Úvod do statistického testování

Co se dá testovat/srovnávat …

- 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.

Hypotéza

- 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: $H_0 : \theta = \theta_0$

- **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: $H_1 : \theta \neq \theta_0$ - oboustranná alternativa

  $H_1 : \theta < \theta_0$ - jednostranná alternativa

  $H_1 : \theta > \theta_0$ - jednostranná alternativa

Proč nulová hypotéza vyjadřuje nepřítomnost sledovaného 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.

Co se při rozhodování o platnosti hypotéz může stát

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

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

- **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 neviny = 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 nezamítneme	**správné rozhodnutí** **P = 1 – α**	**chyba II. druhu** **P = β**
H_0 zamítneme	**chyba I. druhu** **P = α**	**správné rozhodnutí** **P = 1 – β**

- 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 α=0,05, tedy 5%, nebo α=0,01, tedy 1%) a za této podmínky minimalizujeme β → tedy zvyšujeme 1-β, 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.“*

Testovací statistika

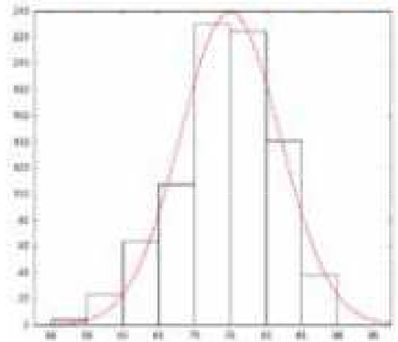
- funkce, odvozena na základě předpokládaných vlastností výběrů, které chceme porovnávat

$$\text{Testová statistika} = \frac{\text{Pozorovaná hodnota – Očekávaná hodnota}}{\text{Variabilita dat}} * \sqrt{\text{Velikost vzorku}}$$

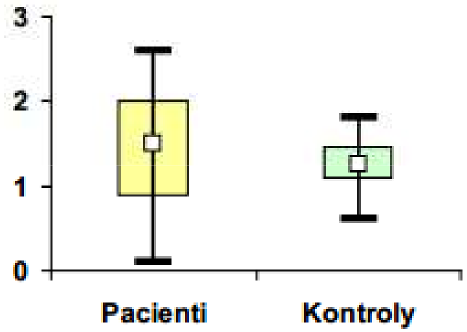
- funkce *pravděpodobnostní*

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

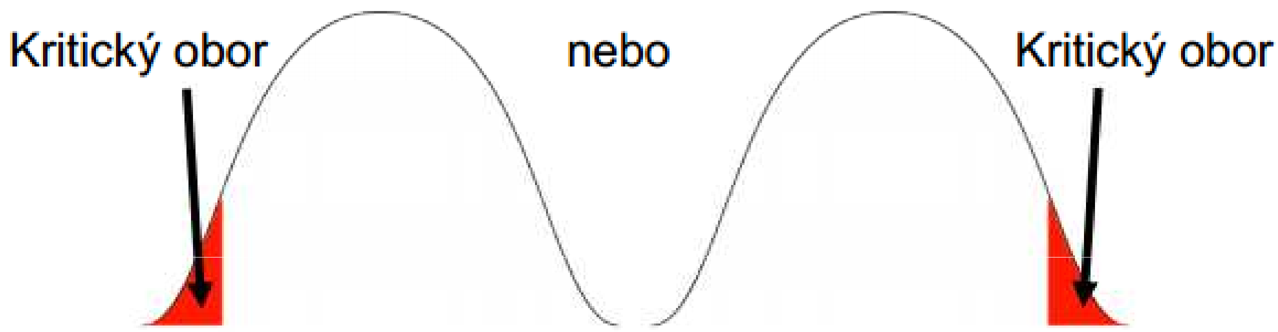
- **Parametrické testy:**
  - Mají předpoklady o rozdělení vstupních dat (např. předpoklad normálního rozdělení), protože se zabývají testováním tvrzení o neznámých parametrech rozdělení (např. střední hodnoty).
  - Mají větší sílu než neparametrické testy.

- **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.

