Přednáška 5

Provádění odhadů

Bodové a intervalové odhady

Význam intervalu spolehlivosti

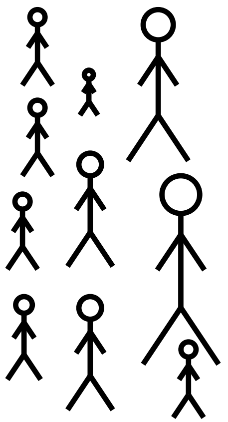
Anotace

- Dva základní přístupy statistického hodnocení jsou popis dat a testování hypotéz.
- Při popisu dat je třeba si uvědomit, že popisné statistiky získané ze vzorku nejsou skutečnou hodnotou v cílové populaci, ale pouze jejím odhadem.
- Přesnost odhadu závisí jednak na variabilitě dat, jednak na velikosti vzorku, při vzorkování celé cílové populace by výsledná popisná statistika již byla přesnou hodnotou, nikoliv odhadem.
- Odhady a s nimi související intervaly spolehlivosti jsou univerzálním statistickým postupem a je možné je dopočítat k libovolné popisné statistice.

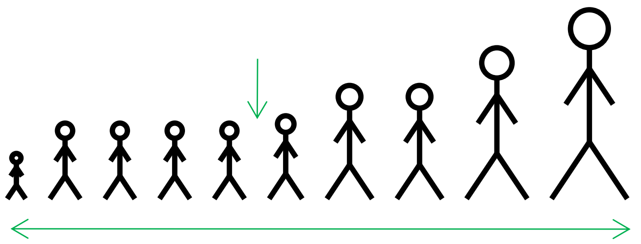
Práce s variabilitou v analýze dat

- V analýze dat existují tři hlavní přístupy k práci s variabilitou

Variabilita dat

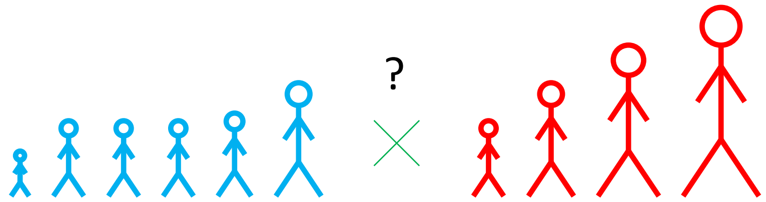


Popisná analýza: <u>popis variability</u>

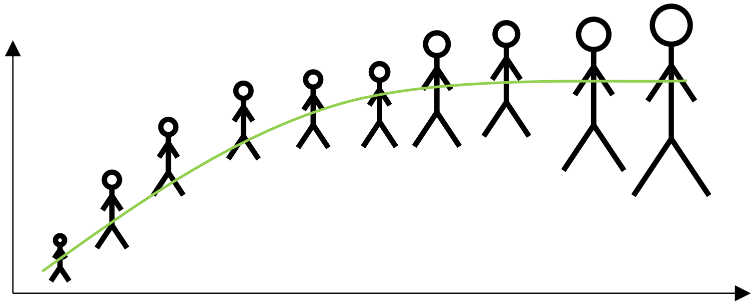


<u>Odhady popisné statistiky</u>

Testování hypotéz: <u>vysvětlení variability</u>

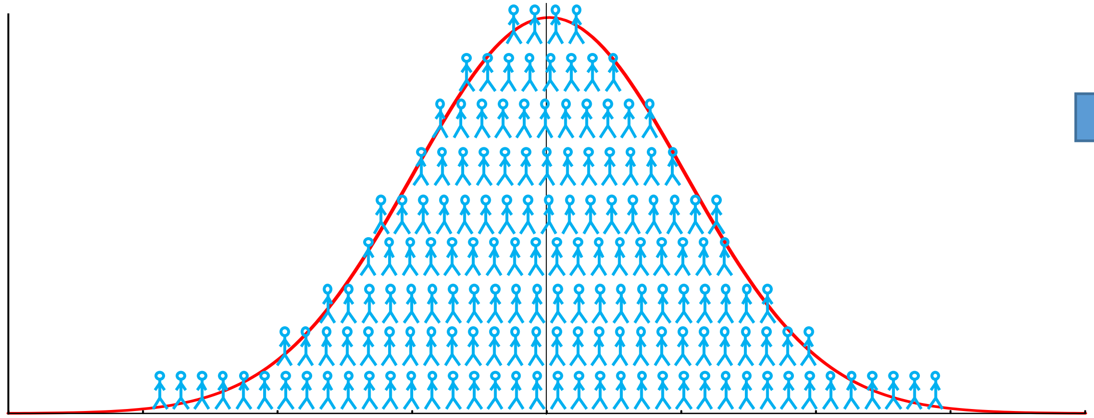


Stochastické modelování: <u>predikce chování systému</u>



Bodový odhad popisné statistiky

- Výpočtem popisné statistiky vzorku získáme tzv. bodový odhad



Bodový odhad průměru, směrodatné odchylky

Je to dostatečné?

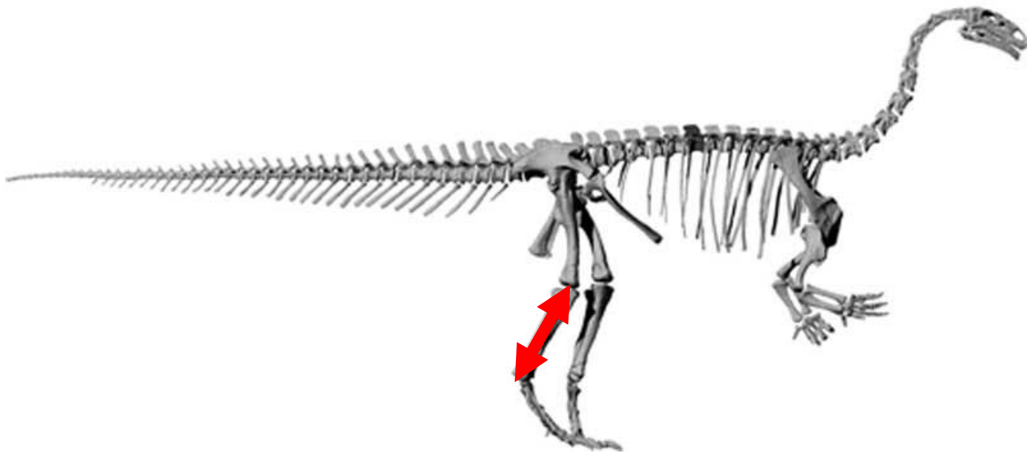
Není, nezohledňujeme vliv náhody, která se uplatnila při vzorkování !!!

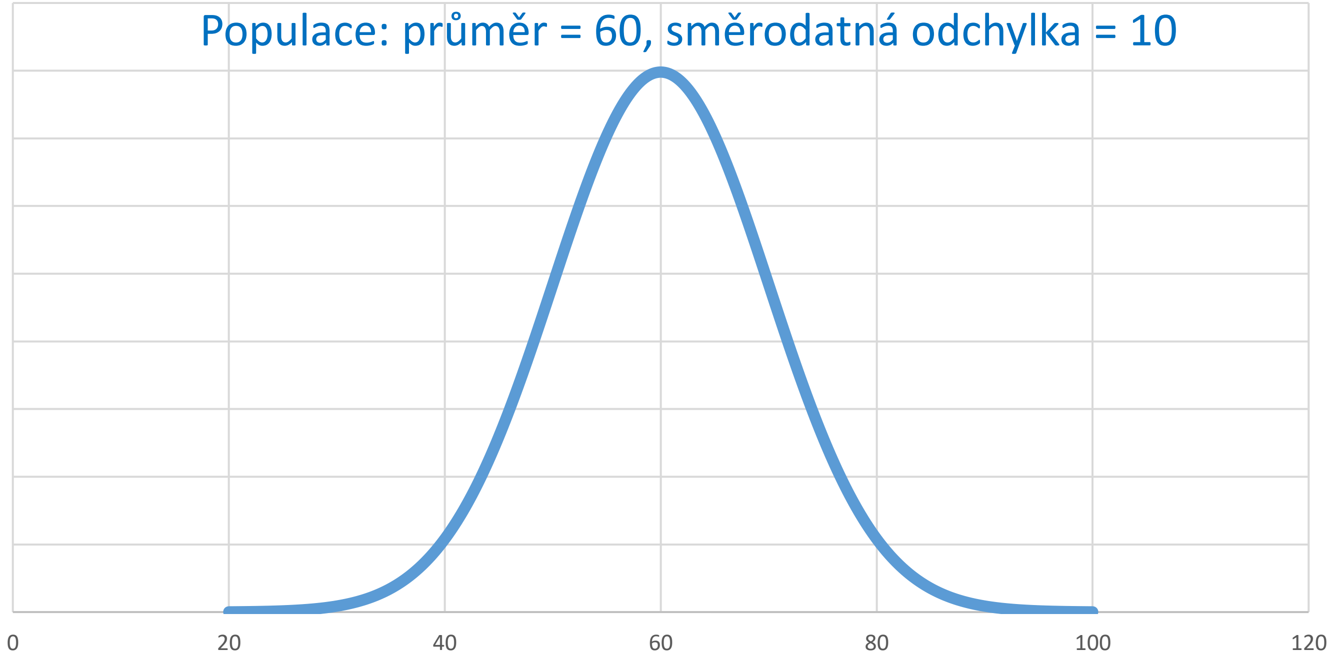
Intervalový odhad

- Bodový odhad je prvním krokem ve statistickém popisu dat.
- Co nám říká jedno číslo? Studie 1 může publikovat číslo x1, studie 2 číslo x2. Které je správnější, lepší, přesnější?
- **Bodový odhad je sám o sobě nedostatečný pro popis parametru rozdělení pravděpodobnosti náhodné veličiny.**
- Zajímá nás přesnost (spolehlivost) bodového odhadu.

Jaký je význam intervalového odhadu a jeho spolehlivosti?

- Provádíme vzorkování populace živočichů a chceme odhadnout průměrnou hodnotu sledované proměnné
- Průměrná délka v populaci = 60, směrodatná odchylka = 10 (tyto hodnoty ve skutečnosti neznáme)

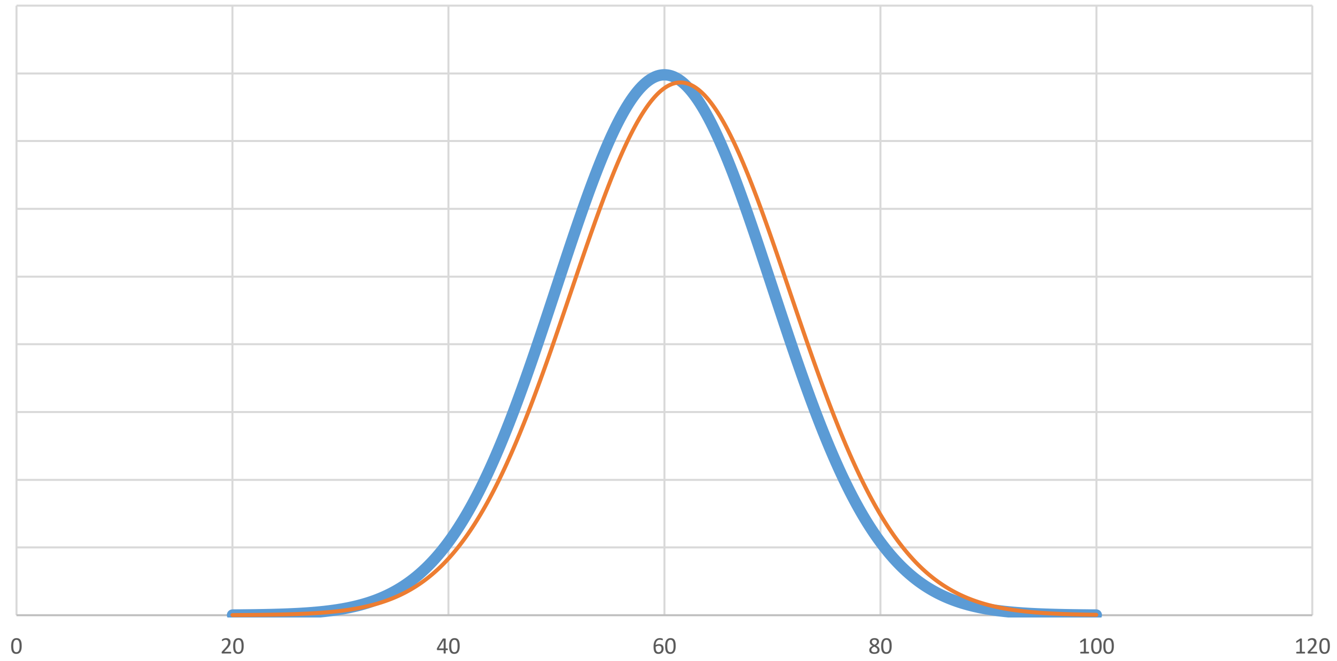




Provedeme vzorkování o velikosti N = 100.

Jedno vzorkování

- Je pouze nízká pravděpodobnost, že vzorek zcela přesně odpovídá sledované populaci



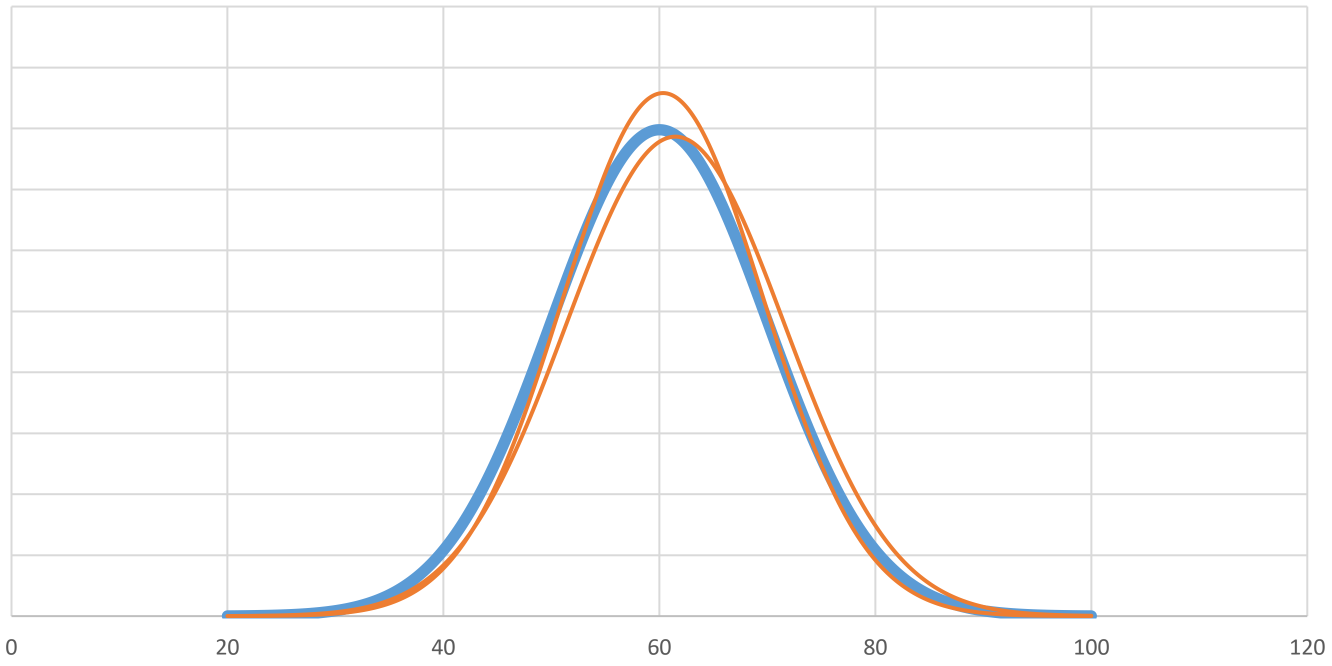
Populace: průměr = 60, směrodatná odchylka = 10

Vzorek 1: průměr = 61.5, směrodatná odchylka = 10.1

Jak by dopadlo další vzorkování?

