Přednáška VIII.
Testování hypotéz o kvantitativních proměnných

- Úvodní poznámky
- Testy o parametrech 1 rozdělení
- Testy o parametrech 2 rozdělení
- Permutační testy

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Opakování – hypotézy

- Co jsou to hypotézy a jak je stanovujeme?

- **Nulová hypotéza**
- **Alternativní hypotéza**

Opakování – co se při rozhodování může stát

- Popište možné výsledky testování hypotéz a uveďte, jak označujeme jejich pravděpodobnosti.

Rozhodnutí	Skutečnost	
	H_0 platí	**H_0 neplatí**
H_0 nezamítneme	A	B
H_0 zamítneme	C	D

Opakování – z-test pro jeden výběr

- Při populačním epidemiologickém průzkumu se zjistilo, že průměrný objem prostaty u mužů je 32,73 ml (*SD* = 18,12 ml). Na hladině významnosti testu α = 0,05 chceme ověřit, jestli se muži nad 70 let liší od celé populace. Máme náhodný výběr o velikosti *n* = 100 a výběrový průměr 36,60 ml.
- Chceme ověřit platnost $H_0: \mu = 32{,}73$ proti $H_1: \mu \neq 32{,}73$
- Platí-li H_0, pak $\bar{X} \sim N(\mu = 32{,}73, \sigma/\sqrt{n} = 1{,}812)$ (předpokládáme, že známe σ)
- Z CLV víme, že by mělo platit: $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$
- Pokud tedy výběrový průměr patří do rozdělení $N(\mu = 32{,}73, \sigma/\sqrt{n} = 1{,}812)$ neměla by jeho hodnota být vzhledem k tomuto rozdělení nijak extrémní.

1. Úvodní poznámky

Spojité × diskrétní náhodné veličiny

- Budeme se zabývat hodnocením spojitých náhodných veličin (mohou nabývat jakýchkoliv hodnot v určitém rozmezí).
- **Příklady**: výška, váha, vzdálenost, čas, teplota.

- Uvedené testy lze ale použít i pro hodnocení diskrétních náhodných veličin – ale **musí to být odůvodnitelné** (např. velký počet možných hodnot).
- **Příklady**: počet krevních buněk, počet hospitalizací, počet krvácivých epizod za rok.

Parametrické a neparametrické testy

- **Parametrické testy** – zabývají se testováním tvrzení o neznámých parametrech rozdělení pravděpodobnosti, kterým se řídí uvažovaná náhodná veličina . Vyžadují různé předpoklady, minimálně specifikaci rozdělení.

- **Neparametrické testy** – tyto procedury jsou nezávislé (nebo téměř nezávislé) na konkrétním rozdělení pravděpodobnosti náhodné veličiny. Vyžadují méně předpokladů – např. symetrii rozdělení. Na druhou stranu mají menší sílu („no free lunch“).

- Testování v případě chybně určeného rozdělení pravděpodobnosti testové statistiky může vést k mylným závěrům z důvodu nerelevantní *p*-hodnoty, respektive *p*-hodnoty stanovené chybnou úvahou.

Postup při statistickém testování

1. Formulujeme **nulovou hypotézu** H_0.
2. Formulujeme **alternativní hypotézu** H_1. Alternativní hypotéza u parametrických testů může být oboustranná nebo jednostranná.
3. **Zvolíme testovou statistiku** jako kritérium pro rozhodnutí o nulové hypotéze (statistiku volíme tak, abychom byli schopni odvodit rozdělení pravděpodobnosti této statistiky při platnosti nulové hypotézy).
4. Hodnotu testové statistiky **vypočítáme na základě pozorovaných hodnot**: $x_1, x_2, \dots, x_n$.
5. Na základě rozdělení testové statistiky **určíme kritický obor** (obor hodnot, kdy zamítáme H_0).
6. Zjistíme, zda **hodnota testové statistiky leží v oboru kritických hodnot**: pokud ano, zamítáme nulovou hypotézu, pokud ne, nezamítáme nulovou hypotézu. Alternativně můžeme zjistit *p*-hodnotu výsledku.

2. Testy o parametrech 1 rozdělení

O co jde?

- **Chceme srovnat sledovanou charakteristiku náhodné veličiny s předem danou hodnotou (konstantou, předpokladem).**

- Test o průměru při známém rozptylu – *z*-test
- Test o průměru při neznámém rozptylu – *t*-test
- Neparametrický test pro 1 výběr – Wilcoxonův test
- Test o rozdílu párových (závislých) pozorování – párový *t*-test
- Test o rozptylu normálního rozdělení

- **Spolu s výsledkem testu by měly být reportovány i intervaly spolehlivosti pro sledovanou charakteristiku (průměr/rozptyl).**

Test o průměru při známém rozptylu – *z*-test

- Předpokládáme realizaci náhodného výběru o rozsahu *n*: $x_1, x_2, \dots, x_n$.
- **Předpokládáme normalitu dat:** $X_i \sim N(\mu, \sigma^2)$ **- velmi silný předpoklad** (silnější než CLV, neřeší totiž *n* jdoucí do nekonečna).