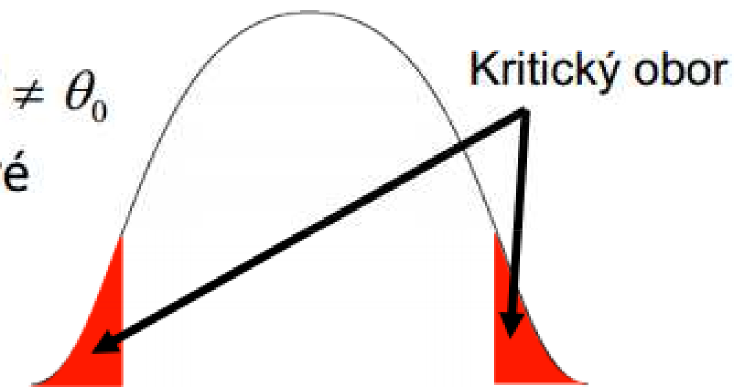
- Testování v případě chybně určeného rozdělení pravděpodobnosti testové statistiky může vést k mylným závěrům z důvodu nerelevantní p-hodnoty → používání neparametrických testů je „bezpečnější“.

Jednostranné a oboustranné testy

- Souvislost s jednostranou a oboustrannou alternativní hypotézou.
- **Jednostranné („One-Tailed“) testy:**
  - Jednostranná alternativní hyp.: $H_1 : \theta < \theta_0$ $H_1 : \theta > \theta_0$
  - Např. testujeme, zda je objem mozkové struktury menší u žen než u mužů či zda je průměrná spotřeba tišících léků větší u pacientů než je populační průměr apod.



- **Oboustranné („Two-Tailed“) testy:**
  - Oboustranná alternativní hyp.: $H_1 : \theta \neq \theta_0$
  - Např. testujeme, zda se objem mozkové struktury liší u žen a mužů apod.



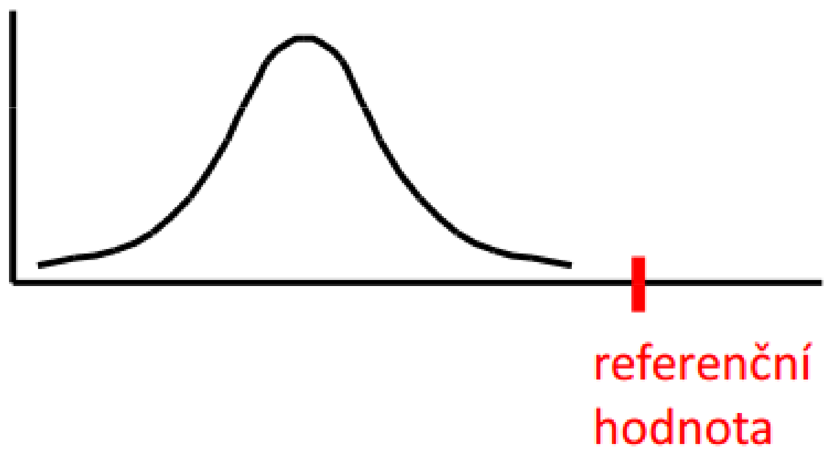
Parametrické testy

Jednovýběrový test

Jednovýběrový test

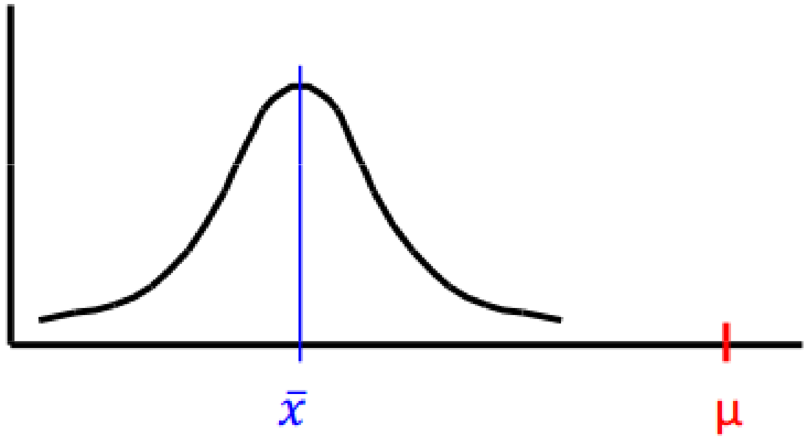
- **Srovnávají jeden vzorek („one sample") s referenční hodnotou** (popřípadě se statistickým parametrem cílové populace).
- V testu je tedy srovnáváno rozložení hodnot (vzorek) s jediným číslem (referenční hodnota, hodnota cílové populace).
- Otázka položená v testu může být vztažena k průměru, rozptylu, 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 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: $T = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

STATISTICA – jednovýběrový t-test

Je délka karet obrovských na Marshallových ostrovech srovnatelná
s celosvětovou populací?

