MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁRSKA PRÁCA

Kristína Turzová

Problémy mnohonásobného testovania

Katedra pravděpodobnosti a matematické štatistiky

Vedúci bakalárskej práce:	RNDr. Matúš Maciak, Ph.D.
Študijný program:	Matematika
Študijný odbor:	Obecná matematika

Praha 2019

Vyhlasujem, že som túto bakalársku prácu vypracovala samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Sb., autorského zákona v platnom znení, najmä skutočnosť, že Univerzita Karlova má právo na uzavretie licenčnej zmluvy o použití tejto práce ako školského diela podľa §60 odst. 1 autorského zákona.

V .... dňa .... Podpis autora

Poďakovanie ...

Názov práce: Problémy mnohonásobného testovania

Autor: Kristína Turzová

Katedra: Katedra pravděpodobnosti a matematické štatistiky

Vedúci bakalárskej práce: RNDr. Matúš Maciak, Ph.D., Katedra pravděpodobnosti a matematické štatistiky

Abstrakt: Testovanie štatistických hypotéz sa využíva pri práci s experimentálne získanými dátami. Táto práca sa zaoberá mnohonásobným testovaním, čo je simultánne testovanie väčšieho počtu hypotéz, a problémami s hladinou významnosti, ktoré pri ňom nastávajú. Objavujúce sa problémy sú popísané, pričom sú definované chyby FWER (Familywise Error Rate) a FDR (False Discovery Rate). Vybrané korekcie využívané pri mnohonásobnom testovaní sú podrobne predstavené a porovnávané pomocou simulácií z hľadiska hladiny a sily testu. Všetky korekcie kontrolujú definované chyby testovania.

Kľúčové slová: mnohonásobné testovanie, štatistický test, hladina testu, nulová hypotéza

Title: Multiple testing problems

Author: Kristína Turzová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Matúš Maciak, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Statistical hypothesis testing is used while analyzing experimental data. This thesis is focused on multiple testing, which means testing many hypotheses simultaneously, and multiple testing problems occurring while running multiple hypotheses tests. These multiple testing problems are described and two errors, FWER (Family-Wise Error Rate) and FDR (False Discovery Rate), are defined. Selected multiple testing corrections are introduced and compared in detail using simulations regarding significance level and power. All of the discussed corrections control for the problem of multiple testing.

Keywords: multiple testing, statistical test, significance level, null hypothesis

Obsah

Úvod

Testovanie hypotéz umožňuje posúdiť, či získané dáta z experimentu vyhovujú vopred určenému predpokladu. Pri testovaní hypotéz je dôležité správne určiť nulovú a alternatívnu hypotézu. Tieto dve hypotézy sú disjunktné tvrdenia, pričom chceme overiť, ktoré z nich platí. Testujeme vždy nulovú hypotézu, ktorú buď zamietneme v prospech alternatívy, alebo nezamietneme. V druhom prípade nevieme určiť, ktorá hypotéza je platná. Test však nemusí vždy rozhodnúť správne. Existujú dva typy chýb vyskytujúcich sa pri testovaní − zamietnutie platnej hypotézy, chyba 1. druhu, a nezamietnutie neplatnej hypotézy, chyba 2. druhu. Za závažnejšiu je považovaná chyba 1. druhu, preto chceme, aby pravdepodobnosť zamietnutia platnej hypotézy bola čo najmenšia. Táto pravdepobnosť sa nazýva hladina testu a je potrebné ju určiť na začiatku každého testovania. Okrem hladiny testu nás zaujíma taktiež sila testu, čo je pravdepodobnosť zamietnutia neplatnej hypotézy. Platí vzťah, že čím menšia je pravdepobnosť nezamietnutia neplatnej hypotézy, tým väčšia je sila testu.

Mnohonásobné testovanie je simultánne testovanie väčšieho počtu hypotéz. Dôvodom testovania viacerých hypotéz môže byť závislosť hypotéz, ktoré chceme testovať. V tomto prípade sa opäť snažíme zamietnuť čo najmenej platných hypotéz, avšak so zvyšujúcim sa počtom hypotéz sa zvyšuje pravdepodobnosť chyby 1. druhu. Aby bola dodržaná hladina testu, je potrebné upraviť hladinu jednotlivých testov, prípadne upraviť kritériá na zamietanie hypotéz.

Cieľom práce je porovnať vybrané korekcie mnohonásobného testovania, konkrétne Bonferroniho, Šidákovu, Holmovu, Simesovu, Hochbergovu a Benjamini-Hochbergovu korekciu. V práci budeme predpokladať čitateľovu znalosť základných pojmov z pravdepodobnosti a matematickej štatistiky. V prvej kapitole vysvetlíme niektoré pojmy z matematickej štatistiky, ktoré budú kľúčové v celej práci. Vysvetlíme, čo je mnohonásobné testovanie a podrobnejšie ukážeme problém s hladinou významnosti, ktorý pri ňom nastáva. Taktiež definujeme niektoré chyby, ktoré nastanú zamietnutím platných hypotéz. Je viacero spôsobov, ako sa dá na tieto chyby pozerať. My vyberáme tie z nich, ktoré sú pri mnohonásobnom testovaní kontrolované najčastejšie a ukážeme vzťah medzi nimi. Druhá kapitola sa zaoberá spomínanými korekciami, pričom pre každú z nich je vysvetlený postup zamietania hypotéz a overenie, či kontrolujú definované chyby. V tretej kapitole budeme porovnávať korekcie pomocou simulácií z hľadiska hladiny a sily testu.

1. Mnohonásobné testovanie

V tejto práci budeme predpokladať, že čitateľ pozná základné pojmy z matematickej štatistiky ako pravdepodobnostný a parametrický priestor, náhodný výber, model, testová štatistika a kritický obor. V prvej kapitole definujeme pojmy z matematickej štatistiky, ktoré pre nás budú v tejto práci klúčové, vysvetlíme ako funguje mnohonásobné testovanie a objasníme aké problémy nastávajú pri jeho používaní. Definujeme chyby, ktoré budeme kontrolovať pri mnohonásobnom testovaní.

1.1 Základné pojmy

V tejto podkapitole budeme predpokladať, že máme náhodný výber

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$$

z rozdelenia $F \in \mathcal{F}$, kde $\mathcal{F} = \{F(\theta),\ \theta \in \Theta\}$ je model. Teda Θ je množina všetkých možných hodnôt parametru θ v modele $\mathcal{F}$. Skutočný parameter budeme značiť θ_X. Označme Θ_H a Θ_A dve disjunktné podmnožiny parametrického priestoru Θ.

Definícia 1. *Množinu Θ_H nazývame nulová hypotéza a Θ_A alternatívna hypotéza. Množinu všetkých rozdelení z modelu $\mathcal{F}$, ktorých parametre splňajú nulovú hypotézu, budeme značiť $\mathcal{F}_H$, podobne množinu rozdelení s parametrami splĺňajúcimi alternatívnu hypotézu značíme $\mathcal{F}_A$.*

Pri testovaní skutočného parametru θ_X budeme zapisovať nulovú hypotézu $H : \theta_X \in \Theta_A$ proti alternatíve $A : \theta_X \in \Theta_A$.

Testovanie hypotéz sa vyhodnocuje pomocou štatistického testu, na jeho definovanie potrebujeme nasledujúce pojmy. Testová štatistika $\mathcal{S} : \mathbb{R}^n \mapsto \Theta$ je merateľná funkcia dát náhodného výberu $\mathbf{X}$, ktorú volíme tak, aby sme vedeli určiť jej presné alebo asymptotické rozdelenie. Následne podľa tohto rozdelenia určíme kritický obor $\mathcal{C}$, pomocou ktorého vyhodnotíme test.

Nech $\mathcal{S}$ je testová štatistika a $\mathcal{C}$ je kritický obor. Ak $\mathcal{S} \in \mathcal{C}$, zamietame nulovú hypotézu v prospech alternatívnej hypotézy. Ak $\mathcal{S} \notin \mathcal{C}$, nemôžeme zamietnuť nulovú hypotézu v prospech alternatívnej hypotézy.

Štatistický test nemusí v každom prípade rozhodnúť správne. Preto nastávajú štyri rôzne situácie vzhľadom k platnosti nulovej hypotézy a vyhodnotenia testu, ktoré môžeme prehľadne vidieť v Tabuľke **??**.

	Platí H	Neplatí H
Zamietame H	Chyba 1.druhu	OK
Nezamietame H	OK	Chyba 2.druhu

Tabuľka 1.1: Všetky možnosti, ktoré nastávajú pri testovaní hypotéz

V prípade zamietnutia platnej hypotézy hovoríme o chybe 1. druhu. Nezamietnutie neplatnej hypotézy sa nazýva chyba 2. druhu. Test volíme tak, aby chyba 1. druhu bola závažnejšia. Preto budeme kontrolovať jej pravdepodobnosť a budeme chcieť, aby bola čo najmenšia.