Dvě vzorkování

- Je pouze nízká pravděpodobnost, že vzorek zcela přesně odpovídá sledované populaci



Populace: průměr = 60, směrodatná odchylka = 10

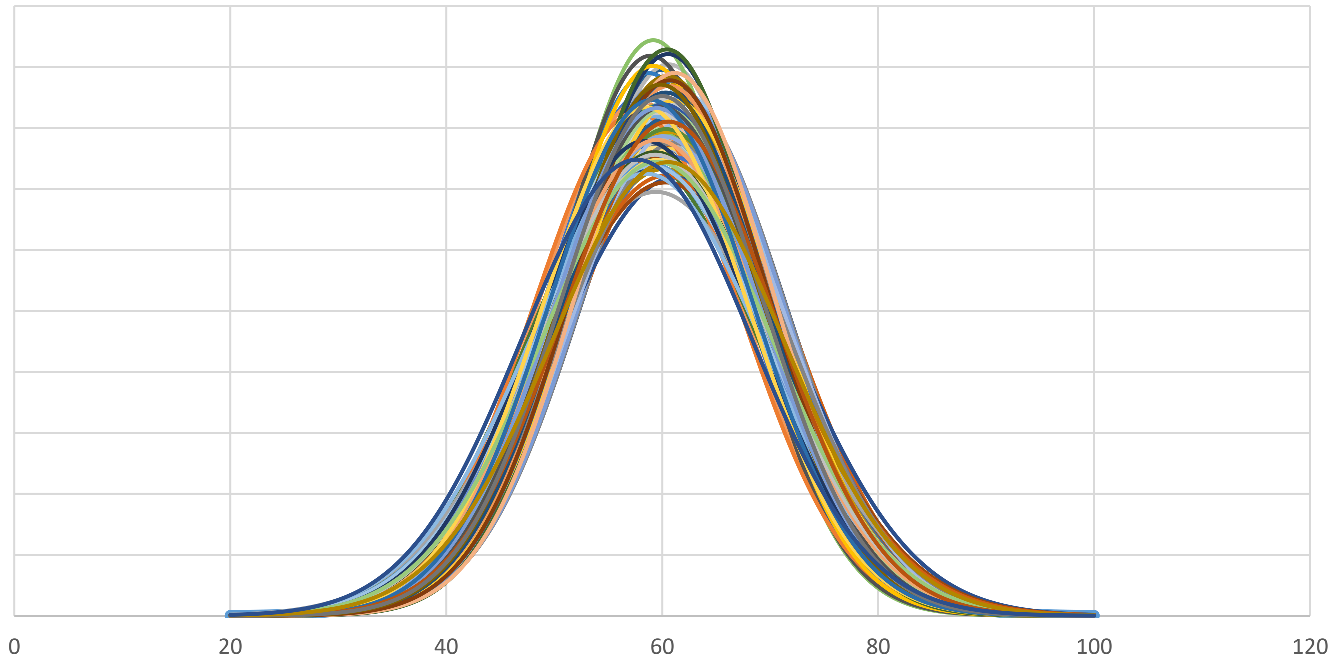
Vzorek 1: průměr = 61.5, směrodatná odchylka = 10.1

Vzorek 2: průměr = 60.4, směrodatná odchylka = 9.3

Jak by dopadlo další vzorkování?

Sto vzorkování

- Je pouze nízká pravděpodobnost, že vzorek zcela přesně odpovídá sledované populaci



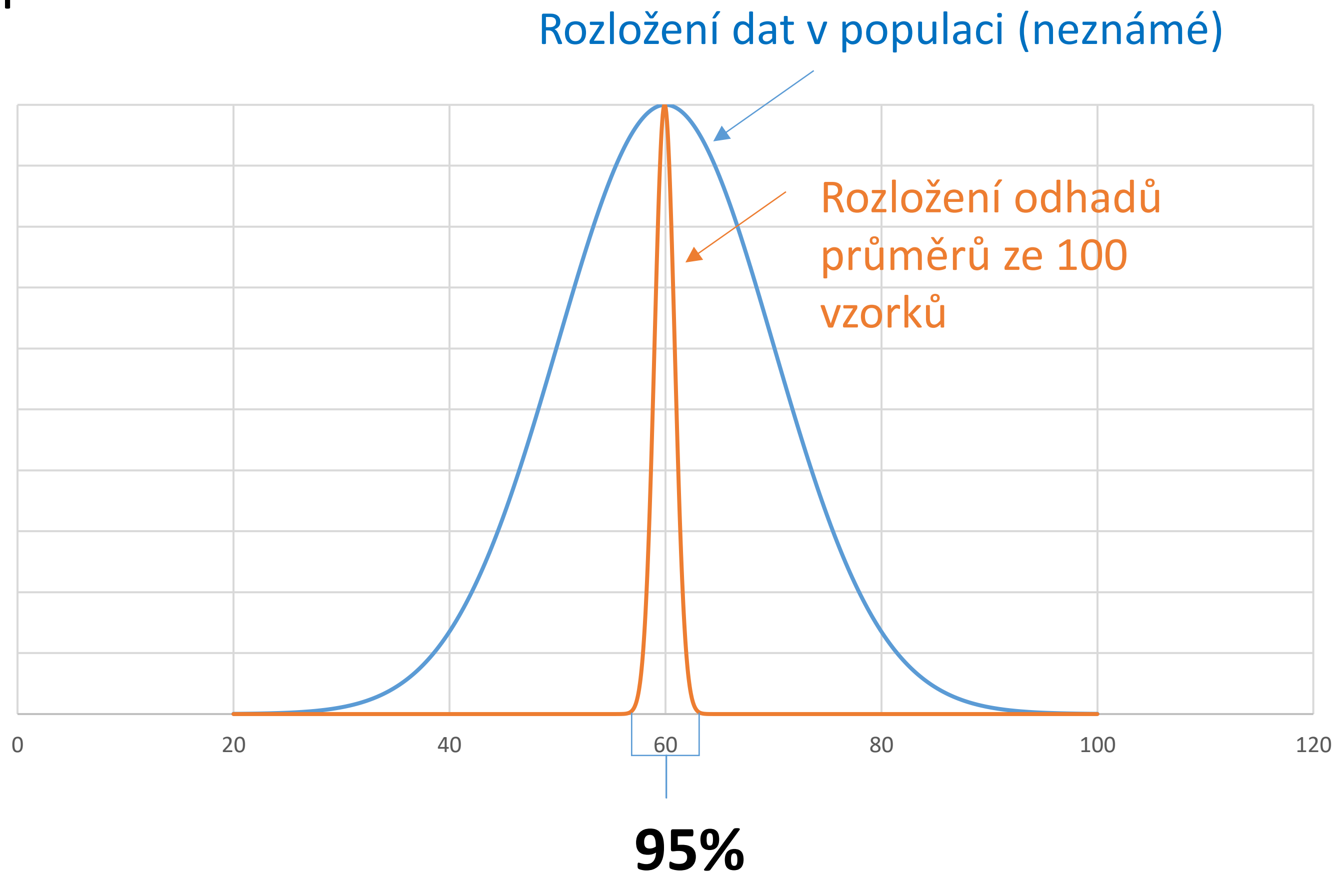
Populace: průměr = 60, směrodatná odchylka = 10

Opakovaným vzorkováním jsme získali různé varianty bodového odhadu simulující jak by při dané velikosti vzorku dopadlo různé vzorkování populace.

Jak by dopadlo další vzorkování?
Jsme schopni jej popsat z pohledu pravděpodobnosti = odhad při dalším vzorkování skončí s určitou pravděpodobností v určitém rozsahu hodnot?

Interval spolehlivosti odhadu I

- Odhady průměru z jednotlivých vzorků vytváří rozložení odhadu průměrů
- Pokud známe rozložení jsme snadno určit rozsah, v němž leží zadané procento hodnot = pravděpodobnost s níž při vzorkování narazíme na odhad průměru v tomto rozmezí
- Nejběžněji se používá 95% rozsah = **95% interval spolehlivosti**
- Jak jej spočítat?



Populace: průměr = 60, směrodatná odchylka = 10

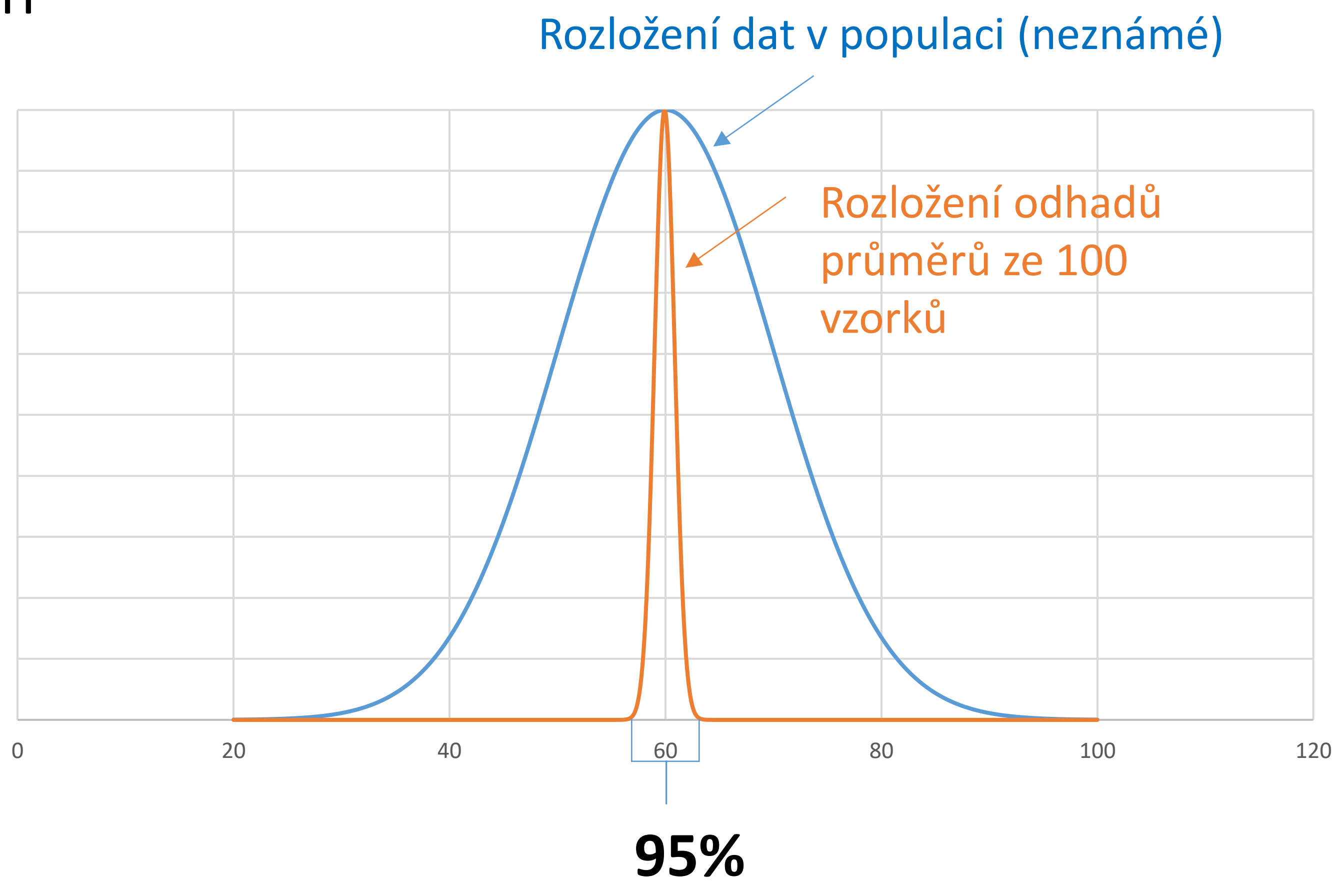
Vzorky (N = 100): průměr = 59.9, směrodatná odchylka odhadů průměru= 0.93

???

Interval spolehlivosti odhadu II

- Jak jej spočítat?
- Empiricky: 2,5% a 97,5% kvantil
- Dle modelového rozdělení:
    - Odhady průměrů mají normální rozdělení
    - Středních 95% hodnot ohraničuje průměr ± 1,96*směrodatná odchylka

- Poznámka: popsaný způsob výpočtu intervalu spolehlivosti se používá pouze v počítačových simulacích, ne při reálném vzorkování (zde z výukových důvodů)



Populace: průměr = 60, směrodatná odchylka = 10

Vzorky (N = 100): průměr = 59.9, směrodatná odchylka odhadů průměru= 0.93

Střední chyba odhadu průměru (standard error, s.e., SE, $s_{\bar{x}}$)

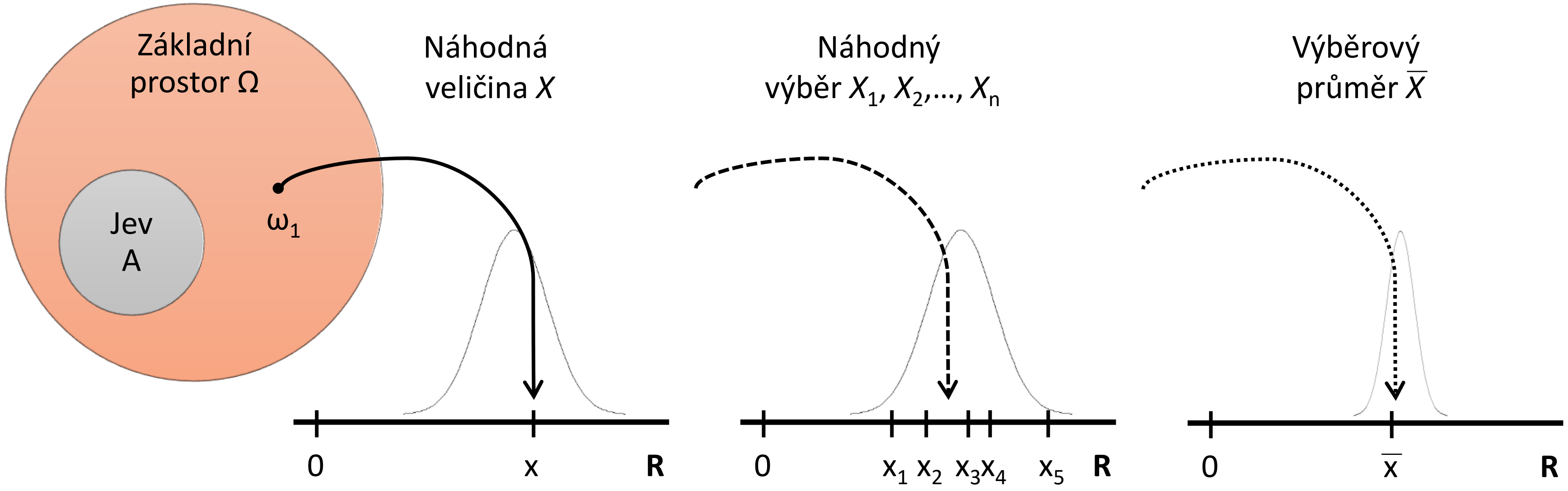
Pravděpodobnostní chování náhodné veličiny

- V klasických statistických výpočtech je interval spolehlivosti odvozen z jednoho vzorku na základě znalosti modelového rozdělení odhadů dané statistiky (např. průměru)
- Dvě charakteristiky odráží vlastnosti rozdělení jedním číslem: střední hodnota a rozptyl. Odmocnina z rozptylu je směrodatná odchylka (SD).