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0 \qquad H_1 : \mu > \mu_0 \qquad H_1 : \mu < \mu_0$$

- Testujeme, zda data náhodného výběru pochazí z rozdělení se stejnou střední hodnotou jako je předpokládaná hodnota μ_0 (konstanta).
- **Předpokládáme, že známe parametr *σ*.**

- Víme, že za platnosti H_0 platí: $\bar{X} \sim N(\mu_0, \sigma^2/n)$
- Testová statistika: $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

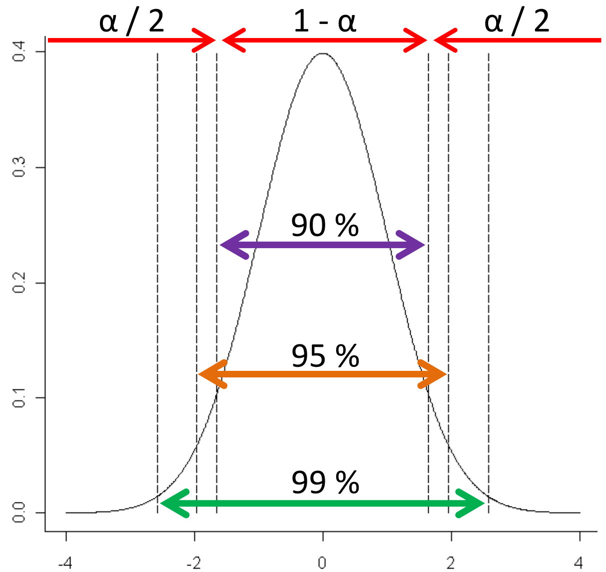
Test o průměru při známém rozptylu – *z*-test

- Nulovou hypotézu zamítáme na hladině významnosti α, když výsledná hodnota *Z* statistiky je větší (nebo menší) než kritická hodnota (příslušný kvantil) rozdělení *N*(0,1).
- „Větší nebo menší“ závisí na předem zvolené alternativě.

- Alternativa $H_1 : \mu \neq \mu_0$
- Zamítáme H_0 když $|Z| > z_{1-\alpha/2}$

- Alternativa $H_1 : \mu > \mu_0$
- Zamítáme H_0 když $Z > z_{1-\alpha}$

- Alternativa $H_1 : \mu < \mu_0$
- Zamítáme H_0 když $Z < z_{\alpha}$



$z_{0,005} = -2,58$ $2,58 = z_{0,995}$

$z_{0,025} = -1,96$ $1,96 = z_{0,975}$

$z_{0,050} = -1,64$ $1,64 = z_{0,950}$

Test o průměru při neznámém rozptylu – *t*-test

- Předpokládáme realizaci náhodného výběru o rozsahu *n*: $x_1, x_2, \dots, x_n$.
- **Předpokládáme normalitu dat:** $X_i \sim N(\mu, \sigma)$ **- velmi silný předpoklad** (silnější než CLV, neřeší totiž *n* jdoucí do nekonečna).

$$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0 \qquad H_1: \mu > \mu_0 \qquad H_1: \mu < \mu_0$$

- Testujeme, zda data náhodného výběru pochazí z rozdělení se stejnou střední hodnotou jako je předpokládaná hodnota μ_0 (konstanta).
- **Neznáme hodnotu parametru σ – musíme ho odhadnout pomocí výběrové směrodatné odchylky (*s*).**
- Víme, že za platnosti H_0 platí: $\bar{X} \sim N(\mu_0, \sigma^2/n)$ $\qquad Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$
- Dále využijeme statistiku *K*: $K = \left(\frac{n-1}{\sigma^2}\right)s^2 \sim \chi^2(n-1)$
- Testová statistika:

$$T = \frac{Z}{\sqrt{K/(n-1)}} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \sim t(n-1)$$

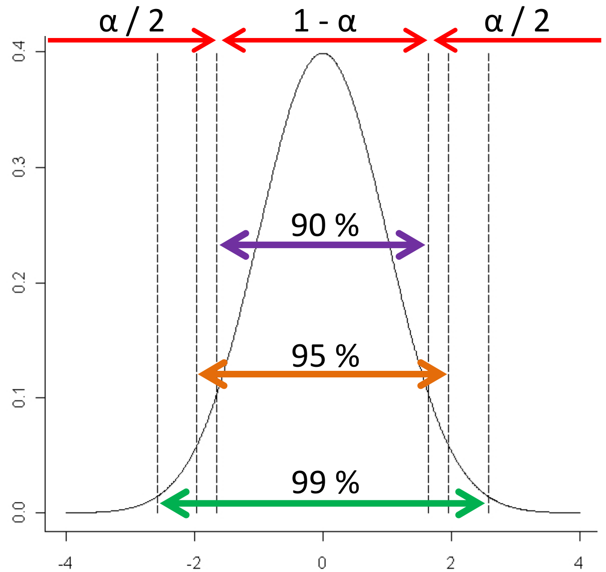
Test o průměru při neznámém rozptylu – *t*-test

- Nulovou hypotézu zamítáme na hladině významnosti α, když výsledná hodnota *T* statistiky je větší (nebo menší) než kritická hodnota (příslušný kvantil) rozdělení *t*(*n* -1).
- „Větší nebo menší“ závisí na předem zvolené alternativě.

