MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Maryna Sychova

Mnohonásobné porovnávání s kontrolami

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zdeněk Hlávka, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční matematika

Praha 2018

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne ....

Poděkování.

Děkuji doc. RNDr. Zdeňku Hlávkovi, Ph.D. za odborné vedení bakalářské práce, cenné rady a připomínky a především čas, který mi věnoval.

Název práce: Mnohonásobné porovnávání s kontrolami

Autor: Maryna Sychova

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. RNDr. Zdeněk Hlávka, Ph.D.

Abstrakt: Hlavním tématem práce je popis metod mnohonásobného porovnávání, které slouží k porovnání dvojic středních hodnot, resp. medianů.

Na začatku je definované mnohonásobné testování a jsou rozebrané metody, které udržují pravděpodobnost chyby prvního druhu na úrovni α. Podrobněji je odvozená Šidákova metoda a předpoklady potřebné pro její použití. Práce dále zahrnuje stručný popis analýzy rozptylu a přehled některých metod mnohonásobného porovnávání. Navíc je uvedená metoda mnohonásobného porovnávání s kontrolou, jeji modifikace a praktická implementace.

Klíčová slova: Mnohonásobné porovnávání; analýza rozptylu; k-výběrový problém

Title: Multiple comparison with controls

Author: Maryna Sychova

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zdeněk Hlávka, Ph.D.

Abstract: The main theme of the diploma thesis is description of multiple comparison methods, which are used to compare pairs of means or medians.

At the beggining we define multiple testing and describe methods that control the probability of first type error at level α. The Šidák method and the prerequisites required for its use are described in detail. The work also includes a brief description of analysis of variance and an overview of several methods of multiple comparison. Additionally, the method of multiple comparison with control, its modifications and practical implementation is presented.

Keywords: Multiple comparison; analysis of variance; k-sample problem

Obsah

Úvod

Hlavním cílem této práce je poskytnout přehled o metodách mnohonásobného porovnávání pro nalazení odlišnosti mezi dvojicemi středních hodnot.

V první kapitole je definované mnohonásobné testování a chyby, které při něm mohou vzniknout, navíc jsou rozebrané metody, která kontrolují pravděpodobnost chyby prvního druhu na úrovni α. Ve druhé kapitole je stručně popsána konstrukce analýzy rozptylu jednoduchého třídění a jsou představeny vybrané metody mnohonásobného porovnávání. V další kapitole je popsaná metoda mnohonásobného porovnávání s jednou kontrolou a jeji praktická implementace. V závěru je popsaná metoda mnohonásobného porovnávání s více kontrolami.

1. Mnohonásobné testování

Ve statistice k mnohonásobnému testování dochází, když současně potřebujeme testovat více než jednu hypotézu. V této kapitole budeme uvažovat m nulových hypotéz H_1,..,H_m.

Při testování mohou vzniknout chyby 1. a 2. druhu. Chyba 1. druhu nastává, pokud zamítáme platnou nulovou hypotézu, zatímco chyba 2. druhu vzniká, pokud nezamítáme neplatnou hypotézu. Jinými slovy, pravděpodobnost chyby 1. druhu je podmíněná pravděpodobnost, že zamítneme nulovou hypotézu H_0 za předpokladu, že H_0 platí $P(T \geq t|H_0)$, kde T označuje testovou statistiku, t je kritická hodnota.

Tabulka 1.1 ukazuje možné výsledky při mnohonásobném testování, kde m_0 je celkový počet platných hypotéz, V označuje počet chyb 1. druhu, R je celkový počet zamítnutých hypotéz, T je počet chyb 2. druhu, W označuje celkový počet nezamítnutých hypotéz. Navíc R a W jsou pozorované hodnoty, zatímco U, T, V, S jsou nepozorované hodnoty, m a m_0 jsou pevné hodnoty a m_0 je neznámá.

Hypotézy	Nezamítáme	Zamítáme	Celkem
Platí	U	V	m_0
Neplatí	T	S	$m - m_0$
Celkem	W	R	m

Tabulka 1.1: Chyby 1. a 2. druhu při mnohonásobném testování

Standardní přístup při testování jednoduchých hypotéz ($m = 1$) je volba vhodného testu, který udržuje chybu 1. druhu na předem stanovené hladině významnosti α. Pokud budeme rozhodovat o zamítnutí $m > 1$ hypotéz, za předpokladu nezávislosti testů, pravděpodobnost toho, že alespoň 1 rozhodnutí bude chybné se rovná:

P(Uděláme alespoň 1 chybu 1. druhu) $= 1 - P$(Neuděláme žádnou chybu) $= 1 - (1 - \alpha)^m$, což je dost velká hodnota i při malých m.

Zobecněním chyby 1. druhu při testování rodiny hypotéz je FWER (familywise error rate), která je definována jako pravděpodobnost, že nastane alespoň 1 chyba 1. druhu, neboli $P(V \geq 1)$. Kontrola FWER na hladině významnosti α znamená, že platí nerovnost FWER $\leq \alpha$. Existuje několik metod pro kontrolu FWER.

1.1 Bonferroniho metoda

Bonferroniho metoda tvrdí, že pro kontrolu FWER na úrovni α stačí zamítnout všechny hypotézy H_i pro které $p_i \leq \frac{\alpha}{m}$, kde p_i je p-hodnota získaná při testování hypotézy H_i.

Z Booleho nerovnosti vyplývá, že pro konečný soubor náhodných jevů A_1,..,A_n pravděpodobnost, že nastane alespoň jedna z nich není větší než součet pravděpodobností jednotlivých událostí. Formálně lze psát:

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i). \tag{1.1}$$

V tomto případě pro celou rodinu hypotéz je úroveň významnosti fixovaná na úrovni α, tedy každý jednotlivý test bude kontrolován na úrovni významnosti $\frac{\alpha}{m}$. Použitím vzorce (1.1) máme:

$$\begin{aligned}\text{FWER} = P(V \geq 1) = P\left\{\bigcup_{i=1}^{m_0}\left(p_i \leq \frac{\alpha}{m}\right)\right\} \leq \\ \leq \sum_{i=1}^{m_0} P\left(p_i \leq \frac{\alpha}{m}\right) \leq m_0\frac{\alpha}{m} \leq m\frac{\alpha}{m} = \alpha\end{aligned}$$

S ohledem na mnohonásobné porovnávání upravíme p-hodnoty a budeme zamítat nulové hypotézy H_i pokud $p_i \leq \frac{\alpha}{m} \Leftrightarrow p_i m \leq \alpha$, kde hodnota $p_i m$ je p-hodnota upravená pomocí Bonferroniho metody a budeme ji značit p^{BONF}.

Tento postup umožňuje omezení FWER, ale jeho síla je nízká, protože každá z hypotéz je kontrolována na úrovni významnosti $\frac{\alpha}{m}$, tedy s větším m se pravděpodobnost zamítnutí nulové hypotézy snižuje.

1.2 Šidákova metoda

Nechť je dáno m různých nulových hypotéz, při použiti Šidákovy metody budeme zamítat nulovou hypotézu, pokud její p-hodnota je menší než $\alpha_S = 1 - (1-\alpha)^{\frac{1}{m}}$.

Chceme vyjasnit, za jakých předpokladů tato metoda kontroluje FWER na úrovni α.

Metoda je založená na větách 1 a 2, které nyní uvedeme. V učebnicích se uvádí s odkazem na Šidák (1967).

