

Analýza dat pro Neurovědy



RNDr. Eva Janoušová
doc. RNDr. Ladislav Dušek, Dr.

Jaro 2013

Blok 2

Jak medicínská data správně
testovat.

Osnova

1. Formulování hypotéz nad medicínskými daty
2. Hladina významnosti a síla testu
3. p-hodnota
4. Vhodná volba typu testu v různých situacích
5. Jednovýběrové testy
6. Párové testy

1. Formulování hypotéz nad medicínskými daty

Statistické testování

- **Cíle:**

1. Chceme srovnávat.
 - 1 náhodný výběr s předpokládanou hodnotou
 - 2 náhodné výběry mezi sebou
 - Více náhodných výběrů mezi sebou
2. Chceme hodnotit změnu náhodné veličiny vzhledem k vnějšímu zásahu.
3. Chceme rozhodovat o nezávislosti dvou náhodných veličin.
4. Chceme rozhodovat o charakteru rozdělení náhodné veličiny.

- **Postup:**

1. Máme danou hypotézu k ověření (např. pacienti a kontroly se liší v hodnotách MMSE skóre).
2. Provedli jsme výběr z populace.
3. Aplikujeme statistický test.
4. Hypotézu prohlásíme za statisticky platnou nebo neplatnou.

Hypotézy

- Hypotéza je tvrzení, které lze na základě pozorovaných dat ohodnotit ze statistického hlediska.
- **Nulová hypotéza** („null hypothesis“) – tvrzení, že se něco nestalo nebo neprojevilo (např. nepřítomnost rozdílu mezi sledovanými skupinami, nepřítomnost efektu léčby apod.) – tzn. tvrzení, že efekt je nulový – je to opak toho, co chceme experimentem prokázat.
- Nulová hypotéza má tvar: $H_0 : q = q_0$
- **Alternativní hypotéza** („alternative hypothesis“) – tvrzení, které popírá platnost nulové hypotézy – tzn. tvrzení, že efekt není nulový. Vymezuje, jaká situace nastává, když nulová hypotéza neplatí.
- Alternativní hypotéza má tvar:
 - $H_1 : q \neq q_0$ - oboustranná alternativa
 - $H_1 : q < q_0$ - jednostranná alternativa
 - $H_1 : q > q_0$ - jednostranná alternativa

Hypotézy – příklady

1. Je objem mozkových komor u **pacientů s Alzheimerovou chorobou větší než u zdravých lidí**? Označme střední hodnotu objemu komor u pacientů symbolem θ_1 a střední hodnotu objemu komor u zdravých lidí θ_2 .
Nulová hypotéza: $H_0 : q_1 = q_2$ Není rozdíl v objemu komor u pacientů a kontrol.
Alternativní hypotéza: $H_1 : q_1 > q_2$ Objemu komor větší u pacientů než u kontrol.
2. Je **průměrná hodnota MMSE skóre u pacientů s Alzheimer. chorobou menší než průměrná hodnota celé populace**? Označme střední hodnotu MMSE u pacientů symbolem θ_1 a u celé populace symbolem θ_0 .
Nulová hypotéza: $H_0 : q_1 = q_2$
Alternativní hypotéza: $H_1 : q_1 < q_0$
3. Liší se objem hipokampu u **pacientů s Alzheimer. chorobou (AD), pacientů s mírnou kognitivní poruchou (MCI) a zdravých lidí (CN)**? Označme střední hodnotu objemu hipokampu u jednotlivých skupin symboly θ_{AD} , θ_{MCI} , θ_{CN} .
Nulová hypotéza: $H_0 : q_{AD} = q_{MCI} = q_{CN}$
Alternativní hypotéza: $H_1 : \text{Nejméně jedno } \theta \text{ je odlišné od ostatních.}$

Proč nulová hypotéza vyjadřuje nepřítomnost efektu?

- Nulová hypotéza je formulována jako opak toho, co chceme experimentem prokázat, proto, že je vždy jednodušší zamítnout hypotézu (na to stačí jeden případ, že hypotéza neplatí) než potvrdit hypotézu.
- Pokud se nám nepodaří nulovou hypotézu vyvrátit (tedy zamítnout), **mluvíme o nezamítnutí nulové hypotézy, ne o přijetí nulové hypotézy!!!**
- Platnost nulové hypotézy ověřujeme pomocí **statického testu** – rozhodovací pravidlo, které pozorovaným datům přiřadí právě jedno ze dvou možných rozhodnutí: nulovou hypotézu H_0 na základě dat nezamítáme nebo nulovou hypotézu H_0 zamítáme.

2. Hladina významnosti a síla testu

Co se při rozhodování o platnosti H_0 může stát

- Máme čtyři možnosti výsledku rozhodovacího procesu o platnosti nulové hypotézy:

Rozhodnutí	Skutečnost	
	H_0 platí	H_0 neplatí
H_0 nezamítáme	správné přijetí platné nulové hypotézy	chyba II. druhu
H_0 zamítáme	chyba I. druhu	správné zamítnutí neplatné nulové hypotézy

- Při rozhodování se můžeme mýlit**, můžeme se dopustit dvou chybných úsudků:
 - **chyba I. druhu** – falešně pozitivní závěr testu – tzn. nesprávné zamítnutí nulové hypotézy (ve skutečnosti není rozdíl mezi skupinami, ale náš závěr z dat je opačný)
 - **chyba II. druhu** – falešně negativní závěr testu – tzn. nerozpoznání neplatné nulové hypotézy (rozdíl mezi skupinami skutečně existuje, my ho ale nejsme schopni na základě dat statisticky prokázat)

Analogie se soudním procesem

- Ctíme presumpci nevinny = předpokládáme, že nulová hypotéza platí.
- Požadujeme **důkaz pro prokázání viny** = na základě dat chceme ukázat, že nulová hypotéza neplatí.

Rozhodnutí	Skutečnost	
	H_0 platí	H_0 neplatí
H_0 nezamítneme	správné přijetí platné nulové hypotézy	chyba II. druhu
H_0 zamítneme	chyba I. druhu	správné zamítnutí neplatné nulové hypotézy

- Když nám bude stačit málo důkazů, zvýší se procento odsouzených nevinných = **chyba I. druhu**, ale zároveň se zvýší i procento odsouzených, kteří jsou skutečně vinni = **správné zamítnutí neplatné nulové hypotézy**.
- Když budeme požadovat hodně důkazů, zvýší se procento nevinných, kteří budou osvobozeni = **správné přijetí platné nulové hypotézy**, ale zároveň se zvýší i procento vinných, kteří budou osvobozeni = **chyba II. druhu**.