- Platí následující:
  - Jednotlivé realizace náhodné veličiny vykazují variabilitu (dle SD).
  - Jakákoliv statistika (např. průměr) je jako transformace náhodných veličin také náhodnou veličinou. Má tedy i rozdělení pravděpodobnosti.
  - Jednotlivé realizace statistiky nad různými náhodnými výběry také vykazují variabilitu (opět úměrnou SD).
    - S.E. – standard error – střední chyba odhadu

Příklad – výběrový průměr

- V případě průměru jsou jeho odhady popsatelné modelem normálního rozdělení
- Normální rozdělení je popsáno průměrem (vlastní odhad průměru) a směrodatnou odchylkou odhadů (pro odlišení od směrodatné odchylky vzorku se v tomto případě nazývá střední chyba odhadu průměru)



SD a SE

- Směrodatná odchylka (SD) není směrodatná chyba popisné statistiky (SE)!
- Směrodatná odchylka (SD) je odrazem variability náhodné veličiny ve sledované populaci.
- Směrodatná chyba (SE) je odrazem přesnosti popisné statistiky jako odhadu střední hodnoty náhodné veličiny.
- Pozor na rozdíl mezi SD a SE v článcích a knihách – tabulkách a grafech!
- **Na čem závisí velikost SE (a tedy i šířka intervalu spolehlivosti?)**

SD a SE

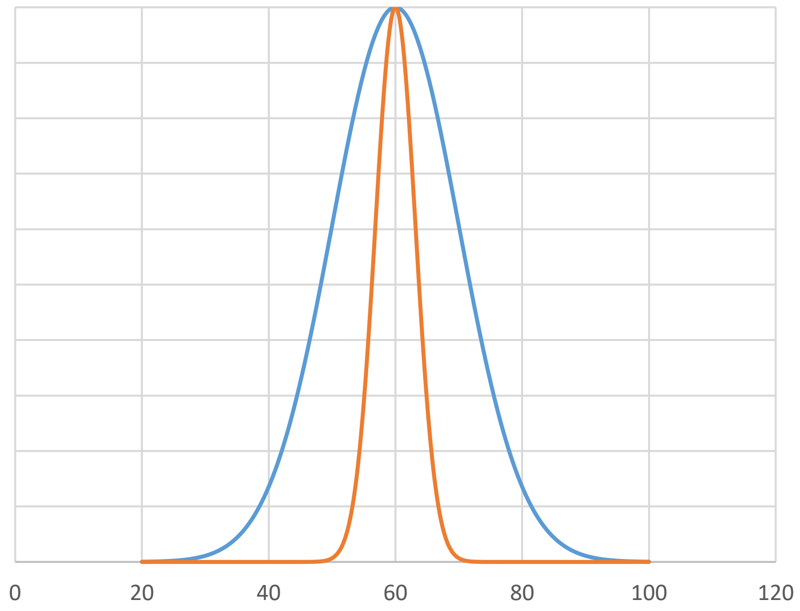
- Směrodatná odchylka (SD) není směrodatná chyba popisné statistiky (SE)!
- Směrodatná odchylka (SD) je odrazem variability náhodné veličiny ve sledované populaci.
- Směrodatná chyba (SE) je odrazem přesnosti popisné statistiky jako odhadu střední hodnoty náhodné veličiny.
- Pozor na rozdíl mezi SD a SE v článcích a knihách – tabulkách a grafech!
- **Na čem závisí velikost SE (a tedy i šířka intervalu spolehlivosti?)**
    - Na velikosti vzorku
    - Variabilitě (směrodatné odchylce) hodnocené proměnné v populaci

- SD populace je daná realitou, ale velikost vzorku je v našich rukou = **změnou velikosti vzorku můžeme měnit šíři intervalu spolehlivosti !!!!**

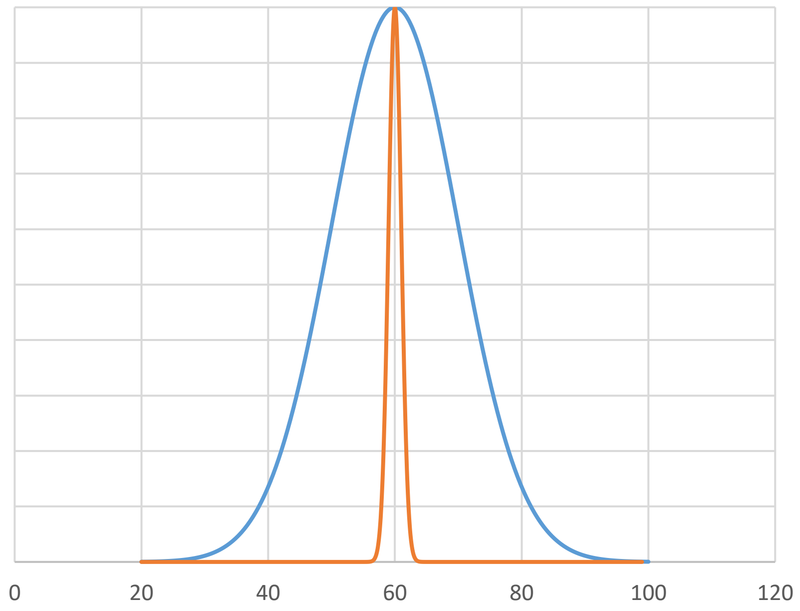
Příklad – interval spolehlivosti při různých velikostech vzorku

- Provádíme vzorkování populace živočichů a chceme odhadnout průměrnou hodnotu sledované proměnné – zkoušíme různé velikosti vzorku
- Průměrná délka v populaci = 60, směrodatná odchylka = 10 (tyto hodnoty ve skutečnosti neznáme)

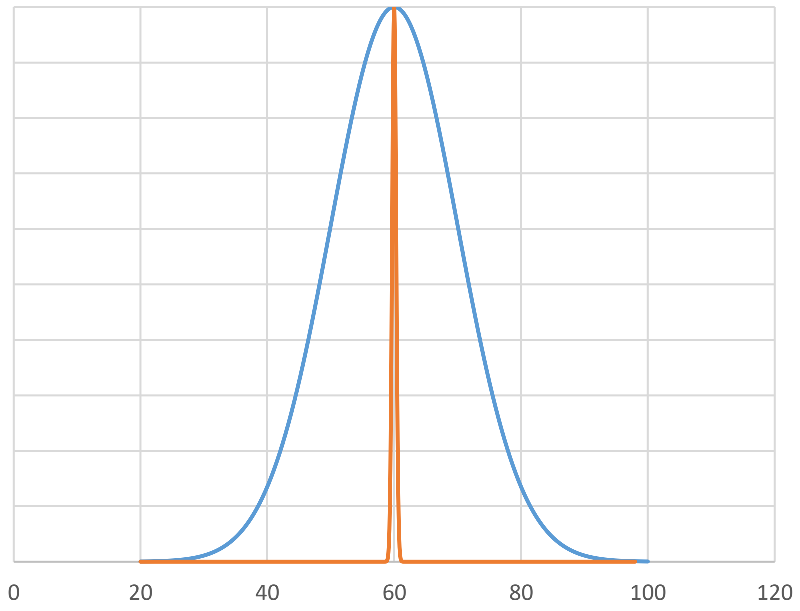
N = 10



N = 100



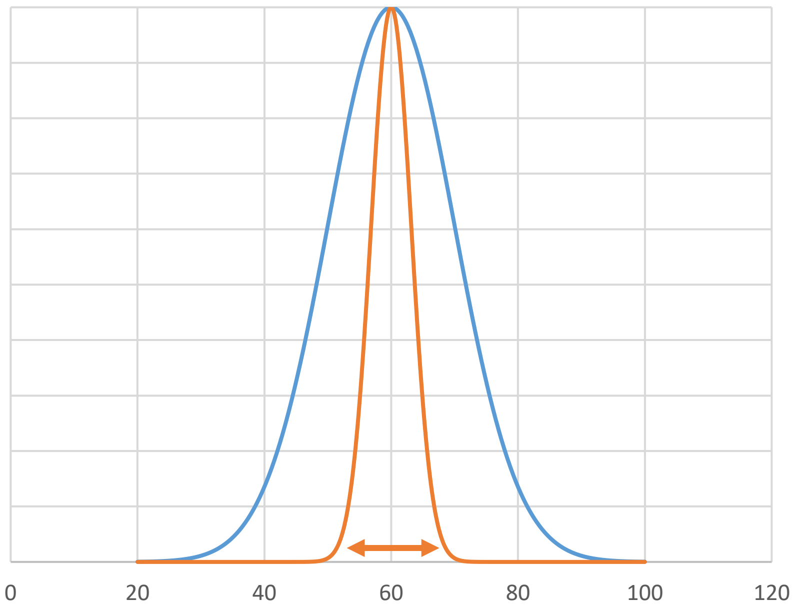
N = 1000



Příklad – interval spolehlivosti při různých velikostech vzorku

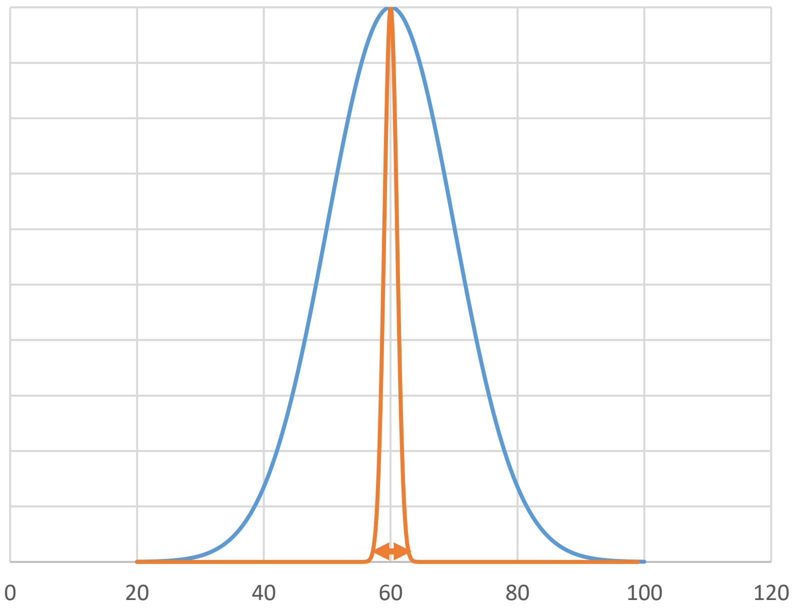
- Provádíme vzorkování populace živočichů a chceme odhadnout průměrnou hodnotu sledované proměnné – zkoušíme různé velikosti vzorku
- Průměrná délka v populaci = 60, směrodatná odchylka = 10 (tyto hodnoty ve skutečnosti neznáme)

N = 10



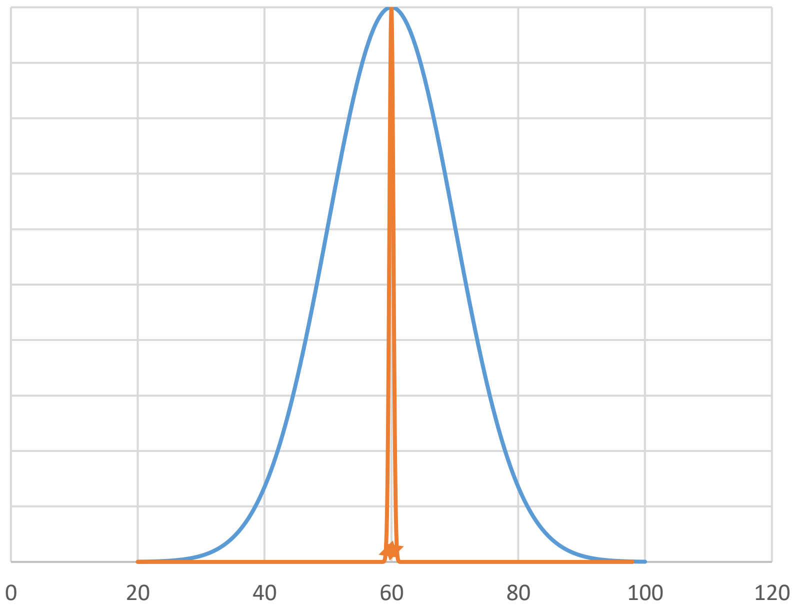
95% IS = 53,8 – 66,2

N = 100



95% IS = 58,0 – 62,0

N = 1000



95% IS = 59,4 – 60,6

Obecný vzorec výpočtu intervalu spolehlivosti

- Interval spolehlivosti lze spočítat pro odhad jakékoliv popisné statistiky (průměr, směrodatná odchylka, procento, korelační koeficient, regresní koeficient, odds ratio atd.)
- Pro danou popisnou statistiku musíme znát odpovídající modelové rozdělení jejího odhadu
- Obecná rovnice pro výpočet hranic intervalu spolehlivosti (v některých případech může být složitější – asymetrické intervaly spolehlivosti, různá rovnice pro dolní a horní hranici):

Bodový odhad ± kvantil modelového rozdělení * střední chyba odhadu

Např. průměr vzorku

V případě průměru a 95% intervalu spolehlivosti to je 2.5% a 97.5% kvantil normálního rozdělení = ± 1.96

V případě průměru je vypočtena jako:

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

Výpočet odhadu průměru

- Bodový odhad průměru daného vzorku $\bar{x}$
- Střední chyba odhadu průměru $s_{\bar{x}} = \frac{s}{\sqrt{N}}$

- Interval spolehlivosti

$$\bar{x} - t_{1-\alpha/2}^{\upsilon=N-1} \frac{s}{\sqrt{N}} \leq \mu \leq \bar{x} + t_{1-\alpha/2}^{\upsilon=N-1} \frac{s}{\sqrt{N}}$$

$$\mu: \bar{x} \pm t_{1-\alpha/2}^{\upsilon=N-1} \frac{s}{\sqrt{N}}$$

$$\mu: \bar{x} \pm t_{1-\alpha/2}^{\upsilon=N-1} s_{\bar{x}}$$

t – Studentovo rozdělení (používáno namísto normálního při malé velikosti vzorku)

ν – stupně volnosti, zde počítány jako N-1

Co je ? $t_{1-\alpha/2}^{\upsilon=N-1}$

Kvantil modelového rozdělení, α znamená zastoupení případů, které do intervalu nechceme zahrnout, zde pro 95% interval spolehlivosti je α = 5%, hledáme tedy 97.5% kvantil studentova rozdělení

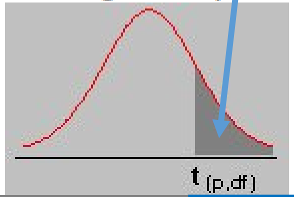
Statistické tabulky t-rozdělení

- Na rozdíl od tabulek normálního rozdělení musíme zohlednit i stupně volnosti
- Z tohoto důvodu je tabulka konstruována jen pro vybrané hodnoty pravděpodobnosti

William Sealy Gosset
Publikace pod pseudonymem Student
t rozdělení na základě experimentů s kvasinkami

Hledáme hodnotu **t** (= kvantil rozdělení) pro danou plochu **(pravděpodobnost)** a **stupně volnosti**

t table with right tail probabilities



Pravděpodobnost (plocha pod křivkou), nejběžněji 0.025 (2*0.025=0.05)

df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869

Stupně volnosti

Odhad průměru a jeho intervalu spolehlivosti – příklad 1

- Provádíme vzorkování populace živočichů a chceme odhadnout průměrnou hodnotu sledované proměnné
- Vzorek: N = 10, průměr (bodový odhad) 61,5, směrodatná odchylka 10,1