Na kontrolu chyby 1. druhu sa používa hladina testu. Niekedy budeme používať takisto výraz hladina významnosti.

Definícia 2. *Nech $\alpha \in (0,1)$ je vopred dané číslo, nech máme test s testovou štatistikou $\mathcal{S}$ a kritickým oborom $\mathcal{C}$. Hladina testu je rovná α, ak je splnená podmienka*

$$\sup P_H(\mathcal{S} \in \mathcal{C}) = \alpha,$$

kde P_H označuje pravdepodobnosť za predpokladu platnosti nulovej hypotézy H a supremum uvažujeme vzhľadom k všetkým nulovým hypotézam. V niektorých prípadoch je hladina testu dosiahnutá len asymptoticky, potom musí pre $n \longrightarrow \infty$ platiť upravená podmienka

$$\sup \lim_{n\to\infty} P_H(\mathcal{S} \in \mathcal{C}) = \alpha.$$

Hladina testu je pravdepodobnosť zamietnutia nulovej hypotézy, ktorá v skutočnosti platí. Je to teda pravdepodobnosť chyby 1. druhu. Pri testovaní vždy na začiatku určíme hladinu významnosti daného testu. V tejto práci budeme väčšinou voliť hladinu testu $\alpha = 0.05$.

Nech $\beta(A) = P_A(\mathcal{S} \in \mathcal{C})$, kde P_A označuje pravdepodobnosť za predpokladu platnosti alternatívnej hypotézy. Hodnotu $\beta(A)$ nazývame sila testu a je to pravdepodobnosť zamietnutia neplatnej hypotézy. Pri testovaní chceme, aby sila testu bola čo najvyššia. V štvrtej kapitole budeme porovnávať korekcie mnohonásobného testovania vzhľadom k ich sile.

1.2 Definícia mnohonásobného testovania

V tejto práci budeme predpokladať, že máme K hypotéz, ktoré chceme testovať, nulové hypotézy budeme značiť H_i a k nim príslušné testové štatistiky $\mathcal{S}^{(i)}$ a kritické obory $\mathcal{C}^{(i)}$, $i \in \{1, \ldots, K\}$. Množinu všetkých nulových hypotéz budeme značiť $\mathcal{H} = \{H_1, \ldots, H_K\}$. Všetky nulové hypotézy budeme testovať simultánne a nezávisle na sebe, tento typ testovania sa nazýva mnohonásobné testovanie. V celej práci budeme predpokladať, že testované hypotézy sú navzájom nezávislé.

Ako sme spomínali v predchádzajúcej podkapitole, pri testovaní hypotéz je potrebné kontrolovať hladinu testu, ktorú určíme na začiatku. Pre mnohonásobné testovanie zvolíme hladinu testu α a každú hypotézu H_i budeme testovať na tejto hladine. To znamená, že pravdepodobnosť zamietnutia platnej nulovej hypotézy H_i bude rovná α pre každé $i \in \{1, \ldots, K\}$

$$P_{H_i}\left(\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right) = \alpha.$$

Takisto vieme vyjadriť pravdepodobnosť nezamietnutia platnej hypotézy, teda pravdepodobnosť, že nespravíme chybu 1. druhu

$$P_{H_i}\left(\mathcal{S}^{(i)} \notin \mathcal{C}^{(i)}\right) = 1 - \alpha.$$

1.3 Problémy s hladinou testu

??? Pri mnohonásobnom testovaní nestačí kontrolovať hladinu jednotlivých testov. Namiesto toho budeme kontrolovať pravdepodobnosť zamietnutia aspoň jednej platnej hypotézy, teda pravdepodobnosť zjednotenia všetkých zamietnutí hypotéz za predpokladu platnosti všetkých nulových hypotéz. Túto pravdepodobnosť budeme označovať α_K a formálne to môžeme zapísať nasledovným spôsobom

$$P_{\mathcal{H}_0}\left(\bigcup_{i=1}^{K}\left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right) = \alpha_K,$$

kde $\mathcal{H}_0 = \bigcap_{i=1}^{K} H_i$ označuje skutočnosť, že všetky nulové hypotézy sú platné. Podobne ako sme opísali v predchádzajúcej podkapitole, môžeme vyjadriť pravdepodobnosť, že nespravíme chybu 1. druhu ani pri jednom testovaní hypotézy H_i, kde $i \in \{1, \ldots, K\}$

$$P_{\mathcal{H}_0}\left(\bigcap_{i=1}^{K}\left[\mathcal{S}^{(i)} \notin \mathcal{C}^{(i)}\right]\right) = \prod_{i=1}^{K} P_{H_i}\left(\mathcal{S}^{(i)} \notin \mathcal{C}^{(i)}\right) = (1-\alpha)^K.$$

Pravdepodobnosť zamietnutia aspoň jednej platnej hypotézy sa dá zapísať nasledujúcim spôsobom

$$P_{\mathcal{H}_0}\left(\exists i \in \{1, \ldots, K\} : \left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right) = 1 - P_{\mathcal{H}_0}\left(\bigcap_{i=1}^{K}[\mathcal{S}^{(i)} \notin \mathcal{C}^{(i)}]\right) = 1-(1-\alpha)^K.$$

Potom platí

$$\begin{aligned}\alpha_K = P_{\mathcal{H}_0}\left(\bigcup_{i=1}^{K}\left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right) &= P_{\mathcal{H}_0}\left(\exists i \in \{1, \ldots, K\} : \left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right) \\ &= 1-(1-\alpha)^K > \alpha.\end{aligned}$$

Teda α_K je väčšia ako α pre každé $K > 1$ a $\alpha \in (0{,}1)$. Z toho vyplýva, že pri testovaní viacerých hypotéz je pravdepodobnosť zamietnutia aspoň jednej platnej hypotézy väčšia ako hladina významnosti α, ktorú sme určili na začiatku. Aby bola splnená celková hladina testu α, je potrebné upraviť hladinu jednotlivých testov. Ukážeme to takisto pomocou obrázkov. Na testovanie hypotéz budeme používať p-hodnotu, ktorú teraz definujeme.

Definícia 3. *Nech $\mathcal{S}$ je testová štatistika, $\mathcal{C}$ je kritický obor a nech s je hodnota testovej štatistiky $\mathcal{S}$ spočítaná z napozorovaných dát, ktoré chceme testovať. P-hodnotu definujeme ako*

- $p = \sup P_H(\mathcal{S} \leq s)$, *ak* $\mathcal{C} = \langle c_U, \infty)$ *pre nejaké* $c_U \in \mathbb{R}$;
- $p = \sup P_H(\mathcal{S} \geq s)$, *ak* $\mathcal{C} = (-\infty, c_L\rangle$ *pre nejaké* $c_L \in \mathbb{R}$;
- $p = \sup 2\min\{P_H(\mathcal{S}) \leq s), P_F(\mathcal{S}) \geq s)\}$, *ak* $\mathcal{C} = (-\infty, c_L\rangle \cup \langle c_U, \infty)$ *pre nejaké* $c_L, c_U \in \mathbb{R}$, $c_L < c_U$ *a zároveň je splnená podmienka* $\sup P_H(\mathcal{S}) \leq c_L) = \sup P_H(\mathcal{S}) \geq c_U) = \frac{\alpha}{2}$.

Supremum uvažujeme vzhľadom k všetkým nulovým hypotézam.

P-hodnota sa niekedy nazýva dosiahnutá hladina testu. Podmienka na konci definície musí byť splnená, aby celková hladina testu bola rovná α. P-hodnota je takisto kľučový pojem na definovanie korekcií, ktoré uvedieme v nasledujúcej kapitole.

Ako sme zmienili vyššie, pomocou p-hodnoty vieme rozhodnúť, či máme hypotézu zamietnuť alebo nie. Budeme k tomu potrebovať nasledujúce tvrdenie, ktorého dôkaz viď (**?**, Tvrdenie 4.1).

Tvrdenie 1. *Nech $\mathcal{S}$ je testová štatistika so spojitým rozdelením a p je p-hodnota. Uvažujme test hypotézy H proti alternatíve A daný pravidlom*

$$\begin{aligned} H \text{ zamietame} &\iff p \leq \alpha, \\ H \text{ nezamietame} &\iff p > \alpha. \end{aligned}$$

Potom má tento test hladinu α.

Na začiatku ukážeme aká je skutočná hladina významnosti v prípade testovania jednej hypotézy. Všetky hypotézy budeme testovať na hladine $\alpha = 0.05$. V celej podkapitole budeme predpokladať, že máme dva nezávislé náhodné výbery

$$\begin{aligned} \mathbf{X}_1 &= (X_{11}, X_{12}, \dots, X_{1n})^T, \\ \mathbf{X}_2 &= (X_{21}, X_{22}, \dots, X_{2n})^T, \end{aligned}$$

z rozdelení $N(\mu_1, \sigma^2)$ a $N(\mu_2, \sigma^2)$, teda budeme pracovať s modelom

$$\mathcal{F} = \{N(\mu, \sigma^2),\ \mu \in \mathbb{R},\ \sigma^2 > 0\}.$$

Rozsah náhodných výberov bude $n = 100$. Oba náhodné výbery vygenerujeme z normovaného normálneho rozdelenia.

