9. SEMINÁŘ

INDUKTIVNÍ STATISTIKA

2. TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

- **Statistická hypotéza** = výrok o statistickém souboru, např.:
  - že sledovaná veličina má normální rozdělení,
  - že dva náhodné výběry pocházejí z jednoho základního souboru,
  - že dvě veličiny jsou na sobě nezávislé apod.
- **Platnost statistických hypotéz OVĚŘUJEME** na základě údajů zjištěných ve výběrovém souboru.
- Základem je zobecnění z výběrových charakteristik zkoumaného VS na parametry ZS - jde o **induktivní soud**.
- K ověření (testování) hypotézy se používá tzv. **testů významnosti**, které rozhodují mezi:
  - **nulovou (testovanou) hypotézou H_0**
  - **hypotézou alternativní (opačnou) H_A**

NULOVÁ HYPOTÉZA

Při testování hypotéz začínáme tím, že předpokládáme např. určitou hodnotu parametru základního souboru a potom učiníme závěr týkající se výběrového statistického ukazatele.

Statistická hypotéza – např.:

- Pravděpodobnost výskytu krevní skupiny 0 v české populaci je 38%.
- Dva výběrové průměry pocházejí z jednoho základního souboru.
- Veličiny nejsou lineárně závislé.

Nulová hypotéza H_0 - testovaná

Obvykle je formulována tak, že **předpokládá nulový rozdíl** (rozdíl blízký nule):

- $\pi = 0{,}38$; $\pi = p$; $\pi - p = 0$
- $\mu_1 = \mu_2 = \mu$; $\mu_1 - \mu_2 = 0$
- $\rho = 0$; $\rho = r$; $\rho - r = 0$

ALTERNATIVNÍ HYPOTÉZA

Alternativní hypotéza H_A - opačná

Předpokládá opak, obvykle tedy předpokládá nenulový rozdíl (rozdíl dostatečně vzdálený od nuly). Musí vždy obsahovat všechny zbývající možnosti, které nejsou obsaženy v hypotéze nulové:

- $\pi \neq 0{,}38$; $\pi \neq p$; $\pi - p \neq 0$
- $\mu_1 \neq \mu_2$; $\mu_1 - \mu_2 \neq 0$
- $\rho \neq 0$

STATISTICKÉ HYPOTÉZY

- **Hypotézy o rozdílech (mezi parametrem a výběrovou charakteristikou, mezi dvěma parametry)**
  - **Lze** prokázat (nebo neprokázat) rozdíl.
  - **Nelze** prokázat shodu (že rozdíl neexistuje).

- **Hypotézy o závislosti**
  - **Lze** prokázat (nebo neprokázat) závislost.
  - **Nelze** prokázat nezávislost (že závislost neexistuje).

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4,57$ $s_1 = 0,70$ $SE_1 = 0,10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5,42$ $s_2 = 0,85$ $SE_2 = 0,11$

- **Věcná** (klinická) významnost
- **Statistická** významnost

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Muži 20-30 let: $n_1 = 50$ $m_1 = 4{,}57$ $s_1 = 0{,}70$ $SE_1 = 0{,}10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5{,}42$ $s_2 = 0{,}85$ $SE_2 = 0{,}11$

Statistickou významnost lze odhadnout pomocí intervalů spolehlivosti:

1. Pokud se **intervaly spolehlivosti, které vytvoříme kolem bodových odhadů m_1 a m_2** překrývají, pak rozdíl mezi nimi není statisticky významný. Naopak, pokud se nepřekrývají, je rozdíl statisticky významný.

   $m_1 = 4{,}57$ **95% CI (4,37; 4,77)** $m_2 = 5{,}42$ **95% CI (5,18; 5,66)**

2. Pro řešení úlohy bychom mohli použít i **intervalový odhad rozdílu průměrů** – pokud CI neobsahuje nulu, je rozdíl statisticky významný.