Pravděpodobnost výsledků rozhodovacího procesu

Rozhodnutí	Skutečnost	
	H_0 platí	H_0 neplatí
H_0 nezamítáme	správné rozhodnutí $P = 1 - \alpha$	chyba II. druhu $P = \beta$
H_0 zamítáme	chyba I. druhu $P = \alpha$	správné rozhodnutí $P = 1 - \beta$

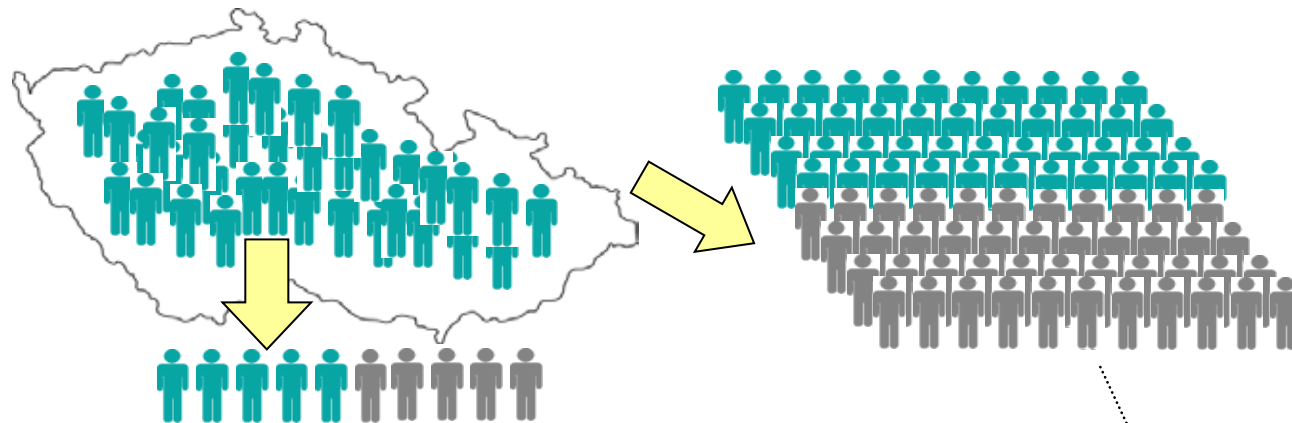
- Jak je vidět z analogie se soudním procesem, nelze zároveň minimalizovat α i β . V praxi je nutné více hlídat α → předem stanovíme maximální hranici pro α (**hladina významnosti testu**, „level of significance“ – většinou $\alpha=0,05$, tedy 5%, nebo $\alpha=0,01$, tedy 1%) a za této podmínky minimalizujeme β → tedy zvyšujeme $1-\beta$, což je tzv. **síla testu** („power of the test“).
- Proč hlídat spíše α než β ?

Benjamin Franklin: „*It is better that 100 guilty persons should escape than that one innocent person should suffer.*“

Síla testu ($1-\beta$)

- Pravděpodobnost, že zamítneme H_0 ve chvíli, kdy H_0 opravdu neplatí – tzn. prokážeme rozdíl tam, kde skutečně existuje.
- Snažíme se sílu testu optimalizovat při zachování hladiny významnosti testu α → princip **výpočtu velikosti experimentálního vzorku před provedením studie** (tzn. snažíme se zjistit, kolik je třeba experimentálních subjektů (pozorování) k tomu, aby měl výsledný test dostatečnou sílu k zamítnutí nulové hypotézy, bude-li tato hypotéza skutečně neplatná).
- Proč je důležité optimalizovat velikost vzorku před provedením studie?
 - **etické aspekty** – nelze zbytečně léčit lidi
 - **ekonomické aspekty** – zbytečné plýtvání prostředky
 - **statistické vlastnosti** – při velkém N lze prokázat cokoliv
- Rizika neplánovaného počtu subjektů ve studii:
 - **malý vzorek** – ztráta času, nemožnost prokázat rozdíl mezi srovnávanými skupinami pacientů
 - **velký vzorek** – ztráta času a prostředků, průkaz klinicky nevýznamného rozdílu mezi srovnávanými skupinami pacientů

Vliv velikosti vzorku na výsledky testování



Dvě skupiny pacientů s nepatrným rozdílem v dané charakteristice, který ale není klinicky významný.

$n_1 = 10, n_2 = 10$
 $p = 0.797$

$n_1 = 100, n_2 = 100$
 $p = 0.140$

$n_1 = 1000, n_2 = 1000$
 $p < 0.001$

Statistická významnost způsobená velkým N

Faktory ovlivňující sílu testu

- Vychází se z výpočtu intervalu spolehlivosti:

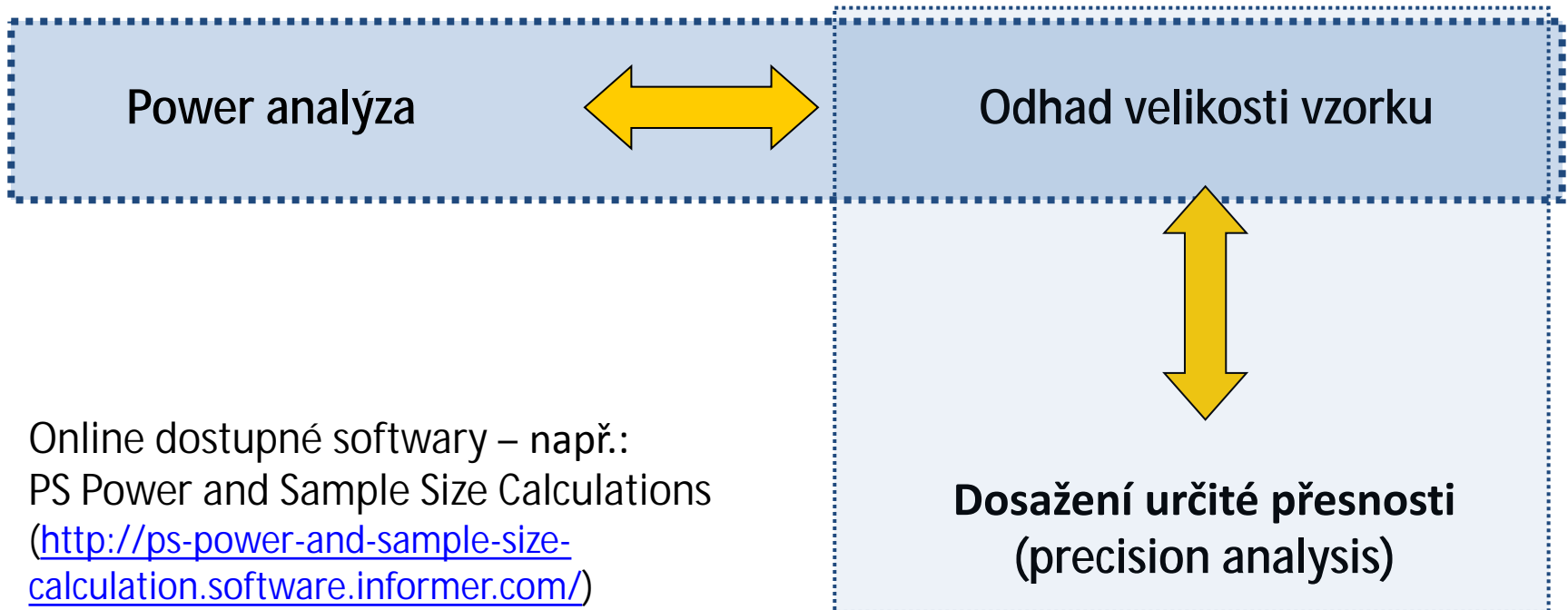
$$\text{Odhadovaný parametr} \pm \text{Kvantil modelového rozložení pro } (1-\alpha/2) * \frac{s}{\sqrt{n}}$$

Chyba odhadu

- **Velikost vzorku:** čím více pozorování (informace o platnosti nulové hypotézy), tím větší má test sílu. Síla testu roste s odmocninou z n .
- **Velikost efektu (účinku):** velikost rozdílu v neznámých parametrech také ovlivňuje sílu testu. Vždy je jednodušší identifikovat jako významný velký efekt, např. velký rozdíl ve středních hodnotách objemu prostaty dvou populací. Naopak je těžší prokázat jako významný menší efekt (menší rozdíl).
- **Variabilita dat:** variabilita dat zvyšuje variabilitu odhadů a ztěžuje tak rozhodnutí o H_0 . Čím více jsou pozorované hodnoty variabilní, tím více dat bude potřeba pro přesný odhad velikosti účinku (rozdílu).
- **Hladina významnosti:** snížíme-li hladinu významnosti testu (např. zvolíme 0,01 místo 0,05), bude obtížnější H_0 zamítnout → sníží se síla testu.