Věta 1. *Nechť $Y = (Y_1,..,Y_m)$ je vektor náhodných veličin z m-rozměrným normálním rozdělením s nulovými středními hodnotami, rozptyly $\sigma_1^2,\sigma_2^2,..,\sigma_m^2$ a pozitivně definitní korelační maticí $R = \{\rho_{ij}\}$. Potom pro libovolná kladná čísla $c_1,c_2,..,c_m$ platí:*

$$\begin{aligned}&P(|Y_1| \leq c_1,|Y_2| \leq c_2,..,|Y_m| \leq c_m) \\ &\qquad \geq P(|Y_1| \leq c_1)P(|Y_2| \leq c_2,..,|Y_m| \leq c_m).\end{aligned}$$

Důkaz. Nechť náhodné veličiny $Y_1,Y_2,..,Y_m$ mají sdruženou hustotu $f(y_1,y_2,..,y_m)$. Podobně, nechť $f(y_1)$ a $f(y_2,..,y_m)$ definují odpovídající marginální hustoty a $f(y_2,..,y_m|y_1)$ je podmíněná hustota pro $Y_1 = y_1$. Položíme

$$\begin{aligned}F(c_1) = &\int_{-c_1}^{c_1}\int_{-c_2}^{c_2}..\int_{-c_m}^{c_m} f(y_1,y_2,..,y_m)dy_1dy_2..dy_m \\ &-\int_{-c_1}^{c_1}\int_{-c_2}^{c_2}..\int_{-c_m}^{c_m} f(y_1)f(y_2,..,y_m)dy_1dy_2..dy_m.\end{aligned}$$

Z toho plyne, že

$$\begin{aligned}F(c_1) &= 2\int_0^{c_1}\int_{-c_2}^{c_2}..\int_{-c_m}^{c_m}[f(y_1,y_2,..,y_m) - f(y_1)f(y_2,..,y_m)]dy_1dy_2..dy_m \\ &= 2\int_0^{c_1} f(y_1)\{\int_{-c_2}^{c_2}..\int_{-c_m}^{c_m}[f(y_2,..,y_m|y_1) - f(y_2,..,y_m)]dy_2..dy_m\}dy_1.\end{aligned}$$

Proto

$$\frac{\delta F(c_1)}{\delta c_1} = 2f(c_1)\int_{-c_2}^{c_2} .. \int_{-c_m}^{c_m} [f(y_2,..,y_m|c_1) - f(y_2,..,y_m)]dy_2..dy_m.$$

Dále se zkoumá chování funkce $F(c_1)$ a dokazuje se, že $F(c_1) \geq 0$ pro všechna $0 \leq c_1 \leq \infty$.

Podrobný důkaz věty lze nalézt v Šidák (1967).

□

Důsledek 1. Za předpokladů věty 1 dostáváme:

$$P(|Y_1| \leq c_1,..,|Y_m| \leq c_m) \geq \prod_{i=1}^{m} P(|Y_i| \leq c_i). \tag{1.2}$$

Pokud budeme uvažovat náhodný výběr n vektorů $Y_j = (Y_{1j},..,Y_{mj})$, $j = 1,..,n$, kde Y_j má m-rozměrné normální rozdělení s neznámými středními hodnotami $\mu_1,..,\mu_m$, známými rozptyly $\sigma_1^2,..,\sigma_m^2$ a positivně definitní korelační maticí $R = \{\rho_{ij}\}$, pak veličiny $X_i = \sqrt{n}\frac{\overline{Y_i}-\mu_i}{\sigma_i}$, $i = 1,..,m$, kde $\overline{Y_i} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ splňují předpoklady věty 1 a důsledku 1.

Pokud chceme najít intervaly spolehlivosti pro $\mu_1,..,\mu_m$ s hladinou spolehlivosti $1-\alpha$, stanovíme hodnoty $c_1,..,c_m$ tak, aby pravá strana (1.2) byla rovna $1-\alpha$ a požadovaný interval spolehlivosti je tedy roven:

$$P(\mu_1 \in I_1,..,\mu_m \in I_m, \text{ kde } I_i = \langle \overline{Y_i} - \frac{c_i\sigma_i}{\sqrt{n}}; \overline{Y_i} + \frac{c_i\sigma_i}{\sqrt{n}} \rangle, i = 1,..,m) \geq 1-\alpha.$$

Tedy chceme aby platilo

$$1-\alpha = (1-\alpha_S)^m \Leftrightarrow (1-\alpha)^{1/m} = 1-\alpha_S \Leftrightarrow 1-(1-\alpha)^{1/m} = \alpha_S$$

Ukazuje se, že k dosažení dané hladiny α musíme přizpůsobit hodnoty α_S použité pro každý konfidenční interval. Můžeme položit $c_1 = c_2 = .. = c_m = c_\alpha$ tak, aby c_α splňovala $\Phi(c_\alpha) = \frac{1}{2}[1-(1-\alpha)^{(1/m)}]$, kde Φ definuje standardní normální rozdělení. Tím jsme ukázali, že Šidákova metoda funguje pro normální rozdělení se známými rozptyly.

Následující věta je rozšíření na situaci kdy máme odhady rozptylů.

Věta 2. *Uvažujme následující dvě pravděpodobnostní rozdělení P, P_1 náhodných veličin $Z_1,..,Z_m,s$. Při P, nechť vektor $Z = (Z_1,..,Z_m)$ má m-rozměrné normální rozdělení s nulovými středními hodnotami, rozptyly $\sigma_1^2,..,\sigma_m^2$ a korelační matici $R = \{\rho_{ij}\}$. Při P_1, nechť vektor $Z = (Z_1,..,Z_m)$ má stejné rozdělení, ale navíc Z_1 je nezávislé na $Z_2,..,Z_m$(jinými slovy $\rho_{12} = \rho_{13} = .. = \rho_{1m} = 0$). Navíc předpokládejme, že s je kladná náhodná veličina nezávislá na $Z_1,..,Z_m$ a má stejná rozdělení při P a P_1. Pak pro libovolná kladná čísla $c_1,..,c_m$ platí:*

$$P(\frac{|Z_1|}{s} \leq c_1, \frac{|Z_2|}{s} \leq c_2,.., \frac{|Z_m|}{s} \leq c_m) \geq P_1(\frac{|Z_1|}{s} \leq c_1, \frac{|Z_2|}{s} \leq c_2,.., \frac{|Z_m|}{s} \leq c_m)$$
$$\geq P(\frac{|Z_1|}{s} \leq c_1) P(\frac{|Z_2|}{s} \leq c_2,.., \frac{|Z_m|}{s} \leq c_m).$$

Důkaz. Pomocí věty 1 můžeme získat následující podmíněné pravděpodobností

$$\begin{aligned}&P(|Z_1| \le c_1 s, |Z_2| \le c_2 s, .., |Z_m| \le c_m s | s)\\ &\qquad \ge P_1(|Z_1| \le c_1 s, |Z_2| \le c_2 s, .., |Z_m| \le c_m s | s).\end{aligned}$$

Vzhledem k tomu, že s má stejná rozdělení při P a P_1, můžeme vzít střední hodnoty s ohledem na toto rozdělení na obou stranách, z čehož plyne první nerovnost věty.