   95% CI (0,56; 1,14)

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4{,}57$ $s_1 = 0{,}70$ $SE_1 = 0{,}10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5{,}42$ $s_2 = 0{,}85$ $SE_2 = 0{,}11$

Statistickou významnost lze objektivně určit testováním statistické hypotézy o rozdílu průměrů $m_1 – m_2$.

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. **Stanovíme nulovou a alternativní hypotézu**
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Nulová hypotéza (testovaná)

- vždy předpokládá, že jde o dva náhodné výběry z jednoho základního souboru (není rozdíl mezi průměrnými hodnotami cholesterolu u mladších a starších mužů).

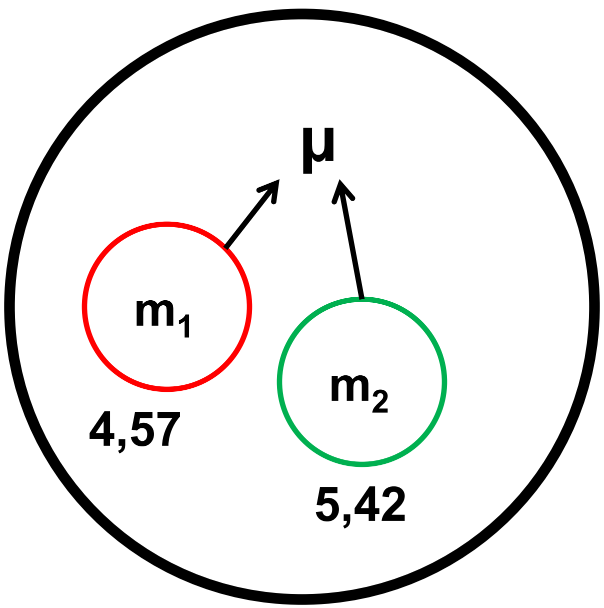
$$H_0: \mu_1 = \mu_2 = \mu$$

$$\mu_1 - \mu_2 = 0$$

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Nulová hypotéza (testovaná)

- vždy předpokládá, že jde o dva náhodné výběry z jednoho základního souboru (není rozdíl mezi průměrnými hodnotami cholesterolu u mladších a starších mužů).



$$H_0: \mu_1 = \mu_2 = \mu$$

$$\mu_1 - \mu_2 = 0$$

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Alternativní hypotéza (opačná)

- předpokládá opak, tj. že jde o dva výběry ze dvou různých základních souborů s rozdílnými průměry (rozdíl mezi průměry je statisticky významný)

$$H_A: \mu_1 \neq \mu_2$$

$$\mu_1 - \mu_2 \neq 0$$

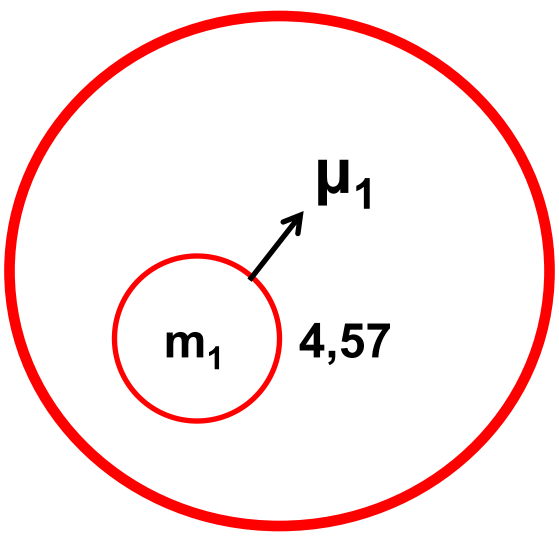
Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