Power analýza a optimalizace velikosti vzorku

- Power analýza (analýza síly testu) a optimalizace velikosti vzorku (sample size estimation) jsou dvě strany téže mince.
- Obě vycházejí z testování hypotéz, jednou však máme jako předpoklad požadovanou sílu testu a chceme optimalizovat n , podruhé jsme limitováni n a ptáme se, jaké jsme v našich podmínkách schopni dosáhnout síly testu.



Online dostupné softwary – např.:
PS Power and Sample Size Calculations
(<http://ps-power-and-sample-size-calculation.software.informer.com/>)

Plánování klinického hodnocení fáze I - IV

COMPARISON OF CLINICAL TRIAL PHASES

	PHASE I	PHASE II	PHASE III	PHASE IV
OBJECTIVES:	Determine the metabolic and pharmacological actions and the maximally tolerated dose	Evaluate effectiveness, determine the short-term side effects and identify common risks for a specific population and disease	Obtain additional information about the effectiveness on clinical outcomes and evaluate the overall risk-benefit ratio in a demographically diverse sample	Monitor ongoing safety in large populations and identify additional uses of the agent that might be approved by the FDA
FACTORS TO BE IDENTIFIED:	<ul style="list-style-type: none"> -Bioavailability -Bioequivalence -Dose proportionality -Metabolism -Pharmacodynamics -Pharmacokinetics 	<ul style="list-style-type: none"> -Bioavailability -Drug-disease interactions -Drug-drug interactions -Efficacy at various doses -Pharmacodynamics -Pharmacokinetics -Patient safety 	<ul style="list-style-type: none"> -Drug-disease interactions -Drug-drug interactions -Dosage intervals -Risk-benefit information -Efficacy and safety for subgroups 	<ul style="list-style-type: none"> -Epidemiological data -Efficacy and safety within large, diverse populations -Pharmacoeconomics
DATA FOCUS:	<ul style="list-style-type: none"> -Vital signs -Plasma and serum levels -Adverse events 	<ul style="list-style-type: none"> -Dose response and tolerance -Adverse events -Efficacy 	<ul style="list-style-type: none"> -Laboratory data -Efficacy -Adverse events 	<ul style="list-style-type: none"> -Efficacy -Pharmacoeconomics -Epidemiology -Adverse events
DESIGN FEATURES:	<ul style="list-style-type: none"> -Single, ascending dose tiers -Unblinded -Uncontrolled 	<ul style="list-style-type: none"> -Placebo controlled comparisons -Active controlled comparisons -Well-defined entry criteria 	<ul style="list-style-type: none"> -Randomized -Controlled -2-3 treatment arms -Broader eligibility criteria 	<ul style="list-style-type: none"> -Uncontrolled -Observational
DURATION:	Up to 1 month	Several months	Several years	Ongoing (following FDA approval)
POPULATION:	Healthy volunteers or individuals with the target disease (such as cancer or HIV)	Individuals with target disease	Individuals with target disease	Individuals with target disease, as well as new age groups, genders, etc.
SAMPLE SIZE:	20 to 80	200 to 300	Hundreds to thousands	Thousands
EXAMPLE:	Study of a single dose of Drug X in normal subjects	Double-blind study evaluating safety and efficacy of Drug X vs. placebo in patients with hypertension	Study of Drug X vs. standard treatment in hypertension study	Study of economic benefit of newly-approved Drug X vs. standard treatment for hypertension

3. p-hodnota

p-hodnota („p-value“, „p-level“)

- Neboli dosažená hladina významnosti testu.
- Značka: p
- Je to pravděpodobnost, s jakou bychom mohli obdržet pozorovaná data nebo data stejně, či ještě více odporující nulové hypotéze, za předpokladu, že je nulová hypotéza pravdivá.
- Čím menší je p , tím neudržitelnější čili méně důvěryhodná je nulová hypotéza.
- Hodnocení, kdy je výsledek testu statisticky významný:
 - Máme zvolenu hladinu významnosti testu (např. $\alpha=0,05$).
 - Dvě možné situace:
 1. $p < \alpha$ – **zamítáme H_0** – statisticky významný výsledek testu
 2. $p \geq \alpha$ – **nezamítáme H_0**

Důležité poznámky k testování hypotéz

- **Nezamítnutí nulové hypotézy neznamena automaticky její přijetí!** Může se jednat o situaci, kdy pro zamítnutí nulové hypotézy nemáme dostatečné množství informace.
- **Dosažená hladina významnosti testu (ať už 0,05, 0,01 nebo 0,10) nesmí být slepě brána jako hranice pro existenci/neexistenci testovaného efektu.** Neexistuje jasná hranice pro významnost či nevýznamnost – často je velmi malý rozdíl mezi p-hodnotou 0,04 a p-hodnotou 0,06.
- **Malá p-hodnota nemusí znamenat velký efekt.** Hodnota testové statistiky a odpovídající p-hodnota může být ovlivněna velkou velikostí vzorku a malou variabilitou pozorovaných dat.
- **Výsledky testování musí být nahlíženy kriticky** – jedná se o závěr založený „pouze“ na jednom výběrovém souboru.


Statistická vs. klinická významnost

- Výsledky studie nemusí odpovídat realitě a skutečnosti. Statistická významnost jednoduše nemusí znamenat, že pozorovaný rozdíl je významný i ve skutečnosti!
- Statistická významnost pouze indikuje, že pozorovaný rozdíl není náhodný (ve smyslu stanovené hypotézy). Lze ji ovlivnit velikostí vzorku.
- Stejně důležitá je i praktická významnost, tedy významnost z hlediska lékaře nebo biologa.

		Praktická významnost	
		ANO	NE
Statistická významnost	ANO	OK, praktická i statistická významnost je ve shodě.	Významný výsledek je statistický artefakt, prakticky nevyužitelný.
	NE	Výsledek může být pouhá náhoda, neprůkazný výsledek.	OK, praktická i statistická významnost je ve shodě.