Pro důkaz druhé nerovnosti, použijeme nerovnost uvedenou v Kimball (1951), která říká, že $E\{F(\xi)G(\xi)\} \ge E\{F(\xi)\}E\{G(\xi)\}$, což platí pro nezáporné rostoucí funkce F,G náhodné veličiny ξ. Z toho plyne, že

$$\begin{aligned}&P_1(|Z_1| \le c_1 s, |Z_2| \le c_2 s, .., |Z_m| \le c_m s)\\ &\quad = E\{P_1(|Z_1| \le c_1 s, |Z_2| \le c_2 s, .., |Z_m| \le c_m s | s)\}\\ &\quad = E\{P(|Z_1| \le c_1 s | s) P(|Z_2| \le c_2 s, .., |Z_m| \le c_m s | s)\}\\ &\quad \ge E\{P(|Z_1| \le c_1 s | s)\} E\{P(|Z_2| \le c_2 s, .., |Z_m| \le c_m s | s)\}\\ &\quad = P(|Z_1| \le c_1 s) P(|Z_2| \le c_2 s, .., |Z_m| \le c_m s).\end{aligned}$$

□

Důsledek 2. Nechť rozdělení P vektoru $Z = (Z_1, .., Z_m)$ je definováno stejně jako ve větě 2. Nechť při rozdělení P_m vektor $Z = (Z_1, .., Z_m)$ má m-rozměrné normální rozdělení s nulovými středními hodnotami, stejnými rozptyly $\sigma_1^2 = .. = \sigma_m^2 = \sigma^2$ a všechny $Z_1, .., Z_m$ jsou nezávislé. Navíc předpokládejme, že s je kladná náhodná veličina nezávislá na $Z_1, .., Z_m$ a má stejná rozdělení při P a P_m. Pak pro libovolná kladná čísla $c_1, .., c_m$ platí:

$$\begin{aligned}&P(\frac{|Z_1|}{s} \le c_1, \frac{|Z_2|}{s} \le c_2, .., \frac{|Z_m|}{s} \le c_m) \ge P_m(\frac{|Z_1|}{s} \le c_1, \frac{|Z_2|}{s} \le c_2, .., \frac{|Z_m|}{s} \le c_m)\\ &\qquad \ge \prod_{i=1}^{m} P(\frac{|Z_i|}{s} \le c_i).\end{aligned}$$

Nyní se vrátíme k intervalům spolehlivosti odvozeným za důsledkem 1, ale neznámé rozptyly nahradíme jejich odhady. Tedy budeme předpokládat, že $\sigma_1^2 = \sigma_2^2 = .. = \sigma_m^2 = \sigma^2$, kde σ^2 je neznámé. Můžeme odhadnout σ_i^2 pomoci

$$s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_{ij} - \overline{Y_i})^2,$$

kde i je nějaký pevný index vybraný z $1, .., m$. Pak veličiny $Z_i = \sqrt{n}(\overline{Y_i} - \mu_i)$, $i = 1, .., m$ a $s^2 = \frac{1}{m} \sum_{i=1}^{m} s_i^2$ splňují předpoklady věty 2 a důsledku 2. Interval spolehlivosti pro $\mu_1, .., \mu_m$ je roven:

$$P(\mu_1 \in I_1, .., \mu_m \in I_m, \text{ kde } I_i = \langle \overline{Y_i} - \frac{c_i s}{\sqrt{n}}; \overline{Y_i} + \frac{c_i s}{\sqrt{n}} \rangle, i = 1, .., m)$$

a můžeme učinit závěr o hladině spolehlivosti obdobný tomu, pro známé rozptyly.

Uvedená metoda je ve článku Šidák (1967) odvozená pouze pro výběry z normálního rozdělení se známými rozptyly nebo s odhady rozptylů.

Tedy, při testování sady m nulových hypotéz H_i s ohledem na mnohonásobné porovnávání upravíme příslušné p-hodnoty a budeme zamítat nulové hypotézy H_i pokud: $p_i \leq \alpha_S = 1-(1-\alpha)^{1/m} \Leftrightarrow 1-(1-p_i)^m \leq \alpha$, kde hodnota $1-(1-p_i)^m$ je p-hodnota upravená pomocí Šidákovy metody a budeme ji značit p^{SID}.

Tato metoda kontroluje FWER na úrovni α, pokud všechny testové statistiky T_i splňují předpoklady věty 1 nebo věty 2.

$$P(|T_1| \leq t_1,..,|T_m| \leq t_m) \geq \prod_{i=1}^{m} P(|T_i| \leq t_i), \forall t_1,..,t_m.$$

Porovnání Bonferroniho a Šidákovy metody

Chceme zjistit, jaká hodnota p^{BONF} nebo p^{SID} je větší, tím pádem která metoda je slabší. Definujme rozdíl těchto dvou hodnot: $f_m(p) = p^{BONF} - p^{SID} = p_i m - 1 + (1-p_i)^m$, kde $m \geq 1$ je pevné, $p_i \in \langle 0;1\rangle$ a chceme se dozvědět, jestli tento rozdíl je větší nebo menší než 0. Abychom zjistili průběh dané funkce, zderivujeme ji, položíme derivace rovnou 0 a podíváme jak se tato funkce chová.

$$\begin{aligned}
&\frac{\partial}{\partial p_i} : m - m(1-p_i)^{m-1} = 0 \\
&1 - p_i = \sqrt[m-1]{\frac{m}{m}} \\
&p_i = 0
\end{aligned}$$

Našli jsme, že funkce $f_m(p)$ má jediný extrémní bod pro $p_i = 0$ a díky tomu, že daná funkce je spojitá, může neomezeně růst nebo klesat napravo od tohoto bodu. Abychom dozvěděli, jestli tato funkce roste nebo klesá pro $p_i > 0$, zvolíme libovolnou hodnotu p_i, např. $p_i = 0.4$, dosadíme ji do naší derivace a zkoumáme chování dané funkce pro všechna $m \geq 1$.

$$\begin{aligned}
m = 1 &: \; 1 - 1 \times 0.6^0 = 0 \\
m = 2 &: \; 2 - 2 \times 0.6^1 > 0 \\
m = 3 &: \; 3 - 3 \times 0.6^2 > 0 \text{ a t.d}
\end{aligned}$$

Z toho plyne, že pro $m = 1$ náš rozdíl je nulový, tedy $p^{BONF} = p^{SID}$, a pro všechna $m > 1 : p^{BONF} > p^{SID}$.

V statistickém softwaru R, který se v této práci používá pro numerické výpočty, úpravě pomoci Bonferroniho metody odpovídá funkce
`p.adjust(p,"bonferroni",n=lenght(p))`, kde p je vektor p-hodnot, n je počet porovnání a musí být nejméně roven délce p. Aby upravená p-hodnota nepřesahovala 1, tato funkce je definovaná následujícím způsobem:

$$p.bonf = fuction(p)\{pmin\{p * length(p),1\}\}$$

Protože pro Šidákovu metodu v R není definovaná funkce `p.adjust`, pro účely teto práce byla definovaná nová funkce:

$$p.sidak = function(p)\{1-(1-p)^{length(p)}\}.$$

1.3 Holmova metoda

Holmova metoda je silnější než Bonferroniho metoda a řeší problém poklesu síly s rostoucím počtem hypotéz.

