, zda daný vzorek pochází z testované cílové populace.

Shrnutí statistických testů

Typ srovnání	Nulová hypotéza	Parametrický test	Neparametrický test
1 výběr dat vs. referenční hodnota	Střední hodnota je rovna zvolené referenční hodnotě.	jednovýběrový t-test / z-test	Jednovýběrový Wilcoxonův test
2 nezávislé skupiny dat (test shody středních hodnot)	Střední hodnoty se mezi skupinami neliší.	nepárový t-test	Mannův-Whitneyho test
2 nezávislé skupin dat (test shody rozptylů = homoskedasticity)	Rozptyl obou skupin je shodný.	F-test	Levenův test
2 párově závislé výběry dat	Rozdíl (diference) párových hodnot je nulový.	párový t-test	Wilcoxonův test; znaménkový test
Shoda rozdělení výběru s teoretickým rozdělením	Rozdělení dat odpovídá teoretickému (vybranému) rozdělení.	test dobré shody (χ^2 test)	Shapirův-Wilkův test; Kolmogorovův-Smirnovův test; Lilieforsův test
3 a více skupin nepárově (test shody středních hodnot)	Střední hodnoty se mezi skupinami neliší.	ANOVA	Kruskalův-Wallisův test
Korelace	Neexistuje vztah mezi hodnotami dvou výběrů.	Pearsonův korelační koeficient	Spearmanův korelační koeficient

Základní rozhodování o výběru statistických testů



Parametrické jednovýběrové testy

- Předpoklad: **normalita dat**
- **Jednovýběrový z-test** (porovnání základního a výběrového souboru, známe střední hodnotu a rozptyl základního souboru)
- **Studentův jednovýběrový t-test** (testování rozdílů dvou středních hodnot) - (porovnání základního a výběrového souboru, známe střední hodnotu ale neznáme rozptyl základního souboru; nahrazujeme jej výběrovým rozptylem našich dat)
- **Chi-kvadrát test** (testování rozdílu cílová vs. výběrová populace)

Jednovýběrový z a t test

- V případě jednovýběrových testů jde o srovnání výběru dat (tedy one sample) s cílovou populací. Pro parametrické testy musí mít datový soubor normální rozložení.
- Rozdíl mezi jednovýběrovým z-testem a t-testem spočívá ve znalosti rozptylu základního souboru (z-test) nebo jeho nahrazení výběrovým rozptylem našich dat (t-test)

z-test:

$$z = \frac{\bar{x} - \mu}{\rho} \sqrt{N}$$

H_0	H_A	Testová statistika	Kritická hodnota
$\bar{x} \leq \mu$	$\bar{x} > \mu$	z / t	$z > z_{1-\alpha} / t > t_{1-\alpha}^{N-1}$

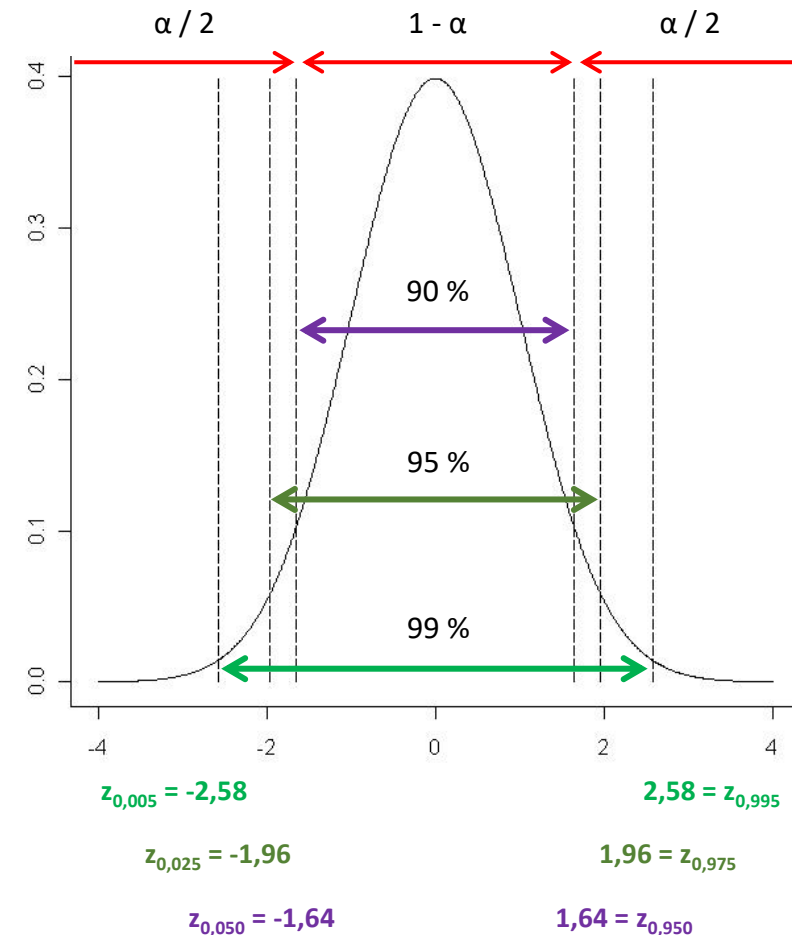
t-test:

$$t = \frac{\bar{x} - \mu}{s} \sqrt{N}$$

$\bar{x} \geq \mu$	$\bar{x} < \mu$	z / t	$z < z_{\alpha} / t < t_{\alpha}^{N-1}$
$\bar{x} = \mu$	$\bar{x} \neq \mu$	z / t	$ z > z_{1-\alpha/2} / t > t_{1-\alpha/2}^{N-1}$