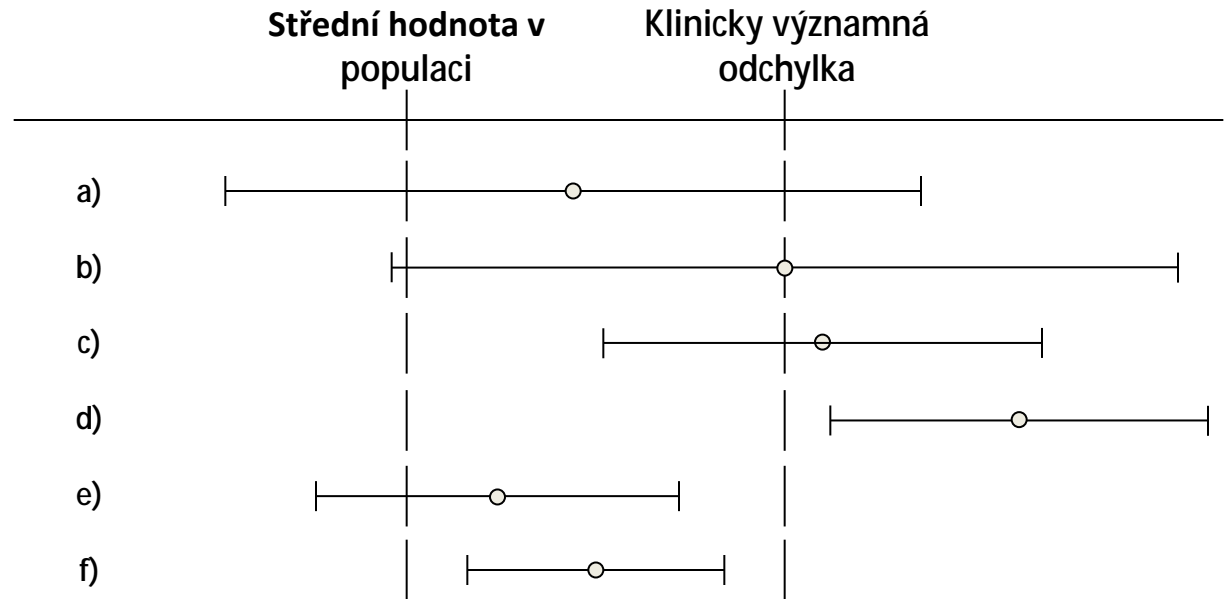
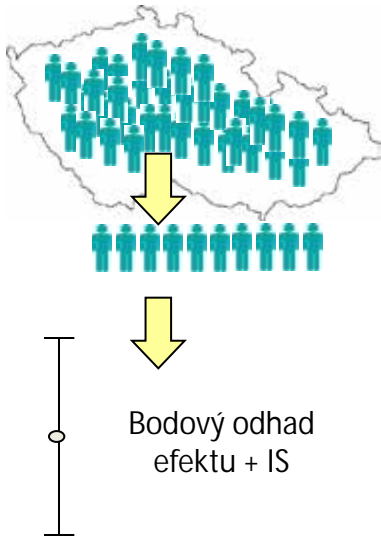
Statistická vs. klinická významnost

		Praktická významnost	
		ANO	NE
Statistická významnost	ANO	OK, praktická i statistická významnost jsou ve shodě.	Významný výsledek je statistický artefakt, prakticky nevyužitelný.
	NE	Výsledek může být pouhá náhoda, neprůkazný výsledek.	OK, praktická i statistická významnost jsou ve shodě.



Statisticky nevýznamný výsledek neznamená, že pozorovaný rozdíl ve skutečnosti neexistuje! Může to být způsobeno nedostatečnou informací v pozorovaných datech!

Statistická vs. klinická významnost



Možnost	Statistická významnost	Klinická významnost
a)	ne	možná
b)	ne	možná
c)	ano	možná
d)	ano	ano
e)	ne	ne
f)	ano	ne

Hodnocení velikosti účinku

- absolutní velikost účinku při srovnání dvou skupin – rozdíl odhadů průměrů: $\bar{x}_1 - \bar{x}_2$
- koeficienty velikosti účinku:
 - dosažený efekt standardizují a jsou tak využitelné pro srovnávání různých experimentů (uplatnění v metaanalýzách)
 - např. Cohenův koeficient d : $d = \frac{\bar{x}_1 - \bar{x}_2}{s}$, kde $s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
 - velký efekt: $d > 0,8$
 - střední efekt: $0,5 < d \leq 0,8$
 - malý efekt: $0,2 < d \leq 0,5$
 - zanedbatelný efekt: $d \leq 0,2$
- korelační koeficienty (hodnocení míry vztahu dvou proměnných)

Shrnutí klíčových pojmů analýzy dat

- **Významnost** – viz. předcházející slidy.
- **Zkreslení výsledků** („biased results“) – zkreslení způsobené starým nebo nenakalibrovaným měřidlem („technical bias“), zkreslení nevhodným výběrem subjektů („selection bias“), sledování zavádějícího faktoru namísto faktoru, který je pravou příčinou sledovaného výsledku.
- **Reprezentativnost** – experimentální vzorek musí svými charakteristikami odpovídat cílové populaci.
- **Srovnatelnost** – pokud chceme srovnávat skupiny mezi sebou, musí být skupiny srovnatelné. Pokud nemůžeme provést randomizaci (tzn. náhodné rozdělení subjektů do skupin), musíme hlídat, aby skupiny byly srovnatelné. Pokud nejsou, můžeme vytvořit podskupiny a ty srovnávat mezi sebou, nebo se snažíme odstranit vliv „nechtěných“ faktorů.
- **Spolehlivost** – sumarizace sledované proměnné jedním číslem (např. průměrem) není dostatečná, protože nepostihujeme variabilitu dat – průměr vypočítaný z dat 10 lidí bude určitě méně přesný (spolehlivý) než průměr vypočítaný z dat 1000 lidí → průměr doplníme o interval spolehlivosti.

4. Vhodná volba testu v různých situacích

Statistické testování - opakování

- **Cíle:**

1. Chceme srovnávat:
 - 1 náhodný výběr s předpokládanou hodnotou
 - 2 náhodné výběry mezi sebou
 - Více náhodných výběrů mezi sebou
2. Chceme hodnotit změnu náhodné veličiny vzhledem k vnějšímu zásahu.
3. Chceme rozhodovat o nezávislosti dvou náhodných veličin.
4. Chceme rozhodovat o charakteru rozdělení náhodné veličiny.

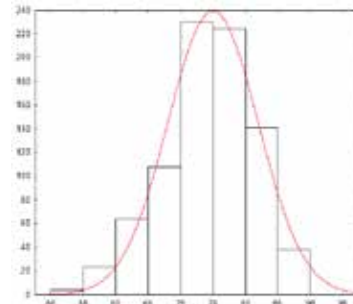
- **Postup:**

1. Máme danou hypotézu k ověření (např. pacienti a kontroly se liší v hodnotách MMSE skóre).
2. Provedli jsme výběr z populace.
3. **Aplikujeme statistický test.**
4. Hypotézu prohlásíme za statisticky platnou nebo neplatnou.

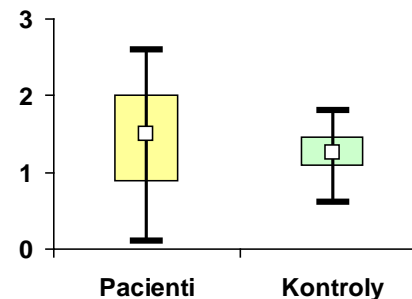
Statistické testy – příklady předpokladů

- **Typ dat** – pokud je předepsáno, že se test má použít na ordinální či nominální data, nemůžeme ho použít na hodnocení spojitých hodnot.

- **Normalita** rozdělení dat – předpoklad u mnoha parametrických testů.



- **Homogenita rozptylu** srovnávaných skupin – tzn. předpoklad, aby byl rozptyl ve skupinách přibližně stejný.



- **Vyrovnané počty subjektů** ve srovnávaných skupinách – nutné z důvodu, aby byly odhady ve srovnávaných skupinách podobně přesné a spolehlivé. Pokud to experimentální situace dovoluje, měly by být přibližně stejné počty opakování standardem.

Parametrické a neparametrické testy

- **Parametrické testy:**
 - Mají předpoklady o rozdělení vstupních dat (např. předpoklad normálního rozdělení), protože se zabývají testováním tvrzení o neznámých parametrech rozdělení (např. střední hodnoty).
 - Mají větší sílu než neparametrické testy.
- **Neparametrické testy:**
 - Nemají předpoklady o rozdělení vstupních dat, je tedy možné je použít při asymetrickém rozdělení nebo odlehlých hodnotách.
 - Mají menší sílu, protože dochází k redukci informační hodnoty původních dat z důvodu, že neparametrické testy nevyužívají původní hodnoty, ale nejčastěji pouze jejich pořadí („rank“).
 - Menší sílu testu je možné vykompenzovat větší velikostí vzorku.
- Testování v případě chybně určeného rozdělení pravděpodobnosti testové statistiky může vést k mylným závěrům z důvodu nerelevantní p-hodnoty → používání neparametrických testů je „bezpečnější“.