Nechť $p_{(1)} \leq ,..,\leq p_{(m)}$ jsou úrovně významnosti p_i, seřazené od nejmenších po největší a $H_{(1)},.., H_{(m)}$ jsou odpovídající nulové hypotézy. Holmova metoda je definovaná následovně:

1. Je-li $p_{(1)} \geq \frac{\alpha}{m}$, nezamítáme nulové hypotézy $H_{(1)},.., H_{(m)}$ a zastavíme se. V opačném případě, pokud $p_{(1)} < \frac{\alpha}{m}$, zamítneme nulovou hypotézu $H_{(1)}$ a pokračujeme v kontrole zbývajících hypotéz na úrovni významnosti $\frac{\alpha}{m-1}$.
2. Je-li $p_{(2)} \geq \frac{\alpha}{m-1}$, nezamítáme nulové hypotézy $H_{(2)},.., H_{(m)}$ a zastavíme se. V opačném případě, pokud $p_{(2)} < \frac{\alpha}{m-1}$, zamítneme nulovou hypotézu $H_{(2)}$ a pokračujeme v kontrole zbývajících hypotéz na úrovni významnosti $\frac{\alpha}{m-2}$ a t.d.

Věta 3. *Holmova metoda kontroluje FWER na hladině významnosti α.*

Důkaz. Nechť I_0 je množina indexů odpovídajících (neznámým) platným nulovým hypotézám, kterých je celkově m_0. Předpokládejme, že nesprávně zamítáme platnou nulovou hypotézu. Musíme prokázat, že pravděpodobnost této události je maximálně α. Nechť H_h je první zamítnutá platná hypotéza. Pak $H_{(1)},..,H_{(h-1)}$ jsou zamítnuté neplatné nulové hypotézy a $h - 1 \leq m - m_0$. Tedy platí $\frac{1}{m-h+1} \leq \frac{1}{m_0}$ (1.3). Protože zamítáme h, platí $p_{(h)} < \frac{\alpha}{m-h+1}$ podle definice metody. Z (1.3) plyne, že pravá strana je maximálně rovna $\frac{\alpha}{m_0}$. Tedy, pokud jsme chybně zamítli platnou hypotézu, musí existovat platná hypotéza s p-hodnotou nejvýše rovnou $\frac{\alpha}{m_0}$.

Definujme $A = \left\{p_i \leq \frac{\alpha}{m_0} \text{ pro některá } i \in I_0\right\}$. Pro libovolný soubor platných nulových hypotéz I_0 platí, že $P(A) \leq \alpha$ (podle Bonferroniho nerovnosti). Proto pravděpodobnost zamítnutí platné hypotézy je nejvýše α.

□

1.4 Benjamini-Hochbergova metoda

Další hodnota, která zobecňuje chyby 1. druhu při mnohonásobném testování rodiny hypotéz je FDR (False Discovery rate), navržená v článku Benjamini a Hochberg (1995). Postupy kontroly FDR poskytují méně přesné kontroly chyb 1. druhu ve srovnání s postupy kontroly FWER. Postupy kontroly FDR mají vyšší sílu, a to za cenu zvýšeného počtu chyb 1. druhu. FDR je definována jako střední hodnota podílu chyb mezi zamítnutými hypotézami. Definujme Q jako poměr počtu chybně zamítnutých hypotéz V ke všem zamítnutým hypotézám R: $Q = \frac{V}{R}$, hodnoty V a R jsou definovane v tabulce 1.1:

$$\text{FDR} = Q_{\text{e}} = E[Q] = E\left[\frac{V}{V+S}\right] = E\left[\frac{V}{R}\right], \text{při } R > 0.$$

Kontrola FDR na úrovni α znamená, že $E\left[\frac{V}{R}\right] \leq \alpha$. Jednou z metod pro kontrolu FDR je Benjamini-Hochbergova metoda.

Jedná se o postup shora dolů (podobný Holmově metodě) s následujícími úrovněmi významnosti: $\alpha_1 = \frac{\alpha}{m}$, ..., $\alpha_i = \frac{i\alpha}{m}$, ..., $\alpha_m = \alpha$.

Nechť $p_{(1)} \leq, ..., \leq p_{(m)}$ jsou úrovně významnosti p_i, seřazené od nejmenších po největší a $H_{(1)}, ..., H_{(m)}$ jsou odpovídající nulové hypotézy. Benjamini-Hochbergova metoda je definovaná následovně:

1. Je-li $p_{(1)} \geq \frac{\alpha}{m}$, přijmeme nulové hypotézy $H_{(1)}, ..., H_{(m)}$ a zastavíme se. V opačném případě, pokud $p_{(1)} < \frac{\alpha}{m}$, zamítneme hypotézu $H_{(1)}$ a pokračujeme v kontrole zbývajících hypotéz na úrovni významnosti $\frac{2\alpha}{m}$.

2. Je-li $p_{(2)} \geq \frac{2\alpha}{m}$, přijmeme nulové hypotézy $H_{(2)}, ..., H_{(m)}$ a zastavíme se. V opačném případě, pokud $p_{(2)} < \frac{2\alpha}{m}$, zamítneme hypotézu $H_{(2)}$ a pokračujeme v kontrole zbývajících hypotéz na úrovni významnosti $\frac{3\alpha}{m}$ a t.d.

Podrobnější popis metody a důkaz lze nalézt v Benjamini a Hochberg (1995).

2. Analýza rozptylu jednoduchého třídění

Cílem analýzy rozptylu je ověřit, zda na střední hodnotu nezávislých náhodných veličin, které pochází z normálního rozdělení se stejným rozptylem, má vliv hodnota jednoho (analýza rozptylu jednoduchého třídění) nebo více faktorů (vícefaktorová analýza rozptylu). Zde se zaměříme pouze na analýzu rozptylu jednoduchého třídění. Pokud zamítneme nulovou hypotézu, budeme hledat výběry, které se od sebe signifikantně liší. To se provádí pomocí metod mnohonásobného porovnávání.

Předpokládáme, že máme náhodné výběry $Y_{i1},Y_{i2}..,Y_{in_i}$, $1 \leq i \leq m$, které jsou nezávislé a mají rozdělení $N(\mu_i,\sigma^2)$. Pokud $n_1 = n_2 = .. = n_m$, říkáme, že jde o vyvážené třídění, jinak jde o nevyvážené třídění. Testujeme nulovou hypotézu $H_0 : \mu_1 = \mu_2 = .. = \mu_m$ proti alternativě H_1 : hypotéza H_0 neplatí. Podrobný popis algoritmu, který nyní stručně uvedeme, lze nalézt v Anděl (1998).

1. Nechť $n = n_1 + n_2 + .. + n_m$ je celkový počet pozorování a m označuje počet výběrů.

2. Utvoříme tabulku dat a pomocných výsledků

data	četnost	součet	průměr
$Y_{11},..,Y_{1n_1}$	n_1	Y_1	$\overline{Y_1}$
..	..	..	..
..			
$Y_{i1},..,Y_{in_i}$	n_i	Y_i	$\overline{Y_i}$
..	..	..	..
..			
$Y_{m1},..,Y_{mn_m}$	n_m	Y_m	$\overline{Y_m}$
celkem	n	$Y_{..}$	$Y_{..}$

Tedy: $Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$ je součet hodnot v i-tém výběru; $Y_{..} = \sum_{i=1}^{m} Y_i$ celkový součet; $\overline{Y_{i.}} = \frac{1}{n_i} Y_{i.}$ odhad střední hodnoty μ_i; $\overline{Y_{..}} = \frac{1}{n} Y_{..}$ odhad střední hodnoty μ.

Potom pro reziduální součet čtverců S_e dostaneme:
$S_e = S_T - S_A = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_{i.}})^2$, kde

$$S_T = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_{..}})^2 \text{ definuje celkový součet čtverců}$$

$$S_A = \sum_{i=1}^{m} n_i (\overline{Y_{i.}} - \overline{Y_{..}})^2 \text{ označuje součet čtverců skupin.}$$

Za platností nulové hypotézy H_0 dostáváme, že hodnota S_A má χ^2 rozdělení s $f_A = m - 1$ stupni volnosti a hodnota S_T má χ^2 rozdělení s $f_T = n - 1$ stupni volnosti.