Příklad: z-test pro jeden výběr 1

- Při populačním epidemiologickém průzkumu se zjistilo, že průměrný objem prostaty u mužů je 32,73 ml (SD = 18,12 ml).
- Na hladině významnosti testu $\alpha = 0,05$ chceme ověřit, jestli se muži nad 70 let liší od celé populace.
- Máme náhodný výběr o velikosti $n = 100$ a výběrový průměr 36,60 ml.
- Chceme ověřit platnost:
 - $H_0: \mu = 32,73$
 - $H_A: \mu \neq 32,73$



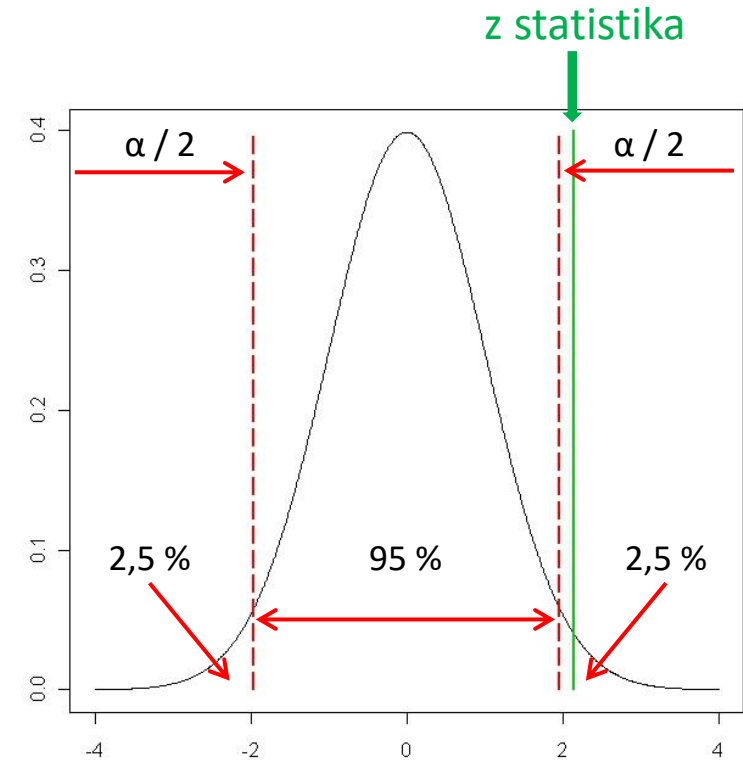
Příklad: z-test pro jeden výběr 2

- Hodnota testové statistiky:
$$z = \frac{\bar{x} - \mu}{\rho} \sqrt{N} = \frac{36,60 - 32,73}{18,12} \sqrt{100} = 2,14$$

- Můžeme zamítnout nulovou hypotézu na hladině významnosti testu $\alpha = 0,05$ nebo ne?

$$z = 2,14 > 1,96 \quad (z_{1-\alpha/2} = z_{0,975})$$

- Nulovou hypotézu o rovnosti objemu prostaty u mužů nad 70 let populační hodnotě 32,73 ml zamítáme na hladině významnosti $\alpha = 0,05$, protože výsledná hodnota z statistiky je větší než kritická hodnota (příslušný kvantil) rozdělení $N(0,1)$.



Příklad: t-test pro jeden výběr



- Určitá linka autobusové městské dopravy má v době dopravní špičky průměrnou rychlost 8 km/hod. Uvažovalo se o tom, zda změna trasy by vedla ke změně průměrné rychlosti. Nová trasa byla proto projeta v deseti náhodně vybraných dnech a byly zjištěny tyto průměrné rychlosti: 8,4; 7,9; 9,0; 7,8; 8,0; 7,8; 8,5; 8,2; 8,2; 9,3. Rozhodněte, zda změna trasy vede ke změně průměrné rychlosti. Předpokládáme normální rozdělení a $\alpha=0,05$.