- **Jaký je 95% interval spolehlivosti?**

- Střední chyba odhadu $s_{\bar{x}} = \frac{s}{\sqrt{N}} = \frac{10{,}1}{\sqrt{10}} = 3{,}207$

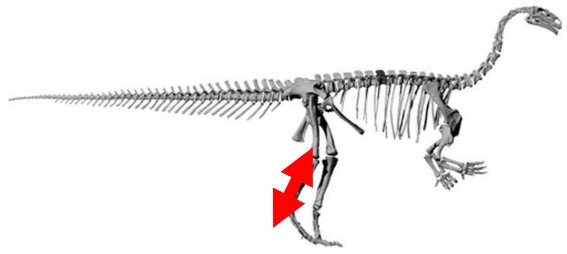
- Kvantil modelového rozdělení pro α=0,05 (1-0,95)

$$t_{1-\alpha/2}^{\upsilon=N-1} = t_{1-0{,}05/2}^{\upsilon=10-1} = t_{0{,}975}^{9}=2{,}262$$

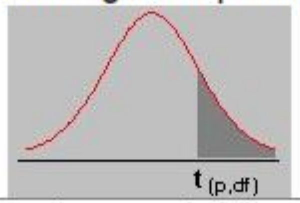
- 95% interval spolehlivosti – výpočet

$$\mu: \bar{x} \pm t_{1-\alpha/2}^{\upsilon=N-1} \frac{s}{\sqrt{N}} = 61{,}5 \pm 2{,}262 * 3{,}207=61{,}5 \pm 7{,}256$$

- 95% interval spolehlivosti - výsledek
  61,5 (54,2 – 68,7)



t table with right tail probabilities



df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809

- **Při opakovaném vzorkování o N=10 bude odhad průměru s pravděpodobností 0,95 ležet v rozsahu (54,2 – 68,7)**

Odhad průměru a jeho intervalu spolehlivosti – příklad 2

- Provádíme vzorkování populace živočichů a chceme odhadnout průměrnou hodnotu sledované proměnné
- Vzorek: N = 100, průměr (bodový odhad) 61,5, směrodatná odchylka 10,1

- **Jaký je 95% interval spolehlivosti?**

- Střední chyba odhadu $s_{\bar{x}} = \frac{s}{\sqrt{N}} = \frac{10,1}{\sqrt{100}} = 1,014$

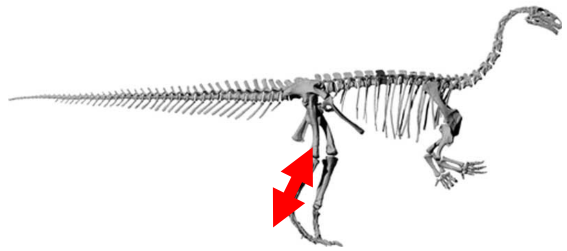
- Kvantil modelového rozdělení pro α=0,05 (1-0,95)

$$t_{1-\alpha/2}^{\upsilon=N-1} = t_{1-0,05/2}^{\upsilon=100-1} = t_{0,975}^{99} = 1,960$$

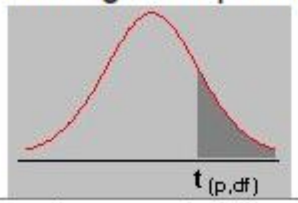
- 95% interval spolehlivosti – výpočet

$$\mu: \bar{x} \pm t_{1-\alpha/2}^{\upsilon=N-1} \frac{s}{\sqrt{N}} = 61,5 \pm 1,960 * 1,014 = 61,5 \pm 1,988$$

- 95% interval spolehlivosti - výsledek
  61,5 (59,5 – 63,5)



t table with right tail probabilities



df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

- **Při opakovaném vzorkování o N=100 bude odhad průměru s pravděpodobností 0,95 ležet v rozsahu (59,5 – 63,5)**

Interval spolehlivosti pro odhad rozptylu

- Příklad asymetrického intervalu spolehlivosti; modelovým rozdělením je Pearsonovo (chi-kvadrát rozdělení)
- **Pro rozptyl**

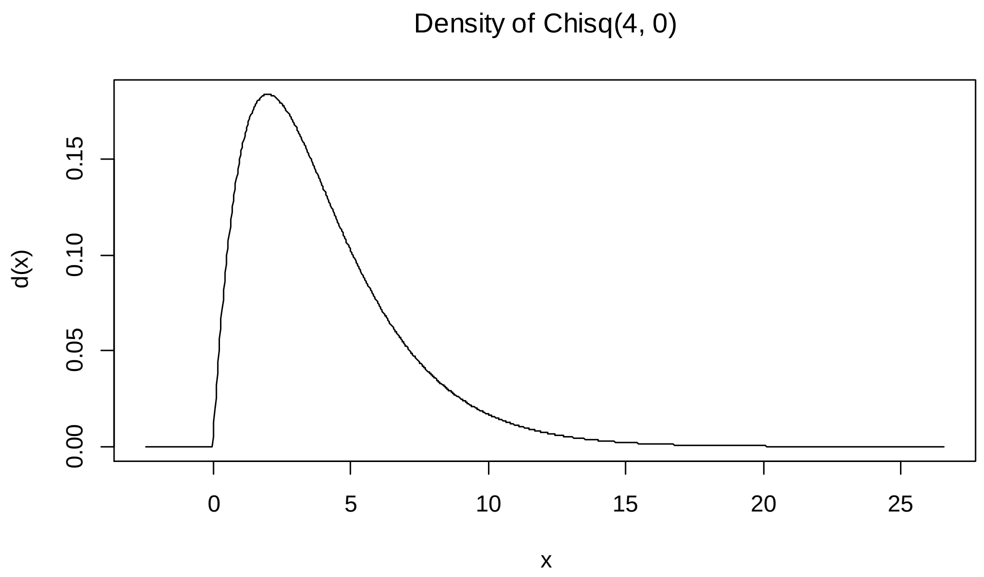
$$\frac{(N-1)s^2}{{x^2}_{\alpha/2}^{\nu=N-1}} \leq \sigma^2 \leq \frac{(N-1)s^2}{{x^2}_{1-\alpha/2}^{\nu=N-1}}$$

- **Pro směrodatnou odchylku**

$$\sqrt{\frac{(N-1)s^2}{{x^2}_{\alpha/2}^{\nu=N-1}}} \leq \sigma \leq \sqrt{\frac{(N-1)s^2}{{x^2}_{1-\alpha/2}^{\nu=N-1}}}$$

- **Pro střední chybu odhadu průměru**

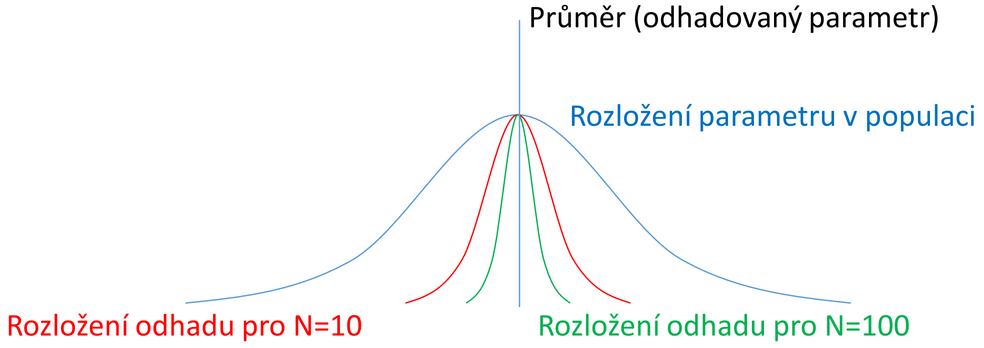
$$\sqrt{\frac{(N-1)s^2}{N{x^2}_{\alpha/2}^{\nu=N-1}}} \leq \frac{\sigma}{\sqrt{N}} \leq \sqrt{\frac{(N-1)s^2}{N{x^2}_{1-\alpha/2}^{\nu=N-1}}}$$



Koncept intervalu spolehlivosti a jeho interpretace: shrnutí

- Při výpočtu odhadu popisné statistiky nás zajímá nejenom její vlastní hodnota (bodový odhad) ale také její rozsah spolehlivosti

- Interval spolehlivosti závisí na:
    - Velikosti vzorku
    - Variabilitě dat
    - Požadované spolehlivosti



- Interval spolehlivosti lze spočítat pro jakoukoliv statistiku (průměr, směrodatná odchylka, korelace, procentuální zastoupení apod.)
- Interval spolehlivosti poskytuje vodítko jak „spolehlivé" jsou naše výsledky a s jakou pravděpodobností jich je možné opakovaně dosáhnout
- 95% interval spolehlivosti je rozsah hodnot do nějž se při opakování studie trefíme s 95% pravděpodobností
- Tvrzení, že v rozsahu 95% intervalu spolehlivosti leží s 95% pravděpodobností skutečný průměr populace není pravdivé, skutečný průměr populace neznáme !!!

Poznámka k intervalu spolehlivosti

- Interval spolehlivosti počítá pouze s variabilitou danou náhodným výběrem, nepočítá se zdroji systematického zkreslení.

- **Příklady:**
- Měření koncentrace polutantu nebo krevního tlaku může být systematicky zkresleno starým měřidlem („technical bias“).
- Měření koncentrace polutantu může být systematicky zkresleno výběrem pouze čistých nebo pouze kontaminovaných lokalit („selection bias“)
- Měření krevního tlaku může být systematicky zkresleno tím, že se do studie přihlásí pouze určitá skupina osob („selection bias“)

Základy testování hypotéz

Princip statistického testování hypotéz

Testová statistika a statistická významnost

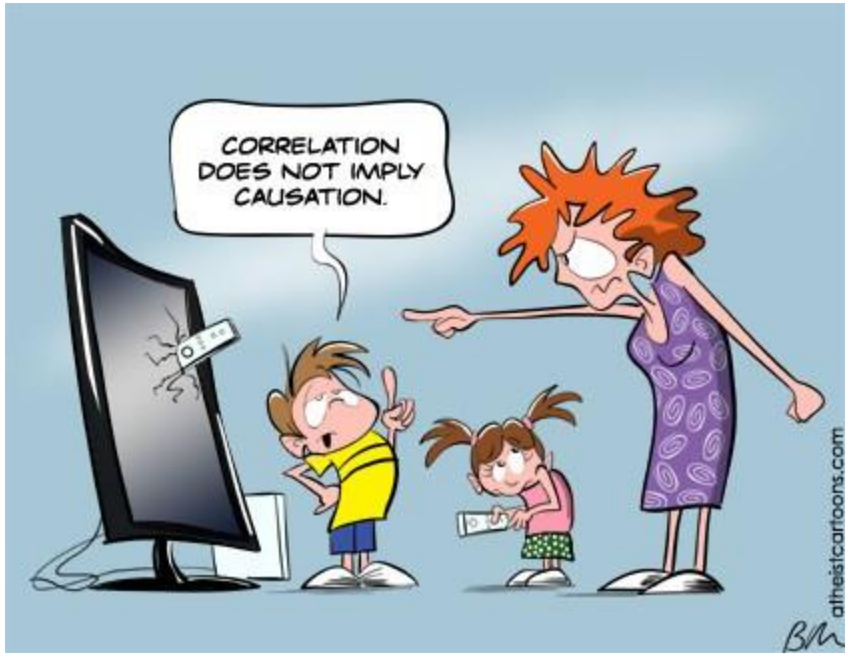
Chyby statistického testování

Anotace

- Testování hypotéz je po popisné statistice druhým hlavním směrem statistických analýz. Při testování pokládáme hypotézy, které se snažíme s určitou pravděpodobností potvrdit nebo vyvrátit.
- Tzv. nulovou hypotézu lze nejlépe popsat jako situaci, kdy předpokládáme vliv náhody (rozdíl mezi skupinami je pouhá náhoda, vztah dvou proměnných je pouhá náhoda apod.), alternativní hypotéza předpokládá vliv nenáhodného faktoru.
- Výsledkem statistického testu je v zásadě pravděpodobnost nakolik je hodnocený jev náhodný nebo ne, při překročení určité hranice (nejčastěji méně než 5% pravděpodobnost, že jev je pouhá náhoda) deklarujeme, že pravděpodobnost náhody je pro nás dostatečně nízká abychom jev prohlásili za nenáhodný
- Statistická významnost je ovlivnitelná velikostí vzorku a tak je pouze indicií k prohlášení např. rozdílu dvou skupin pacientů za skutečně významný. V ideální situaci je nezbytné aby rozdíl byl významný nejenom statisticky (=nenáhodný), ale i prakticky (=nejde pouze o artefakt velikosti vzorku).

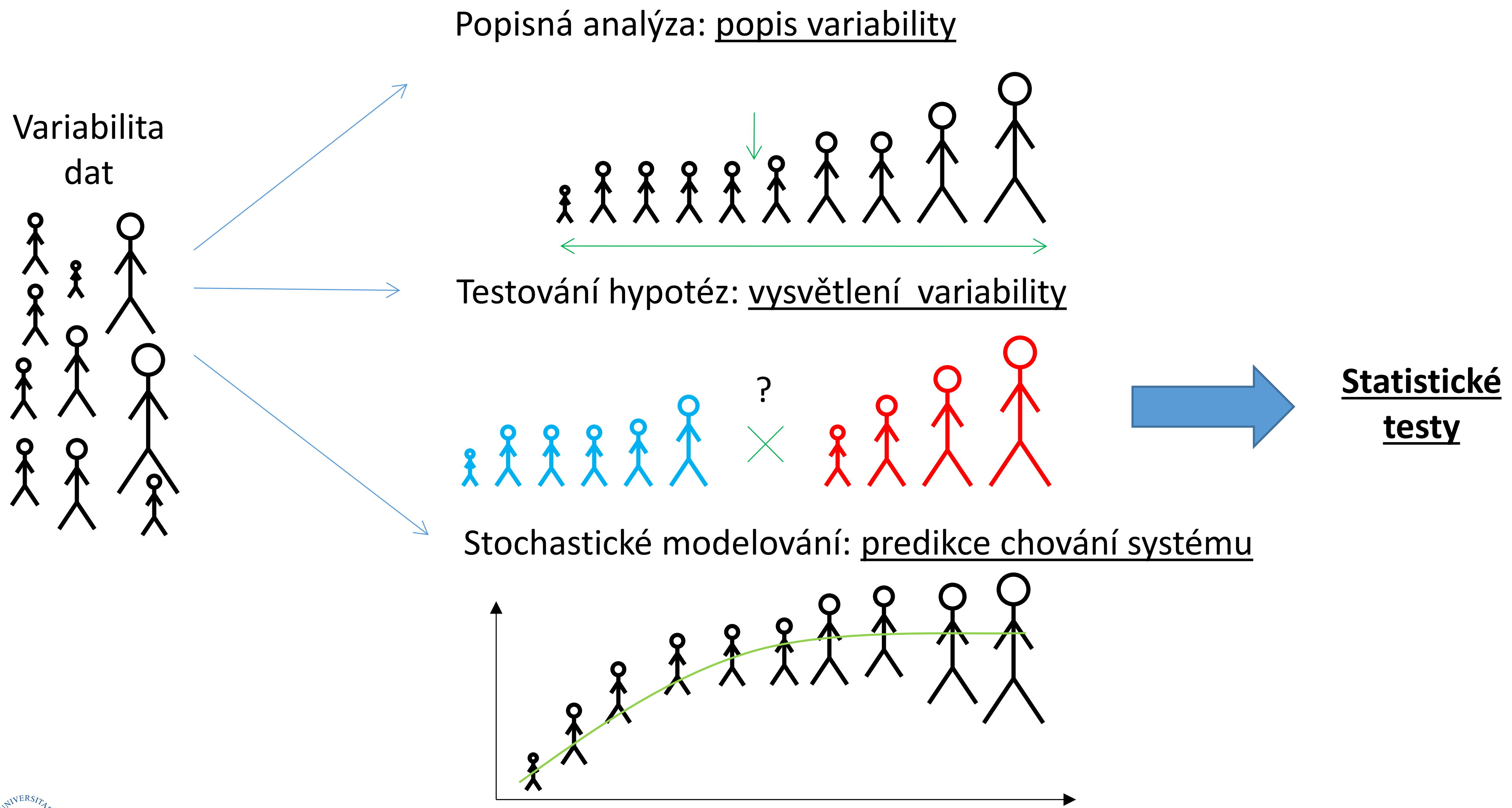
Statistické testování neznamená průkaz kauzality !!!!