3. Vypočteme hodnotu testové statistiky

$$F = \frac{n-m}{m-1}\frac{S_A}{S_e} = \frac{n-m}{m-1}\frac{S_A}{S_T - S_A},$$

   která má rozdělení $F_{m-1,n-m}$.

4. Pokud překročí veličina F kritickou hodnotu $F_{m-1,n-m}(1-\alpha)$ zamítneme H_0 na hladině α.

V případě zamítnutí nás zajímá, pro které dvojice je $\mu_i \neq \mu_j$. K těmto účelům slouží metody mnohonásobného porovnání. Metody, která se nejčastěji používají v úlohách analýzy rozptylu, jsou Tukeyova a Scheffého metoda.

2.1 Tukeyova metoda mnohonásobného porovnávání

Tukeyova metoda mnohonásobného porovnávání je obdobou t-testu a používá se v případě vyváženého třídění, tedy pokud $n_1 = .. = n_m = r$. Testuje se nulová hypotéza $H_0 : \mu_i = \mu_j$, oproti alternativě $H_1 : \mu_i \neq \mu_j$, tj. testujeme hypotézu, že střední hodnoty porovnávaných skupin i a j jsou stejné. Testová statistika má tvar:

$$Q_{ij} = \frac{|\overline{Y_i} - \overline{Y_j}|}{\frac{s}{\sqrt{r}}}, \tag{2.1}$$

kde s^2 je nestranný odhad rozptylu σ^2 a pro shodné počty pozorování r ve skupinách i a j je ve tvaru $s^2 = \frac{S_E}{m(r-1)}$ a testová statistika Q_{ij} má za platnosti H_0 rozdělení, které se nazývá studentizované rozpětí a porovná se s příslušnou kritickou hodnotou $q_{m,n-m}(\alpha)$ (viz Anděl (1998)).

Definice 1. *Nechť $X_1,..,X_m$ jsou nezávislé a stejně rozdělené náhodné veličiny s rozdělením $N(\mu,\sigma^2)$. Nechť s_ν^2 je odhad σ^2 s ν stupni volnosti, který je nezávislý na pozorováních X_i a $\frac{\nu s^2}{\sigma^2} \sim \chi^2$. Pak hodnota $Q = \frac{R}{s}$, kde $R = \max(X_1,..,X_m) - \min(X_1,..,X_m)$ se nazývá studentizované rozpětí.*

Pokud je hodnota testového kritéria Q_{ij} větší než kritická hodnota, potom zamítáme nulovou hypotézu. Tento test musíme provést pro všechny možné dvojice skupin. Tukeyho test patří k nejužívanějším a považuje se také za jeden z nejlepších z hlediska vhodného kompromisu síly testu a možnosti výskytu chyby prvního druhu.

$$1 - \mathrm{FWER} = P(Q_{ij} < q_{m,n-m}, \forall i,j) = P(\frac{|\overline{Y_i} - \overline{Y_j}|}{\frac{s}{\sqrt{r}}} < q_{m,n-m}, \forall i,j)$$

$$= P(\frac{\max \overline{Y_i} - \min \overline{Y_j}}{\frac{s}{\sqrt{r}}} < q_{m,n-m}) = P(\frac{R}{\frac{s}{\sqrt{r}}} < q_{m,n-m}) = P(Q \leq q_{m,n-m}) = 1 - \alpha,$$

kde $\overline{Y_i} \sim N(\mu_i, \frac{\sigma^2}{r})$ a $\frac{s^2}{r}$ je odhad $\frac{\sigma^2}{r}$.

2.2 Schefféhο metoda mnohonásobného porovnávání

Schefféhο metoda umožňuje porovnávat rozdíly mezi středními hodnotami jak u vyváženého, tak u nevyváženého třídění (viz Anděl (2007)). Nulová hypotéza je definovaná stejně jako u Tukeyho testu, který je popsán v sekci 2.1 a testová statistika je definovaná následujícím způsobem:

$$S = \frac{|\overline{Y_i} - \overline{Y_j}|}{s\sqrt{(m-1)(\frac{1}{n_i} + \frac{1}{n_j})}}, \tag{2.2}$$

kde $s^2 = \frac{S_E}{n-m}$.

Kritická hodnota je $\sqrt{F_{m-1,n-m}(1-\alpha)}$, což je kvantil Fisher-Snedecorova rozdělení.

2.3 Dunnettova T3 metoda mnohonásobného porovnávání

Dunnetova T3 metoda se používá, pokud máme nevyvážená data a navíc předpokládáme, ze naše náhodné výběry $Y_{i1}, Y_{i2}..,Y_{in_i}$, $1 \leq i \leq m$ pochází z normálního rozdělení se střední hodnotou μ_i a rozptyly σ_i^2 (viz Dunnett (1980)). Tato metoda je založena na studentizovaném maximálním modulu.

Definice 2. *Nechť $X_1,..,X_m$ jsou nezávislé a stejně rozdělené náhodné veličiny s rozdělením $N(0,\sigma^2)$. Nechť s_ν^2 je odhad σ^2 s ν stupni volnosti, který je nezávislý na pozorováních X_i a $\frac{\nu s^2}{\sigma^2} \sim \chi^2_\nu$. Pak hodnota $M_{m\nu} = \frac{\max|X_i|}{s_\nu}$ se nazývá studentizovaný maximální modul.*

Dunnettova T3 metoda se používá, když jsou velikosti vzorku a počet stupňů volnosti malé, tj. $n_i < 50$ a $\nu < 50$.

Tato metoda testuje nulovou hypotézu $H_0 : \mu_i = \mu_j, \forall i \neq j = 1,..,m$ proti alternativě $H_1 : \mu_i \neq \mu_j$ alespoň pro jeden pár.

Testová statistika je definovaná následujícím způsobem:

$$t_{ij} = \frac{|\overline{Y_i} - \overline{Y_j}|}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}}, \text{ kde } i \neq j = 1,..,m.$$

Budeme zamítat nulovou hypotézu, pokud $t_{ij} \geq M_{\alpha,k^*,\nu}$, kde $k^* = m(m-1)/2$ je počet porovnání a $M_{\alpha,k^*,\nu}$ je horní α kvantil studentizovaného maximálního modulu. Obecně je obtížné vypočítat přesnou hodnotu $M_{\alpha,k^*,\nu}$, jeden ze způsobu je založen na Šidákově nerovnosti, která je popsaná v sekci 1.2. Dunnettova T3 metoda kontroluje FWER na úrovni α (viz Stoline (1981),Field (2009)).

$$\begin{aligned} 1 - \mathrm{FWER} &= P(t_{ij} < M_{\alpha,k^*,\nu}, \forall i,j) = P(\frac{|\overline{Y_i} - \overline{Y_j}|}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}} < M_{\alpha,k^*,\nu}, \forall i,j) \\ &= P(\max_{ij} \frac{|\overline{Y_i} - \overline{Y_j}|}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}} < M_{\alpha,k^*,\nu}) = P(\frac{\max_{ij}|X_{ij}|}{s} < M_{\alpha,k^*,\nu}) = 1 - \alpha, \end{aligned} \tag{2.3}$$

kde $X_{ij} = \frac{\overline{Y_i}-\overline{Y_j}}{\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}}$ a $s^2 = \frac{1}{m}\sum_{i=1}^{m} s_i^2$. Nerovnost 2.3 platí, pokud X_{ij} splňují předpoklady věty 2. Protože X_{ij} jsou závislé, použitím Šidákově nerovnosti dostáváme, že $1 - \text{FWER} = 1 - \alpha$.