- Postup:

1. Na hladině významnosti 0,05 testujeme hypotézu $H_0: \mu = 8$, proti $H_A: \mu \neq 8$

2. Vypočteme aritmetický průměr a rozptyl výběrového souboru.

3. Vypočteme testovou statistiku t:
$$t = \frac{\bar{x} - \mu}{s} \sqrt{N} = \frac{8,310 - 8}{0,507} \sqrt{10} = 1,934$$

4. Vypočtené t porovnáme s kritickou hodnotou:

$$t_{1-\alpha/2}^{N-1} = t_{0,975}^9 = 2,262$$

5. Je-li $|t| \leq t_{1-\alpha/2}^{N-1}$ - > statisticky nevýznamný rozdíl testovaných parametrů při zvolené α ; **nulovou hypotézu nezamítáme**, na hladině významnosti $\alpha=0,05$ se nepodařilo prokázat, že by změna trasy měla za následek změnu průměrné rychlosti.

Typické výstupy SW (Statistica, v jiných obdobné)

Výběrový průměr
(průměr pozorovaných dat)

Rozsah výběru

Standardní chyba

Hodnota testovacího kritéria


Stupeň volnosti

Test of means against reference constant (value) (04_doprava.sta)								
Variable	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
rychlost	8,310000	0,506513	10	0,160174	8,000000	1,935401	9	0,084934

Výběrová směrodatná odchylka
(pozorovaných dat)

Referenční konstanta-předpokládaná velikost střední hodnoty

POZOR: Platí pro oboustranný test!!!



Příklad k řešení: t-test 1

Data - koncentrace antibiotika v cílovém orgánu

- Při 1000 měřeních antibiotika byla zjištěna v cílovém orgánu průměrná koncentrace 202,5 jednotek a směrodatná odchylka 44 jednotek.
- Požadovaná koncentrace antibiotika je 200 jednotek.

Výzkumné otázky

1. Je daný rozdíl 2,5 významný vzhledem k variabilitě znaku na hladině významnosti 5%?
2. Jaká je skutečná hladina významnosti?

Příklad k řešení: t-test 1

Data - koncentrace antibiotika v cílovém orgánu

- Při 1000 měřeních antibiotika byla zjištěna v cílovém orgánu průměrná koncentrace 202,5 jednotek a směrodatná odchylka 44 jednotek.
- Požadovaná koncentrace antibiotika je 200 jednotek.

Výzkumné otázky

1. Je daný rozdíl 2,5 významný vzhledem k variabilitě znaku na hladině významnosti 5%?
2. Jaká je skutečná hladina významnosti?

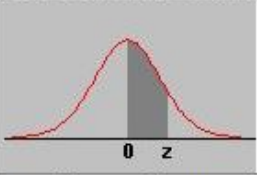
$$t = \frac{\bar{x} - \mu}{s} \sqrt{N} = \frac{202,5 - 200}{44} \sqrt{1000} = 1,797$$

Příklad k řešení: t-test 1

$$t = \frac{\bar{x} - \mu}{s} \sqrt{N} = \frac{202,5 - 200}{44} \sqrt{1000} = 1,797 \sim 1,8$$

$$t_{1-\alpha/2}^{N-1} = t_{0,975}^{999} = 1,960$$

Area between 0 and z



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706

Výzkumné otázky

1. Je daný rozdíl 2,5 významný vzhledem k variabilitě znaku na hladině významnosti 5%?
 - **Nulovou hypotézu nezamítáme**
2. Jaká je skutečná hladina významnosti?
 - **$p = 2 * (1 - 0,4641) = 0,072$**

Příklad k řešení: t-test 2

Data - aktivita enzymu v buňkách

- Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

Výzkumné otázky

1. otázka zní, zda se naměřené hodnoty našeho vzorku liší od výsledků dřívější rozsáhlé studie zaměřené na celou cílovou populaci, kde byla zjištěna průměrná aktivita 2,5 jednotky?
2. otázka – jakou minimální odchylku X od jiné hodnoty bychom zachytili při daných hodnotách?
3. za předpokladu, že z praktického hlediska je významná odchylka již 0,2 jednotky, jaký minimální počet měření musíme provést, abychom ji byli schopni prokázat ?

Příklad k řešení: t-test 2

Data - aktivita enzymu v buňkách

- Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

Výzkumné otázky

1. otázka zní, zda se naměřené hodnoty našeho vzorku liší od výsledků dřívější rozsáhlé studie zaměřené na celou cílovou populaci, kde byla zjištěna průměrná aktivita 2,5 jednotky?

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n} = \frac{3,5 - 2,5}{1} \sqrt{25} = 5$$

$$t_{0,975}^{24} = 2,064 \Rightarrow t > t_{1-\alpha/2}^{24} \Rightarrow H_0 \text{ zamítnuta při } \alpha \leq 0,05$$

Příklad k řešení: t-test 2

Data - aktivita enzymu v buňkách

- Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

Výzkumné otázky

2. otázka – jakou minimální odchylku X od jiné hodnoty bychom zachytili při daných hodnotách?