Najskôr budeme testovať hypotézu H_1, ktorá bude tvaru

$$H_1 : \mu_1 = 0.$$

Presným jednovýberovým t-testom budeme testovať vygenerované dáta. Tento postup budeme opakovať niekoľko krát po sebe, pričom počet testovaní budeme meniť, postupne $100, 1000, 10000, 100000, 1000000$. Potrebujeme zistiť počet zamietnutých platných hypotéz. Všetky hypotézy sú platné a podiel zamietnutých hypotéz k počtu testovaní bude rovný skutočnej hladine testu.

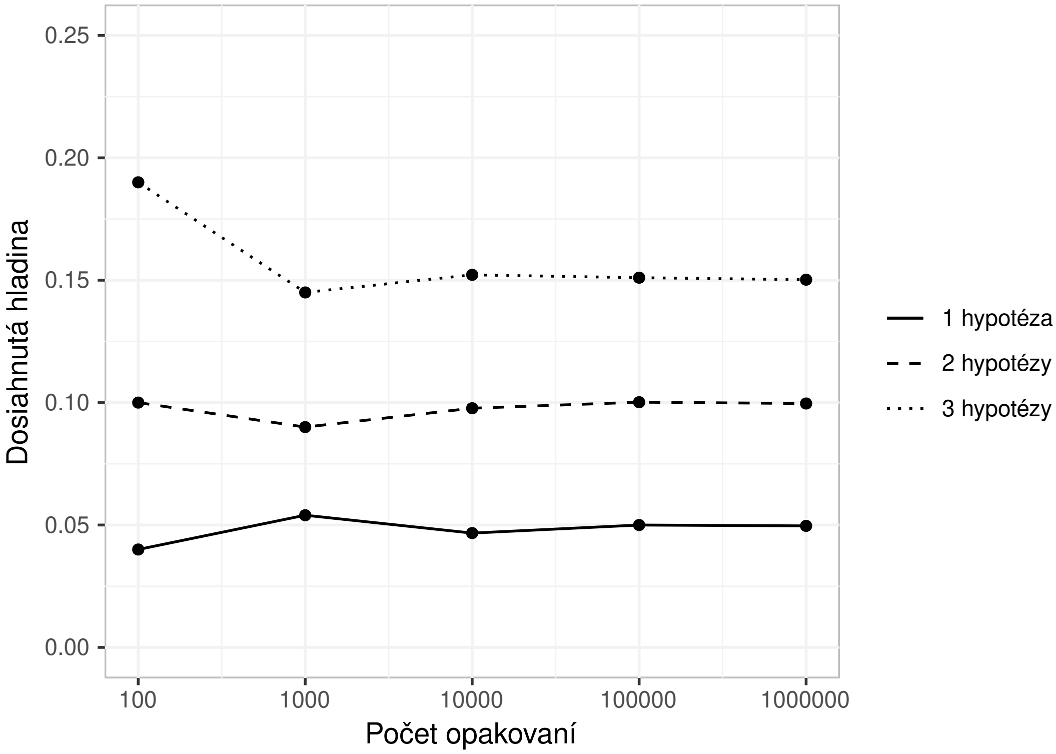
V druhom prípade budeme zároveň testovať dve hypotézy, jedna bude testovať strednú hodnotu a druhá rozptyl náhodného výberu $\mathbf{X}_1$. Hypotézy budú tvaru

$$\begin{aligned} H_1 &: \mu_1 = 0; \\ H_2 &: \sigma^2 = 1. \end{aligned}$$

V treťom prípade budeme testovať naviac jednu hypotézu, rovnosť stredných hodnôt náhodných výberov $\mathbf{X}_1$, $\mathbf{X}_2$. Hypotézy budú mať tvar

$$\begin{aligned} H_1 &: \mu_1 = 0; \\ H_2 &: \sigma^2 = 1; \\ H_3 &: \mu_1 = \mu_2. \end{aligned}$$

Mnohonásobné testovanie zopakujeme niekoľko krát po sebe pre rôzne nagenerované dáta, počet opakovaní bude rovnaký ako predtým. Dáta boli vygenerované tak, aby boli hypotézy platné. Strednú hodnotu budeme testovať presným jednovýberovým t-testom a rozptyl pomocou jednovýberového χ^2-testu. Rovnosť stredných hodnôt otestujeme presným dvojvýberovým t-testom, je splnený predpoklad rovnosti rozptylov.



Obr. 1.1: Skutočná hladina významnosti pri testovaní rôzneho počtu hypotéz pri počtoch opakovaní $100, 1000, 10000, 100000, 1000000$

Na Obrázku **??** je porovnanie skutočnej hladiny významnosti pri testovaní jednej hypotézy a pri mnohonásobnom testovaní dvoch alebo troch hypotéz. Už pri testovaní malého počtu hypotéz je skutočná hladina významnosti výrazne vyššia ako hladina, ktorú sme určili na začiatku.

1.4 Definície chýb FWER a FDR

Ako sme ukázali v predchádzajúcej podkapitole, pri mnohonásobnom testovaní je potrebné kontrolovať chyby, ktoré sú prísnejšie ako je chyba 1. druhu. V tejto podkapitole definujeme Familywise Error Rate a False Discovery Rate, budeme ich kontrolovať v korekciách, ktoré uvedieme v tejto práci.

Nech $\mathcal{H}'$ je podmnožina $\mathcal{H}$. Skutočnosť, že platia všetky hypotézy v $\mathcal{H}'$ označíme $\mathcal{H}'_0$.

Definícia 4. *Nech máme K hypotéz $H_1, \ldots, H_K$ a nech $\mathcal{S}^{(i)}$ sú ich testové štatistiky a $\mathcal{C}^{(i)}$ kritické obory, $i \in \{1, \ldots, K\}$. Familywise Error Rate definujeme ako pravdepodobnosť zamietnutia aspoň jednej platnej hypotézy za predpokladu, že* *platia hypotézy z ktorejkoľvek podmnožiny $\mathcal{H}' \subseteq \mathcal{H} = \{H_1, \ldots, H_K\}$. Budeme ju značiť* FWER. *Teda platí*

$$\text{FWER} = P_{\mathcal{H}_0'}\left(\exists H_i \in \mathcal{H}_0' : \left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right).$$

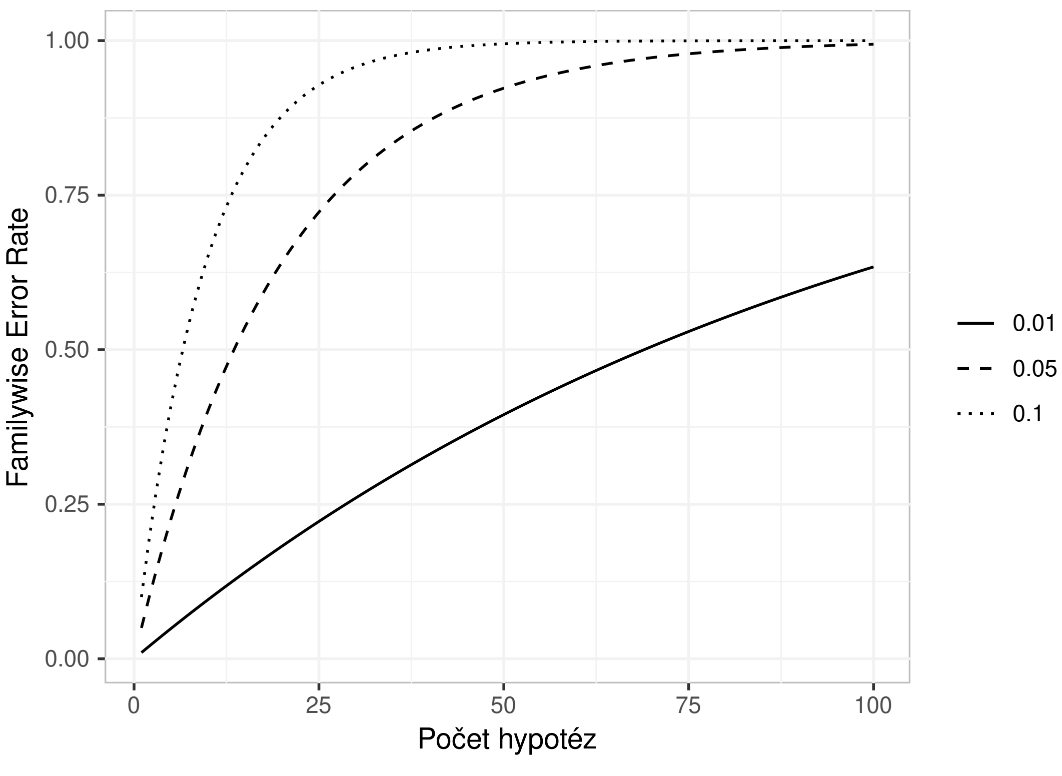
Chyba FWER sa dá kontrolovať dvomi rôznymi spôsobmi, v Definícii **??** sme uviedli silnú kontrolu chyby FWER. Pri druhom spôsobe predpokladáme platnosť všetkých nulových hypotéz, ktoré testujeme, ako sme uviedli v predcházajúcej podkapitole

$$\text{FWER} = P_{\mathcal{H}_0}\left(\exists i \in \{1, \ldots, K\} : \left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right).$$

Nazýva sa to slabá kontrola chyby FWER. V tomto prípade máme zaručenú FWER na hladine α len za platnosti všetkých hypotéz. Preto je výhodnejšie kontrolovať silnú kontrolu chyby FWER.