- Alternativa $H_1 : \mu \neq \mu_0$
- Zamítáme H_0 když $|T| > t_{1-\alpha/2}^{(n-1)}$

- Alternativa $H_1 : \mu > \mu_0$
- Zamítáme H_0 když $T > t_{1-\alpha}^{(n-1)}$

- Alternativa $H_1 : \mu < \mu_0$
- Zamítáme H_0 když $T < t_{\alpha}^{(n-1)}$



Kantily *t* rozdělení závisí kromě α i na velikosti vzorku (*n*-1).

Příklad – *t*-test pro jeden výběr

- Chceme srovnat průměrný energetický příjem skupiny 11 žen ve věku 22 – 30 let s doporučenou hodnotou (7725 kJ). Průměrný energetický příjem skupiny žen byl 6753,6 kJ se směrodatnou odchylkou *s* = 1142,1 kJ.
- Přibližná normalita dat byla ověřena graficky.
- Nulovou a alternativní hypotézu vyjádříme jako:

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$$

- Testová statistika: $T = (\bar{X} - \mu_0)/(s/\sqrt{n}) \sim t(n-1)$
- Její realizace: $t = (6753{,}6 - 7725)/(1142{,}1/\sqrt{11}) = -2{,}821$
- Absolutní hodnotu *t* srovnáme s kvantilem *t* rozdělení s 10 stupni volnosti.

$$|t| = 2{,}821 > 2{,}228 = t_{0{,}975}^{10} = t_{1-\alpha/2}^{n-1} \Longrightarrow \text{Zamítáme } H_0$$

Příklad – interpretace výsledku

$$|t| = 2{,}821 > 2{,}228 = t_{0{,}975}^{10} = t_{1-\alpha/2}^{n-1} \Longrightarrow$$ Zamítáme H_0

- Na hladině významnosti α = 0,05 můžeme říci, že sledovaná skupina žen měla statisticky významně nižší energetický příjem než je doporučená denní hodnota 7725 kJ.

Neparametrický test pro 1 výběr – Wilcoxonův test

- Předpokládáme realizaci náhodného výběru o rozsahu *n*: $x_1, x_2, \dots, x_n$.
- **Předpokládáme symetrii dat** (daleko slabší předpoklad než normalita dat)

  → nulová hypotéza se týká mediánu

  $H_0 : \tilde{x} = x_0 \qquad H_1 : \tilde{x} \neq x_0$

- Princip Wilcoxonova testu je takový, že spočítáme diference $x_1, x_2, \dots, x_n$ od x_0 a podíváme se, jestli je zhruba ½ diferencí kladných a ½ záporných.
- To je ekvivalentní s tím, že zhruba polovina hodnot $x_1, x_2, \dots, x_n$ je menších než x_0 a polovina hodnot $x_1, x_2, \dots, x_n$ je větších než x_0.

- Spočítáme diference (nulové vyhodíme): $y_i = x_i - x_0$
- Diference seřadíme podle velikosti absolutních hodnot: $|y_{(1)}| < |y_{(2)}| < \dots < |y_{(n)}|$

Neparametrický test pro 1 výběr – Wilcoxonův test

- Spočítáme diference (nulové vyhodíme): $y_i = x_i - x_0$
- Diference seřadíme podle velikosti absolutních hodnot: $|y_{(1)}| < |y_{(2)}| < \ldots < |y_{(n)}|$
- Jako R_i označíme pořadí diference y_i.

- Testovací statistika: $\min(S^+, S^-)$

  kde $$S^+ = \sum_{y_i > 0} R_i \text{ a } S^- = \sum_{y_i < 0} R_i$$

- Pro malá n (cca do 30) lze kritickou hodnotu pro statistiku $\min(S^+,S^-)$ odpovídající zvolenému α najít v tabulkách – je-li výsledná hodnota $\min(S^+,S^-)$ menší nebo rovna kritické hodnotě, zamítáme H_0.
- Pro větší n lze rozdělení testové statistiky $\min(S^+,S^-)$ aproximovat normálním rozdělením s parametry: $$E(\min(S^+, S^-)) = n(n+1)/4$$

  $$D(\min(S^+, S^-)) = n(n+1)(2n+1)/24$$

Příklad – Wilcoxonův test pro jeden výběr

- Chceme srovnat průměrný energetický příjem skupiny 11 žen ve věku 22 – 30 let s doporučenou hodnotou (7725 kJ).
- Nulovou a alternativní hypotézu vyjádříme jako: $H_0 : \tilde{x} = x_0 \quad H_1 : \tilde{x} \neq x_0$