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n} = \frac{d}{s} \sqrt{n} \rightarrow d = \frac{t_{1-\alpha/2}^v}{\sqrt{n}} s \rightarrow d = \frac{2,064}{5} 1$$

Příklad k řešení: t-test 2

Data - aktivita enzymu v buňkách

- Při zjišťování aktivity enzymu v buňkách na vzorku 25 měření byl zjištěn průměr 3,5 jednotek a směrodatná odchylka 1.

Výzkumné otázky

3. za předpokladu, že z praktického hlediska je významná odchylka již 0,2 jednotky, jaký minimální počet měření musíme provést, abychom ji byli schopni prokázat ?

$$t = \frac{\bar{x} - \mu}{s} \sqrt{n} = \frac{d}{s} \sqrt{n} \rightarrow n = \left(\frac{t_{1-\alpha/2}^v}{d} s \right)^2$$

Jednovýběrový test pro rozptyl

- V případě one sample testů jde o srovnání výběru dat (tedy one sample) s cílovou populací. Pro parametrické testy musí mít datový soubor normální rozložení.

Chi-kvadrát test:

$$\chi^2 = \frac{(N - 1)s^2}{\sigma^2}$$

H_0	H_A	Testová statistika	Kritická hodnota
$s^2 \leq \sigma^2$	$s^2 > \sigma^2$	χ^2	$\chi^2 > \chi_{1-\alpha}^{2(N-1)}$
$s^2 \geq \sigma^2$	$s^2 < \sigma^2$	χ^2	$\chi^2 < \chi_{\alpha}^{2(N-1)}$
$s^2 = \sigma^2$	$s^2 \neq \sigma^2$	χ^2	$\chi^2 > \chi_{1-\alpha/2}^{2(N-1)}$ nebo $\chi^2 < \chi_{\alpha/2}^{2(N-1)}$

Neparametrické jednovýběrové statistické testy

Jednovýběrový t-test

Jednovýběrový test rozptylu

Parametrické vs. neparametrické testy

Parametrické testy

- Mají předpoklady o rozložení vstupujících dat (např. normální rozložení)
- Při stejném N a dodržení předpokladů mají vyšší sílu testu než testy neparametrické
- **Pokud nejsou dodrženy předpoklady parametrických testů, potom jejich síla testu prudce klesá a výsledek testu může být zcela chybný a nesmyslný**



Neparametrické testy

- Vyžadují méně předpokladů o rozložení vstupujících dat, lze je tedy použít i při asymetrickém rozložení, odlehlých hodnotách, či nedetekovatelném rozložení
- Snížená síla těchto testů je způsobena redukcí informační hodnoty původních dat, kdy neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí
- Souvisí s malou velikostí souboru (nejsme schopni normalitu dat ověřit)

Proč nemusí parametrický a neparametrický test vyjít stejně?

Jednovýběrový Wilcoxonův test

- Předpokladem je symetrické rozdělení dat kolem mediánu.
- Testuje, zda je **medián** jednoho výběru roven hodnotě c (v případě párového designu je $x_{0.5}$ reprezentováno mediánem rozdílu hodnot)

$$H_0: x_{0.5} = c \text{ proti } H_1: x_{0.5} \neq c.$$

Postup:

1. Spočítáme rozdíly hodnot výběru s testovanou hodnotou mediánu.
2. Absolutní hodnoty rozdílů uspořádáme vzestupně a přiřadíme jim pořadí.
3. Spočítáme statistiky S_w^+ a S_w^- , které odpovídají **součtu pořadí kladných (S_w^+) a záporných rozdílů (S_w^-)**. Jako finální hodnotu testové statistiky bereme minimum z S_w^+ a S_w^- . Nulovou hypotézu zamítáme, pokud hodnota testové statistiky menší nebo rovna tabelované kritické hodnotě (při dané hladině významnosti a počtu nenulových rozdílů).

nebo

3. Pro $N > 30$ lze využít asymptotické normality statistiky S_w^+

$$E(S_w^+) = \frac{n(n+1)}{4}$$

$$D(S_w^+) = \frac{n(n+1)(2n+1)}{24}$$

$$Z = \frac{S_w^+ - E(S_w^+)}{\sqrt{D(S_w^+)}} \approx N(0,1)$$

- Pokud $|Z| \geq u_{1-\alpha/2}$ zamítáme nulovou hypotézu, že medián výběru je roven hodnotě c .