Pri mnohonásobnom testovaní budeme väčšinou kontrolovať Familywise Error Rate a budeme chcieť, aby bola táto chyba menšia ako zvolená hladina významnosti. V tejto práci budeme pracovať so silnou kontrolou chyby FWER, pokiaľ explicitne neuvedieme inak.

Problém s mnohonásobným testovaním ukážeme na príklade s konkrétnymi hodnotami. Nech $\alpha = 0.05$ a počet hypotéz $K = 10$. Jednotlivé hypotézy budeme testovať na hladine α. Pravdepodobnosť, že spravíme minimálne jednu chybu pri mnohonásobnom testovaní, bude $\text{FWER} = 1 - (1 - 0.05)^{10} = 0.4012631$.



Obr. 1.2: Familywise Rrror Rate pri rôznych počtoch hypotéz s rôznymi hodnotami α

Na Obrázku **??** vidíme, že už pri nižšom počte hypotéz je veľkosť FWER výrazne vyššia ako hladina významnosti, ktorú sme chceli dodržať.

Aby bola táto chyba menšia alebo rovná ako α, ktoré sme zvolili na začiatku, je nutné testovať jednotlivé hypotézy s menšou hladinou významnosti.

V Tabuľke **??** môžeme vidieť všetky možnosti, ktoré môžu nastať pri mnohonásobnom testovaní, pričom D, E, F, G označujú počty hypotéz H_i v každej možnosti pre $i \in \{1, \ldots, K\}$, kde K je ich súčet. K_0 označuje počet platných hypotéz a Z počet zamietnutých hypotéz. Nás bude najviac zaujímať veľkosť D, pretože ide o počet zamietnutých platných hypotéz.

	Platné H_i	Neplatné H_i	Celkom
Zamietnuté H_i	D	E	Z
Nezamietnuté H_i	F	G	K-Z
Celkom	K_0	K-K_0	K

Tabuľka 1.2: Počet hypotéz v každej možnosti, ktorá nastáva

Familywise Error Rate sa dá zapísať aj pomocou tohto značenia, platí

$$\mathrm{FWER} = P(D \geq 1).$$

Pri niektorých korekciách sa často kontroluje chyba FDR, ktorá nie je taká striktná ako FWER.

Definícia 5. *False Discovery Rate definujeme ako strednú hodnotu z podielu zamietnutých platných hypotéz k zamietnutým hypotézam za predpokladu zamietnutia aspoň jednej hypotézy. Budeme ju značiť* FDR. *Platí*

$$\mathrm{FDR} = E\left(\frac{D}{Z} \;\middle|\; Z > 0\right) P(Z > 0).$$

Ak sú všetky nulové hypotézy pravdivé, FWER a FDR sa rovnajú, dôkaz nájdeme v článku **?**. V ostatných prípadoch platí FDR $\leq$ FWER. Pokiaľ kontrolujeme Familywise Error Rate, kontrolujeme zároveň aj False Discovery Rate. Opačná implikácia neplatí.

2. Korekcie

V predchádzajúcej kapitole sme opísali problém, ktorý nastáva pri mnohonásobnom testovaní. Problém s hladinou významnosti sa dá vyriešiť použitím rôznych korekcií, ktoré korigujú celkovú hladinu významnosti jednotlivých testov.

Nech $p_1, \dots, p_K$ sú postupne p-hodnoty testov nulových hypotéz $H_1, \dots, H_K$. P-hodnoty zoradíme podľa veľkosti a následne označíme v poradí

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(K)}.$$

Nulové hypotézy patriace k zoradeným p-hodnotám označíme

$$H_{(1)}, H_{(2)}, \dots, H_{(K)}.$$

Tento zápis budeme potrebovať na definovanie niektorých korekcií.

Korekcie, ktoré upravujú hladinu testov delíme na dve skupiny:

- simultánne zamietajúce, hypotézy sa zamietajú nezávisle na sebe,
- postupne zamietajúce, hypotézy zamietame pomocou zoradených p-hodnôt.

Korekcie rozlišujeme taktiež podľa toho, ktorú chybu kontrolujú. Väčšina korekcií kontroluje silnejšiu verziu chyby FWER, ktorá zároveň kontroluje aj chybu FDR, ako sme uviedli v predchádzajúcej kapitole. Existujú však aj korekcie kontrolujúce len chybu FDR, teda nie sú až také striktné.

Väčšina korekcií upravujúcich hladinu testu je založená na Booleovej nerovnosti, ktorú uvedieme aj s dôkazom.

Tvrdenie 2. *Nech $B_1, \dots, B_n$ sú náhodné javy, $n \in \mathbb{N}$. Potom platí*

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

Dôkaz. Nerovnosť dokážeme indukciou. Pre $n = 1$ nerovnosť platí. Budeme predpokladať, že nerovnosť platí pre n, teda

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

Dokážeme, že nerovnosť platí aj pre $n+1$, použijeme k tomu rovnosť

$$P\left(\bigcup_{i=1}^{n+1} B_i\right) = P\left(\bigcup_{i=1}^{n} B_i\right) + P(B_{n+1}) - P\left(\bigcup_{i=1}^{n} B_i \cap B_{n+1}\right).$$

Keďže

$$P\left(\bigcup_{i=1}^{n} B_i \cap B_{n+1}\right) \geq 0,$$

platí

$$P\left(\bigcup_{i=1}^{n+1} B_i\right) \leq P\left(\bigcup_{i=1}^{n} B_i\right) + P(B_{n+1}) \leq \sum_{i=1}^{n} P\left(B_i\right) + P(B_{n+1}) = \sum_{i=1}^{n+1} P\left(B_i\right).$$

Druhá nerovnosť vyplýva z indukčného predpokladu.

□

2.1 Bonferroni

Najznámejšou korekciou pri úpravách mnohonásobného testovania je Bonferroniho korekcia. Je založená na Booleovej nerovnosti, ktorú sme ukázali na začiatku kapitoly.

Podľa Tvrdenia **??** zamietame nulovú hypotézu práve vtedy, keď je p-hodnota menšia alebo rovná ako hladina testu. Toto tvrdenie použijeme aj pri odvodení tejto korekcie. Nech $p_1, \ldots, p_K$ sú postupne p-hodnoty hypotéz $H_1, \ldots, H_K$. Bonferroniho korekcia zamieta nulovú hypotézu práve vtedy, keď $p_i \leq \frac{\alpha}{K}$ pre každé $i \in \{1, \ldots, K\}$. Chceme, aby veľkosť FWER bola menšia alebo rovná ako hladina testu α, ktorú určíme na začiatku. Nech $\mathcal{H'}_0$ je množina platných hypotéz, ich počet je K'. Nech I je indexová množina platných hypotéz, teda $I = \{i \in \{1, \ldots, K\}; H_i \in \mathcal{H}'_0\}$. Potom platí

$$\mathrm{FWER} = P_{\mathcal{H}'_0}\left(\exists H_i \in \mathcal{H}'_0 : \left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right) = P_{\mathcal{H'}_0}\left(\bigcup_{i \in I}\left[\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right]\right)$$

$$\leq \sum_{i \in I} P_{H^i}\left(\mathcal{S}^{(i)} \in \mathcal{C}^{(i)}\right) = K'\alpha_0 \leq K\alpha_0 = K\frac{\alpha}{K} = \alpha,$$

kde $\alpha_0 = \frac{\alpha}{K}$ je hladina jednotlivých testov.

Už pri testovaní 10 hypotéz by sme museli použiť na jednotlivé testy hladinu $\alpha_0 = 0.005$. To je veľmi nízka hladina testu a niektoré neplatné hypotézy by nemuseli byť zamietnuté, teda sila testu je taktiež nižšia. Z odvodenia korekcie vidíme, že kontroluje chybu FWER v silnejšom zmysle.