Parametrické a neparametrické testy pro kvantitativní data – přehled

Typ srovnání	Parametrický test	Neparametrický test
1 skupina dat s referenční hodnotou – jednovýběrové testy:	Jednovýběrový t-test, jednovýběrový z-test	Wilcoxonův test
2 skupiny dat párově – párové testy:	Párový t-test	Wilcoxonův test, znaménkový test
2 skupiny dat nepárově – dvouvýběrové testy:	Dvouvýběrový t-test	Mannův-Whitneyův test, mediánový test
Více skupin nepárově:	ANOVA	Kruskalův- Wallisův test

Jednostranné a oboustranné testy

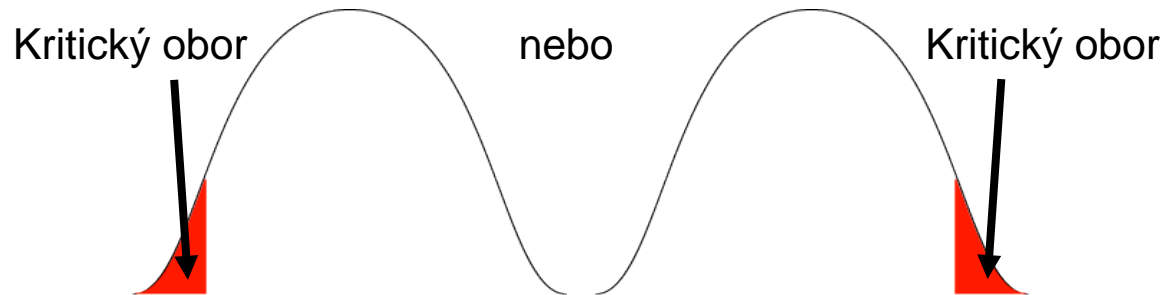
- Souvislost s jednostranou a oboustrannou alternativní hypotézou.

- **Jednostranné („One-Tailed“) testy:**

- Jednostranná alternativní hyp.: $H_1 : q < q_0$ $H_1 : q > q_0$

- Např. testujeme, zda je objem mozkové struktury menší u žen než u mužů či zda je průměrná spotřeba

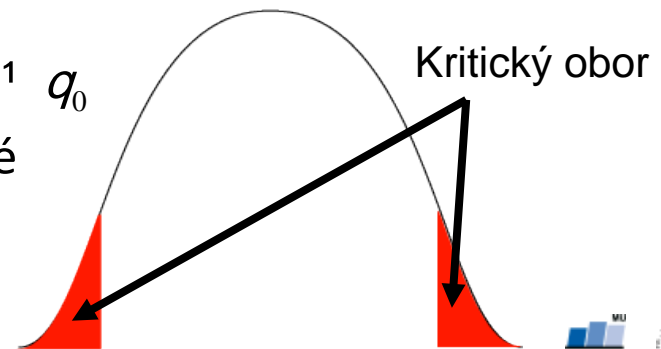
tisících léků větší u pacientů než je populační průměr apod.



- **Oboustranné („Two-Tailed“) testy:**

- Oboustranná alternativní hyp.: $H_1 : q \neq q_0$

- Např. testujeme, zda se objem mozkové struktury liší u žen a mužů apod.



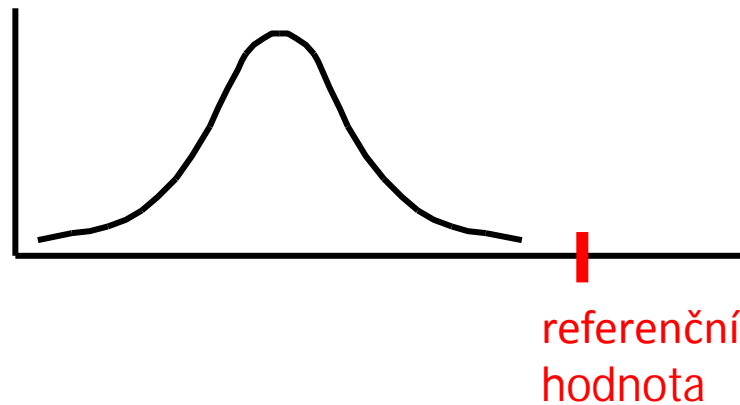
Zásady při testování

1. Znat základní typy **testů** a vědět, pro jaká data se používají.
2. **Ověřit předpoklady testu** – smysl má pouze aplikace „správného“ testu na „správná“ data.
3. Posoudit, zda je výsledek **významný i z klinického hlediska**.
4. Být si vědom toho, že **statistický test není nic víc než matematický vzorec** aplikovaný na data, tedy existuje nenulová pravděpodobnost, že výsledek bude chybný (viz chyba I. a II. druhu). Ovlivnit výsledky testu můžeme například změnou velikosti vzorku.

5. Jednovýběrové testy

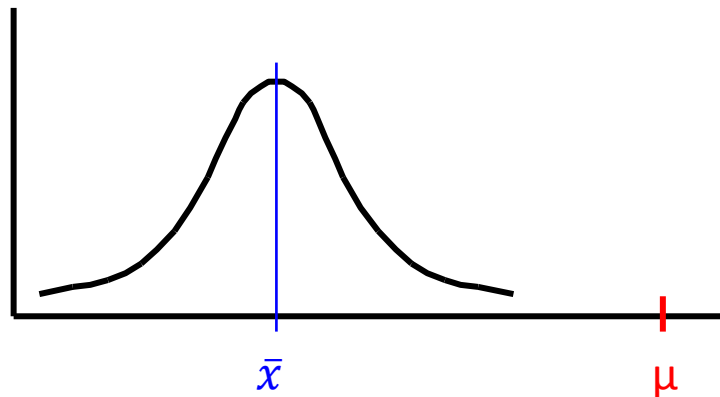
Jednovýběrové („One-Sample“) testy

- Srovnávají jeden vzorek („one sample“) 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.
- Parametrické jednovýběrové testy, kterým se budeme věnovat:
 - **jednovýběrový t-test** (test o střední hodnotě při neznámém rozptylu)
 - **jednovýběrový z-test** (test o střední hodnotě při známém rozptylu)



Jednovýběrový t-test

- Srovnáváme **střední hodnotu jednoho výběru s referenční hodnotou**.
- Jde o test o střední hodnotě při **neznámém rozptylu** – tzn. testujeme, zda se průměr dané proměnné v našem výběru liší od referenční hodnoty (často populačního průměru), přičemž rozptyl dané proměnné počítáme z našeho výběru.



- Předpoklad: **normalita dat**
- Testová statistika:
$$T = \frac{\bar{x} - m}{s / \sqrt{n}}$$

Jednovýběrový t-test

- **Příklad:** Chceme srovnat průměrný objem hipokampu u 406 pacientů s MCI v našem souboru s průměrným objemem hipokampu 6575 mm³ zjištěným při populačním epidemiologickém průzkumu.
- Tzn. hypotézy budou mít tvar: $H_0 : \bar{x} = 6575$ a $H_1 : \bar{x} \neq 6575$
- **Postup:**
 1. Ověření normality – vykreslíme histogram objemu hipokampu pacientů s MCI.
 2. Aplikujeme statistický test – 3 možnosti:
 - I. Testování pomocí intervalu spolehlivosti
 - II. Testování pomocí kritického oboru
 - III. Testování pomocí p-hodnoty
 3. Nulovou hypotézu zamítneme nebo nezamítneme.