- Výsledek statistického testování neznamená kauzální prokázání nebo neprokázání vztahu, jde pouze o indicii k našemu rozhodování.



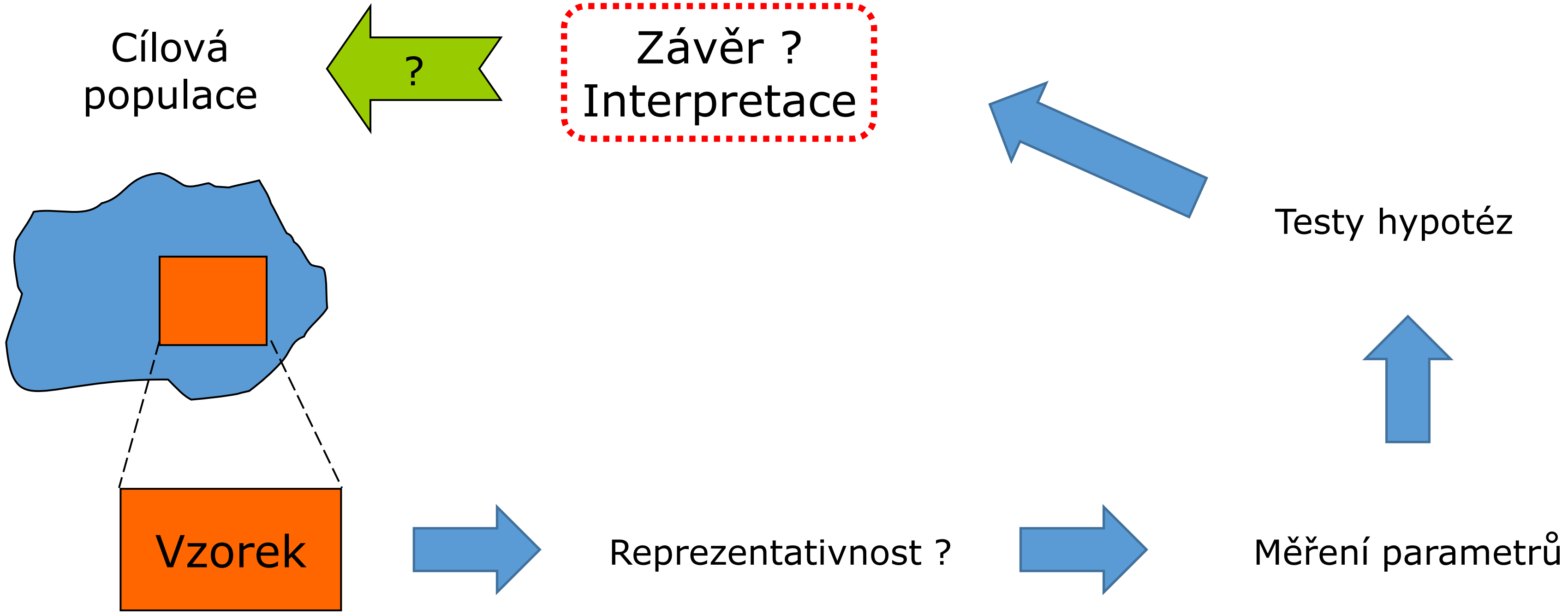
Práce s variabilitou v analýze dat

- V analýze dat existují tři hlavní přístupy k práci s variabilitou



Princip testování hypotéz

- Formulace hypotézy
- Výběr cílové populace a z ní reprezentativního vzorku
- Měření sledovaných parametrů
- Použití odpovídajícího testu → závěr testu
- Interpretace výsledků



Stanovení hypotézy

- **Nulová hypotéza („null hypothesis“)** – tvrzení o neznámých vlastnostech rozdělení pravděpodobnosti sledované náhodné veličiny (znaku, vlastnosti) týkající se cílové populace.
- Nulová hypotéza má tvar: $H_0 : \theta = \theta_0$
- Nulová hypotéza obecně říká, že rozdíl není, popřípadě, že rozdíl je tak malý, že jej můžeme považovat za náhodný -> základní otázkou testování tak je „jak definovat co je pro nás „dostatečně“ náhodné?“

- **Alternativní hypotéza** – tvrzení o neznámých vlastnostech rozdělení pravděpodobnosti sledované náhodné veličiny, které popírá platnost nulové hypotézy. Vymezuje, jaká situace nastává, když nulová hypotéza neplatí.
- Alternativní hypotéza má tvar:

$$H_1 : \theta \neq \theta_0$$

$$H_1 : \theta < \theta_0$$

$$H_1 : \theta > \theta_0$$

Příklady hypotézy

- Liší se lokality poblíž lidských sídel od lokalit v chráněných rezervacích co do míry znečištění?

  Míra znečištění na lokalitách poblíž sídel: θ_1 $H_0 : \theta_1 = \theta_2$

  Míra znečištění na lokalitách v rezervacích: θ_2 $H_1 : \theta_1 \neq \theta_2$

- Je efekt snížení systolického tlaku novým antihypertenzivem stejný u hypertoniků, kteří kouří, jako u hypertoniků, kteří nekouří?

  Střední hodnota efektu u kuřáků: θ_1 $H_0 : \theta_1 = \theta_2$

  Střední hodnota efektu u nekuřáků: θ_2 $H_1 : \theta_1 < \theta_2$

Proč nulová hypotéza vyjadřuje nepřítomnost efektu?

- Nulová hypotéza odráží fakt, že se něco nestalo nebo neprojevilo → je stanovena obvykle jako opak toho, co chceme experimentem prokázat.

- **Nulová hypotéza je postavena tak, abychom ji mohli pomocí pozorovaných hodnot vyvrátit.**

- Pro zamítnutí platnosti nulové hypotézy nám totiž stačí najít jeden příklad, kdy nulová hypotéza neplatí – tím příkladem má být náš náhodný výběr (naše pozorovaná data).

- Zamítnout nulovou hypotézu je jednodušší než nulovou hypotézu potvrdit.

Testování hypotéz

- Testování hypotéz se zabývá rozhodováním o platnosti stanovených hypotéz na základě pozorovaných dat.

- Platnost hypotéz ověřujeme pomocí **statistického testu** – rozhodovacího pravidla, které každému náhodnému výběru přiřadí právě jedno ze dvou možných rozhodnutí – H_0 nezamítáme nebo H_0 zamítáme.

Statistický test

- Testování hypotéz probíhá na základě dat.

- **Testované hypotéze odpovídá statistický test**, respektive testová statistika, která umožní ověřit platnost nulové hypotézy.

- **Testová statistika** je vzorec vycházející z pozorovaných dat s rozdělením pravděpodobnosti, sama tedy má také **rozdělení pravděpodobnosti**. Rozdělení pravděpodobnosti testové statistiky za platnosti H0 se označuje jako „null distribution“.

Postup statistického testování

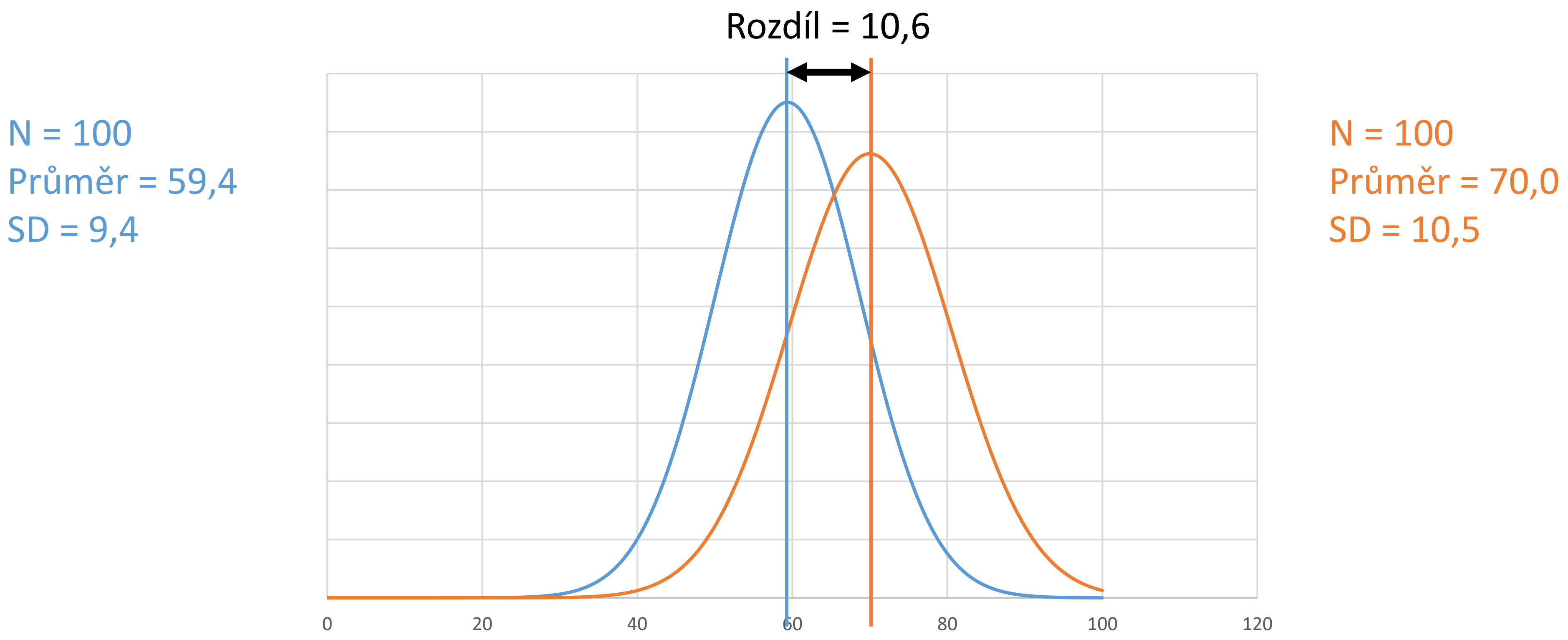
- Formulujeme nulovou hypotézu H_0 (sledovaný efekt je nulový)
- Formulujeme alternativní hypotézu H_A(sledovaný efekt je různý mezi skupinami) Alternativní hypotéza u parametrických testů může být oboustranná nebo jednostranná.

- Hypotéza musí být stanovena tak abychom mohli vybrat a spočítat tzv. testovou statistiku (např. hypotéza o průměrech bude pravděpodobně řešena pomocí t-testu, jehož testová statistika má t rozdělení)

- Hodnotu testové statistiky vypočítáme na základě pozorovaných hodnot
- Vypočtenou testovou statistiku porovnáme s jejím rozdělením (= rozdělení náhodných rozdílů), posoudíme náhodnost rozdílu a vyslovíme závěr o zamítnutí / nezamítnutí H_0

Na čem závisí hodnota testové statistiky?

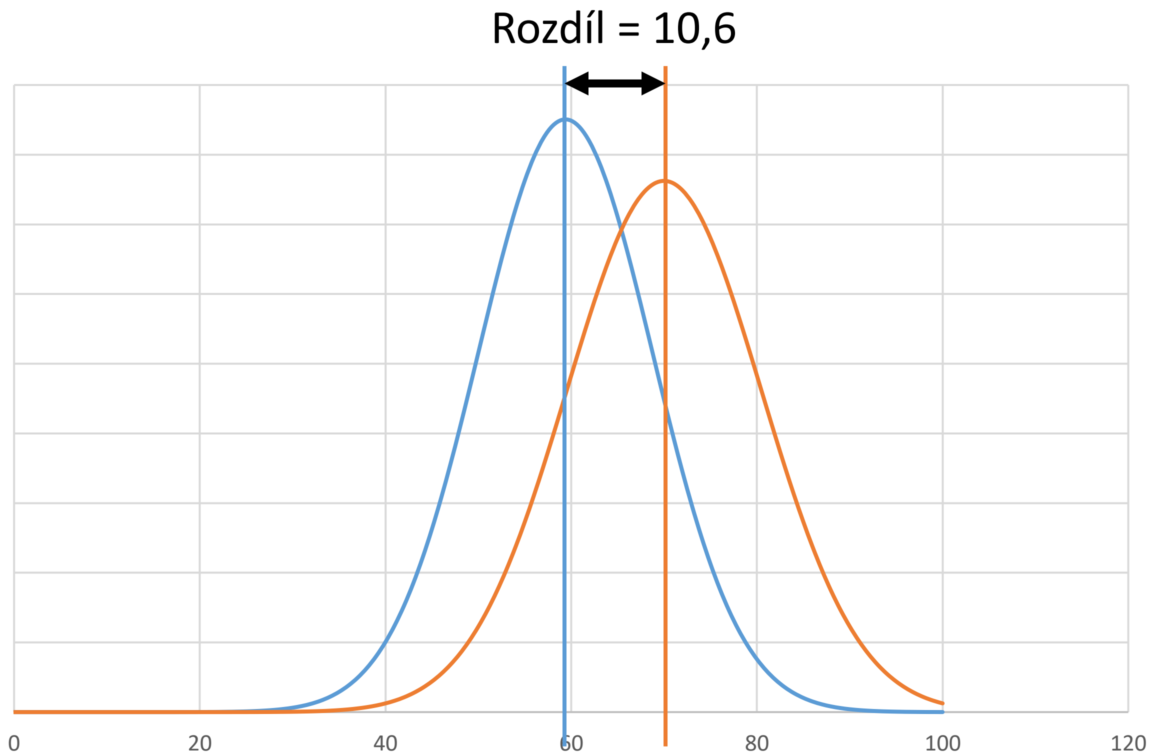
- Máme dvě skupiny hodnot, každá je popsána svojí velikostí, průměrem a směrodatnou odchylkou – co ovlivňuje významnost rozdílu jejich průměrů?



Na čem závisí hodnota testové statistiky?

- Máme dvě skupiny hodnot, každá je popsána svojí velikostí, průměrem a směrodatnou odchylkou – co ovlivňuje významnost rozdílu jejich průměrů?