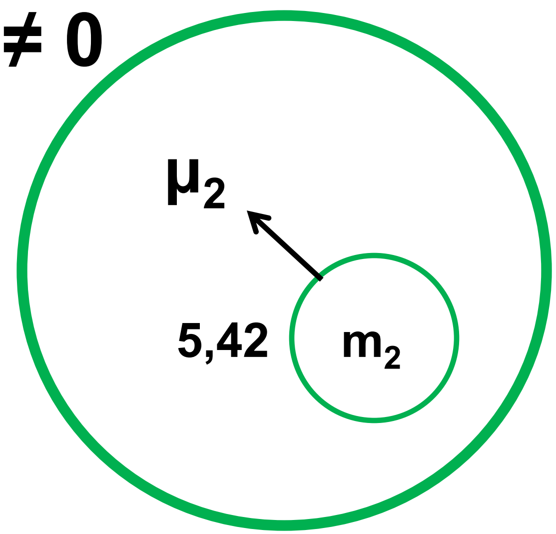
Alternativní hypotéza (opačná)

- předpokládá opak, tj. že jde o dva výběry ze dvou různých základních souborů s rozdílnými průměry (rozdíl mezi průměry je statisticky významný)

$$H_A: \ \mu_1 \neq \mu_2$$

$$\mu_1 - \mu_2 \neq 0$$





Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4{,}57$ $s_1 = 0{,}70$ $SE_1 = 0{,}10$
Muži 40-50 let: $n_2 = 60$ $m_2 = 5{,}42$ $s_2 = 0{,}85$ $SE_2 = 0{,}11$

1. **Stanovení nulové a alternativní hypotézy**

H_0: $\mu_1 = \mu_2 = \mu$; $\mu_1 - \mu_2 = 0$
H_A: $\mu_1 \neq \mu_2$; $\mu_1 - \mu_2 \neq 0$

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. **Zvolíme hladinu významnosti**
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

HLADINA VÝZNAMNOSTI

- Je-li pravděpodobnost nějakého jevu velmi malá, chováme se (většinou) tak, jako by nemohla vůbec nastat.

- Je-li malá pravděpodobnost, že H_0 platí, chováme se tak, jako by neplatila a zamítáme ji.

- Tato malá pravděpodobnost se nazývá **hladina významnosti**, obvykle **α** = 0,05 nebo 0,01. Vyjadřuje **riziko nesprávného zamítnutí H_0**, tzv. chyba 1. druhu

- **β** ozn. chybu 2. druhu, souvisí se silou statistického testu. Nastává, když **H_0 nezamítáme, přestože ve skutečnosti neplatí**.

- **Síla testu** = 1- β: schopnost zamítnout H_0, když neplatí.

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4,57$ $s_1 = 0,70$ $SE_1 = 0,10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5,42$ $s_2 = 0,85$ $SE_2 = 0,11$

2. Hladinu významnosti si zvolíme např. $\alpha = 0,05$.

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. **Vybereme vhodný test**
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

TESTY VÝZNAMNOSTI

- Platnost statistických hypotéz prověřujeme pomocí tzv. **testů významnosti**:
  - **Testy pro hodnoty parametrů**
    (měříme vzdálenost pozorované statistiky od hypotézou stanovené hodnoty parametru)
  - **Srovnávání rozdílů parametrů**
    (např. test významnosti pro rozdíly středních hodnot či pravděpodobností)
  - **Zjišťování typu rozložení četností**
    (test dobré shody, test normality)
  - **Hodnocení závislostí**
    (testy závislosti)

TESTY VÝZNAMNOSTI

Parametrické testy

- Vycházejí ze srovnávání parametrů **μ, σ, π** (zastoupených při srovnávání výběrovými charakteristikami m, p, s).
- Musíme znát typ rozložení testované veličiny, hypotézy se týkají parametrů tohoto rozložení.
- Srovnáváme charakteristiky dvou **nezávislých** výběrů.

TESTY VÝZNAMNOSTI

Neparametrické testy

- Velkou skupinu tvoří např. testy založené na pořadí
- Výhody: jsou početně jednodušší a **nepředpokládají znalost typu rozložení** a lze je použít pro **závislé** výběry
- Nevýhody: mají menší sílu, tzn. mají menší schopnost zamítnout nulovou hypotézu, když ta skutečně neplatí.