2.4 Dunnův test

Bez předpokladu normality lze použít Dunnův test(viz Dunn (1964), Dunn (1961)), který je vhodný po použití Kruskal-Wallisova testu. Kruskal-Wallisův test, který je neparametrickým analogem ANOVA, je testem nulové hypotézy, že žádná z m skupin stochasticky nepřevládá. Jinými slovy, mějme nezávislý náhodný výběr $Y_{i1},..,Y_{in_i}$, který pochází z rozdělení se spojitou distribuční funkcí F_i, pro $i = 1,..,m$. Testuje se hypotéza, že distribuční funkce všech výběrů je shodná, tedy $H_0 : F_1(x) = .. = F_m(x), \forall x$ proti alternativě H_1 : alespoň jedna distribuční funkce se liší. Pokud zamítneme nulovou hypotézu, je třeba rozhodnout, které skupiny se od sebe signifikantně liší.

Označme $n = \sum_{i=1}^{m} n_i$ celkový rozsah všech výběrů Y_{ij}, pro $i = 1,..,m$, $j = 1,..,n_i$. Všech n hodnot uspořádáme do rostoucí posloupnosti a určíme pořadí R_{ij} každé veličiny Y_{ij}. Součet pořadí všech hodnot i-tého výběru, $i = 1,..,m$, označme W_i. Pro kontrolu musí platit, že celkový součet pořadí je $W_1 + .. + W_m = n(n+1)/2$. Vstupem do testu je průměrná hodnota pořadí v jednotlivých skupinách, $\overline{W_i} = \frac{W_i}{n_i}$. Testová statistika pro srovnání dvou skupin je definovaná následujícím vztahem:

$$z_{ij} = \frac{\overline{W_i} - \overline{W_j}}{\sigma_{ij}}, \tag{2.4}$$

kde $\sigma_{ij}^2 = [\frac{n(n+1)}{12}](\frac{1}{n_i} + \frac{1}{n_j})$ je rozptyl $\overline{W_i} - \overline{W_j}$. Pokud se v datech vyskytují shody, tak se obvykle pracuje s průměrným pořadím a výpočet rozptylu se mírně zkomplikuje.

Prohlásíme, že distribuční funkce i-tého a j-tého výběru jsou pro hladinu významnosti α statisticky významné, pokud $z_{ij} \geq z_{1-\alpha/2k^*}$, kde $k^* = m(m-1)/2$ je počet porovnání a $z_{1-\alpha/2k^*}$ je upravený kvantil normálního standardizovaného rozdělení.

3. Mnohonásobné porovnávání s kontrolou

Na rozdíl od předchozích metod, kde jsme porovnávali střední hodnoty všech skupin, při porovnání s kontrolou budeme porovnávat jednotlivé skupiny se skupinou kontrolní.

Mějme celkem (m+1) skupin, z nich jedna je kontrolní a označíme ji jako 0. skupinu. V této kapitole budeme předpokládat vyvážená data, tedy $n = n_1 = .. = n_m$. Nechť Y_{ij}, která odpovídá j-té experimentální jednotce, získané při i-tém ošetření $(j = 0, ..., n)$ má normální rozdělení se střední hodnotou μ_i a rozptylem σ^2. Z toho plyne, ze vzájemně nezávislé aritmetické průměry $\overline{Y_i} = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ mají rozdělení $N(\mu_i, \sigma^2/n)$, pro $i = 0,..,m$. Kromě toho nechť s^2 je nestranným odhadem σ^2, který je nezávislý na všech aritmetických průměrech a $\frac{\nu s}{\sigma^2}$ má χ^2 rozdělení s $\nu = \sum_{i=1}^{m} n - (m+1)$ stupni volnosti.

3.1 Dunnettova metoda

Budeme testovat hypotézy $H_{0_i} : \mu_i \leq \mu_0$ proti alternativě $H_{1_i} : \mu_i > \mu_0$, pro $i = 1,..,m$.

Definujme testovou statistiku následujícím způsobem:

$$t_i = \frac{\overline{Y_0} - \overline{Y_i}}{s\sqrt{\frac{1}{n} + \frac{1}{n_0}}}, \text{ kde } i = 1,..,m. \tag{3.1}$$

Každá testová statistika (3.1) má jednorozměrné t-rozdělení. Tedy vektor testových statistik $T = (t_1,..,t_m)$ má m-rozměrné Studentovo rozdělení s ν stupni volnosti a korelační matici $R = (\rho_{ij})_{ij}$, kde, za předpokladu vyvážených dat, ρ_{ij} je konstantní a pro všechna $i \neq j$ se rovná

$$\begin{aligned}\rho_{ij} &= \frac{cov(\overline{Y_0} - \overline{Y_i}, \overline{Y_0} - \overline{Y_j})}{\sqrt{var(\overline{Y_0} - \overline{Y_i})}\sqrt{var(\overline{Y_0} - \overline{Y_j})}} = \frac{cov(\overline{Y_0}, \overline{Y_0})}{\sqrt{\frac{1}{n_0}\sigma^2 + \frac{1}{n}\sigma^2}\sqrt{\frac{1}{n_0}\sigma^2 + \frac{1}{n}\sigma^2}} \\ &= \frac{var\overline{Y_0}}{(\frac{1}{n_0} + \frac{1}{n})\sigma^2} = \frac{\frac{1}{n_0}\sigma^2}{(\frac{1}{n_0} + \frac{1}{n})\sigma^2} = \frac{n}{n + n_0}\end{aligned} \tag{3.2}$$

Nechť $c_k = t^{\alpha}_{k,\nu,p}$ je horní α-kvantil $\max_{1 \leq i \leq k} t_i$, pro $k = 1,..,m$. Pro $k = 1$ $c_1 = t^{\alpha}_{\nu}$, horní α-kvantil jednorozměrného rozdělení s ν s.v. Stanovíme hodnotu c_k tak, aby FWER$\leq \alpha$.

Single-Step (SS) testing

Používá stejnou kritickou konstantu c_k pro všechny testy a zamítá H_{0_i} pokud $t_i \geq c_k$ pro $i = 1,..,m$ v ostatních případech hypotézu H_{0_i} nezamítáme. Pro kontrolu FWER na hladině α stanovíme kritické hodnoty c_k pomocí Bonferroniho nebo Šidákovy metody, které byly popsány v sekcích 1.1 a 1.2.

Step-Down (SD) testing

Nejdřív setřídíme testové statistiky $t_1 \leq .. \leq t_m$ (a příslušné hypotézy jako H_{0_1},..,H_{0_m}). Začínáme z t_m a pokračujeme t_{m-1} a t.d pokud $t_i \geq c_i$, v takovém případe zamítáme H_{0_i}. Kdy poprvé pozorujeme $t_i < c_i$, např. $i = k$, nezamítáme všechny hypotézy H_{0_1},..,H_{0_k}. Pro kontrolu FWER na hladině α lze hodnotu kritické konstanty c_i stanovit pomoci Holmovy metody, která byla popsaná v sekci 1.3.

Podrobnější popis metody lze nalézt v Dunnett a Tamhane (1992).