Jednovýběrový znaménkový test

- Lze použít v situaci, kdy není splněn předpoklad symetrie rozdělení kolem mediánu.
- Testuje, zda je medián jednoho výběru roven hodnotě c (v případě párového designu je $x_{0,5}$ reprezentováno mediánem rozdílu hodnot)

$$H_0: x_{0,5}=c \text{ proti } H_1: x_{0,5} \neq c.$$

Postup:

1. Spočítáme rozdíly hodnot výběru s testovanou hodnotou mediánu.
 2. Spočítáme statistiku S_z^+ , která odpovídá počtu kladných rozdílů → **test nevyužívá hodnot pořadí původních dat ale pouze informaci, zda se hodnota realizuje nad nebo pod mediánem** → dochází ke snížení síly testu
 3. Nulovou hypotézu zamítáme, pokud statistika S_z^+ realizuje v kritickém oboru hodnot $W=(0,k_1)U(k_2,n)$, kde n odpovídá počtu nenulový rozdílů a hodnoty k_1 a k_2 lze dohledat v matematických tabulkách.
- **nebo**
3. Pro $N > 20$ lze využít asymptotické normality statistiky S_z^+ .

$$E(S_z^+) = \frac{n}{2} \quad D(S_z^+) = \frac{n}{4} \quad Z = \frac{S_z^+ - E(S_z^+)}{\sqrt{D(S_z^+)}} \approx N(0,1)$$

Pokud $|Z| \geq u_{1-\alpha/2}$ zamítáme nulovou hypotézu, že medián výběru je roven hodnotě c .

Příklad: jednovýběrový test


- U 15 náhodně vybraných pacientů byla vyhodnocena doba, kterou museli strávit v čekárně, než byli sestrou pozváni do ordinace. Na 5% hladině významnosti testujte nulovou hypotézu, že medián čekací doby je roven půl hodině.



Příklad: jednovýběrový test – Wilcoxonův test

- U 15 náhodně vybraných pacientů byla vyhodnocena doba, kterou museli strávit v čekárně, než byli sestrou pozváni do ordinace. Na 5% hladině významnosti testujte nulovou hypotézu, že medián čekací doby je roven půl hodině.

Pacient č.	čekací doba (min)	medián	rozdíl	rozdíl	pořadí
1	1	30	-29	29	15
2	45	30	15	15	10
3	25	30	-5	5	3.5
4	15	30	-15	15	10
5	34	30	4	4	2
6	19	30	-11	11	8
7	31	30	1	1	1
8	25	30	-5	5	3.5
9	8	30	-22	22	14
10	12	30	-18	18	12
11	20	30	-10	10	6
12	15	30	-15	15	10
13	40	30	10	10	6
14	20	30	-10	10	6
15	10	30	-20	20	13


$$S_w^+ = 19$$
$$S_w^- = 101$$

$$\min(S_w^+, S_w^-) = 19$$

Kritická hodnota $w_{15}(0,05) = 25$

Hodnota testové statiky je menší než kritická hodnota → **zamítáme H_0**

Příklad: jednovýběrový test – Znaménkový test

- U 15 náhodně vybraných pacientů byla vyhodnocena doba, kterou museli strávit v čekárně, než byli sestrou pozváni do ordinace. Na 5% hladině významnosti testujte nulovou hypotézu, že medián čekací doby je roven půl hodině.

Pacient č.	čekací doba (min)	medián	rozdíl	Větší než medián?
1	1	30	-29	Ne
2	45	30	15	Ano
3	25	30	-5	Ne
4	15	30	-15	Ne
5	34	30	4	Ano
6	19	30	-11	Ne
7	31	30	1	Ano
8	25	30	-5	Ne
9	8	30	-22	Ne
10	12	30	-18	Ne
11	20	30	-10	Ne
12	15	30	-15	Ne
13	40	30	10	Ano
14	20	30	-10	Ne
15	10	30	-20	Ne



$$S_z^+ = 4$$

Kritický obor: $W = (0, 3) \cup (12, 15)$

Hodnota statistiky se realizuje mimo kritický obor hodnot → **nezamítáme H_0**

Příklad: Řešení v softwaru

1) Výstup Wilcoxonova testu

Testová statistika: $\min(S_w^+, S_w^-)$

Wilcoxon Matched Pairs Test (v cekarne.sta)				
Marked tests are significant at p < .05000				
Pair of Variables	Valid N	T	Z	p-value
doba & median	15	19.00000	2.328644	0.019879

Statistika a p-hodnota pro asymptotickou variantu testu (používat pouze pro $N > 30$)

Počet nenulových rozdílů

2) Výstup znaménkového testu

Podíl hodnot menších než testovaný medián

Sign Test (v cekarne.sta)				
Marked tests are significant at p < .05000				
Pair of Variables	No. of Non-ties	Percent v < V	Z	p-value
doba & median	15	73.33333	1.549193	0.121335

Statistika a p-hodnota pro asymptotickou variantu testu (používat pouze pro $N > 20$)