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4{,}57$ $s_1 = 0{,}70$ $SE_1 = 0{,}10$
Muži 40-50 let: $n_2 = 60$ $m_2 = 5{,}42$ $s_2 = 0{,}85$ $SE_2 = 0{,}11$

3. Pro srovnání průměrů zvolíme u-test

- Při dostatečně velkých souborech mají rozdíly výběrových průměrů normální rozdělení.
- **u-test (z-test):**
  - parametrický test
  - normální rozložení
  - Vypočítaná testovací charakteristika **u** (někdy ozn. z) se srovnává s kritickými hodnotami normálního rozložení.
- U malých souborů se pro srovnání průměrů používá t-test.

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. **Ověříme, zda jsou splněny podmínky pro použití testu**
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

PODMÍNKY PRO POUŽITÍ TESTU

Podmínky pro použití u-testu pro srovnávání průměrů:

1. **$n_1 > 30$, $n_2 > 30$**
   – pro menší soubory Studentův t-test (vypočítáme testovací charakteristiku **t** a srovnáme ji s kritickými hodnotami Studentova rozdělení – viz skripta str. 41).
2. **nezávislé výběry** (hodnoty ve srovnávaných souborech se vzájemně neovlivňují)
   – testy pro párované hodnoty
3. **stejné rozptyly**
   – neliší se významně (F-test)

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4,57$ $s_1 = 0,70$ $SE_1 = 0,10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5,42$ $s_2 = 0,85$ $SE_2 = 0,11$

4. Ověření podmínek pro použití u-testu:

1. **50 > 30; 60 > 30**
2. **soubory jsou nezávislé**
3. **předpokládáme stejné rozptyly**

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. **Vypočítáme testovací charakteristiku**
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

TESTOVACÍ CHARAKTERISTIKA

- Testy významnosti rozhodují mezi $\mathbf{H_0}$ a $\mathbf{H_A}$, a to nejčastěji pomocí výpočtu tzv. **testovací charakteristiky**
- Vymezuje **obor hodnot pro zamítnutí a** obor hodnot pro **nezamítnutí $\mathbf{H_0}$**.
- Pro stanovení takových oborů hodnot je nezbytné, aby měla některé ze známých teoretických rozdělení – umožní to stanovení tzv. **kritických hodnot.**
- Kritické hodnoty vymezují **interval spolehlivosti**, jenž je mírou vzdálenosti od 0. Leží-li hodnota testovací charakteristiky mimo tento interval, zamítáme H_0.

VZDÁLENOST OD NULY

- Pokud je rozdíl srovnávaných průměrů **„rozumně blízko nule“**, pak můžeme říct, že rozdíl vznikl náhodou a **nezamítáme nulovou hypotézu.**

- Je-li rozdíl **„hodně vzdálen od nuly“**, dáváme přednost alternativní hypotéze, tj. **zamítáme nulovou hypotézu**.