Testování pomocí intervalu spolehlivosti

Příklad: Chceme srovnat průměrný objem hipokampu u 406 pacientů s MCI v našem souboru s průměrným objemem hipokampu 6575 mm³ zjištěným při populačním epidemiologickém průzkumu.

Výpočet intervalu spolehlivosti:

$$n = 406$$

$$\bar{x} = 6552,6 \text{ mm}^3$$

$$s = 176,2 \text{ mm}^3$$

$$\bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \leq m \leq \bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$

$$6552,6 - \frac{176,2}{\sqrt{406}} t_{1-0,05/2}(406-1) \leq m \leq 6552,6 + \frac{176,2}{\sqrt{406}} t_{1-0,05/2}(406-1)$$

$$6535,4 \leq m \leq 6569,8$$

Protože 95% interval spolehlivosti (6535,4; 6569,8) neobsahuje populační průměr 6575 → **zamítáme** nulovou hypotézu → **Průměrný objem hipokampu u pacientů s MCI v našem souboru se statisticky významně liší od populačního průměru.**

Testování pomocí kritického oboru

Příklad: Chceme srovnat průměrný objem hipokampu u 406 pacientů s MCI v našem souboru s průměrným objemem hipokampu 6575 mm^3 zjištěným při populačním epidemiologickém průzkumu.

Výpočet testové statistiky:

$$n = 406$$

$$\bar{x} = 6552,6 \text{ mm}^3$$

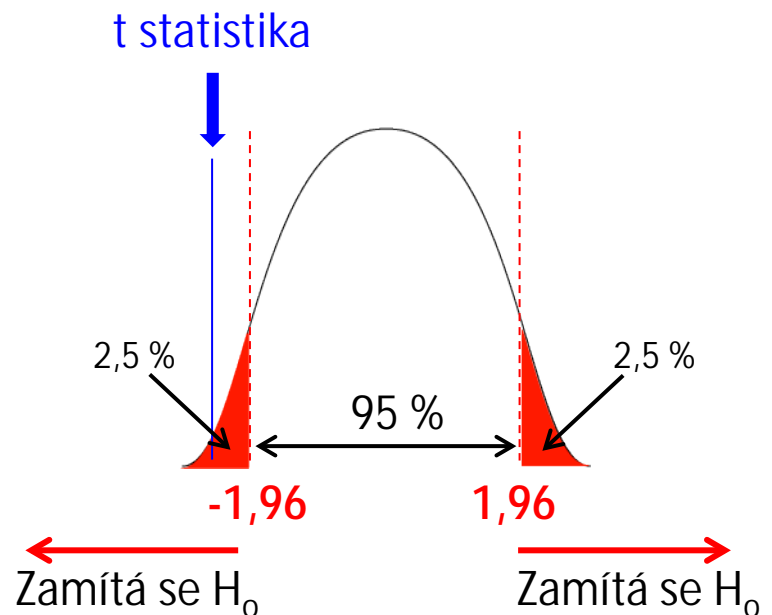
$$s = 176,2 \text{ mm}^3$$

$$t = \frac{\bar{x} - m}{s/\sqrt{n}} = \frac{6552,6 - 6575}{176,2/\sqrt{406}} = -2,56$$

Stanovení kritického oboru:

kritické hodnoty: $t_{\alpha/2}(405) \cong -1,96$

$$t_{1-\alpha/2}(405) \cong 1,96$$



Protože testová statistika $t = -2,56$ leží v kritickém oboru → **zamítáme** nulovou hypotézu → **Průměrný objem hipokampu u pacientů s MCI v našem souboru se statisticky významně liší od populačního průměru.**

Testování pomocí p-hodnoty

Příklad: Chceme srovnat průměrný objem hipokampu u 406 pacientů s MCI v našem souboru s průměrným objemem hipokampu 6575 mm^3 zjištěným při populačním epidemiologickém průzkumu.

Výpočet testové statistiky:

$$n = 406$$

$$\bar{x} = 6552,6 \text{ mm}^3$$

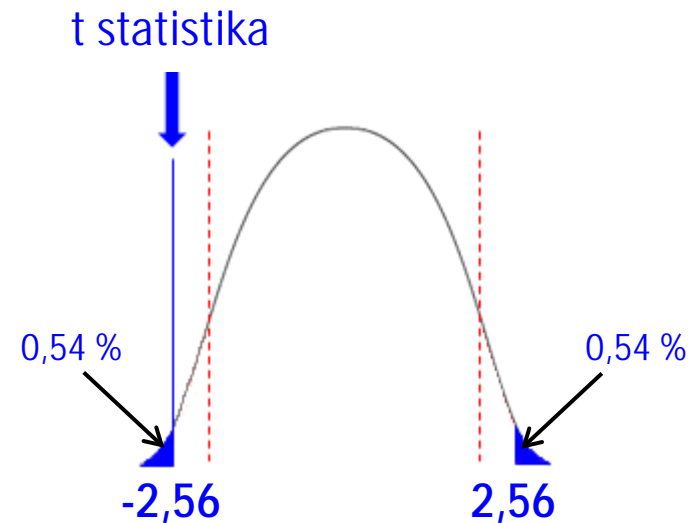
$$s = 176,2 \text{ mm}^3$$

$$t = \frac{\bar{x} - m}{s/\sqrt{n}} = \frac{6552,6 - 6575}{176,2/\sqrt{406}} = -2,56$$

Výpočet p-hodnoty:

$$p = 2 \times (P(T \leq -2,56)) = 2 \times 0,0054 = 0,0108$$

Protože p-hodnota $0,0108 < 0,05 \rightarrow$ **zamítáme** nulovou hypotézu \rightarrow **Průměrný objem hipokampu u pacientů s MCI v našem souboru se statisticky významně liší od populačního průměru.**



Zmenšení N

N = 406

Mean	Std.Dv.	N	Std.Err.	Lower CI	Upper CI	Reference	t-value	df	p
6552,6	176,2	406	8,7	6535,4	6569,8	6575	-2,56	405	0,0108

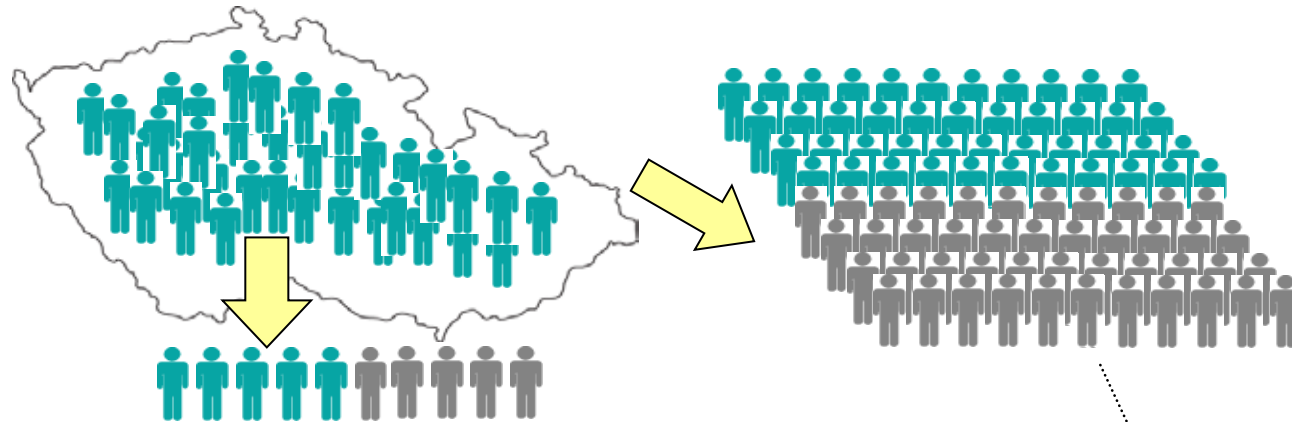
$p=0,0108 < 0,05 \rightarrow$ **zamítáme** nulovou hypotézu

N = 100

Mean	Std.Dv.	N	Std.Err.	Lower CI	Upper CI	Reference	t-value	df	p
6552,2	171,4	100	17,1	6518,2	6586,2	6575	-1,33	99	0,1865

$p=0,1865 > 0,05 \rightarrow$ **nezamítáme** nulovou hypotézu

Vliv velikosti vzorku na výsledky testování - opakování



Dvě skupiny pacientů s nepatrným rozdílem v dané charakteristice, který ale není klinicky významný.