Počet nenulových rozdílů

Parametrické dvouvýběrové statistické testy

Dvouvýběrový nepárový t-test

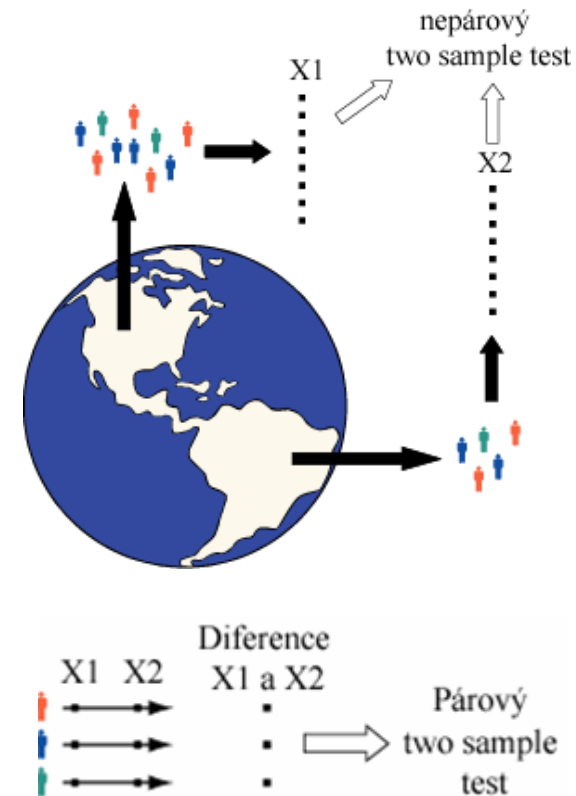
Dvouvýběrový párový t-test

Anotace

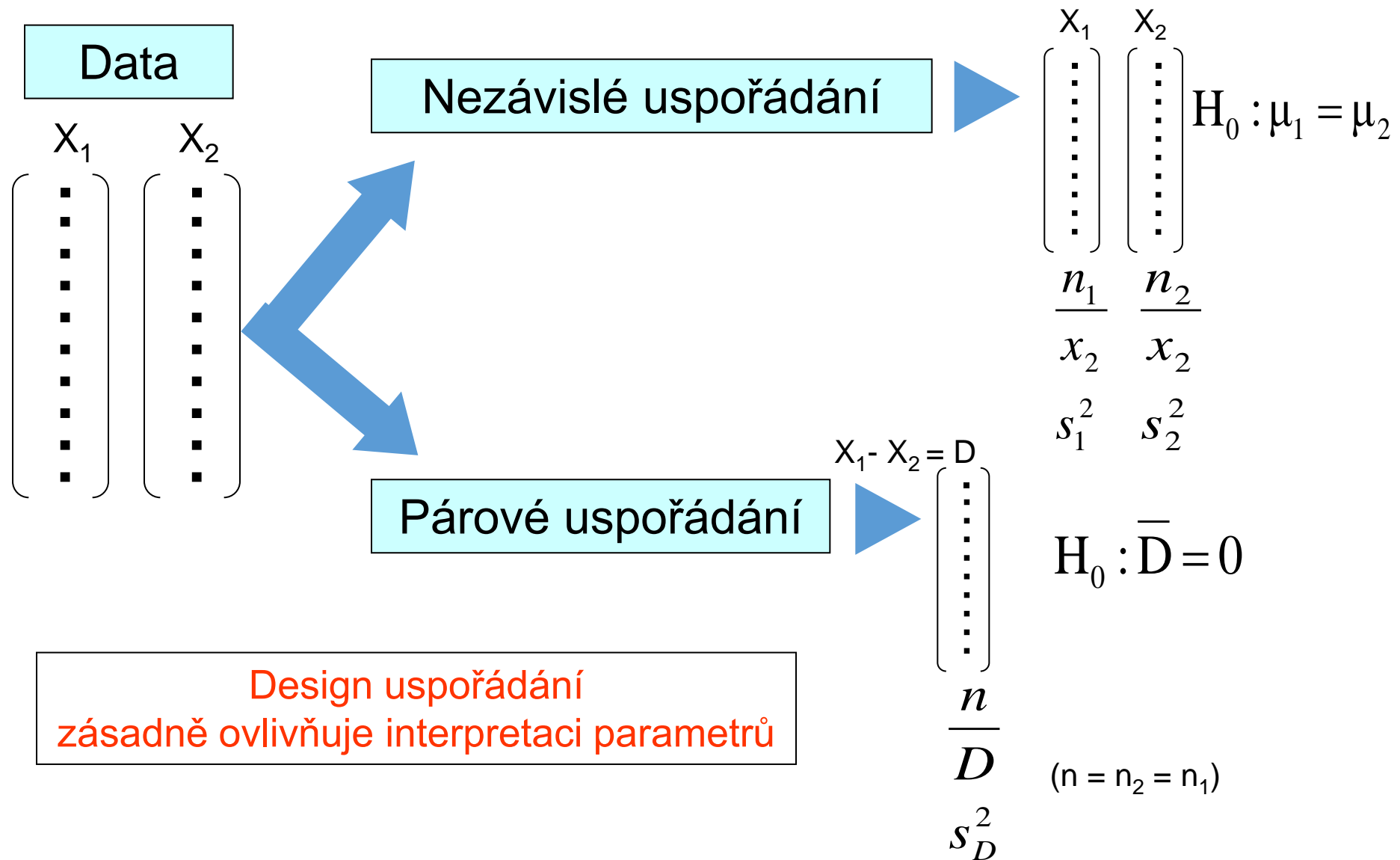
- Jedním z nejčastějších úkolů statistické analýzy dat je srovnání spojitých dat ve dvou skupinách pacientů.
- Na výběr je celá škála testů, výběr konkrétního testu se pak odvíjí od toho, zda je o srovnání párové nebo nepárové a zda je vhodné použít test parametrický (má předpoklady o rozložení dat) nebo neparametrický (nemá předpoklady o rozložení dat, nicméně má nižší vypovídací sílu).
- Nejznámějšími testy z této skupiny jsou tzv. t-testy používané pro srovnání průměrů dvou skupin hodnot