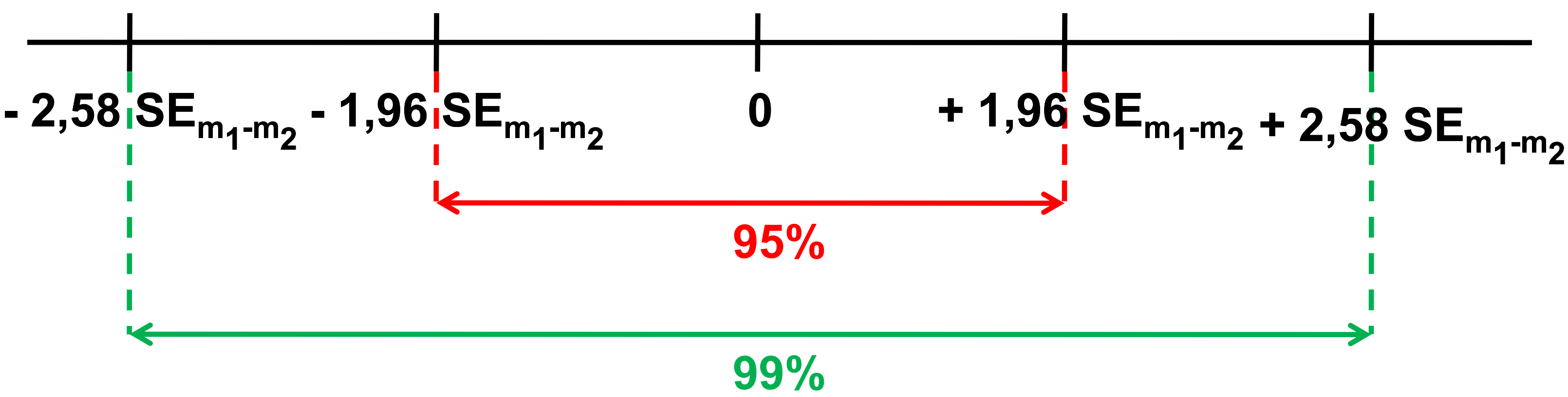
VZDÁLENOST OD NULY

Chyba rozdílu průměrů

- Rozdíly průměrů mají normální rozdělení s parametry **μ** a **σ**; σ odhadujeme pomocí SE

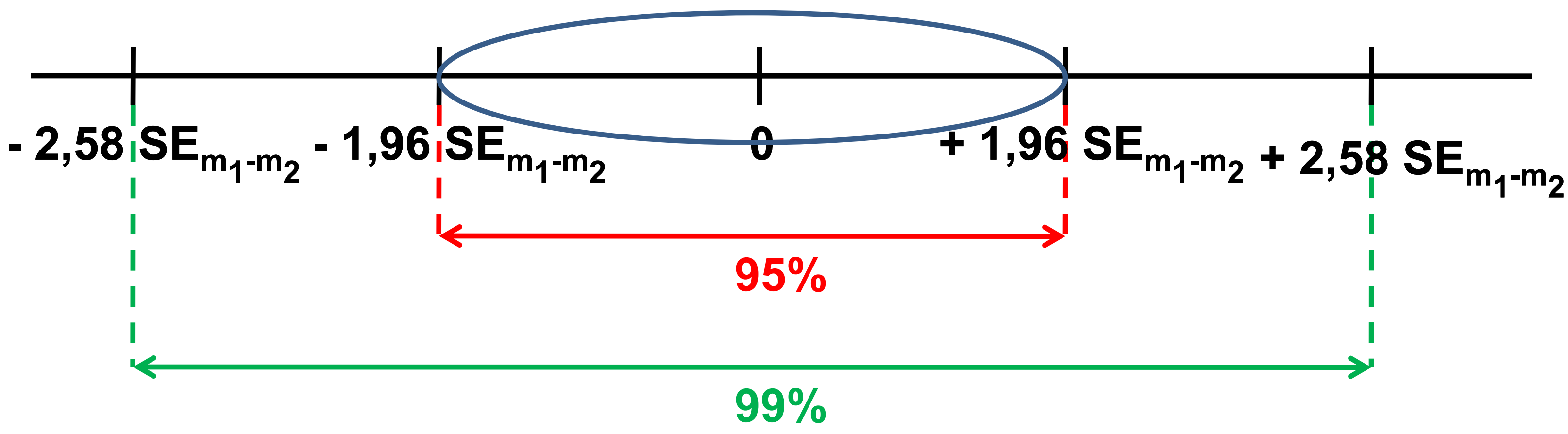
- $\mathbf{SE_{m_1 - m_2}}$ = chyba rozdílu průměrů ($m_1 - m_2$), přičemž pro nezávislé výběry platí:

$$\mathbf{SE^2_{m_1 - m_2} = SE^2_{m_1} + SE^2_{m_2}}$$



VZDÁLENOST OD NULY

- Řeší se pomocí stanovení intervalu, ve kterém se nachází 95% rozdílů průměrů, vypočítaných ve VS, které byly vybrány z jednoho ZS.
- Pokud **H_0** platí **($\mu_1 = \mu_2 = \mu$)**, pak s pravděpodobností 0,95 by se měl rozdíl $m_1 - m_2$ nacházet v 95% intervalu spolehlivosti.