2.2 Šidák

Šidákova korekcia kontroluje taktiež chybu FWER v silnejšom zmysle. Patrí medzi korekcie, ktoré testujú hypotézy simultánne. Jej odvodenie vyplýva z alternatívneho zápisu FWER, ktorý sme uviedli pred Definíciou **??**,

$$\mathrm{FWER} = 1 - (1 - \alpha_0)^K,$$

kde K je počet všetkých hypotéz a α_0 hladina jednotlivých testov. Keď rovnosť prepíšeme iným spôsobom, dostaneme

$$\alpha_0 = 1 - (1 - \mathrm{FWER})^{\frac{1}{K}}.$$

Ak chceme, aby FWER bola menšia alebo rovná ako α, zvolíme

$$\alpha_0 = 1 - (1 - \alpha)^{\frac{1}{K}}.$$

Podrobné odvodenie korekcie a jej dôkaz nájdeme v článku **?**.

Z Tabuľky **??** vidíme, že Šidákova korekcia má v porovnaní s Bonferroniho korekciou väčšie hladiny jednotlivých testov, to znamená, že aj sila testu bude vyššia. Pri vyšších počtoch hypotéz sú upravené hladiny týchto dvoch korekcií podobné. Pre $\alpha = 0.01$ nie je vzhľadom k zaokrúhleniu vidieť rozdiel medzi korekciami.

	BONF	**ŠIDÁK**	**BONF**	**ŠIDÁK**	**BONF**	**ŠIDÁK**
α	0.01		0.05		0.1	
1	0.0100	0.0100	0.0500	0.0500	0.1000	0.1000
2	0.0050	0.0050	0.0250	0.0253	0.0500	0.0513
3	0.0033	0.0033	0.0167	0.0169	0.0333	0.0345
4	0.0025	0.0025	0.0125	0.0127	0.0250	0.0260
5	0.0020	0.0020	0.0100	0.0102	0.0200	0.0209
6	0.0017	0.0017	0.0083	0.0085	0.0167	0.0174
7	0.0014	0.0014	0.0071	0.0073	0.0143	0.0149
8	0.0013	0.0013	0.0063	0.0064	0.0125	0.0131
9	0.0011	0.0011	0.0056	0.0057	0.0111	0.0116
10	0.0010	0.0010	0.0050	0.0051	0.0100	0.0105

Tabuľka 2.1: Veľkosti α_0 pri použití Bonferroniho a Šidákovej korekcie pre rôzne zvolené hladiny α

2.3 Holm

Holmova korekcia je opäť založená na Booleovej nerovnosti. Patrí medzi korekcie s postupným zamietaním hypotéz.

Na definovanie korekcie budeme potrebovať zoradené p-hodnoty $p_{(i)}$ a k nim príslušné hypotézy $H_{(i)}$, ako sme zaviedli na začiatku kapitoly. Odvodenie korekcie nájdeme v článku **?**.
Procedúra :

- Ak $p_{(1)} \leq \frac{\alpha}{K}$, zamietame hypotézu $H_{(1)}$ a pokračujeme ďalším krokom. V opačnom prípade nezamietame hypotézy $H_{(1)}, H_{(2)}, \dots, H_{(K)}$. Zo zoradených p-hodnôt vyplýva $\frac{\alpha}{K} < p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(K)}$, podľa Bonferroniho korekcie nezamietame žiadnu hypotézu.
- Ak $p_{(i)} \leq \frac{\alpha}{K-i+1}$, zamietame hypotézu $H_{(i)}$ a pokračujeme ďalším krokom. V opačnom prípade nezamietame hypotézy $H_{(i)}, H_{(i+1)}, \dots, H_{(K)}$.
- Ak $p_{(K)} \leq \alpha$, zamietame $H_{(K)}$ a procedúra končí. V opačnom prípade nezamietame hypotézu $H_{(K)}$.

Tvrdenie 3. *Holmova korekcia kontroluje, aby chyba* FWER *bola menšia alebo rovná ako α.*

Dôkaz. Nech $\mathcal{H}_0' \subseteq \mathcal{H}$ je množina pravdivých hypotéz, nech I je jej množina indexov, K' je počet prvkov množiny I. Z Booleovej nerovnosti vyplýva nasledujúca nerovnosť

$$\text{FWER} = P_{\mathcal{H}_0'}\left(\exists i \in I : p_i \leq \frac{\alpha}{K'}\right) = P_{\mathcal{H}_0'}\left(\bigcup_{i\in I}\left[p_i \leq \frac{\alpha}{K'}\right]\right)$$

$$\leq \sum_{i\in I} P_{H_i}\left(p_i \leq \frac{\alpha}{K'}\right) = K'\frac{\alpha}{K'} = \alpha.$$

Pokiaľ platí

$$\forall i \in I : p_i > \frac{\alpha}{K'},$$

existuje K' hypotéz, ktoré sú väčšie ako $\frac{\alpha}{K'}$ a určite platí

$$p_{(K)} > \cdots > p_{(K-K'+1)} > \frac{\alpha}{K'}.$$

Procedúra sa zastaví najneskôr v kroku $K - K' + 1$. To implikuje, že množina všetkých hypotéz s p-hodnotou menšou ako $\frac{\alpha}{K'}$ obsahuje aj množinu platných hypotéz.

□

Z dôkazu Tvrdenia **??** vyplýva, že Holmova korekcia kontroluje chybu FWER v silnejšom zmysle.

2.4 Simes

Na definíciu Simesovej korekcie budeme opäť používať zápis zoradených p-hodnôt a k nim príslušných hypotéz. Túto korekciu zaraďujeme medzi simultánne zamietajúce.
Procedúra :

- Ak $p_{(i)} \leq \frac{i\alpha}{K}$ pre každé $i \in \{1, \ldots, K\}$, zamietame $\mathcal{H}$, kde $\mathcal{H}$ označuje množinu všetkých nulových hypotéz.

Tvrdenie 4. *Nech $p_{(1)}, \ldots, p_{(K)}$ sú zoradené štatistiky z K nezávislých náhodných veličín z rovnomerného rozdelenia na* (0,1) *a nech* $\alpha \in (0,1)$. *Potom platí*

$$P\left(p_{(i)} > \frac{i\alpha}{K};\ i \in \{1, \ldots, K\}\right) = 1 - \alpha.$$

Dôkaz Tvrdenia **??** sa nachádza v článku **?**. Táto korekcia kontroluje FWER len v slabšom zmysle, teda zaručuje FWER $\leq \alpha$ iba v prípade, že platia všetky nulové hypotézy.

2.5 Hochberg

V článku **?** je porovnaná Holmova a Simesova korekcia. Keďže Simesova korekcia kontroluje chybu FWER len v slabšom zmysle, Hochberg vytvoril korekciu odvodenú z týchto dvoch korekcií, ktorá bude kontrolovať silnú verziu chyby FWER. Táto korekcia má podobný zápis ako Holmova korekcia, avšak procedúra začína testovaním hypotézy s najväčšou p-hodnotou.
Procedúra :

- Ak $p_{(K)} \leq \alpha$, zamietame hypotézy $H_{(K)}, H_{(K-1)}, \ldots, H_{(1)}$. V opačnom prípade nezamietame hypotézu $H_{(K)}$ a pokračujeme ďalším krokom.

- Ak $p_{(i)} \leq \frac{\alpha}{K-i+1}$, zamietame hypotézy $H_{(K-i+1)}, H_{(K-i)}, \ldots, H_{(1)}$, V opačnom prípade nezamietame hypotézu $H_{(K-i+1)}$ a pokračujeme ďalším krokom.

- Ak $p_{(1)} \leq \frac{\alpha}{K}$, zamietame $H_{(1)}$. V opačnom prípade nezamietame hypotézu $H_{(1)}$.

Tvrdenie 5. *Nech $j \in \{1, \ldots, K\}$. Ak pre každé $i \in \{1, \ldots, j\}$ platí $p_{(i)} \leq \frac{\alpha}{K-i+1}$, tak Simesova korekcia zamieta všetky $H_{(i)}$ pre $i \leq j$.*

Dôkaz. Podľa predpokladu pre každé $i \in \{1, \ldots, j\}$ platí $p_{(i)} \leq \frac{\alpha}{K-i+1}$. Označme $\mathcal{H}' \subseteq \mathcal{H}$ množinu hypotéz, ktoré spĺňajú túto nerovnosť, $\mathcal{H}' = \{H_{(1)}, \ldots, H_{(j)}\}$. Pre každé $i \leq j$ platí nerovnosť $\frac{\alpha}{K-i+1} \leq \frac{i\alpha}{j}$, potom

$$p_{(i)} \leq \frac{\alpha}{K-i+1} \leq \frac{i\alpha}{j}.$$

Je splnená podmienka Simesovej korekcie a zamietame hypotézy $H_{(1)}, \ldots, H_{(j)}$. □

2.6 Benjamini-Hochberg

Podľa článku **?** nie je vždy potrebné kontrolovať chybu FWER, pretože sa pozeráme len na to, či nastala chyba alebo nie. Ich korekcia bude kontrolovať chybu FDR, konkrétne strednú hodnotu podielu neplatných zamietnutých hypotéz k všetkým zamietnutým hypotézam.