Dvouvýběrové testy: párové a nepárové I

- Při použití two sample testů srovnáváme spolu dvě rozložení. Jejich základním dělením je podle designu experimentu na testy párové a nepárové.
- Základním testem pro srovnání dvou nezávislých rozložení spojitých čísel je nepárový two-sample t-test
- Základním testem pro srovnání dvou závislých rozložení spojitých čísel je párový two-sample t-test

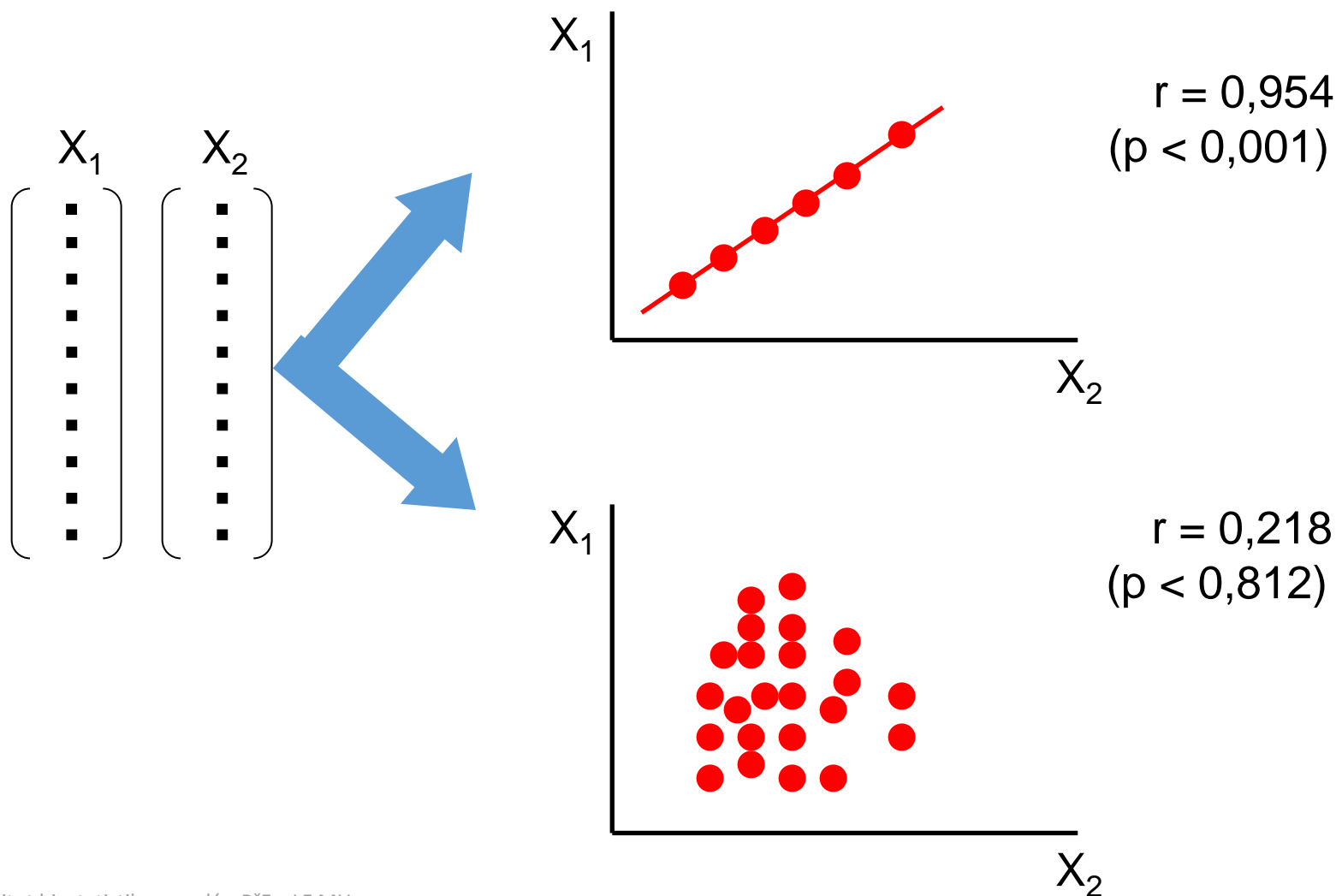


Dvouvýběrové testy: párové a nepárové II



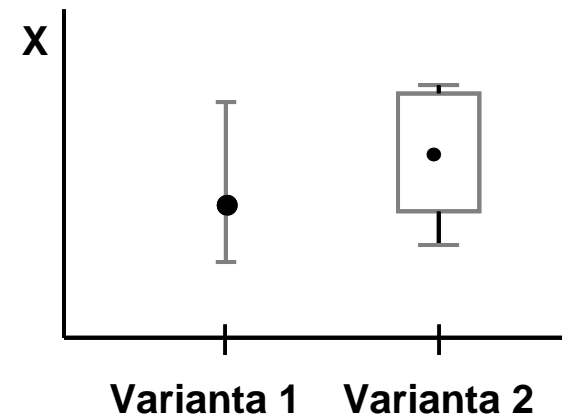
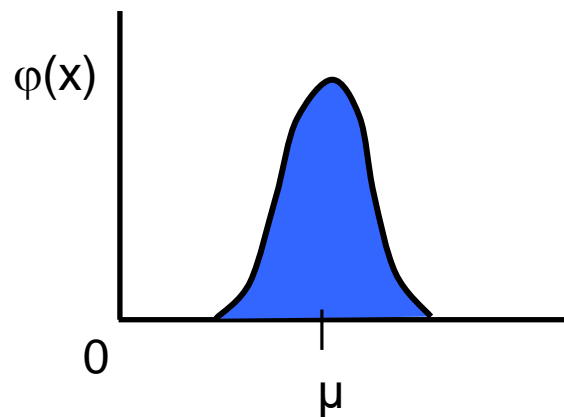
Dvouvýběrové testy: párové a nepárové III

- Identifikace párovitosti (Korelace, Kovariance)



Předpoklady nepárového dvouvýběrového t-testu

- Náhodný výběr subjektů jednotlivých skupin z jejich cílových populací
- Nezávislost obou srovnávaných vzorků
- Přibližně normální rozložení proměnné ve vzorcích, drobné odchylky od normality ovšem nejsou kritické, test je robustní proti drobným odchylkám od tohoto předpokladu, normalita může být testována testy normality
- Rozptyl v obou vzorcích by měl být přibližně shodný (homoscedastic). Tento předpoklad je testován několika možnými testy – Levenův test nebo F-test.
- Vždy je vhodné prohlédnout histogramy proměnné v jednotlivých vzorcích pro okometrické srovnání a ověření předpokladů normality a homogenity rozptylu – nenahradí statistické testy, ale poskytne prvotní představu.



Nepárový dvouvýběrový t-test – výpočet I

- nulová hypotéza: průměry obou skupin jsou shodné, alternativní hypotéza je, že nejsou shodné, two tailed test
- prohlédnout průběh dat, průměr, medián apod. pro zjištění odchylek od normality a nehomogenita rozptylu, provést F –test
- F-test pro srovnání dvou výběrových rozptylů
 - Používá se pro srovnání rozptylu dvou skupin hodnot, často za účelem ověření homogenity rozptylu těchto skupin dat.