$n_1 = 10, n_2 = 10$
 $p = 0.797$

$n_1 = 100, n_2 = 100$
 $p = 0.140$

$n_1 = 1000, n_2 = 1000$
 $p < 0.001$

Statistická významnost způsobená velkým N

Oboustranný vs. jednostranný jednovýběrový t-test

Oboustranný jednovýběrový t-test:

Příklad: Chceme srovnat objem hipokampu u pac. s MCI s populačním průměrem. Tzn. chceme ověřit, zda se objem hipokampu u pac. s MCI v našem souboru **liši** od populačního průměru.

Alternativní hypotéza: $H_1: \bar{x} \neq m$ $\bar{x} = 6552,6 \text{ mm}^3$
 $\mu = 6575 \text{ mm}^3$
 $p = 0,0108$

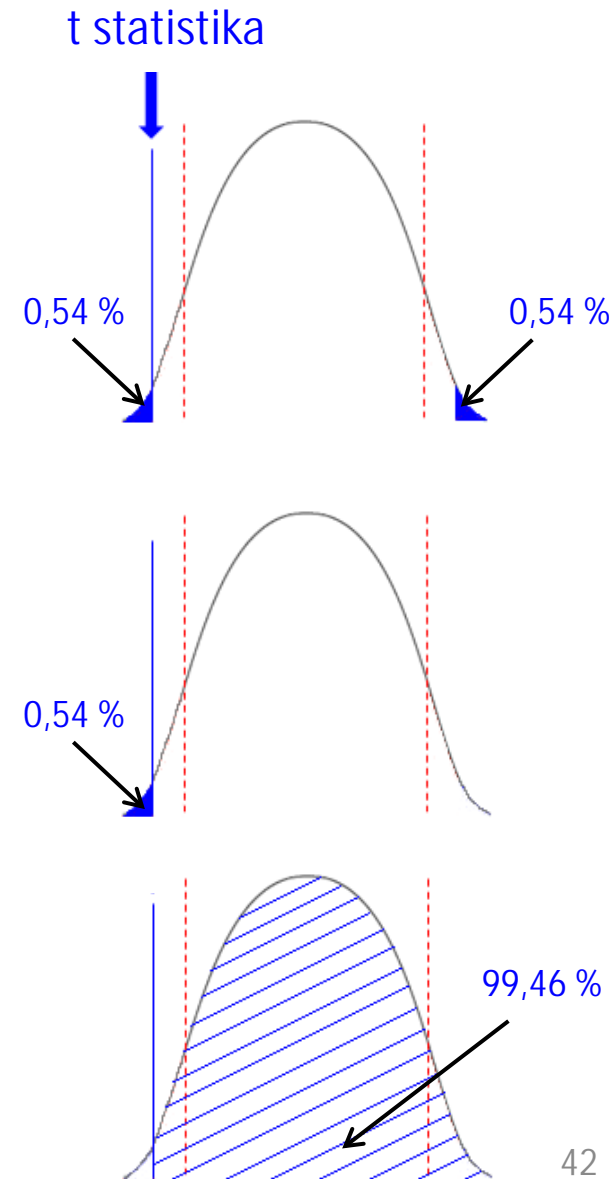
Jednostranný jednovýběrový t-test:

1. **Levostranný** – příklad: Chceme ověřit, zda je objem hipokampu u pac. s MCI v našem souboru **menší** než populační průměr: $H_1: \bar{x} < m$

$p = 0,0108/2 = 0,0054$

2. **Pravostranný** – příklad: Chceme ověřit, zda je objem hipokampu u pac. s MCI v našem souboru **větší** než populační průměr: $H_1: \bar{x} > m$

$p = 1 - 0,0108/2 = 0,9946$

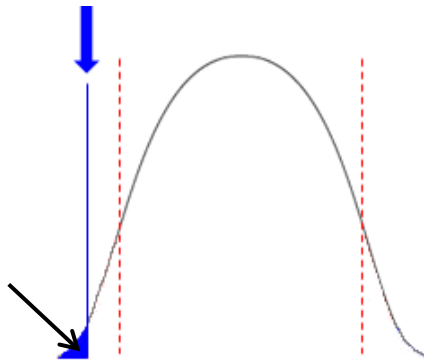


Jednostranný jednovýběrový t-test

Skutečnost: $\bar{x} < \mu$

Levostranný jednovýběrový t-test:

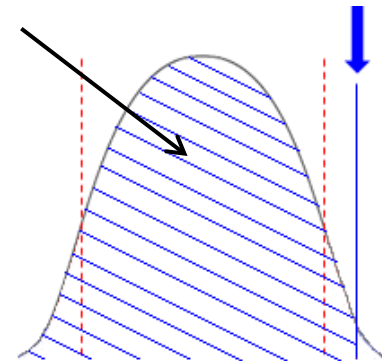
$H_1: \bar{x} < \mu$



Skutečnost: $\bar{x} > \mu$

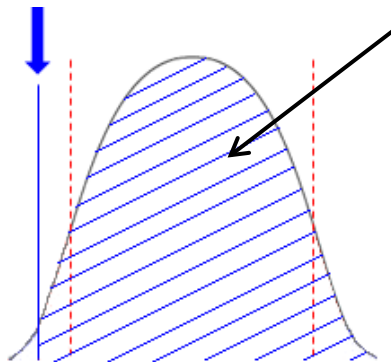
Levostranný jednovýběrový t-test:

$H_1: \bar{x} < \mu$



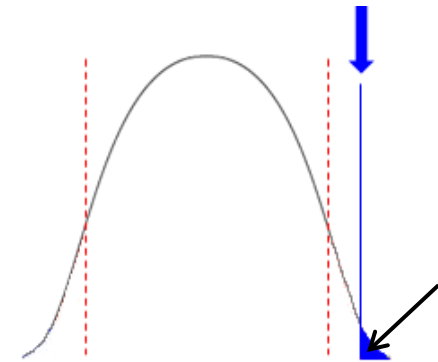
Pravostranný jednovýběrový t-test:

$H_1: \bar{x} > \mu$



Pravostranný jednovýběrový t-test:

$H_1: \bar{x} > \mu$



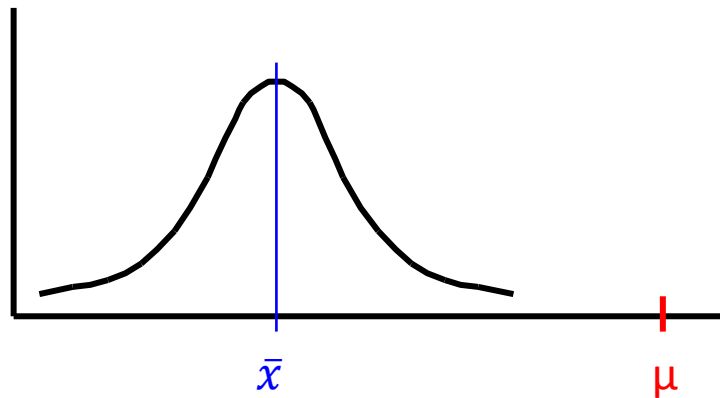
Úkol 1

- **Zadání:** Zjistěte, zda se liší průměrný objem amygdaly u mužů v našem souboru od populačního průměrného objemu 2800 mm³ (nezapomeňte ověřit předpoklady).
- **Řešení:**

Variable	Test of means against reference constant (value) (Data_neuro_vycistena2) Include condition: v5="M"							
	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
Amygdala_volume (mm3)	2786,292	239,3598	482	10,90254	2800,000	-1,25734	481	0,209240

Z-test

- Srovnáváme **střední hodnotu jednoho výběru s referenční hodnotou**.
- Jde o test o střední hodnotě při **známém rozptylu** – tzn. testujeme, zda se průměr dané proměnné v našem výběru liší od referenční hodnoty (často populačního průměru), přičemž známe rozptyl dané proměnné pro celou populaci.