Táto korekcia nie je výhodnejšia, pokiaľ sú všetky nulové hypotézy pravdivé, pretože v tomto prípade sa chyby FWER a FDR rovnajú.

Budeme pracovať so zoradenými p-hodnotami ako pri predchádzajúcich korekciách.
Procedúra :

- Nech j je najväčšie i, pre ktoré platí $p_{(i)} \leq \frac{i\alpha}{K}$. Potom zamietame všetky $H_(i)$ pre $i \in \{1, \ldots, j\}$.

Tvrdenie 6. *Pre nezávislé testové štatistiky a pre akékoľvek zostavenie nepravdivých nulových hypotéz kontroluje Benjamini-Hochbergova korekcia chybu* FDR *na hladine α.*

Dôkaz. Máme K nezávislých p-hodnôt, ktoré patria k nulovým hypotézam. Nech K_0 je počet p-hodnôt príslušných k pravdivým hypotézam a $K_1 = K - K_0$ počet p-hodnôt príslušných k nepravdivým hypotézam. P-hodnoty príslušné k platným hypotézam označíme $p^{(1)}, \ldots, p^{(K_0)}$ a p-hodnoty príslušné k neplatným hypotézam $p^{(K_0+1)}, \ldots, p^{(K)}$. Použijeme značenie ako v Tabuľke **??**, pripomeňme, že FDR $= E\left(\frac{D}{Z}\right)$. Benjamini-Hochbergerova korekcia spĺňa

$$E\left(\frac{D}{Z} \,\middle|\, p_{K_0+1} = p^{(1)}, \ldots, p_K = p^{(K_1)}\right) \leq \frac{K_0}{K}\alpha \leq \alpha.$$

Tvrdenie vyplýva z nerovnosti, jej dôkaz nájdeme v článku **?**.

□

3. Porovnanie korekcií

V tejto kapitole porovnáme korekcie, ktoré sme opísali v predchádzajúcej kapitole. Korekcie porovnáme z dvoch rôznych hladísk. V prvej podkapitole sa budeme pozerať na hladinu významnosti jednotlivých korekcií. Neskôr porovnáme korekcie vzhľadom k ich sile. Na konci kapitoly ukážeme, že korekcie nemajú vplyv na konzistenciu testu.

3.1 Z hľadiska hladiny testu

Predpokladajme, že máme K náhodných výberov s rozsahom $n = 100$.

$$\begin{gathered}\mathbf{X}_1 = (X_{11}, X_{12}, \ldots, X_{1n})\\ \mathbf{X}_2 = (X_{21}, X_{22}, \ldots, X_{1n})\\ \vdots\\ \mathbf{X}_K = (X_{K1}, X_{K2}, \ldots, X_{Kn})\end{gathered}$$

Strednú hodnotu náhodného výberu $\mathbf{X}_i$ označíme μ_i, $i \in \{1, \ldots, K\}$. Budeme testovať K hypotéz v rovnakom čase, ktoré budú tvaru

$$\begin{gathered}H_1 : \mu_1 = 0,\\ \vdots\\ H_K : \mu_K = 0.\end{gathered}$$

Aby sme zistili hladinu významnosti mnohonásobného testovania, dáta vygenerujeme z normovaného normálneho rozdelenia. Teda všetky hypotézy sú platné. Mnohonásobné testovanie prevedieme opakovane 100000-krát na rôzne vygenerovaných dátach, pričom chceme, aby bola dodržaná hladina významnosti $\alpha = 0.05$. Pri každom opakovaní budeme zisťovať koľko hypotéz bolo zamietnutých, pričom počet hypotéz budeme meniť. Následne spočítame koľkokrát z 100000 opakovaní sme zamietli určité počty hypotéz. V práci uvedieme podiel tohto počtu opakovaní k všetkým opakovaniam. V prvom prípade budeme testovať 3 hypotézy.

	BONF	**ŠIDÁK**	**HOLM**	**SIMES**	**HOCH**	**BH**	**-**
1	0.0475	0.0483	0.0468	0.0000	0.0469	0.0458	0.1360
2	0.0009	0.0009	0.0017	0.0000	0.0015	0.0034	0.0073
3	0.0000	0.0000	0.0001	0.0493	0.0000	0.0001	0.0001
$\sum$	0.0484	0.0492	0.0486	0.0493	0.0484	0.0493	0.1434

Tabuľka 3.1: Podiel počtu zamietnutých hypotéz v každom opakovaní k počtu opakovaní, pri testovaní 3 hypotéz rôznymi korekciami s počtom opakovaní 100000

V Tabuľkách **??**, **??**, **??** vidíme podiel počtu zamietnutých hypotéz v každom opakovaní k počtu opakovaní pri testovaní 3, 5, 10 hypotéz. V poslednom riadku je podiel počtu chýb, ktoré sme spravili pri 100000 opakovaniach k počtu všetkých opakovaní. Tým sme získali celkovú hladinu významnosti jednotlivých korekcií. Vidíme, že každá korekcia dodržuje hladinu významnosti $\alpha = 0.05$. Simesova

	BONF	ŠIDÁK	HOLM	SIMES	HOCH	BH	-
1	0.0476	0.0485	0.0472	0.0000	0.0472	0.0457	0.2051
2	0.0009	0.0010	0.0014	0.0000	0.0013	0.0038	0.0213
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0012
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	0.0000	0.0497	0.0000	0.0000	0.0000
$\sum$	0.0485	0.0495	0.0486	0.0497	0.0485	0.0497	0.2276

Tabuľka 3.2: Podiel počtu zamietnutých hypotéz v každom opakovaní k počtu opakovaní, pri testovaní 5 hypotéz rôznymi korekciami s počtom opakovaní 100000

	BONF	ŠIDÁK	HOLM	SIMES	HOCH	BH	-
1	0.0478	0.0486	0.0475	0.0000	0.0475	0.0458	0.3140
2	0.0008	0.0009	0.0011	0.0000	0.0011	0.0306	0.0749
3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0031	0.0103
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0497	0.0000	0.0000	0.0000
$\sum$	0.0486	0.0495	0.0486	0.0497	0.0486	0.0497	0.4002

Tabuľka 3.3: Podiel počtu zamietnutých hypotéz v každom opakovaní k počtu opakovaní, pri testovaní 10 hypotéz rôznymi korekciami s počtom opakovaní 100000

korekcia kontroluje len slabšiu verziu chyby FWER, teda pri platnosti všetkých hypotéz je dodržaná hladina testu. Podľa hodnôt v tabuľkách sa dá spozorovať, že táto korekcia funguje iným spôsobom ako ostatné. Za splnenia určitých podmienok zamieta všetky hypotézy, v opačnom prípade nezamietne žiadnu. Benjamini-Hochbergova korekcia kontroluje chybu FDR, ktorá nie je až taká striktná. V porovnaní s inými korekciami zamieta častejšie väčší počet hypotéz ako 1-2. Ostatné korekcie kontrolujú silnejšiu verziu chyby FWER, skutočná hladina týchto korekcií je nižšia ako hladina Simesovej alebo Benjamini-Hochbergovej korekcie. Teda ich sila by mala byť nižšia. Holmova a Hochbergova korekcia majú podobné výsledky vzhľadom k testovaniu každého počtu hypotéz, ktoré sme uviedli. Takisto si môžeme všimnúť, že počet chýb pri použití Simesovej a Benjamini-Hochbergovej korekcie je rovnaký.

3.2 Z hľadiska sily testu

V tejto podkapitole budeme porovnávať silu korekcií za rôznych podmienok. Sila testu závisí na zvolenej alternatíve, na počte nulových hypotéz a v našom prípade aj na počte hypotéz, ktoré testujeme. Aby sme zistili silu testov, musíme vygenerovať dáta tak, aby neboli všetky nulové hypotézy platné. Z tohto porovnania vylúčime Simesovu korekciu, ktorá kontroluje chybu FWER len v slabšom zmysle, teda nie je zaručená hladina významnosti $\alpha = 0.05$.

Rozoberieme viacero možností, budeme meniť pomer platných a neplatných hypotéz, pričom neplatné hypotézy budeme generovať s rôznymi strednými hodnotami. Silu testov porovnáme pre rôzne počty hypotéz, ktoré budeme testovať.