ROZHODNUTÍ

Testovací charakteristika „u“

- Pokud leží rozdíl mimo interval spolehlivosti, pak **zamítáme** nulovou hypotézu.

$$|m_1 - m_2| \geq 1,96 SE_{m_1-m_2} = \frac{|m_1 - m_2|}{SE_{m_1-m_2}} \geq 1,96$$

u

- Pokud leží rozdíl v intervalu spolehlivosti, pak nulovou hypotézu **nezamítáme**.

$$|m_1 - m_2| < 1,96 SE_{m_1-m_2} = \frac{|m_1 - m_2|}{SE_{m_1-m_2}} < 1,96$$

Nezamítnutí nulové hypotézy neznamená její přijetí!!!

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4{,}57$ $s_1 = 0{,}70$ $SE_1 = 0{,}10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5{,}42$ $s_2 = 0{,}85$ $SE_2 = 0{,}11$

5. Výpočet testovací charakteristiky u:

$m_1 – m_2 = 4{,}57 – 5{,}42 = -0{,}88$

$SE_{m_1-m_2}{}^2 = 0{,}10^2 + 0{,}11^2 = 0{,}0221$

$SE_{m_1-m_2} = 0{,}15$

$u = 0{,}88 : 0{,}15 = 5{,}66$

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. **Srovnáme ji s odpovídajícími kritickými hodnotami**
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. Výsledky interpretujeme

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4,57$ $s_1 = 0,70$ $SE_1 = 0,10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5,42$ $s_2 = 0,85$ $SE_2 = 0,11$

6. Srovnání testovací charakteristiky s kritickou hodnotou:

- **|5,66| >1,96**
- Testovací charakteristika je větší než kritická hodnota pro α = 0,05, tzn. leží mimo 95% CI.
- **Malé soubory – t-test: kritické hodnoty Studentova t-rozdělení (skripta str. 41).** Stupně volnosti $f = (n_1 + n_2 - 2)$.

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. **Zamítneme nebo nezamítneme nulovou hypotézu**
8. Výsledky interpretujeme

ZAMÍTNUTÍ A NEZAMÍNUTÍ H_0

- **Nezamítnutí H_0**– rozdíly nepřesahují velikost rozdílů způsobených náhodou, **ale** mohla nastat tzv. chyba druhého typu.

- **Zamítnutí H_0**– pravděpodobnost, že rozdíl mezi průměry je způsoben pouze náhodou je tak malá, že tuto možnost zamítáme – a přijímáme alternativní hypotézu (**ale** s rizikem chyby prvního typu).

ZAMÍTNUTÍ A NEZAMÍNUTÍ H_0

	SKUTEČNOST	
ZÁVĚR TESTU	H_0 neplatí	H_0 platí
Zamítáme H_0	Správné rozhodnutí	Chyba I. typu
Nezamítáme H_0	Chyba II. typu	Správné rozhodnutí

ZAMÍTNUTÍ A NEZAMÍNUTÍ H_0

- **Nulovou hypotézu nelze potvrdit (přijmout).**
  - Lze ji jen zamítnout
  - Uvádí se pravděpodobnost jejího chybného zamítnutí za předpokladu, že $\mathbf{H_0}$ platí.
  - Tato pravděpodobnost = hladina významnosti.
- **Nezamítnutí H_0 neznamená její přijetí.**
  - Je to rozhodnutí dvojznačné - $\mathbf{H_0}$ může platit, ale nemusí.
  - Dostupná data nestačila k jejímu zamítnutí.
- **Když na zvolené hladině významnosti zamítáme H_0, pak se stejnou pravděpodobností přijímáme H_A (např. rozdíl je).**