Žena	Denní energetický příjem v kJ	Diference od hodnoty 7725 kJ	Pořadí absolutní hodnoty diference
1	5260	-2465	11
2	5470	-2255	10
3	5640	-2085	9
4	6180	-1545	8
5	6390	-1335	7
6	6515	-1210	6
7	6805	-920	4
8	7515	-210	1,5
9	7515	-210	1,5
10	8230	505	3
11	8770	1045	5

Příklad – Wilcoxonův test pro jeden výběr

- Výpočet testové statistiky: $S^{+} = \sum_{y_i > 0} R_i = 8$ a $S^{-} = \sum_{y_i < 0} R_i = 58$

$$\min(S^{+}, S^{-}) = 8$$

- Kritická hodnota z tabulek pro n = 11: $w_n(\alpha) = w_{11}(0{,}05) = 10$
- Výsledná hodnota statistiky min(S^{+},S^{-}) je menší než 10: ⇒ Zamítáme H_0

Poznámka

- **Parametrické a neparametrické testy nemusí vycházet stejně**. Důvody:

1. Nesplněné předpoklady parametrického testu.
2. Malá síla neparametrického testu.

- Je-li však dobře specifikován pravděpodobnostní model a je-li dostatek dat, bude to vycházet stejně.

- Měli bychom preferovat parametrické testy, ALE pouze po důkladném ověření jejich předpokladů!

Párový *t*-test

- Předpokládáme realizaci dvourozměrného náhodného vektoru o rozsahu *n*:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \ldots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

(máme dvojice hodnot, které patří k sobě)

- Předpokládáme dvourozměrné normální rozdělení: $\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} \right)$
- Nulovou a alternativní hypotézu vyjádříme jako:

$$H_0 : \mu_1 - \mu_2 = d_0 \quad H_1 : \mu_1 - \mu_2 \neq d_0 \quad H_1 : \mu_1 - \mu_2 < d_0 \quad H_1 : \mu_1 - \mu_2 > d_0$$

- Párový problém převedeme na případ jednoho výběru – nebudeme počítat s dvojicemi hodnot, ale s rozdíly: $d_i = x_i - y_i$
- Následně testujeme, zda je průměr hodnot d_1, d_2, … , d_n různý od předpokládané hodnoty d_0.

Párový *t*-test

- Dále postupujeme jako při *t*-testu pro jeden výběr. Testová statistika má tvar:

$$T = \frac{\bar{d} - d_0}{s_d / \sqrt{n}} \sim t(n-1)$$

- Nulovou hypotézu zamítáme na hladině významnosti α, když výsledná hodnota *T* statistiky je větší (nebo menší) než kritická hodnota (příslušný kvantil) rozdělení *t*(*n* -1).

- Alternativa $H_1 : \mu_1 - \mu_2 \neq d_0 \iff H_1 : \mu_d \neq d_0$
- Zamítáme H_0 když $|T| > t_{1-\alpha/2}^{(n-1)}$

- Alternativa $H_1 : \mu_1 - \mu_2 > d_0 \iff H_1 : \mu_d > d_0$
- Zamítáme H_0 když $T > t_{1-\alpha}^{(n-1)}$

- Alternativa $H_1 : \mu_1 - \mu_2 < d_0 \iff H_1 : \mu_d < d_0$
- Zamítáme H_0 když $T < t_{\alpha}^{(n-1)}$

Příklad – párový *t*-test

- Wiebe a Bortolotti (2002) zkoumali žluté zbarvení ocasního peří datlů zlatých.
- Všimli si, že někteří ptáci mají jedno ocasní pero jinak zbarvené než ta ostatní → chtěli vědět, jestli je odchylka ve žlutém zbarvení statisticky významná.
- Měřenou veličinou byl yellowness index („index žlutosti“)

Pták	Index pro typické pero	Index pro atypické pero	Rozdíl (*d*)
A	-0.255	-0.324	0.069
B	-0.213	-0.185	-0.028
C	-0.19	-0.299	0.109
D	-0.185	-0.144	-0.041
E	-0.045	-0.027	-0.018
F	-0.025	-0.039	0.014
G	-0.015	-0.264	0.249
H	0.003	-0.077	0.080
I	0.015	-0.017	0.032
J	0.020	-0.169	0.189
K	0.023	-0.096	0.119
L	0.040	-0.330	0.370
M	0.040	-0.346	0.386
N	0.050	-0.191	0.241
O	0.055	-0.128	0.183
P	0.058	-0.182	0.240