Budeme predpokladať, že máme K nezávislých náhodných výberov so strednými hodnotami μ_i s rozsahom $n = 100$, kde $i \in \{1, \dots, K\}$. Dáta vygenerujeme z normovaného normálneho rozdelenia. Budeme uvažovať niekoľko prípadov. Počet náhodných výberov K bude postupne 4, 8, 16, 32, 64, pričom budeme testovať K hypotéz. Pomer neplatných hypotéz ku všetkým hypotézam budeme meniť, rozdelíme ich na 4 rôzne prípady, počet nepatných hypotéz bude postupne K, $\frac{3}{4}K$, $\frac{1}{2}K$, $\frac{1}{4}K$. Ako sme spomenuli, sila testu často závisí na rozdiele nulovej hypotézy a testovaných dát. Preto budeme uvažovať 3 možnosti vo voľbe stredných hodnôt neplatných hypotéz.

Budeme testovať strednú hodnotu náhodných výberov, testujeme K hypotéz tvaru

$$\begin{gathered} H_1 : \mu_1 = 0, \\ \vdots \\ H_K : \mu_K = 0. \end{gathered}$$

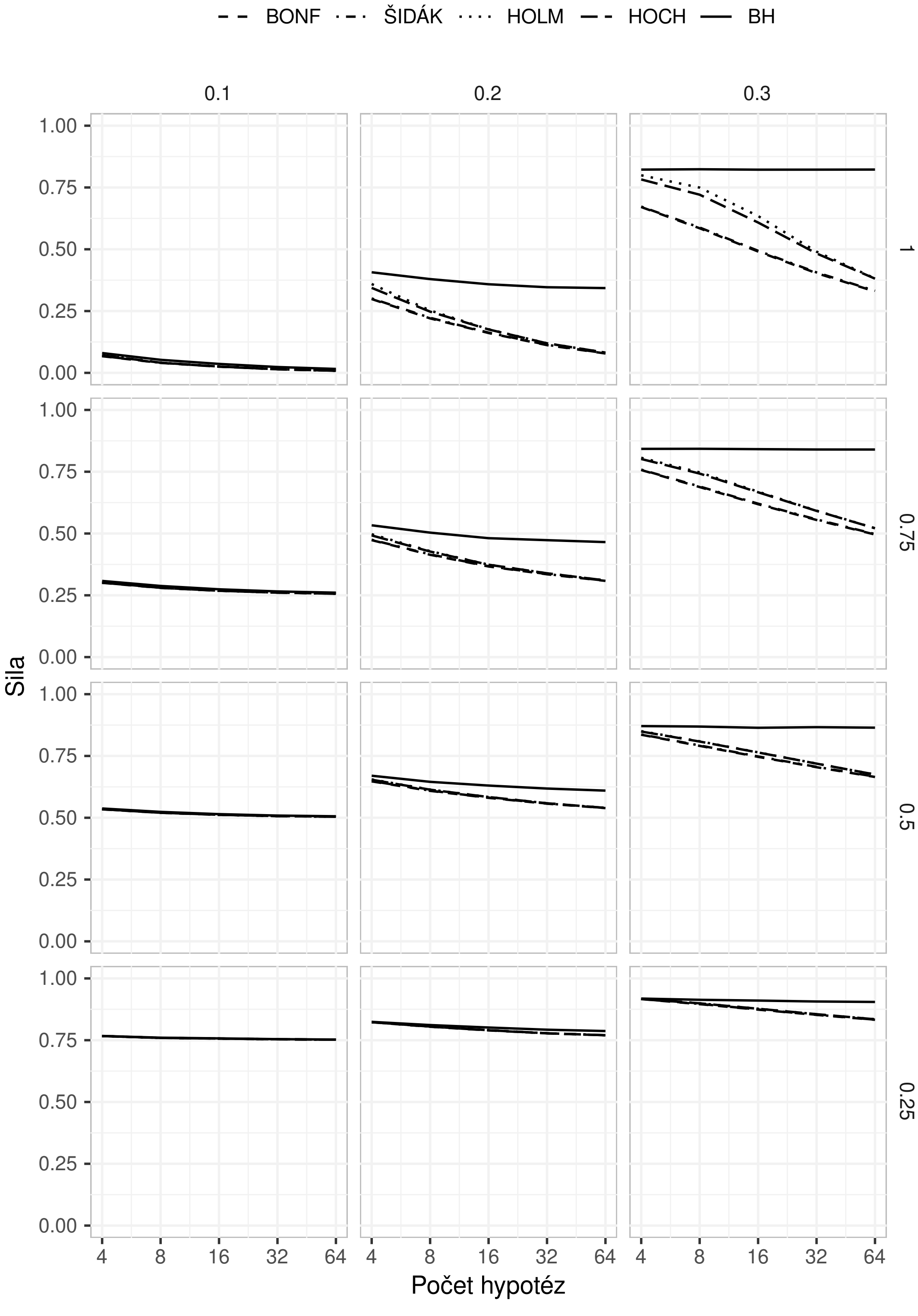
Platné hypotézy vygenerujeme so strednou hodnotou rovnou 0. Neplatné hypotézy budeme generovať postupne so strednými hodnotami rovnými 0.1, 0.2, 0.3.

Na Obrázku **??** vidíme silu mnohonásobného testovanie s použitím rozličných korekcií. Na horizontálnej osy je počet hypotéz, ktoré sme testovali. Nad obrázkami je stredná hodnota generovaných dát s neplatnými hypotézami, napravo vidíme pomer neplatných hypotéz ku všetkým hypotézam.

Pre neplatné hypotézy s akoukoľvek strednou hodnotou platí čím menší je počet neplatných hypotéz, tým väčšia je sila testu. Pri voľbe neplatnej hypotézy so strednou hodnotou, ktorá je blízko strednej hodnoty z nulovej hypotézy, je sila testu výrazne nižšia ako pri iných stredných hodnotách. Je to z dôvodu, že test nevie vždy správne vyhodnotiť, či daná hypotéza platí, keďže sú dané stredné hodnoty blízko seba.

Pri voľbe strednej hodnoty 0.1 pre neplatné hypotézy nie je medzi korekciami vidieť žiadny rozdiel pre všetky zvolené pomery hypotéz. Pre ostatné zvolené stredné hodnoty sú rozdiely vidieteľnejšie pri vyššom počte neplatných hypotéz. Bonferroniho a Šidákova majú veľmi podobnú silu testu vo všetkých prípadoch. Holmova korekcia sa vo väčšine prípadoch správa podobne ako Hochbergova, avšak v prípade neplatnosti všetkých hypotéz má pri nižšom počte hypotéz väčšiu silu. Najvyššiu silu vo všetkých prípadoch má Benjamini-Hochbergova korekcia. Pri zvyšujúcom počte neplatných hypotéz je sila tejto korekcie výrazne vyššia ako sily iných korekcií. Je to kvôli kontrolovaniu chyby FDR, ktorá nie je až taká striktná ako FWER, ktorú kontrolujú všetky ostatné korekcie.

Na silu testu sa môžeme pozerať aj iným spôsobom. V predchádzajúcej podkapitole sme pomocou simulácii ukázali, že korekcie korigujú celkovú hladinu testovania. V tomto prípade by sme chceli overiť, že korekcie mnohonásobného testovanie nemajú vplyv na konzistenciu testu, teda chceme, aby sila testu konvergovala k 1 pre rastúci rozsah náhodných výberov.



Obr. 3.1: Sila testovania pri použití rôznych korekcií so strednými hodnotami neplatných hypotéz 0.1, 0.2, 0.3 a podielom neplatných hypotéz 1, $\frac{3}{4}$, $\frac{1}{2}$, $\frac{1}{4}$

Predpokladajme, že máme dva náhodné výbery

$$\mathbf{X}_1 = (X_{11}, X_{12}, \ldots, X_{1n})^T,$$

$$\mathbf{X}_2 = (X_{21}, X_{22}, \ldots, X_{2n})^T,$$

z rozdelení $N(\mu_1, \sigma_1^2)$ a $N(\mu_2, \sigma_2^2)$. Aby sme ukázali konzistenciu, budeme meniť rozsah náhodných výberov n, postupne 50, 100, 150, 200, 250. Budeme testovať tri hypotézy, ktoré budú mať tvar

$$
\begin{aligned}
H_1 &: \mu_1 = 0; \\
H_2 &: \mu_2 = 0; \\
H_3 &: \sigma_1^2 = \sigma_2^2.
\end{aligned}
$$