ZAMÍTNUTÍ A NEZAMÍNUTÍ H_0

P-value

- udává pravděpodobnost, že hodnocený rozdíl je způsoben náhodou
- pokud je menší než zvolená hladina významnosti, nulovou hypotézu zamítáme, pokud je větší nulovou hypotézu nezamítáme
- Např.: α = 5% (pravděpodobnost platnosti H_0)

  p-value = 0,00073, zamítáme H_0

  p-value = 0,07300, nezamítáme H_0

Příklad 1: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4,57$ $s_1 = 0,70$ $SE_1 = 0,10$
Muži 40-50 let: $n_2 = 60$ $m_2 = 5,42$ $s_2 = 0,85$ $SE_2 = 0,11$

7. Zamítnutí nebo nezamítnutí nulové hypotézy:

$|5,66| > 1,96$

Na 5% hladině významnosti **zamítáme nulovou hypotézu** a přijímáme hypotézu alternativní, tj. rozdíl mezi mladšími a staršími muži je statisticky významný.

TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

1. Stanovíme nulovou a alternativní hypotézu
2. Zvolíme hladinu významnosti
3. Vybereme vhodný test
4. Ověříme, zda jsou splněny podmínky pro použití testu
5. Vypočítáme testovací charakteristiku
6. Srovnáme ji s odpovídajícími kritickými hodnotami
7. Zamítneme nebo nezamítneme nulovou hypotézu
8. **Výsledky interpretujeme**

Příklad: SROVNÁVÁNÍ PRŮMĚRŮ

Jsou rozdíly v průměrné hladině cholesterolu v různých věkových skupinách tak velké, že je pro její hodnocení vhodné používat různé normy?

Muži 20-30 let: $n_1 = 50$ $m_1 = 4{,}57$ $s_1 = 0{,}70$ $SE_1 = 0{,}10$

Muži 40-50 let: $n_2 = 60$ $m_2 = 5{,}42$ $s_2 = 0{,}85$ $SE_2 = 0{,}11$

8. Interpretace výsledků:

Na 5% hladině významnosti jsme prokázali, že existuje statisticky významný rozdíl v průměrných hodnotách cholesterolu u dvou srovnávaných věkových skupin.

Tzn., že při zjištěné variabilitě znaku může být tak velký rozdíl jen zřídka způsoben pouze náhodou. Můžeme tak předpokládat vedle náhody i vliv jiných faktorů (např. věku).

SHRNUTÍ PŘÍKLADU

1. H_0: $\mu_1 = \mu_2 = \mu$; $\mu_1 - \mu_2 = 0$
   H_A: $\mu_1 \neq \mu_2$; $\mu_1 - \mu_2 \neq 0$

2. $\alpha = 0{,}05$

3. u-test

4. $n_1 > 30$; $n_2 > 30$; nezávislé soubory; stejné rozptyly

5. $|u| = 5{,}66$

6. $5{,}66 > 1{,}96$

7. Zamítáme nulovou hypotézu a přijímáme H_A.

8. Rozdíl mezi průměrnými hodnotami cholesterolu je statisticky významný na pětiprocentní hladině významnosti.

Příklad 2:
SROVNÁNÍ PRAVDĚPODOBNOSTÍ

Byl sledován výskyt alergií u studentů LF.

Muži: $n_1 = 105$ $k_1 = 21$ $p_1 = 0{,}20$ (20%)

Ženy: $n_2 = 195$ $k_2 = 19$ $p_2 = 0{,}097$ (9,7%)

Otázka: Je rozdíl ve výskytu alergie u mužů a u žen způsoben náhodou, anebo lze odvodit, že alergie postihují muže častěji?

Příklad 2: SROVNÁNÍ PRAVDĚPODOBNOSTÍ

- **Podmínka pro použití u-testu:**

  $n_1 \cdot p_1 \cdot (1 - p_1) > 9$

  $n_2 \cdot p_2 \cdot (1 - p_2) > 9$

- **Standardní chyba rozdílu pravděpodobností**

$$SE^2 = SE_1^2 + SE_2^2 = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}$$

- **Testovací charakteristika u;** $u = \dfrac{p_1 - p_2}{SE}$