Příklad – párový *t*-test

- Pracovní hypotéza: „Je odchylka ve žlutém zbarvení statisticky významná?“.
- Nulová hypotéza a alternativa: $H_0 : \mu = 0$ $H_1 : \mu > 0$
- Za platnosti H_0 předpokládáme: $\bar{d} \sim N(0, \sigma^2/n)$
- Vypočtené statistiky: $\bar{d} = 0{,}137$ a $s_{\bar{d}} = 0{,}135$
- Testová statistika: $t = \dfrac{\bar{d} - d_0}{s/\sqrt{n}} = \dfrac{0{,}137 - 0}{0{,}135/\sqrt{16}} = 4{,}06$
- Absolutní hodnotu *t* srovnáme s kvantilem *t* rozdělení s 15 stupni volnosti.

$$|t| = 4{,}06 > 1{,}75 = t_{0{,}95}^{15} = t_{1-\alpha}^{n-1} \Rightarrow$$ Zamítáme H_0

3. Testy o parametrech 2 rozdělení

Testy pro dva výběry

- **Chceme srovnat sledovanou charakteristiku náhodné veličiny ve dvou nezávislých skupinách.**

- Test o rozdílu průměru dvou nezávislých výběrů – *t*-test pro dva výběry (při stejných rozptylech)
- Test o shodnosti rozptylů dvou nezávislých výběrů – *F*-test
- Welchova korekce pro *t*-test při nestejných rozptylech
- Neparametrický test pro 2 výběry – Mann-Whitneyho test

- **Spolu s výsledkem testu by měly být reportovány i intervaly spolehlivosti pro pozorované rozdíly v průměrech/mediánech či podíl rozptylů.**

T-test pro dva výběry při stejných rozptylech

- Máme realizaci 1. náhodného výběru o rozsahu n_1: $x_1, x_2, \dots, x_{n_1}$ a na ní nezávislou realizaci 2. náhodného výběru o rozsahu n_2: $y_1, y_2, \dots, y_{n_2}$.
- **Předpokládáme normalitu dat:** $X_i \sim N(\mu_1, \sigma^2)$

  … a stejný rozptyl (i když neznámý) $Y_i \sim N(\mu_2, \sigma^2)$

- Testujeme, zda náhodné výběry pochazí z rozdělení se středními hodnotami, které se liší o předpokládanou hodnotu *c* (konstanta).

$$H_0: \mu_1 - \mu_2 = c \qquad H_1: \mu_1 - \mu_2 \neq c \qquad H_1: \mu_1 - \mu_2 < c \qquad H_1: \mu_1 - \mu_2 > c$$

- Neznáme hodnotu parametru σ^2, ale předpokládáme, že je stejný pro oba výběry – parametr musíme odhadnout pomocí váženého průměru odhadů rozptylu v jednotlivých výběrech:

$$s_*^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

T-test pro dva výběry při stejných rozptylech

- Víme, že za platnosti H_0 platí: $\overline{X} - \overline{Y} \sim N(c, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$
- Testová statistika: $T = \dfrac{\overline{X} - \overline{Y} - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$
- „Větší nebo menší" závisí na předem zvolené alternativě.
- Alternativa $H_1 : \mu_1 - \mu_2 \neq c$
  - Zamítáme H_0 když $|T| > t_{1-\alpha/2}^{(n_1+n_2-2)}$
- Alternativa $H_1 : \mu_1 - \mu_2 > c$
  - Zamítáme H_0 když $T > t_{1-\alpha}^{(n_1+n_2-2)}$
- Alternativa $H_1 : \mu_1 - \mu_2 < c$
  - Zamítáme H_0 když $T < t_{\alpha}^{(n_1+n_2-2)}$

Příklad – *t*-test pro dva výběry

- Máme pacienty se špatně kontrolovanou hypertenzí – sledujeme účinek ACE inhibitoru (ACE-I) a antagonisty pro angiotensin II receptor (AIIA) na snížení diastolického tlaku (TKd) těchto pacientů po 6 měsících od zahájení léčby.

- Nulová a alternativní hypotéza: $H_0 : \mu_1 - \mu_2 = 0 \qquad H_1 : \mu_1 - \mu_2 \neq 0$
- Nulová hypotéza vyjadřuje stejný účinek obou léků na snížení TKd.

- Pacienti léčení ACE-I: $n_1 = 1926 \qquad \bar{x} = 12{,}7 \text{ mmHg} \qquad s_1 = 9{,}96 \text{ mmHg}$
- Pacienti léčení AIIA: $n_2 = 1887 \qquad \bar{y} = 12{,}8 \text{ mmHg} \qquad s_2 = 9{,}79 \text{ mmHg}$