Želvy

$n = 48$

$\bar{x} = 124{,}7 cm$

$s = 20{,}5 cm$

$\mu = 120 cm$

$H_0 =$

3 možnosti testování

- pomocí intervalu spolehlivosti
- pomocí kritické hodnoty
- pomocí p-hodnoty

H0 = není rozdíl mezi délkou želv na Marshallových ostrovech a délkou celé populace karet obrovských

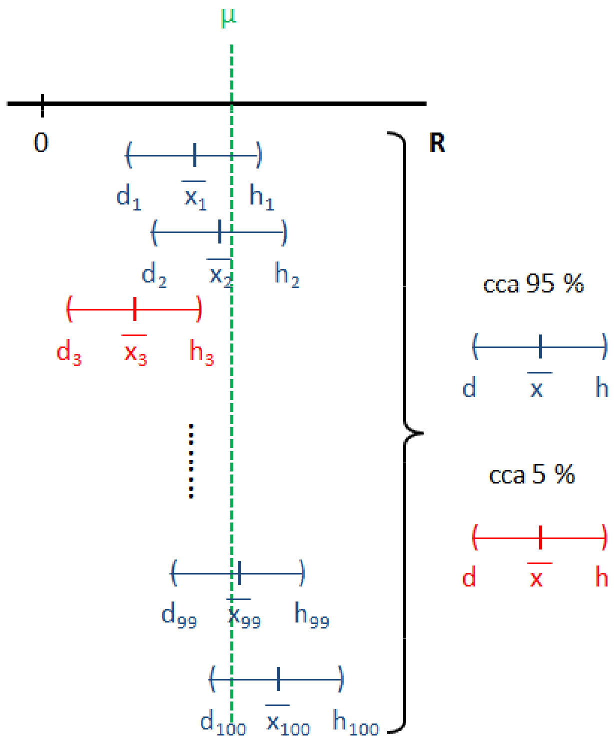
H1 = je rozdíl mezi délkou karet obrovských na MO a celé populace

Želvy interval spolehlivosti

- Pokud hladina významnosti α = 0,05
  - 95% IS pro výběrový průměr

    - Poloha neznámého parametru je konstantní.
    - 95% interval spolehlivosti má následující interpretaci:

    Pokud bychom opakovaně vybírali skupiny subjektů o stejné velikosti (*n*) a počítali výběrový průměr s 95% IS, pak 95 % těchto intervalů spolehlivosti neznámý parametr obsahuje a 5 % ho neobsahuje. Tedy 95% IS obsahuje neznámý parametr s rizikem α.



Želvy interval spolehlivosti

- IS pro jedovýběrový t-test

$$\bar{x} - \tfrac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \le \mu \le \bar{x} + \tfrac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$

kvantil studentova rozložení   počet stupňů volnosti

http://en.wikipedia.org/wiki/Student's_t-distribution

$n = 48$

$\bar{x} = 124{,}7cm$

$s = 20{,}5cm$

$\mu = 120cm$

- $$124{,}7 - \frac{20{,}5}{\sqrt{48}}.2{,}02 \le \mu \le 124{,}7 + \frac{20{,}5}{\sqrt{48}}.2{,}02$$