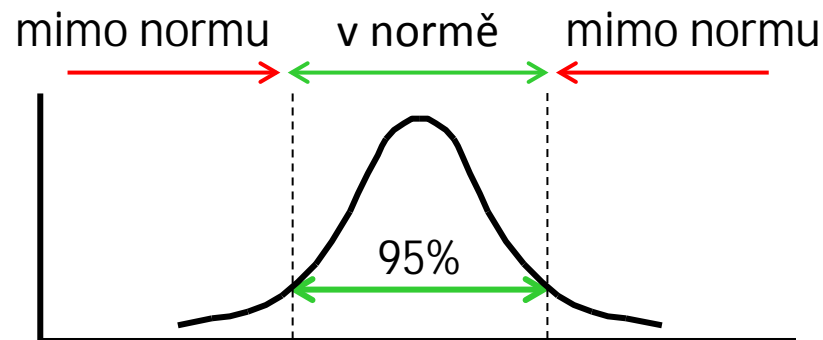
- Předpoklad: **normalita dat**
- Testová statistika:
$$Z = \frac{\bar{x} - m}{s / \sqrt{n}}$$

Z-test

- **Příklad:** Při populačním průzkumu bylo zjištěno, že průměrná hodnota MMSE skóre je 27,5 (SD = 4). Chceme zjistit, zda se průměrná hodnota MMSE skóre u 406 pacientů s MCI v našem souboru liší od populační průměrné hodnoty.
- Tzn. hypotézy budou mít tvar: $H_0 : \bar{x} = 27,5$ a $H_1 : \bar{x} \neq 27,5$
- **Postup:**
 1. Ověření normality – vykreslíme histogram MMSE skóre u pacientů s MCI, abychom ověřili, že průměr je dobrý ukazatel středu hodnot.
 2. Aplikujeme statistický test – vypočítáme p-hodnotu:
 - v Excelu:
`=2*MIN(Z.TEST(A1:A406;27,5;4);1-Z.TEST(A1:A406;27,5;4))`
 - v Matlabu: `[H,P] = ztest(X,27.5,4)`
 3. Nulovou hypotézu zamítneme nebo nezamítneme:
 $p=0,013 < 0,05 \rightarrow$ **zamítáme** nulovou hypotézu \rightarrow **Průměrná hodnota MMSE skóre u pacientů s MCI v našem souboru se statisticky významně liší od populačního průměru.**

Z-skóre

- Odečtení populačního průměru (μ) a vydělení populační směrodatnou odchylkou (σ): $u_i = \frac{x_i - m}{s}$
- Souvislost se standardizací: $u_i = \frac{x_i - \bar{x}}{s}$
- Často při hodnocení různých skóre – určuje se, kteří lidé jsou mimo normu.

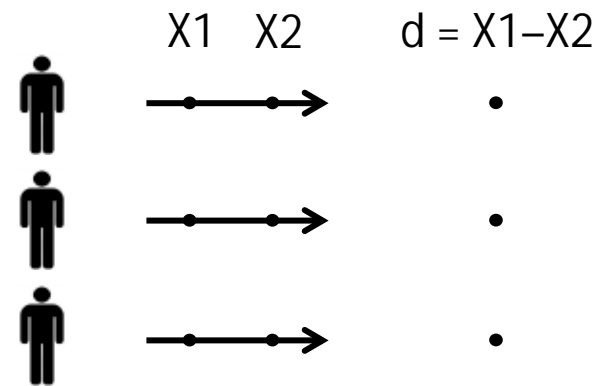


6. Párové testy

Párový t-test

- Srovnáváme **dvě skupiny dat**, které ale na **sobě nejsou nezávislé** – mezi objekty existuje vazba (např. člověk před a po operaci, stejný kmen krysy)
- Příklady: srovnání objemu hipokampu na začátku léčby a 1 rok po zahájení léčby, srovnání kognitivního výkonu pacientů před a po léčbě

- Test je v podstatě prováděn na **diferencích skupin** (rozdílech původních hodnot), nikoliv na původních datech → **obě skupiny tedy musí mít shodný počet hodnot** (všechna měření v jedné skupině musí být spárována s měřením v druhé skupině!)



- Předpoklad: **normalita diferencí** (rozdílů původních hodnot)
- Testová statistika: $T = \frac{\bar{d} - d_0}{s_d / \sqrt{n}}$, kde \bar{d} je průměrný rozdíl, d_0 je referenční hodnota (většinou 0), s_d je směrodatná odchylka rozdílů

Párový t-test

- **Příklad:** Chceme srovnat, zda se liší objem hipokampu u pacientů s Alzheimerovou chorobou při vstupu do studie a 2 roky po zahájení studie (tzn. chceme zjistit, zda došlo ke změně objemu hipokampu).
- Tzn. hypotézy budou mít tvar: $H_0 : \bar{d} = 0$ a $H_1 : \bar{d} \neq 0$
- **Postup:**
 1. Ověření existence vazby mezi oběma skupinami dat pomocí tečkového grafu.
 2. Ověření normality rozdílů – vytvoříme novou proměnnou, která bude obsahovat rozdíly objemů hipokampu, a vykreslíme histogram.
 3. Aplikujeme statistický test (v softwaru STATISTICA: t-test, dependent samples).
 4. Nulovou hypotézu zamítneme nebo nezamítneme:
 $p < 0,001 < 0,05 \rightarrow$ **zamítáme** nulovou hypotézu \rightarrow **Rozdíl v objemu hipokampu u pacientů s AD při vstupu do studie a 2 roky po zahájení studie je statisticky významný.**
- **Poznámka:** Stejně výsledky dostaneme, pokud použijeme jednovýběrový t-test a jako vstupní proměnnou vezmeme proměnnou s rozdílem objemů.

Úkol 2

- **Zadání:** Zjistěte, zda se liší MMSE skóre u kontrolních subjektů (CN) při vstupu do studie a dva roky po zahájení studie (nezapomeňte ověřit předpoklady).
- **Řešení:**

T-test for Dependent Samples (Data_neuro_vycistena2)										
Marked differences are significant at $p < ,05000$										
Include condition: v3=1										
Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
MMSE	29,19333	0,910024								
MMSE_24	28,98000	1,206637	150	0,213333	1,256322	2,079713	149	0,039265	0,010637	0,416029

Poděkování...

Příprava výukových materiálů předmětu „DSAN01 Analýza dat pro Neurovědy“ je finančně podporována prostředky projektu FRVŠ č. 942/2013 „Inovace materiálů pro interaktivní výuku a samostudium předmětu Analýza dat pro Neurovědy“