- Vážený odhad parametru σ²: $s_*^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(1926-1)9{,}96^2 + (1887-1)9{,}79^2}{1926+1887-2} = 97{,}54$

$$s_* = 9{,}88$$

Příklad – *t*-test pro dva výběry

- Víme, že za platnosti H_0 platí: $\overline{X} - \overline{Y} \sim N(0, \sigma^2(\frac{1}{1926} + \frac{1}{1887}))$
- Testová statistika: $t = \dfrac{\overline{x} - \overline{y} - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{12{,}7 - 12{,}8 - 0}{9{,}88 \sqrt{\frac{1}{1926} + \frac{1}{1887}}} = -0{,}31$
- Absolutní hodnotu *t* srovnáme s kvantilem *t* rozdělení s 3811 stupni volnosti (zde již klidně můžeme použít kvantil rozdělení *N*(0,1)).

$$|t| = 0{,}31 < 1{,}96 = z_{0{,}975} = z_{1-\alpha/2}$$

Nezamítáme H_0

- Na hladině významnosti α = 0,05 nelze prokázat rozdíl mezi ACE-I a AIIA ve snížení diastolického tlaku u pacientů se špatně kontrolovanou hypertenzí.

Předpoklady *t*-testu pro dva výběry

- **Normalita pozorovaných hodnot obou náhodných výběrů** – velmi silný předpoklad.
- **Nutno otestovat nebo alespoň graficky ověřit** (histogram, box plot).

- **Stejný rozptyl náhodné veličiny v obou srovnávaných skupinách** – také silný předpoklad.
- Opět **nutno otestovat nebo alespoň graficky ověřit** (histogram, box plot).

Ověření předpokladu o stejných rozptylech – *F*-test

- Máme realizaci 1. náhodného výběru o rozsahu n_1: $x_1, x_2, \dots, x_{n_1}$ a na ní nezávislou realizaci 2. náhodného výběru o rozsahu n_2: $y_1, y_2, \dots, y_{n_2}$.
- **Předpokládáme normalitu dat:** $X_i \sim N(\mu_1, \sigma_1^2)$

  (střední hodnoty neznáme) $Y_i \sim N(\mu_2, \sigma_2^2)$

- Testujeme, zda náhodné výběry pochazí z rozdělení se stejným rozptylem.

$H_0 : \sigma_1^2 = \sigma_2^2$ $\quad H_1 : \sigma_1^2 \neq \sigma_2^2$ $\quad H_1 : \sigma_1^2 < \sigma_2^2$ $\quad H_1 : \sigma_1^2 > \sigma_2^2$

- Testová statistika: $F = \dfrac{s_1^2}{s_2^2}$

- Za platnosti H_0 má F statistika Fisherovo rozdělení se stupni volnosti $(n_1 - 1)$ a $(n_2 - 1)$.

Ověření předpokladu o stejných rozptylech – *F*-test

- Víme, že za platnosti H_0 platí: $F \sim F(n_1 - 1, n_2 - 1)$
- Hodnotu *F* statistiky tedy srovnáváme s kvantily $F_{\alpha/2}^{(n_1-1,n_2-1)}$ a $F_{1-\alpha/2}^{(n_1-1,n_2-1)}$

- „Větší nebo menší" závisí na předem zvolené alternativě.
- Alternativa $H_1 : \sigma_1^2 \neq \sigma_2^2$
  - Zamítáme H_0 když $F < F_{\alpha/2}^{(n_1-1,n_2-1)}$ nebo $F > F_{1-\alpha/2}^{(n_1-1,n_2-1)}$
- Alternativa $H_1 : \sigma_1^2 > \sigma_2^2$
  - Zamítáme H_0 když $F > F_{1-\alpha}^{(n_1-1,n_2-1)}$
- Alternativa $H_1 : \sigma_1^2 < \sigma_2^2$
  - Zamítáme H_0 když $F < F_{\alpha}^{(n_1-1,n_2-1)}$

Příklad – *F*-test

- Máme **dvě skupiny dětí s hypotyreózou**: první skupina jsou děti s mírnými symptomy, druhá skupina jsou děti s výraznými symptomy.
- **Chceme srovnat hladinu tyroxinu v séru.**
- Můžeme si dovolit použít *t*-test pro dva výběry?

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 < \sigma_2^2$$

Hladina tyroxinu v séru (nmol/l)	Mírné symptomy (n_1 = 9)	Výrazné symptomy (n_2 = 7)
	34	5
	45	8
	49	18
	55	24
	58	60
	59	84
	60	96
	62	
	86	
Průměr	56,4	42,1
SD	14,22	37,48

Příklad – *F*-test

Hladina tyroxinu v séru (nmol/l)	Mírné symptomy (n_1 = 9)	Výrazné symptomy (n_2 = 7)
Průměr	56,4	42,1
SD	14,22	37,48

- Testová statistika:

$$F = \frac{s_1^2}{s_2^2} = \frac{(14{,}22)^2}{(37{,}48)^2} = 0{,}144$$

- Hodnotu *F* srovnáme s α kvantilem *F* rozdělení s 8 a 6 stupni volnosti.

$$F = 0{,}144 < 0{,}279 = F_{0{,}05}^{(8,6)} = F_{\alpha}^{(n_1-1,n_2-1)} \Longrightarrow$$ Zamítáme H_0

Stejné rozptyly?