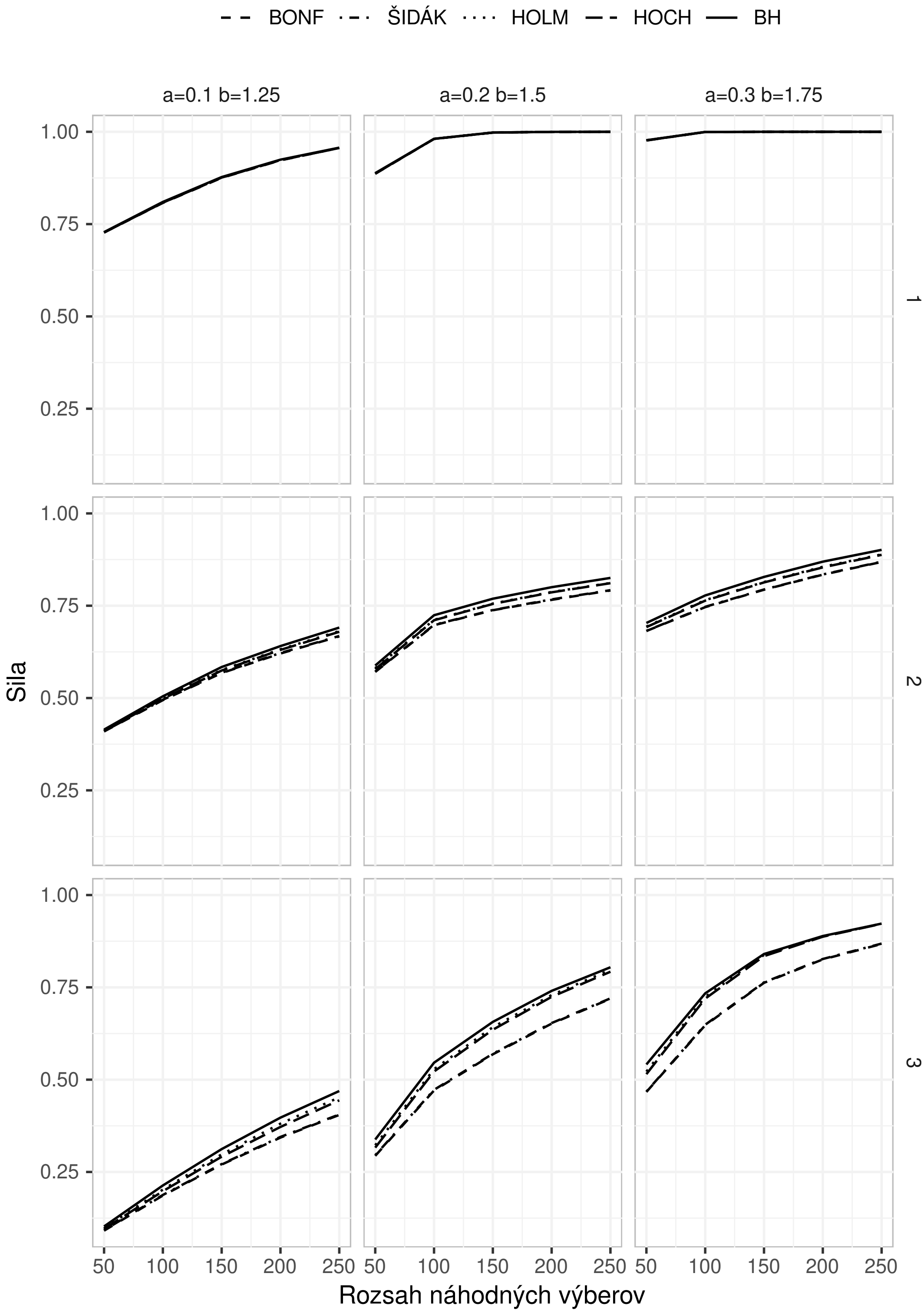
Na zistenie sily testu, potrebujeme vygenerovať dáta, pre ktoré nebudú všetky hypotézy platné. Náhodné výbery, pre ktoré budú platiť všetky hypotézy, budeme generovať z normovaného normálneho rozdelenia. V tomto prípade uvedieme tri možnosti pre náhodné výbery, kde $a, b \in \mathbb{R}$ budú nami zvolené parametre:

1. $\mathbf{X}_1 \sim N(0{,}1)$ a $\mathbf{X}_2 \sim N(0{,}b)$, v tejto možnosti budú hypotézy H_1, H_2 platné a H_3 neplatná;
2. $\mathbf{X}_1 \sim N(0{,}1)$ a $\mathbf{X}_2 \sim N(a{,}b)$, hypotéza H_1 je platné, hypotézy H_2, H_3 sú neplatné;
3. $\mathbf{X}_1 \sim N(a{,}1)$ a $\mathbf{X}_2 \sim N(a{,}b)$, v tomto prípade sú všetky hypotézy neplatné.

Poradie možností budeme používať v grafe. Keďže veľkosť sily závisí od zvolenej alternatívy, parametre a, b budeme postupne meniť, uvedieme tri možnosti voľby týchto parametrov:

- $a = 0.1$, $b = 1.25$;
- $a = 0.2$, $b = 1.5$;
- $a = 0.3$, $b = 1.75$.

Na Obrázku **??** je na horizontálnej osi uvedený rozsah náhodných výberov, na vertikálnej osi sila testu. Nad obrázkami vidíme zvolené parametre a napravo poradie možnosti ohľadne voľby generovaných náhodných výberov, ako sme uviedli vyššie. Vidíme, že pre každú voľbu parametrov a, b a taktiež pre každú uvedenú možnosť voľby platných a neplatných hypotéz, sa sila testu zvyšuje s rastúcim rozsahom náhodného výberu. Pomocou Obrázka **??** sme ukázali, že sila testu konverguje k 1 pri rastúcom rozsahu náhodných výberov. Teda korekcie mnohonásobného testovania nemajú vplyv na konzistenciu testov.



Obr. 3.2: Sila testovania pri použití rôznych korekcií s troma možnosťami voľby platných a neplatných hypotéz, pre rôzne parametre rozdelení a, b

Záver

Táto práca sa zaoberala problémom s hladinou testu pri mnohonásobnom testovaní. Na začiatku práce sme definovali kľúčové pojmy z matematickej štatistiky, ako p-hodnota, hladina a sila testu.

Podrobne sme objasnili problém s hladinou testu nastávajúci pri mnohonásobnom testovaní. Pomocou obrázkov sme ukázali rozdiely medzi testovaním jednej a viacerých hypotéz. Definovali sme chyby FWER a FDR, ktoré sa kontrolujú pri testovaní a opísali sme aký je medzi nimi rozdiel.

Táto práca sa zaoberala porovnávaním rozličných korekcií mnohonásobného testovania, vysvetlili sme ako fungujú vybrané korekcie. Pred definovaním korekcií sme uviedli Booleovu nerovnosť, ktorá je základným tvrdením pre Bonferroniho korekciu a taktiež niektoré ďalšie odvíjajúce sa z nej. Najskôr sme opísali simultánne zamietajúce korekcie, Bonferroniho a Šidákovu. Porovnali sme upravenú hladinu testov pre tieto dve korekcie. Ostatné opísané korekcie patria medzi postupne zamietajúce. Definovali sme Holmovu a Simesovu korekciu, z ktorých bola odvodená Hochbergova korekcia. Ako poslednú korekciu sme uviedli Benjamini-Hochbergovu odlíšujúcu sa od ostatných kontrolou chyby FDR narozdiel od kontroly FWER. Pre každú korekciu sme uviedli základné predpoklady na jej definíciu a opísali sme postup zamietania hypotéz. Pri väčšine z nich sme uviedli dôkaz, že korekcie naozaj kontrolujú dané chyby.

V praktickej časti práce sme porovnali korekcie najskôr z hľadiska hladiny testu. Pomocou simulácii sme určili počet zamietnutých platných hypotéz na určitý počet opakovaní pre testovanie rôzneho počtu hypotéz. Následné sme pomocou týchto hodnôt zistili celkovú hladinu testu pre každú korekciu. Podľa výsledkov simulácii sme porovnali korekcie a vysvetlili v čom sa odlišujú. Ďalšie hľadisko porovnávania bola sila testu. Keďže sila testu závisí od rôznych faktorov, v simuláciach sme menili podmienky pre testovanie, konkrétne pomer neplatných a platných hypotéz a strednú hodnotu vygenerovaných dát neplatných hypotéz. Simulácie sme previedli pre rôzne počty testujúcich hypotéz. V niektorých prípadoch sme nepostrehli rozdiely v korekciách, avšak pri väčšom počte neplatných hypotéz bola sila testu Benjamini-Hochbergovej korekcie výrazne vyššia ako sila testu pri použítí iných korekcií. Je to práve z dôvodu kontroly chyby FDR, ktorá nie je až striktná ako kontrola FWER. Na konci praktickej časti sme pomocou simulácii ukázali silu testu pre testovanie troch hypotéz s rôznymi rozsahmi náhodného výberu. Uvažovali sme rôzne možnosti platných a neplatných hypotéz a takisto sme menili parametre rozdelení náhodných výberov, pre ktoré hypotézy nie sú platné. Touto simuláciou sme ukázali, že sila testu sa v každom prípade zvyšuje pre rastúci rozsah náhodných výberov. To znamená, že korekcie nemajú vplyv na konzistenciu testu.