3.2 Příklad

Nechť naše pozorování patří do dvou disjunktních skupin, které jsou pak rozdělená do 13 kategorií a nechť pozorování X_{ji} v j-té skupině a i-té kategorii jsou shrnuté pomocí jejich vyběrových průměrů $\overline{X_{ji}}$ a rozptylů $\hat{\sigma}_{ji}$, $j \in \{1{,}2\}$, $i = 1,..13$. Tabulka 3.1 ukazuje pozorované průměrné maximální rychlosti při výskoku a směrodatné odchylky pro chlapce a dívky ve 13 věkových kategorií.

Věkové	Dívky			Chlapci		
kategorie	$\overline{X_1}$	$\hat{\sigma}_1$	n_1	$\overline{X_2}$	$\hat{\sigma}_2$	n_2
6-7	1.89	0.17	33	1.87	0.18	19
7-8	2.00	0.21	43	1.98	0.20	38
8-9	2.01	0.21	33	2.06	0.21	38
9-10	2.06	0.18	42	2.14	0.18	29
10-11	2.19	0.22	42	2.17	0.19	45
11-12	2.23	0.15	30	2.31	0.23	37
12-13	2.26	0.13	41	2.35	0.23	40
13-14	2.30	0.22	32	2.53	0.21	36
14-15	2.28	0.23	31	2.66	0.19	20
15-16	2.37	0.17	29	2.72	0.22	26
16-17	2.33	0.19	17	2.83	0.28	9
17-18	2.35	0.18	25	2.76	0.16	13
18-19	2.33	0.17	34	2.87	0.10	14

Tabulka 3.1: Pozorované průměrné maximální rychlosti při výskoku a směrodatné odchylky pro chlapce a dívky ve 13 věkových kategoriích.

Pro účely dalších výpočtů definujeme hodnotu $\overline{Y_i} = \overline{X_{1_i}} - \overline{X_{2_i}}$, $i = 1,..,13$, která vyjadřuje rozdíl v průměrných rychlostech skoku mezi chlapci a dívkami ve stejných věkových kategorií.

Předpokládejme, že data jsou realizací nezávislých náhodných výběrů, které pochází z normálního rozdělení se stejným rozptylem, který odhadneme následujícím způsobem:

$$s^2 = \frac{\sum_{k=1}^{2}\sum_{i=1}^{13}\hat{\sigma}_{k_i}^2(n_{k_i}-1)}{\sum_{k=1}^{2}\sum_{i=1}^{13}n_{k_i}-26} = 0.039$$

Nyní aplikujeme několik metod na naše data.

Dvouvýběrový t-test

Budeme testovat nulovou hypotézu $H_{0_i} : \overline{Y_i} = 0$, pro $i = 1,..13$ proti alternativě H_{1_i} : hypotéza H_i neplatí.

Testová statistika je definovaná předpisem:

$$t_i = \frac{\overline{Y_i}}{s\sqrt{\frac{1}{n_{1_i}} + \frac{1}{n_{2_i}}}}, i = 1,..13$$

P-hodnoty spočítáme podle vzorce: $p_i = 2(1 - F(|t_i|))$, kde t_i je pozorovaná hodnota testové statistiky a F je distribuční funkce t-rozdělení s $\sum_{k=1}^{2}\sum_{i=1}^{13} n_{k_i} - 26$ stupni volnosti.

Výsledné p-hodnoty, uvedené v tabulce 3.2, upravíme pomocí metod, které jsou popsané v sekcích 1.1 až 1.4.

Věkové kategorií	t-test p-hodnoty	Bonferroni	Šidák	Holm	BH
6-7	0.724	1.000	1.000	0.724	0.724
7-8	0.648	1.000	1.000	1.000	0.701
8-9	0.285	1.000	0.987	1.000	0.370
9-10	0.092	1.000	0.715	0.552	0.141
10-11	0.635	1.000	1.000	1.000	0.701
11-12	0.098	1.000	0.737	0.552	0.141
12-13	0.040	0.514	0.408	0.277	0.073
13-14	0.000	0.000	0.000	0.000	0.000
14-15	0.000	0.000	0.000	0.000	0.000
15-16	0.000	0.000	0.000	0.000	0.000
16-17	0.000	0.000	0.000	0.000	0.000
17-18	0.000	0.000	0.000	0.000	0.000
18-19	0.000	0.000	0.000	0.000	0.000

Tabulka 3.2: p-hodnoty dvouvýběrového t-testu, jejich upravené hodnoty pomoci Bonferroniho, Šidákovy, Holmove a Benjamini-Hochbergove metody

Z tabulky 3.2 je vidět, že dvouvýběrový t-test bez kontroly FWER prokázal statisticky významný rozdíl v rychlosti výskoků mezi chlapci a dívky od 12 let (věková kategorie 12-13), zatím co p-hodnoty upravené pomoci Bonferroniho, Šidákovy, Holmovy a Benjamini-Hochbergovy metody prokazují statisticky významný rozdíl v rychlosti výskoku mezi chlapci a dívkami od 13 let (věková skupina 13-14).

Dunnetova metoda

Nyní aplikujeme Dunnetovu metodu, která je popsaná v sekci 3.1, na naše data. V původním článku Hlávka a Hušková (2017) se uvádí, že statisticky významný rozdíl v rychlosti výskoků mezi chlapci a dívky pozorujeme v 11,5 let. Za kontrolní skupinu zvolíme rozdíl ve věkových kategorií od 6-7 až 10-11 pro chlapce a dívky a příslušná data shrneme pomocí vážených průměrů, které definujeme následujícím způsobem:

$$\overline{Y_0} = \overline{K_1} - \overline{K_2} = 0.0002642794,$$

kde $\overline{K_1} = \frac{\sum_{i=1}^{5}\overline{X_{1_i}}*n_{1_i}}{\sum_{i=1}^{5}n_{1_i}} = 2.037306$ vyjadřuje vážený průměr věkových kategorií 6-7 až 10-11 pro dívky a $\overline{K_2} = \frac{\sum_{i=1}^{5}\overline{X_{2_i}}*n_{2_i}}{\sum_{i=1}^{5}n_{2_i}} = 2.037041$ vyjadřuje vážený průměr věkových kategorií 6-7 až 10-11 pro chlapce.

Budeme testovat hypotézu $H_{0_i} : \overline{Y_i} = \overline{Y_0}$ proti alternativě $H_{1_i} : \overline{Y_i} > \overline{Y_0}$, $i = 6,..,13$.

Testová statistika je ve tvaru:

$$t_i = \frac{\overline{Y_0} - \overline{Y_i}}{s\sqrt{\frac{1}{n_{1_i}} + \frac{1}{n_{2_i}} + \frac{1}{\sum_{j=1}^{5} n_{1_j}} + \frac{1}{\sum_{j=1}^{5} n_{2_j}}}}, i = 6,..,13.$$

Výsledné p-hodnoty, spolu z p-hodnotami upravenými pomoci metod, které jsou popsané v sekcích 1.1 až 1.4, jsou uvedené v tabulce 3.3.

Věkové	Dunnet				
kategorií	p-hodnoty	Bonferroni	Šidák	Holm	BH
6-7					
7-8					
8-9					
9-10					
10-11					
11-12	0.127	1.000	0.662	0.127	0.127
12-13	0.062	0.496	0.401	0.124	0.071
13-14	0.000	0.000	0.000	0.000	0.000
14-15	0.000	0.000	0.000	0.000	0.000
15-16	0.000	0.000	0.000	0.000	0.000
16-17	0.000	0.000	0.000	0.000	0.000
17-18	0.000	0.000	0.000	0.000	0.000
18-19	0.000	0.000	0.000	0.000	0.000

Tabulka 3.3: p-hodnoty Dunnetova testu a p-hodnoty upravené pomocí Bonferroniho, Šidákovy, Holmovy a Benjamini-Hochbergovy metody

Z tabulky 3.3 je vidět, že Dunnetova metoda jak bez kontroly chyby 1. druhu, tak i p-hodnoty upravené pomocí Bonferroniho, Šidákovy, Holmovy a Benjamini-Hochbergovy metody prokázují statisticky významný rozdíl v rychlosti výskoků mezi chlapci a dívky od 13 let (věková kategorie 13-14).