- Myslíte si, že jsou stejné rozptyly obou souborů v praxi časté?
- Pokud ne, zkuste vymyslet příklad…

Welchova korekce pro nestejné rozptyly

- Welch (1937) navrhl korekci pro výpočet *T* statistiky se zohledněním nestejných rozptylů.
- Víme, že za platnosti H_0 platí: $\bar{X} - \bar{Y} \sim N(c, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$
- Testová statistika: $T = \dfrac{\bar{X} - \bar{Y} - c}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(\nu)$
- **Počet stupňů volnosti NENÍ roven $n_1 + n_2 - 2$, ale třeba ho stanovit následovně:**

$$\nu = \frac{[(s_1^2 / n_1) + (s_2^2 / n_2)]^2}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$$

- Kritické hodnoty pro zamítnutí H_0 lze odvodit stejně, jako v případě *t*-testu pro dva výběry se stejným rozptylem.

Neparametrický test pro 2 výběry – Mann-Whitneyho test

- Máme realizaci 1. náhodného výběru o rozsahu n_1: $x_1, x_2, \dots, x_{n_1}$ a na ní nezávislou realizaci 2. náhodného výběru o rozsahu n_2: $y_1, y_2, \dots, y_{n_2}$.

$$X_i \sim F(x) \qquad Y_i \sim F(y)$$

- **Předpokládáme stejné rozdělení dat v obou souborech** (slabší předpoklad než normalita dat) → nulová hypotéza se týká distribučních funkcí.

$$H_0 : F(x) = F(y) \qquad H_1 : F(x) \neq F(y)$$

- **Pointa Mann-Whitneyho testu**: pokud x_i a y_j pochází ze stejného rozdělení, pak by pravděpodobnost $P(x_i > y_j)$ měla být zhruba 50 %.
- To je ekvivalentní tomu, že při srovnání všech dvojic x_i a y_j bude v případě cca 50 % dvojic menší x_i a naopak.

Neparametrický test pro 2 výběry – Mann-Whitneyho test

- Pro výpočet nejprve seřadíme všechna pozorování podle velikosti (jako by byly z jednoho vzorku) a přiřadíme jednotlivým hodnotám jejich pořadí.
- Statistikou T_1 označíme součet pořadí v 1. skupině.
- **Testové statistiky**: $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T_1 \qquad U' = n_1 n_2 - U$
- Větší z hodnot U a U' následně srovnáme s kritickou hodnotou z tabulek (v případě oboustranného testu). Je-li kritická hodnota menší, H_0 zamítáme. Pro jednostranný test uvažujeme dle nulové hypotézy pouze buď statistiku U nebo U'.
- Pro vzorky s $n_1 > 10$ a $n_2 > 10$ lze rozdělení statistiky U aproximovat normálním rozdělením s charakteristikami: $E(U) = n_1 n_2 / 2$

$$D(U) = n_1 n_2 (n_1 + n_2 + 1) / 12$$

Příklad – Mann-Whitneyho test

- Máme **dvě skupiny dětí s hypotyreózou**: první skupina jsou děti s mírnými symptomy, druhá skupina jsou děti s výraznými symptomy.
- **Chceme srovnat hladinu tyroxinu v séru** (*t*-test pro dva výběry není vhodný)

$H_0 : F(x) = F(y)$ $\qquad$ $H_1 : F(x) \neq F(y)$

Hladina tyroxinu v séru (nmol/l)	Mírné symptomy (n_1 = 9)	Výrazné symptomy (n_2 = 7)
	34	5
	45	8
	49	18
	55	24
	58	60
	59	84
	60	96
	62	
	86	
Průměr	56,4	42,1
SD	14,22	37,48

Příklad – Mann-Whitneyho test

- Seřadíme všechna pozorování podle velikosti a přiřadíme jednotlivým hodnotám jejich pořadí. Součet pořadí v 1. skupině: T_1 = 84,5.

Skupina n_1 = 9	Skupina n_2 = 7	Pořadí
	5	1
	8	2
	18	3
	24	4
34		5
45		6
49		7
55		8
58		9
59		10
	60	11,5
60		11,5
62		13
	84	14
86		15
	96	16

$$U = 9 * 7 + \frac{9(9+1)}{2} - 84{,}5 = 63 + 45 - 84{,}5 = 23{,}5$$

$$U' = 9 * 7 - 23{,}5 = 39{,}5$$

- max(U,U') = 39,5.
- Srovnáme s kritickou hodnotou z tabulek (*pozor na správné tabulky*):

$$\max(U, U') = 39{,}5 < 51 = U_{0{,}05(2)}^{(9,7)} = U_{\alpha(1/2)}^{(n_1,n_2)}$$

→ Nezamítáme H_0

Příklad – Mann-Whitneyho test

- Zdá se vám ten výsledek správný?
- Pokud ne, čemu to lze přisoudit?