N = 100
Průměr = 59,4
SD = 9,4

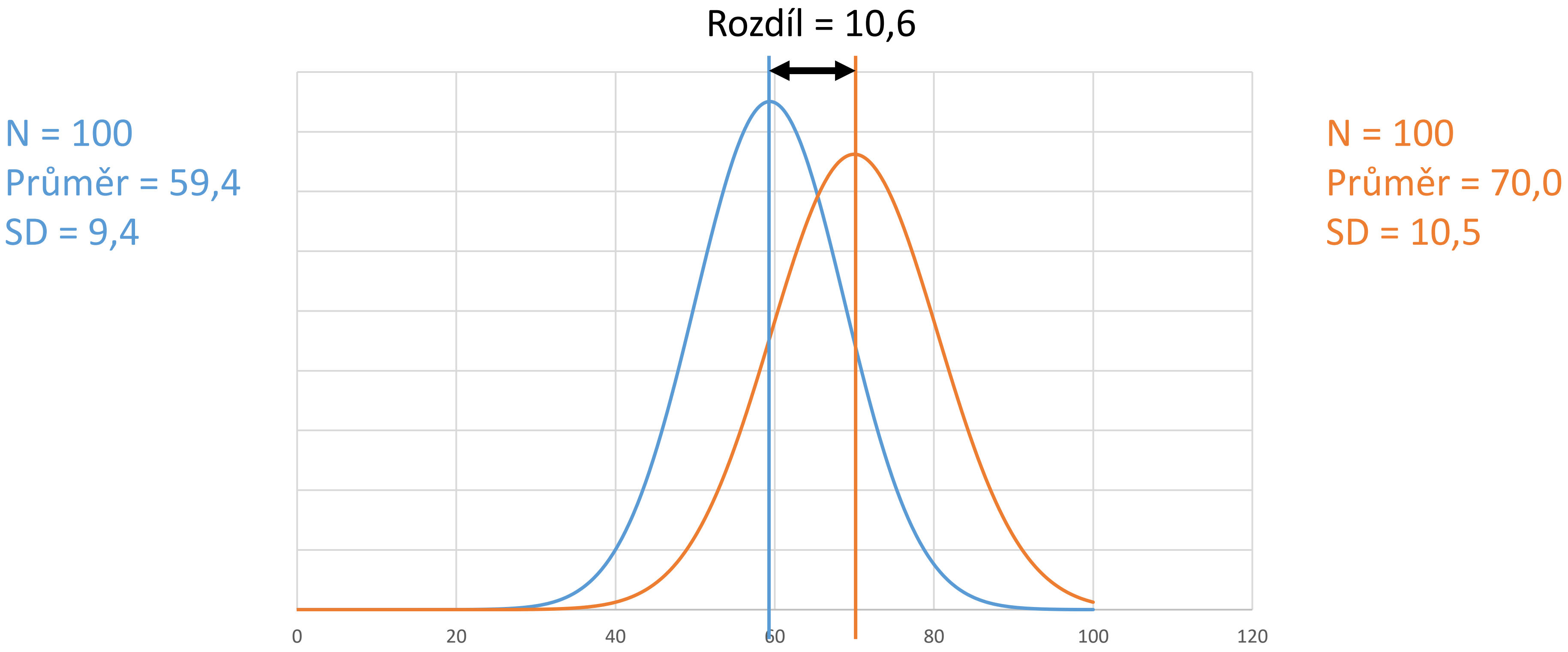


N = 100
Průměr = 70,0
SD = 10,5

- Na velikosti vzorku (větší vzorek = větší významnost) a směrodatné odchylce (větší variabilita = menší významnost) - ovlivňují spolehlivost s jakou odhadujeme srovnávané průměry
- Na velikosti rozdílu mezi srovnávanými průměry (větší rozdíl = větší významnost)

Testová statistika

- Testová statistika kombinuje velikost rozdílu s dalšími charakteristikami dat (velikost vzorku, variabilita atd.), jde vlastně o rozdíl vážený dalšími charakteristikami
- Hodnota testové statistiky je ve vazbě na významnost rozdílu
- Pro finální rozhodnutí o významnosti rozdílu je nezbytné testovou statistiku porovnat s jejím rozdělením náhodných rozdílů (= jaké by bylo rozdělení této statistiky, kdyby byl rozdíl náhodný)



Dva způsoby získání rozdělení testové statistiky

- Testová statistika představuje rozdělení náhodných rozdílů, lze ji získat dvěma způsoby

- Aproximací na modelové rozdělení
  - „standardní“ postup, výhodou je snadný výpočet, citlivé na nedodržení předpokladů o rozložení dat
  - Různé testy mají své rozdělení náhodných rozdílů popsány různými mdolovými rozděleními (např. t-test pomocí t-rozdělení, test dobré shody pomocí Pearsonova (chi-kvadrát rozdělení)

- Permutační metody
  - Rozdělení náhodných rozdílů je získáno pomocí počítačové simulace buď všech možných nebo zadaného počtu náhodných situací
  - Vhodné pro malé velikosti vzorku nebo situace, kdy není možná aproximace na modelová rozdělení
  - Náročné na výpočetní výkon (v současnosti stále menší problém)
  - Výukově názorné

Způsoby testování

- Testování H0 proti HA na hladině významnosti α můžeme provést třemi různými způsoby:

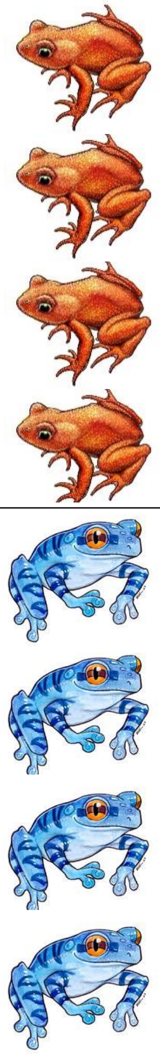
1. Kritický obor (označení W) neboli obor zamítnutí H0 ,
2. Interval spolehlivosti,
3. P-hodnota.

Příklad: permutační testování

Hodnotíme velikost dvou druhů žab, od každého druhu jsme vzorkovali 100 jedinců.

N=100

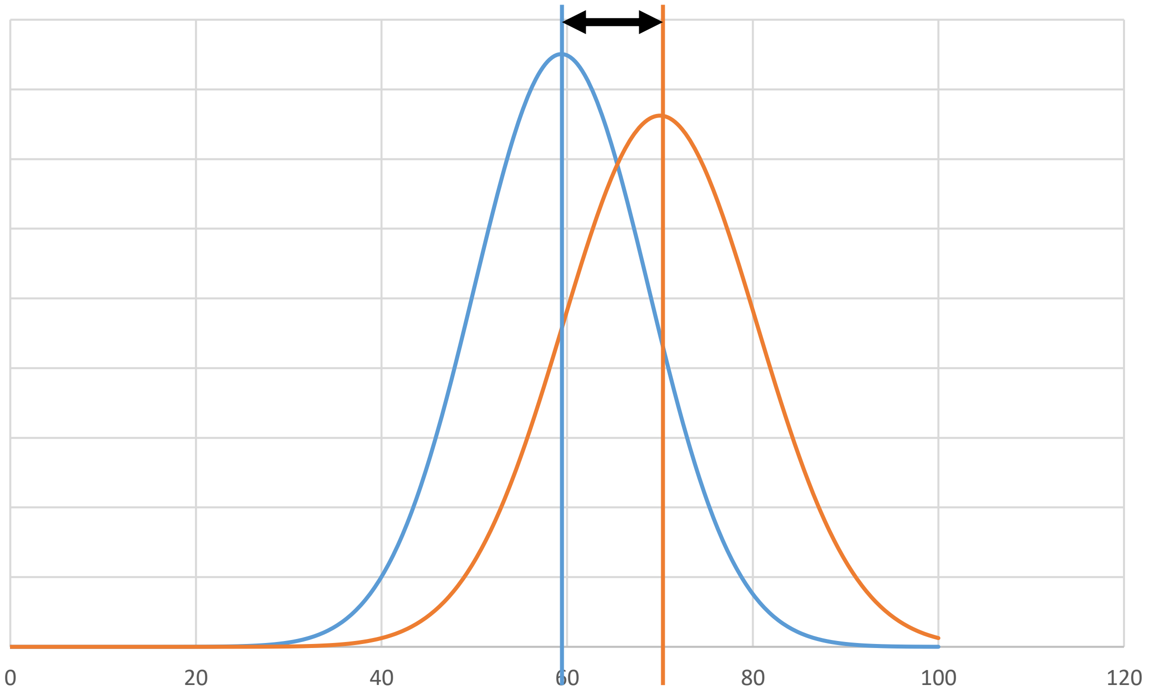
N=100



Rozdíl ???

Rozdíl = 10,6

N = 100
Průměr = 59,4
SD = 9,4



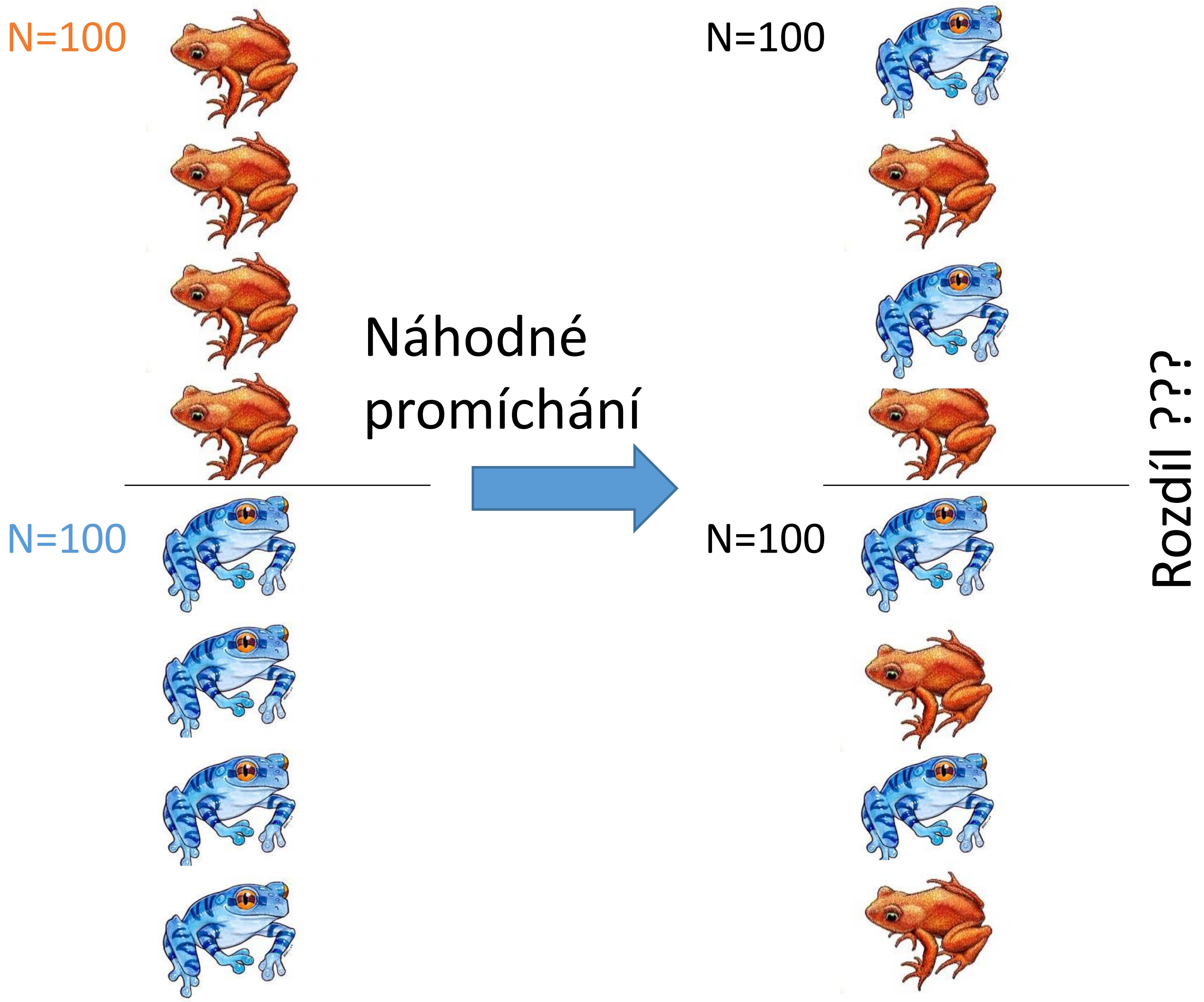
N = 100
Průměr = 70,0
SD = 10,5

Jak zjistit, zda pozorovaný rozdíl je daný pouhou náhodou?

Nasimulujeme si ho !!!!

Příklad: permutační testování

Hodnotíme velikost dvou druhů žab, od každého druhu jsme vzorkovali 100 jedinců.

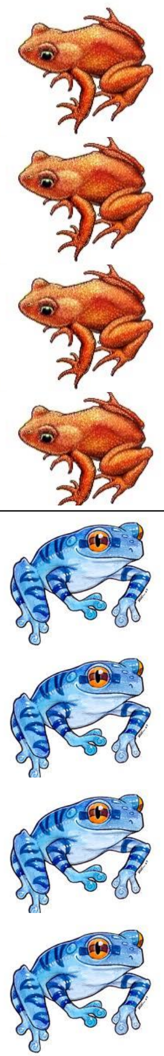


Jaký je nejpravděpodobnější rozdíl mezi skupinami po náhodném promíchání?

Příklad: permutační testování

Hodnotíme velikost dvou druhů žab, od každého druhu jsme vzorkovali 100 jedinců.

N=100



N=100

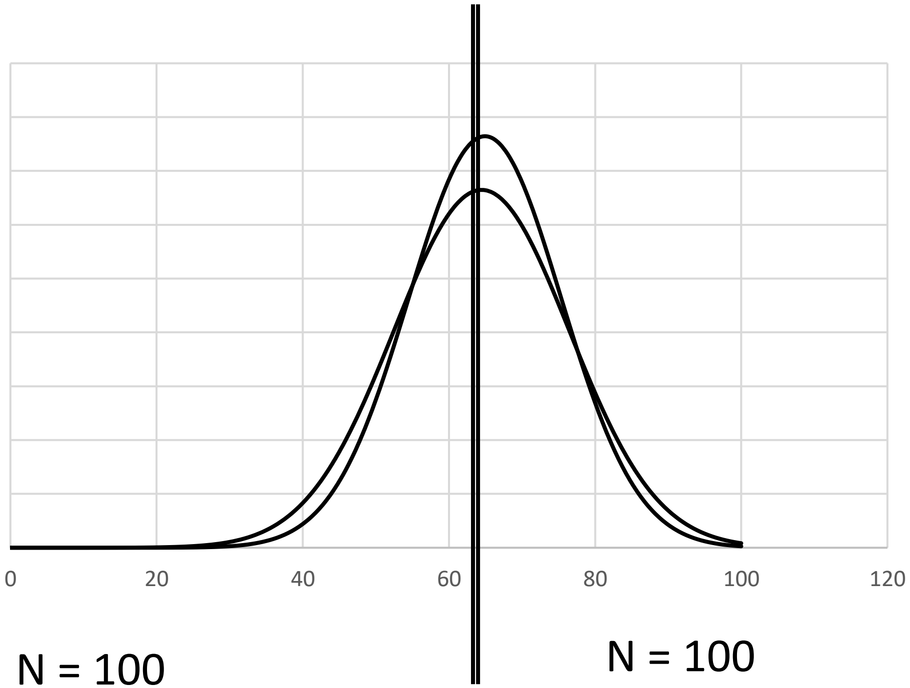
Náhodné promíchání

N=100

N=100

Rozdíl ???