Modifikace Dunnetovy metody

Nyní modifikujeme předchozí metodu tak, že budeme testovat hypotézu H_{0_i} : $\overline{Y_i} = 0$ proti alternativě $H_{1_i} : \overline{Y_i} > 0$, $i = 6,..,13$.

Testová statistika je ve tvaru:

$$t_i = \frac{\overline{Y_i}}{s\sqrt{\frac{1}{n_{1_i}} + \frac{1}{n_{2_i}}}}, i = 6,..,13,$$

kde

$$s^2 = \frac{\sum_{k=1}^{2}\sum_{i=6}^{13}\hat{\sigma}_{k_i}^2(n_{k_i} - 1)}{\sum_{k=1}^{2}\sum_{i=6}^{13} n_{k_i} - 16} = 0.196.$$

Výsledné p-hodnoty, spolu z p-hodnotami upravenými pomocí metod, které jsou popsané v sekcích 1.1 až 1.4, jsou uvedené v tabulce 3.4.

| Věkové | | Dunnet (modifikace) | | | |
kategorií	p-hodnoty	Bonferroni	Šidák	Holm	BH
6-7					
7-8					
8-9					
9-10					
10-11					
11-12	0.097	0.775	0.557	0.097	0.097
12-13	0.039	0.313	0.273	0.078	0.045
13-14	0.000	0.000	0.000	0.000	0.000
14-15	0.000	0.000	0.000	0.000	0.000
15-16	0.000	0.000	0.000	0.000	0.000
16-17	0.000	0.000	0.000	0.000	0.000
17-18	0.000	0.000	0.000	0.000	0.000
18-19	0.000	0.000	0.000	0.000	0.000

Tabulka 3.4: p-hodnoty modifikovaného Dunnetova testu a p-hodnoty upravené pomocí Bonferroniho, Šidákovy, Holmovy a Benjamini-Hochbergovy metody

Z tabulky 3.4 je vidět, že modifikovaná Dunnetova metoda bez kontroly chyby 1. druhu spolu z p-hodnotami upravenými pomocí Benjamini-Hochbergovy metody prokázují statisticky významný rozdíl v rychlosti výskoků mezi chlapci a dívkami od 12 let (věková kategorie 12-13), zatím co p-hodnoty upravené pomocí Bonferroniho, Šidákovy a Holmovy metody prokazují statisticky významný rozdíl v rychlosti výskoku mezi chlapci a dívkami od 13 let (věková skupina 13-14).

3.3 Porovnávání s více kontrolami

V odborné literature existují další vylepšení. Napřiklad, v článku Dunnett a Tamhane (1997) je popsána modifikace předchozí metody. Cílem je klasifikovat nové ošetření s ohledem na několik norem: specifikovat normy, kterým je nové ošetření nadřazené, ty, kterým je nové ošetření ekvivalentní a ty, pro které nelze prokazat ani nadřazenost ani rovnocennost. Důvodem současného testování nadřazenosti a rovnocennosti je skutečnost, že v mnoha případech si výzkumník přeje, aby nejprve zjistil, zda je nové ošetření lepší než standard, v takovém případě se stává možným kandidátem na nahrazení tohoto standardu. Za druhé, jestliže nadřazenost nemůže být stanovena, pokud nové ošetřeni má ekvivalentní účinnost jako norma, pak se stává možným kandidátem k použití jako alternativní metoda ošetření. Nemožnost prokázat buď převahu či rovnocennost nového ošetřeni naznačují, že je nelze doporučit pro použití, protože ve skutečnosti může mít nižší účinnost než standard.

Označme μ_i neznámou střední účinnost pro i-té ošetření ($i = 0,..,m$), kde 0 označuje nové ošetření.

Porovnání nového ošetření se standardní metodou je založeno na testování následujících dvojic hypotéz:

$$H_{0_i} : \mu_0 - \mu_i \leq 0 \text{ proti alternativě } H_{1_i} : \mu_0 - \mu_i > 0$$
$$\text{a}$$
$$H'_{0_i} : \mu_0 - \mu_i \leq -\delta \text{ proti alternativě } H'_{1_i} : \mu_0 - \mu_i > -\delta,$$

kde $\delta > 0$ označuje klinicky nevýznamný rozdíl v účinnosti. Zamítnutí hypotézy H_{0_i} znamená, že účinnost testovaného ošetření je lepší než účinnost i-tého standardu, zatímco nezamítnutí H_{0_i} společně se zamítnutím H'_{0_i} zjistí, že ošetření nemůže být horší než standard o více než o δ, a proto je z definice klinický ekvivalentní. Nezamítnutí obou nulových hypotéz H_{0_i} a H'_{0_i} znamená, že jsme neprokázali, že nová léčba je buď nadřazená, nebo ekvivalentní s i-tým standardem a proto ji nemůžeme doporučit jako náhradu.

Závěr

V dané práce jsme definovali mnohonásobné testování a rozebrali metody, které udržují pravděpodobnost chyby prvního druhu na úrovni α. Podroněji jsme se zabývali odvozením Šidákovy metody a předpokladů potřebných pro její použití. Hlavním cílem bylo seznámit čtenáře s vybranými metodami mnohonásobného porovnávání, pro nalezení rozdílu mezi dvojicemi středních hodnot. Dalším cílem bylo rozšířit přehled těchto metod, který je v mnoha statistických učebnicích omezen pouze na Scheffého a Tukeyovu metodu. Hlavní část práce se věnuje porovnání s kontrolou a jeji praktické implementace.

Seznam použité literatury

Anděl, J. (1998). *Statistické metody.* Druhé vydání. Matfyzpress, Praha. ISBN 80-85863-27-8.

Anděl, J. (2007). *Základy matematické statistiky.* Druhé vydání. Matfyzpress, Praha. ISBN 80-7378-001-1.

Benjamini, Y. a Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **59**(1), 289–300.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, **56**(293), 52–64.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, **6** (3), 241–252.

Dunnett, C. W. (1980). Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, **75**(372), 796–800.

Dunnett, C. W. a Tamhane, A. C. (1992). A step-up multiple test procedure. *Journal of the American Statistical Association*, **87**(417), 162–170.

Dunnett, C. W. a Tamhane, A. C. (1997). Multiple testing to establish superiority/equivalence of a new treatment compared with k standard treatments. *Statistics in Medicine*, **16**(21), 2489–2506.

Field, A. (2009). *Discovering statistics using SPSS.* Sage publications.

Hlávka, Z. a Hušková, M. (2017). Two-sample gradual change analysis. *REVSTAT–Statistical Journal*, **15**(3), 355–372.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, **62**(318), 626–633.

Kimball, A. W. (1951). On dependent tests of significance in the analysis of variance. *The Annals of Mathematical Statistics*, **22**(4), 600–602.

Stoline, M. R. (1981). The status of multiple comparisons: simultaneous estimation of all pairwise comparisons in one-way anova designs. *The American Statistician*, **35**(3), 134–141.