4. Permutační testy

Princip permutačních testů

- Permutační testy jsou neparametrickými testy, ale místo pořadí pracují s pozorovanými hodnotami.

- Principem permutačního testování je srovnání pozorované testové statistiky s testovými statistikami, které by bylo možno teoreticky získat ze stejného datového souboru, když by přiřazení jednotlivých pozorovaných hodnot do sledovaných skupin bylo náhodné.

- Permutační test je tedy založen na výpočtu všech možných hodnot testové statistiky, které lze získat opakovaným přeskupením původního souboru dat tak, že v rámci každého opakování **zůstane zachován jak celkový počet pozorování (celkové n), tak počet pozorování náležících do jednotlivých skupin (např. n_1 a n_2)**.

Výpočet permutačních testů

- Výslednou *p*-hodnotu pak odhadneme jako podíl počtu testových statistik, které byly v absolutní hodnotě větší než původní pozorovaná testová statistika (tedy představují extrémnější výsledky experimentu), k celkovému počtu provedených permutací.

- Tedy **odhad *p*-hodnoty** lze vyjádřit následovně:

$$p = \frac{\# t_i : |t_i| \geq t}{M} = \frac{m}{M}, \quad i = 1, \ldots, M$$

- Permutační testy jsou velmi oblíbené v hodnocení genomických a proteomických dat.

Příklad – permutační test pro dva výběry

- Srovnání hmotnosti dvou skupin pacientů.

$n_1 = 7 \qquad \bar{x}_A = 74{,}5\ \text{kg} \qquad s_A = 9{,}49\ \text{kg}$

$n_2 = 8 \qquad \bar{x}_B = 87{,}1\ \text{kg} \qquad s_B = 6{,}95\ \text{kg}$

$$H_0 : \mu_1 - \mu_2 = c \qquad H_1 : \mu_1 - \mu_2 \neq c$$

- Pro permutační test použijeme *T* statistiku pro dva výběry.
- Zvolíme hladinu významnosti testu: $\alpha = 0{,}05$.
- Pro $n_1 = 7$ a $n_2 = 8$ je možnost provést celkem 6435 jedinečných permutací.

Kategorie pacienta	**Hmotnost pacienta (kg)**
A	91,5
A	79,8
A	66,2
A	70,7
A	63,4
A	77,7
A	71,9
B	83,9
B	92,2
B	85,4
B	99,2
B	77,5
B	80,8
B	91,6
B	86,2

Příklad – permutační test pro dva výběry

Kategorie pacienta	Hmotnost pacienta (kg)	Pořadí permutace 1	2	3	…	6435
A	91,5	A	B	B	…	B
A	79,8	B	B	B	…	B
A	66,2	A	A	A	…	A
A	70,7	A	B	A	…	B
A	63,4	B	B	A	…	A
A	77,7	B	B	B	…	A
A	71,9	B	A	A	…	B
B	83,9	A	B	A	…	A
B	92,2	B	B	A	…	A
B	85,4	A	A	B	…	A
B	99,2	A	A	B	…	A
B	77,5	A	A	A	…	B
B	80,8	B	A	B	…	B
B	91,6	B	B	B	…	B
B	86,2	B	A	B	…	B
Testová statistika	**2,900**	**0,429**	**0,341**	**3,106**	**…**	**0,798**

Příklad – permutační test pro dva výběry

- Srovnání hmotnosti dvou skupin pacientů: A a B.
- Pro výpočet *p*-hodnoty permutačního testu je potřeba následující:
  1. Hodnota původní testové statistiky: ***t* = 2,900**
  2. Celkový počet provedených permutací: ***M* = 6435**
  3. Počet permutací, kdy je absolutní hodnota testové statistiky t_i, $i = 1, \ldots, M$, větší nebo rovna původní testové statistice $t = 2{,}900$. Zde je ***m* = 59**.

- Pak *p*-hodnotu můžeme odhadnout následovně:

$$p = \frac{m}{M} = \frac{59}{6435} = 0{,}009$$

Výsledná *p*-hodnota je menší než zvolená hladina významnosti testu $\alpha = 0{,}05$. → Zamítáme H_0

Permutační test pro dva výběry

- Interpretace výsledné *p*-hodnoty je zde stejná jako pro klasický *t*-test.
- Velkou výhodou permutačního testování je fakt, že jej lze použít pro jakoukoliv testovou statistiku.

- Klíčovým předpokladem je zaměnitelnosti pozorovaných hodnot v obou srovnávaných skupinách – oba soubory by neměly mít výrazně odlišnou variabilitu (proto bychom neměli permutační test použít na příklad s hypotyreózou).

- **Při malém *n* (cca 10 – 20) je poměrně malý také počet dostupných permutací, což může vést k nepřesnému odhadu *p*-hodnoty.**
- Při 1000 permutacích je nejmenší dosažitelná *p*-hodnota 0,001, 100 000 permutací umožňuje dosáhnout *p*-hodnoty až 0,00001.