Rozdíl = 0,4



N = 100
Průměr = 64,9
SD = 10,4

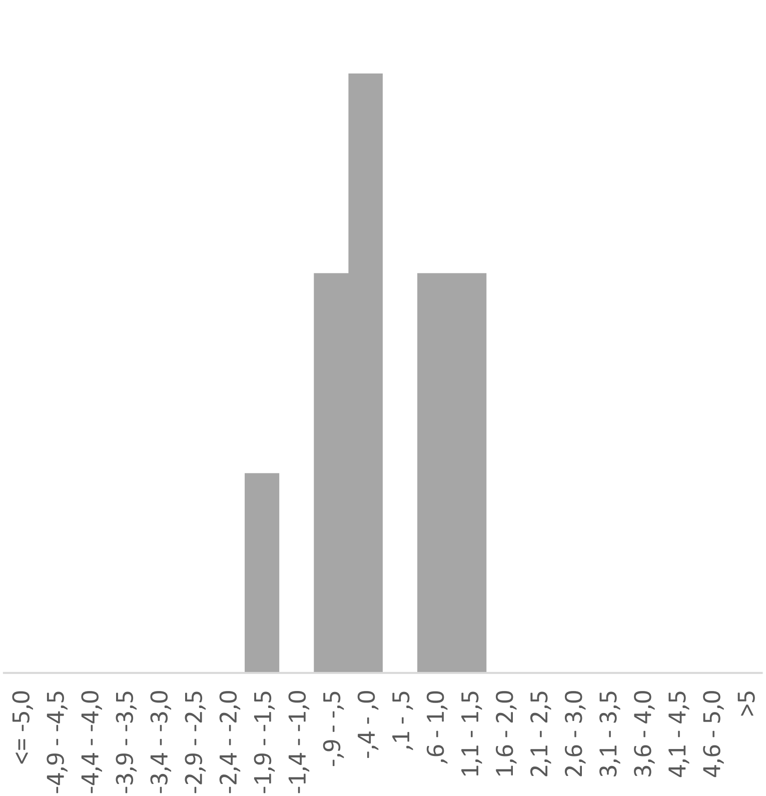
N = 100
Průměr = 64,5
SD = 12,0

Pro stabilizaci výsledku potřebujeme velký počet permutací.

Výsledky při různém počtu permutací

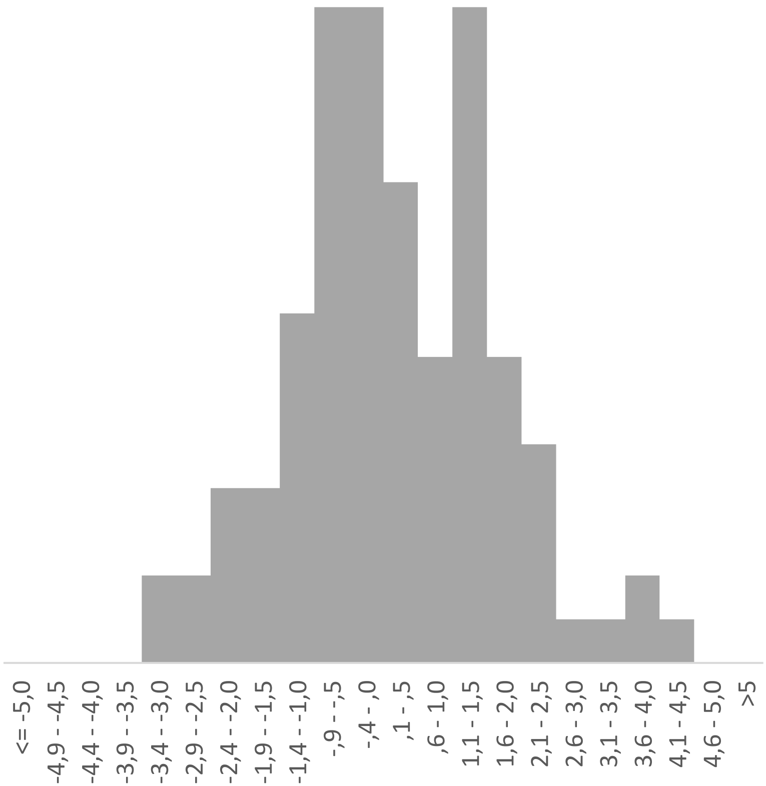
- Se zvyšujícím počtem permutací pozorujeme vytváření rozdělení náhodných rozdílů

N = 10



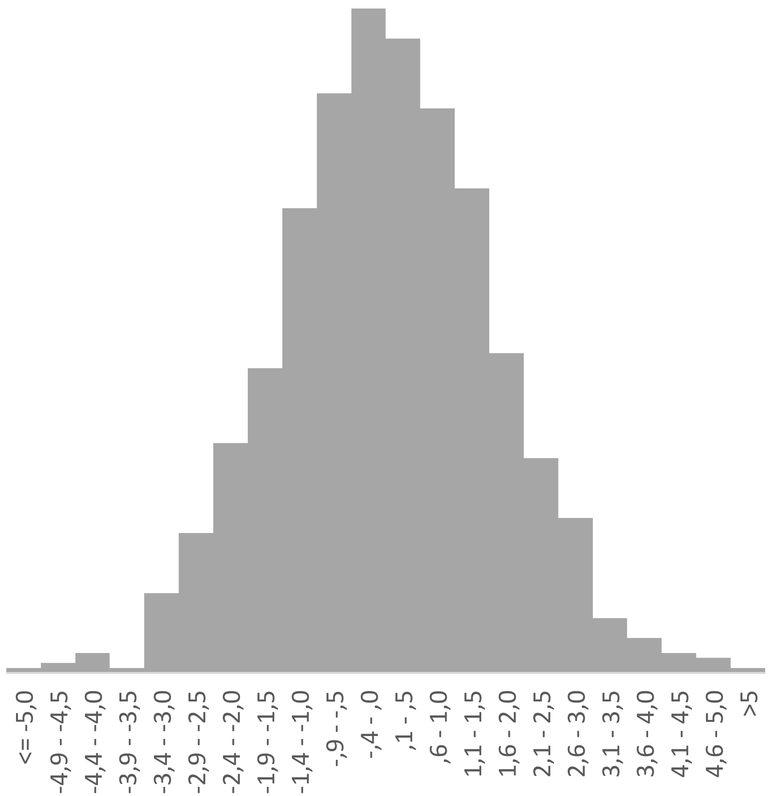
Náhodné rozdíly

N = 100



Náhodné rozdíly

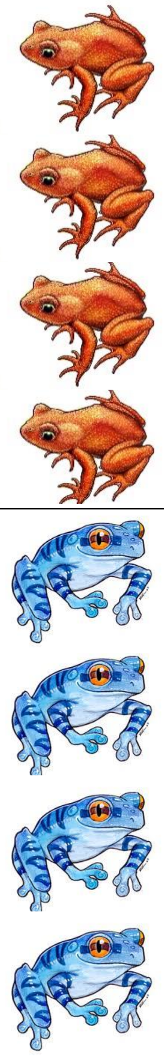
N = 1000



Náhodné rozdíly

Náhodné rozdíly vs. pozorovaný rozdíl

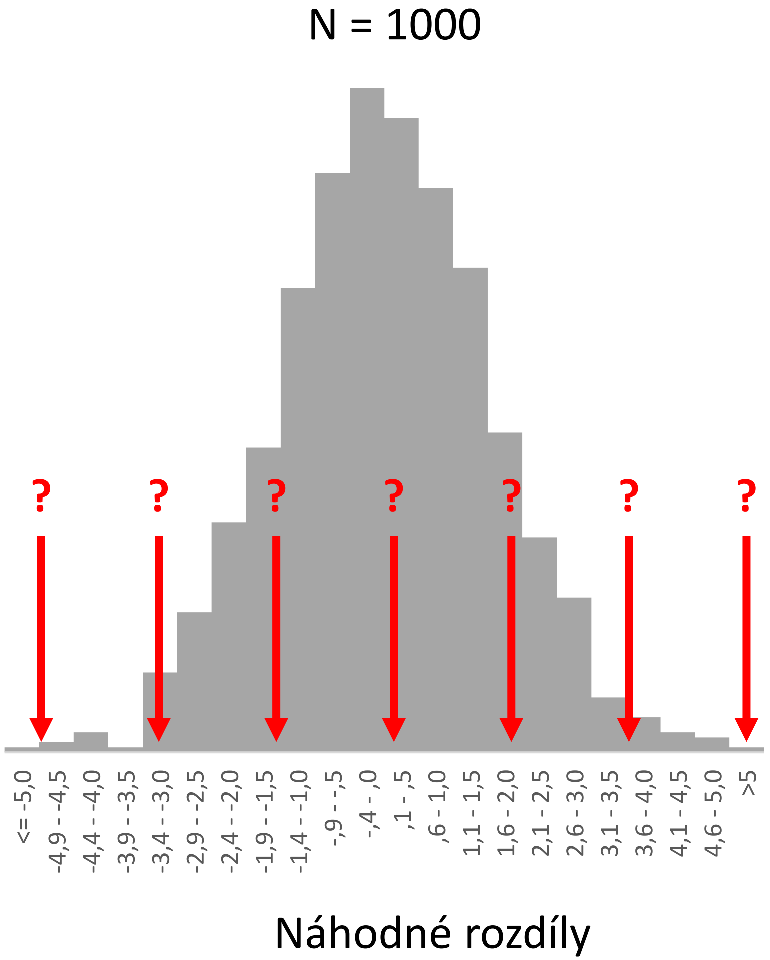
N=100



N=100

Rozdíl = 10,6

- Reálný rozdíl porovnáme s rozložením náhodných rozdílů

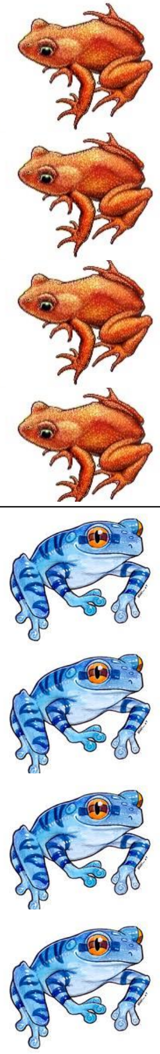


Rozložení náhodných rozdílů a jeho využití pro testování

- Stanovíme si kritický obor testové statistiky = s jakou pravděpodobností náhodného vzniku pozorovaného rozdílu jsme schopni se smířit při zamítnutí nulové hypotézy (tedy prohlášení, že rozdíl nepovažujeme za náhodný)
- Nejběžněji se používá kritický obor testové statistiky vedoucí k pravděpodobnosti náhodného rozdílu 0.05 nebo 0.01 (tzv. hladina statistické významnosti, nejde o přírodní zákon, pouze o domluvu)
- Náš skutečný rozdíl porovnáme s rozložením náhodných rozdílů a stanoveným kritickým oborem této statistiky
- Pokud skutečný rozdíl leží v kritickém oboru, říkáme, že na dané hladině významnosti zamítáme nulovou hypotézu
- Pro danou hodnotu testové statistiky jsme schopni určit i přesnou pravděpodobnost s jakou existují náhodné rozdíly větší než je náš pozorovaný rozdíl = pravděpodobnost, že námi pozorovaný rozdíl je pouhá náhoda

Statistická významnost pozorovaného rozdílu

N=100



N=100

Rozdíl = 10,6

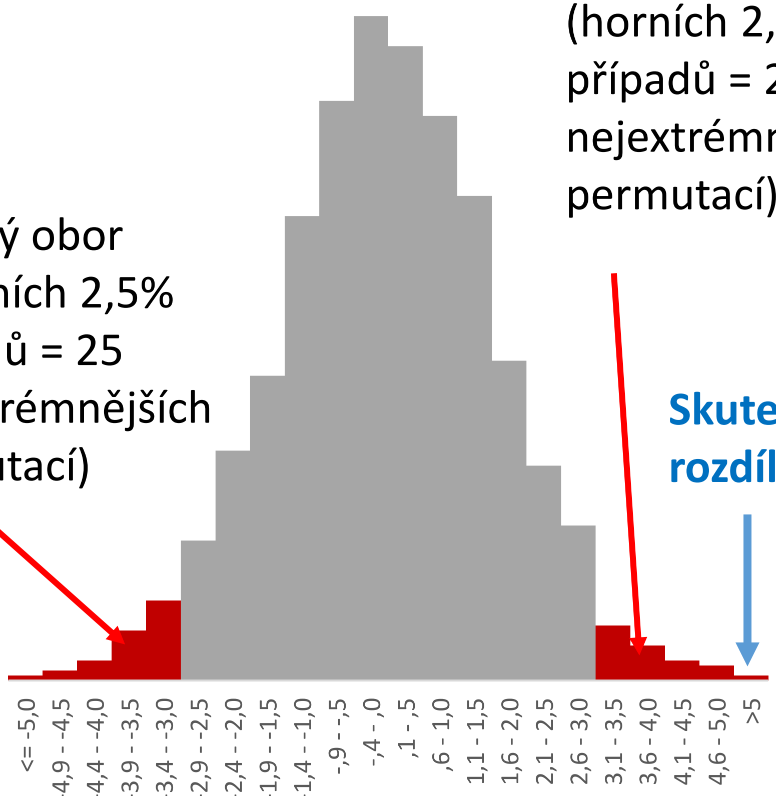
- Jako hladinu statistické významnosti budeme uvažovat 0.05 (5%)

N = 1000

Kritický obor (horních 2,5% případů = 25 nejextrémnějších permutací)

Kritický obor (spodních 2,5% případů = 25 nejextrémnějších permutací)

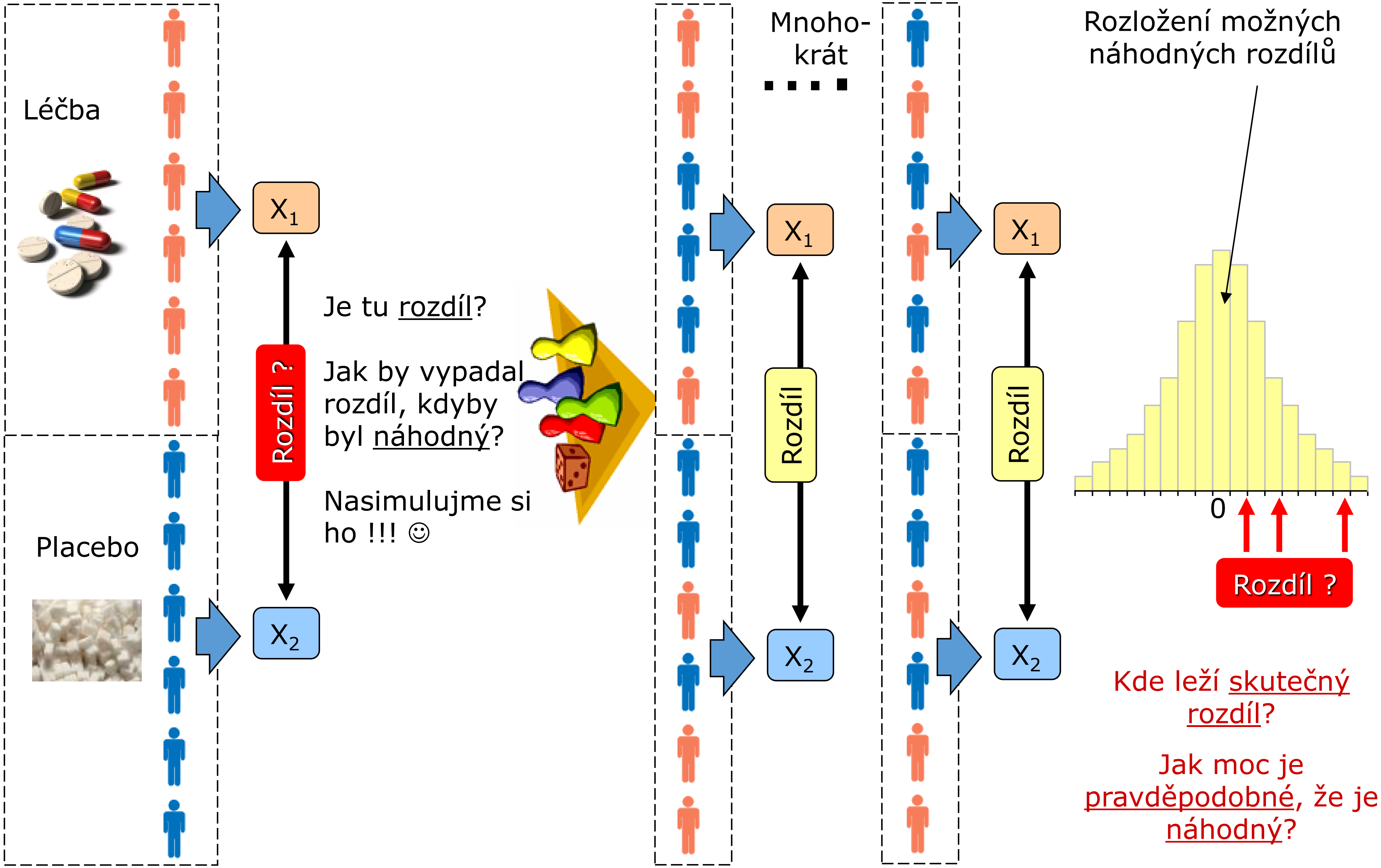
Skutečný rozdíl = 10,6



Náhodné rozdíly

1. Skutečný rozdíl leží v kritickém oboru testové statistiky = **zamítáme nulovou hypotézu o shodě průměru obou skupin**
2. Existuje pouze jeden náhodný rozdíl vzniklý permutacemi větší než je skutečný rozdíl = pravděpodobnost, že pouhou náhodou existuje větší rozdíl než je námi pozorovaný je 1/1000 = 0,001 = **statistická významnost námi pozorovaného rozdílu je p=0,001.**

Co znamená náhodný rozdíl? Shrnutí.