  $$118{,}7 \le \mu \le 130{,}7$$

- 120 je v IS – nezamítáme nulovou hypotézu o rovnosti délek

Želvy – kritický obor

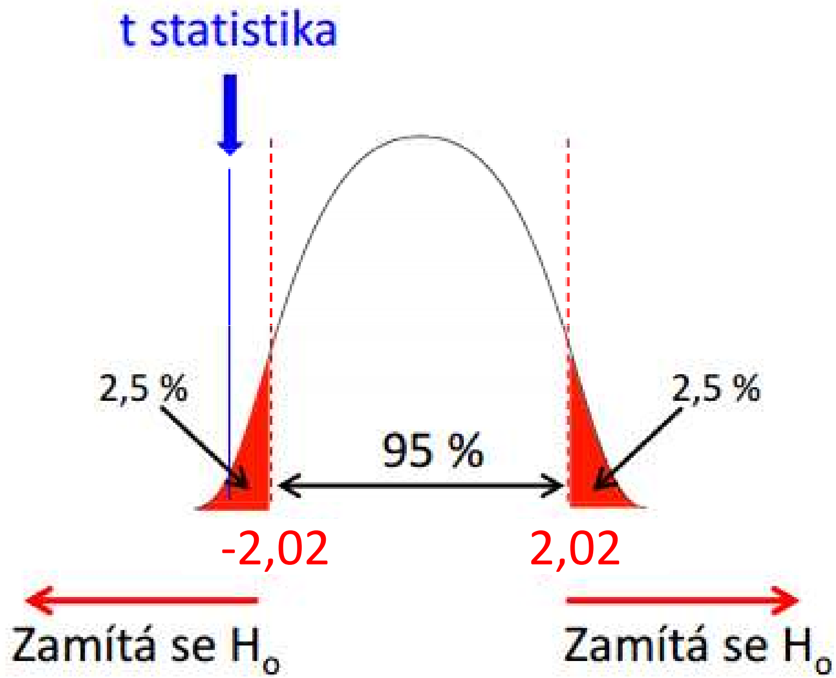
$n = 48$

$\bar{x} = 124{,}7cm$

$s = 20{,}5cm$

$\mu = 120cm$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = 1{,}59$$



t-hodnota není v kritickém oboru hodnot – NEZAMÍTÁME NULOVOU HYPOTÉZU

p-hodnota

- 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. α=0,05).
  - Dvě možné situace:
    1. $p < \alpha$ – **zamítáme H_0** – statisticky významný výsledek testu
    2. $p \geq \alpha$ – **nezamítáme H_0**

Želvy – p-hodnota

$$n = 48$$

$$\bar{x} = 124{,}7cm$$

$$s = 20{,}5cm$$

$$\mu = 120cm$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = 1{,}59$$

$$p = 2.(P(T \leq 1{,}58) = 2.0{,}061 = 0{,}122$$

<u>p-hodnota menší než 0,05 – NEZAMÍTÁME NULOVOU HYPOTÉZU</u>

Testy o rovnosti průměru - váhy

Systematická chyba měřicího přístroje se eliminuje nastavením přístroje a měřením etalonu,
jehož správná hodnota je μ = 10,00.
Nezávislými měřeními za stejných podmínek byly získány hodnoty: 10,24 10,12 9,91 10,19 9,78 10,14 9,86 10,17 10,05, které považujeme
za realizace náhodného výběru rozsahu 9 z rozložení $N(\mu,\sigma^2)$.

Je možné při riziku 0,05
vysvětlit odchylky od hodnoty 10,00 působením náhodných vlivů?

*mereni_etalonu.sta

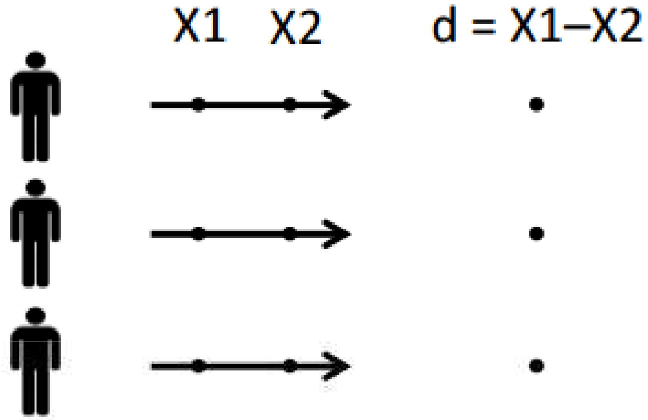
Testy o rovnosti průměru – automat

Při kontrole balicího automatu, který má plnit cukrem balíčky
o hmotnosti 1000 g, byly při přesném převážení
5 balíčků zjištěny tyto odchylky (v gramech)
od požadované hodnoty: 3, -2, 2, 0, 1.

Na hladině významnosti 0,05 testujte hypotézu, že automat nemá systematickou odchylku od požadované hodnoty.

Dvouvýběrový t-test - <u>párový</u>

- **Srovnáváme dvě skupiny dat, které ale na sobě nejsou nezávislé – mezi objekty existuje vazba** (např. člověk před a po operaci, stejný kmen krys)
- Příklady: srovnání objem 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