- V případě ověření homogenity je testována hypotéza shody rozptylů (two tailed); v případě shodných rozptylů je vše v pořádku a je možné pokračovat ve výpočtu t-testu, v opačném případě není vhodné test počítat.

H_0	H_A	Testová statistika
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{s_1^2}{s_2^2}$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$F = \frac{s_2^2}{s_1^2}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$F = \frac{\max(s_1^2; s_2^2)}{\min(s_1^2; s_2^2)}$

Nepárový dvouvýběrový t-test – výpočet II

- Výpočet testové statistiky (stupně volnosti jsou $\nu = n_1 + n_2 - 2$):

$$t = \frac{\text{Rozdíl} - \text{průměr}}{SE(\text{rozdíl průměrů})} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
 vážený odhad rozptylu

- výsledné t srovnáme s tabulární hodnotou t pro dané stupně volnosti a α (obvykle $\alpha=0,05$)
- Lze spočítat interval spolehlivosti pro rozdíl průměrů (např. 95%), počet stupňů volnosti a s^2 odpovídají předchozím vzorcům

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0,975} SE(\bar{x}_1 - \bar{x}_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{0,975} \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Nepárový dvouvýběrový t-test – výpočet shrnutí

- Nulová hypotéza: průměry obou skupin jsou shodné
 - Alternativní hypotéza je, že nejsou shodné.
- Prohlédnout průběh dat, průměr, medián apod.
 - Ověřit normalitu dat (např. Shapiro-Wilk test)
 - Ověřit homogenitu rozptylů (F-test)
 - V případě ověření homogeneity je testována hypotéza shody rozptylů; v případě shodných rozptylů je vše v pořádku a je možné pokračovat ve výpočtu t-testu, v opačném případě není vhodné test počítat.
- Vypočítat hodnotu testové statistiky a p-hodnotu. Když je vypočítaná p-hodnota menší než 0,05, zamítáme nulovou hypotézu.