Kde leží skutečný rozdíl?

Jak moc je pravděpodobné, že je náhodný?

Zamítnutí / nezamítnutí nulové hypotézy

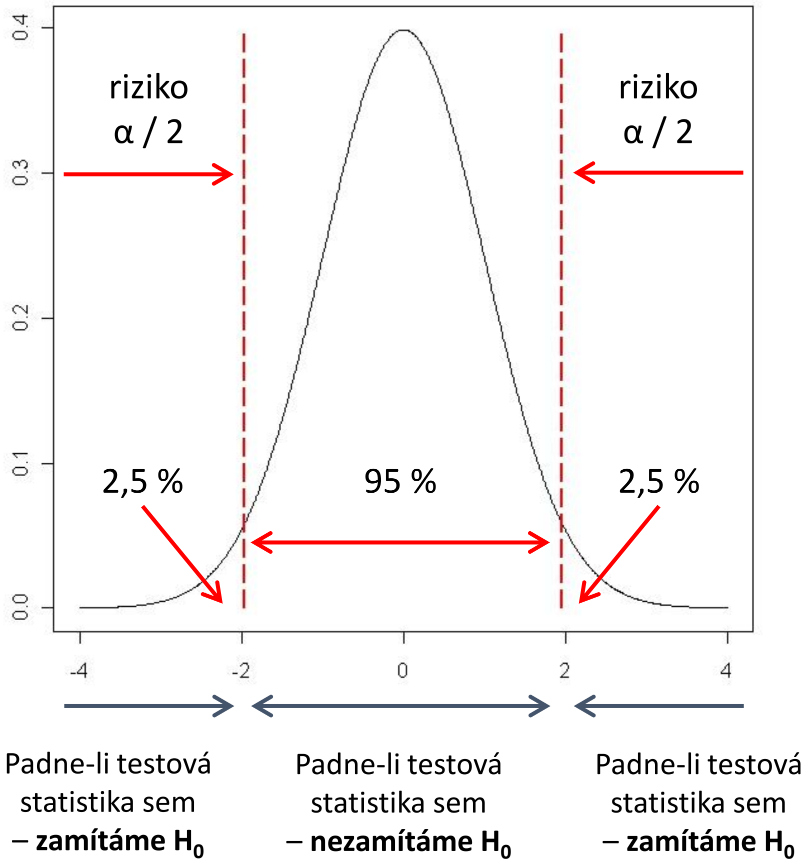
- Hodnotu testové statistiky srovnáme s kvantilem (kritickou hodnotou) jejího rozdělení odpovídajícím zvolené hladině významnosti testu α.

- Představuje-li pozorovaná hodnota testové statistiky extrémnější (méně pravděpodobnou) hodnotu v rámci rozdělení odpovídajícího nulové hypotéze než je kritická hodnota (kvantil) odpovídající zvolenému riziku α, pak nulovou hypotézu zamítáme.

Zamítnutí / nezamítnutí nulové hypotézy

Oboustranný test při α = 0,05

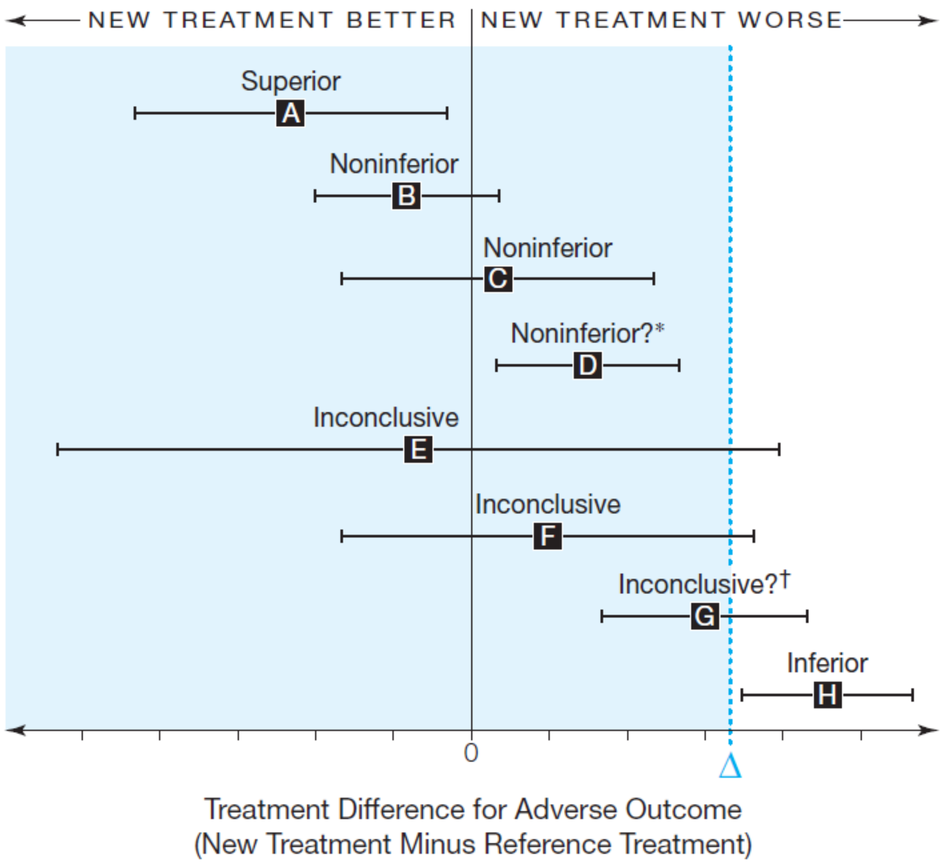
$$H_0 : \theta_1 = \theta_2 \qquad H_1 : \theta_1 \neq \theta_2$$



Rozdělení náhodných rozdílů:
- Buď příslušné modelové rozdělení
- Nebo výsledek simulace

Zamítnutí nulové hypotézy:
- Naše testová statistika spadá do kritického oboru
- Odvozená přesná hodnota p je menší než s kritickým oborem spjaté p

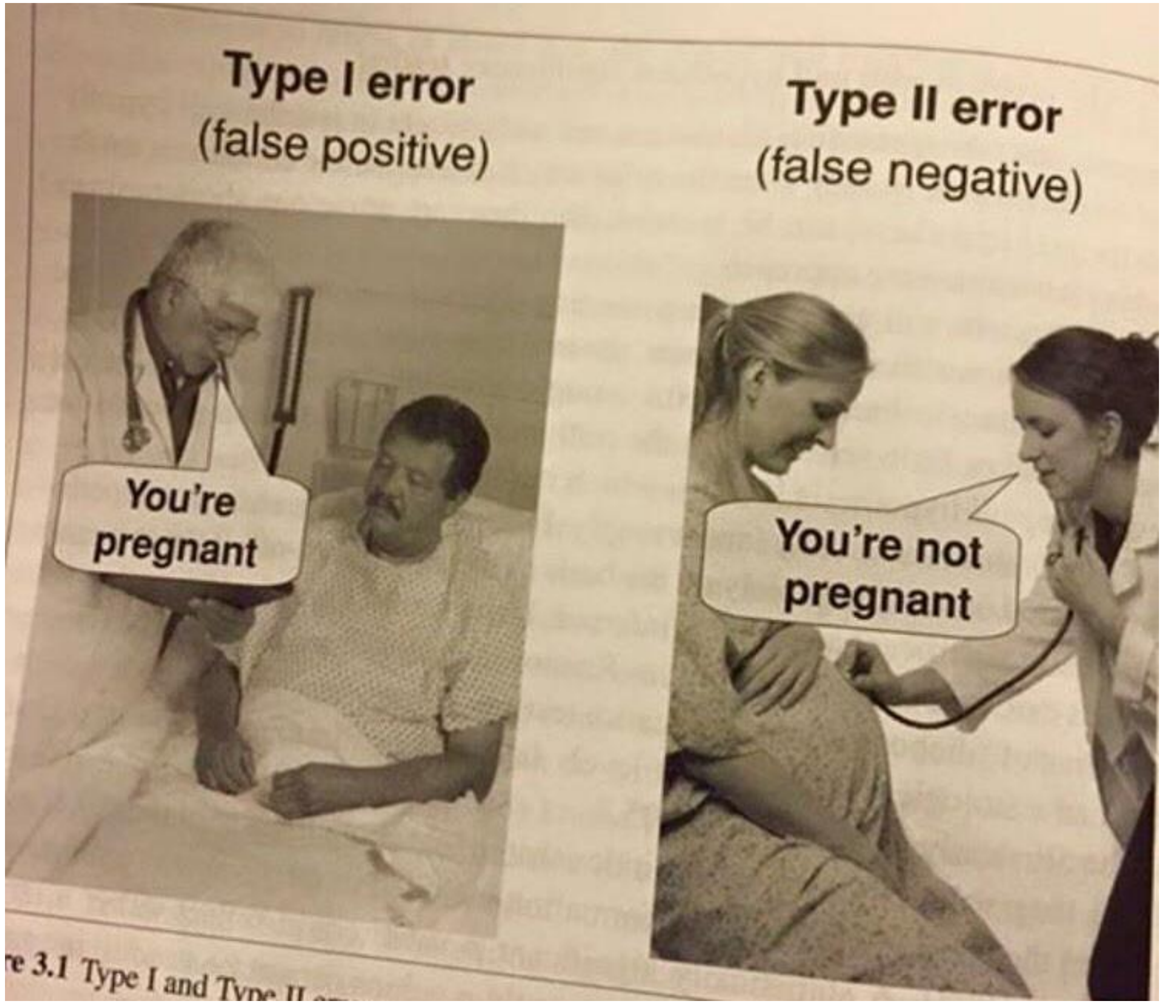
Testování pomocí intervalů spolehlivosti



- Principem testování pomocí intervalů spolehlivosti je výpočet intervalu spolehlivosti pro daný rozdíl nebo míru vztahu proměnných a porovnání s referenční hodnotou (např. 0 v případě rozdílu).
- Pokud interval neobsahuje tuto referenční hodnotu, jde o ekvivalent prokázání statistické významnosti rozdílu na dané hladině významnosti (95% interval spolehlivosti je ekvivalentní hladině významnosti 0.05)

Source: Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA. 2006 Mar 8;295(10):1152-60.
Statistics and Informatics Services Group, Department of Reproductive Health and Research, World Health Organization, Geneva.

Možné chyby při testování hypotéz



Co se při rozhodování může stát

- Vzhledem k nulové hypotéz máme čtyři možnosti výsledku rozhodovacího procesu:

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.

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.**

- 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** $\boldsymbol{P = 1 - \alpha}$	**chyba II. druhu** $\boldsymbol{P = \beta}$
H_0 zamítneme	**chyba I. druhu** $\boldsymbol{P = \alpha}$	**správné rozhodnutí** $\boldsymbol{P = 1 - \beta}$

- 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“) a za této podmínky minimalizujeme β.

Co znamená „padnutí testové statistiky“

- Je-li hodnota testové statistiky větší než kvantil příslušný riziku α, pak mohly nastat dvě situace:

1. **buď H0 platí a my jsme pozorovali málo pravděpodobný jev**
2. **nebo H0 neplatí**

- My pracujeme s rizikem α, tedy málo pravděpodobné jevy jsou součástí našeho rizika, proto v tomto případě volíme možnost 2 a zamítáme H0.

Chyby statistického testu jako důsledek našeho rozhodnutí

- Samotná statistická významnost znamená pouze pravděpodobnost toho, že námi pozorovaný rozdíl nebo vztah proměnných je daný pouhou náhodou
- V okamžiku, kdy na základě této pravděpodobnosti provedeme rozhodnutí o neplatnosti nulové hypotézy, smiřujeme se s pravděpodobností (odpovídající dané statistické významnosti), že toto rozhodnutí je chybné a ve skutečnosti nulová hypotéza platí (rozdíl je daný pouhou náhodou)

- Každé naše rozhodnutí o zamítnutí nulové hypotézy v sobě skrývá hada chyby I. druhu

P-hodnota

- P-hodnota vyjadřuje pravděpodobnost za platnosti H_0, s níž bychom získali stejnou nebo extrémnější hodnotu testové statistiky (samozřejmě vzhledem k jednostrannosti nebo oboustrannosti testu).

- Platí tedy, že čím nižší p-hodnota testu je, tím menší nám tento test indikuje pravděpodobnost, že platí nulová hypotéza. Jinak řečeno, vyjde-li nám při vyhodnocení statistického testu p-hodnota „blízká nule" (standardně jsou opět přijímány dvě hranice: 5 % a 1 %), znamená to, že naše nulová hypotéza má velmi malou oporu v pozorovaných datech a můžeme ji zamítnout.

P-hodnota

- Výslednou p-hodnotu tedy srovnáme se zvolenou hladinou významnosti α s tím, že nulová hypotéza je zamítána ve chvíli, kdy p-hodnota testu klesne pod tuto hladinu.

- Dá se tedy říci, že ve chvíli, kdy riziko falešně pozitivního výsledku v souvislosti se zamítnutím nulové hypotézy klesne pod vybranou hladinu (např. 5 % nebo 1 %), pak ji zamítáme.

- P-hodnotu lze chápat jako číselný indikátor platnosti nebo neplatnosti nulové hypotézy vyjádřený na pravděpodobnostní škále. A jako každý indikátor, může i p-hodnota indikovat špatný výsledek, neboť si stále musíme uvědomovat, že nám hrozí jak chyba I. druhu, tak chyba II. druhu.

Síla testu

- Pravděpodobnost chyby II. druhu značíme β.
- 1 – β se nazývá síla testu a vyjadřuje pravděpodobnost, že zamítneme H_0 ve chvíli, kdy H0 opravdu neplatí.

- Snažíme se sílu testu optimalizovat při zachování hladiny významnosti testu α → princip výpočtu velikosti experimentálního vzorku před provedením studie

- Optimalizovat sílu testu a velikost vzorku předem není triviální, můžeme narazit na spoustu problémů – biologické limity, etické limity, finanční limity.

Faktory ovlivňující sílu testu

- **Velikost vzorku:** čím více pozorování (informace o platnosti nulové hypotézy), tím větší má test sílu. Stejně jako u intervalů spolehlivosti, 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.