



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

## **BAKALÁŘSKÁ PRÁCE**

Dominik Rózsahegyi

# **Sila dvojvýberových testov**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Matúš Maciak, Ph.D.

Studijní program: Matematika

Studijní obor: Obecná matematika

Praha 2020

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) 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 Praze dne 30.7.2020

Podpis autora

Chcel by som sa poďakovať vedúcemu bakalárskej práce RNDr. Matúšovi Maciakovi, Ph.D, za odborné vedenie práce, trpezlivosť, venovaný čas a cenné rady, ktoré mi poskytoval. Ďalej by som sa chcel poďakovať rodine a priateľom za podporu.

Název práce: Sila dvojjvýberových testov

Autor: Dominik Rózsahegyí

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Matúš Maciak, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Jedným z často používaných štatistických nástrojov sú dvojjvýberové testy, pomocou ktorých rozhodujeme o tom či experimentálne získané dáta z rôznych populácií spĺňajú vopred zadaný predpoklad. Vhodnosť ich použitia môžeme určiť napríklad pomocou sily testu, vyjadrujúcej pravdepodobnosť zamietnutia neplatného tvrdenia. V úvode sa čitateľ oboznámi s pojmami testovania hypotéz, ktoré sú potrebné pre zavedenie jednotlivých testov. Testom, ktoré môžeme použiť v prípade, že sú analyzované dáta kompletné sa venuje druhá kapitola. V prípade, že niektoré pozorovania v čase analýzy dát nepoznáme presne, ide o cenzorované pozorovania a je vhodnejšie použiť príslušné testy uvedené v tretej kapitole. Pomocou simulácií odhadneme empirickú silu testov a pozorujeme jej správanie za rôznych podmienok.

Klíčová slova: štatistický test, nulová hypotéza, alternatíva, sila testu, dvojjvýberový problém

Title: The power of two sample tests

Author: Dominik Rózsahegyí

Department: Department of Probability and Mathematical Statistics

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

Abstract:

Two-sample tests are one of commonly used statistical tools with which we make decisions if experimentally obtained data from different populations satisfy pre-specified statement. Suitability of using them could be dedicated by the power of test, which states for probability of rejection of invalid statement. The reader gets to know with terms of hypothesis testing which are necessary for introduction of tests. The second chapter introduce tests which could be used when analysed data are complete. If some observations are not exactly known, we call them censored and it is more suitable to use tests listed in the third chapter. We estimate the power for tests in simulations and observe its behavior with different conditions.

Keywords: statistical test, null hypothesis, alternative hypothesis, test power, two-sample problem

# Obsah

Úvod	2
<b>1 Základné pojmy a definície</b>	<b>3</b>
1.1 Testovanie hypotéz . . . . .	3
<b>2 Dvojjvýberové testy pre úplné dáta</b>	<b>5</b>
2.1 Presný t-test . . . . .	5
2.2 Asymptotický z-test . . . . .	7
2.3 Wilcoxonov test . . . . .	9
<b>3 Dvojjvýberové testy pre cenzorované dáta</b>	<b>12</b>
3.1 Coxov-Mantelov test . . . . .	13
3.2 Logrank test . . . . .	13
3.3 Gehanov test . . . . .	14
<b>4 Simulácie</b>	<b>16</b>
4.1 Porovnanie skutočných a empirických silofunkcií . . . . .	16
4.2 Porovnanie sily testov pre úplné dáta . . . . .	17
4.3 Porovnanie sily testov pre cenzorované dáta . . . . .	20
<b>Záver</b>	<b>26</b>
<b>Zoznam použitej literatúry</b>	<b>27</b>

# Úvod

Dvojvýberové testy sú dôležitým štatistickým nástrojom na rozhodovanie o zamietnutí alebo nezamietnutí vopred zadaného tvrdenia o charakteristikách dvoch populácií. Jednou z vlastností testov, ktorá nás zaujíma je sila testu vyjadrujúca pravdepodobnosť, že test zamietne nepravdivé tvrdenie. Cieľom práce je porovnať zvolené testy na základe ich sily.

V prvej kapitole uvedieme základne pojmy testovania hypotéz. Pomocou testovania hypotéz posudzujeme či experimentálne získané dáta odpovedajú vopred zadanému tvrdeniu. V druhej kapitole uvedieme niektoré základné dvojvýberové testy a odvodíme ich skutočnú silu.

V praxi sa často stáva, že hodnota niektorých pozorovaní je známa iba čiastočne, napríklad z dôvodu obmedzeného časového horizontu zberu dát. V takom prípade ide o cenzorované pozorovania, ktoré môžeme vynechať, no vhodnejšie je zvoliť testy, ktoré ich využijú. Testy pre cenzorované dáta uvedieme v tretej kapitole.

Na záver, v simulačnej štúdií, odhadneme empirickú silu uvedených testov. Budeme uvažovať, že analyzované dáta sú kompletne alebo obsahujú aj cenzorované pozorovania a na základe empirickej sily porovnáme testy za rôznych predpokladov.

# 1. Základné pojmy a definície

Testovanie hypotéz slúži k zisteniu, či experimentálne získané dáta spĺňajú vopred zadaný predpoklad, formulovaný ako hypotéza. O pravdivosti hypotézy rozhodujeme pomocou štatistického testu. V tejto kapitole budú uvedené základné pojmy z testovania hypotéz pre prípad dvojvýberového problému. Ten sa zaoberá porovnávaním určitých charakteristík dvoch rozdelení na základe dvoch nezávislých náhodných výberov  $\mathbf{X} = (X_1, \dots, X_n)^T$  a  $\mathbf{Y} = (Y_1, \dots, Y_m)^T$ .

## 1.1 Testovanie hypotéz

Pri testovaní hypotéz testujeme tvrdenie, ktoré nazývame nulová hypotéza. Budeme ju značiť  $H_0$ . Tvrdenie, ktoré je v rozpore s nulovou hypotézou nazývame alternatívna hypotéza, značíme  $H_1$ . Záverom testu je zamietnutie alebo nezamietnutie nulovej hypotézy. Postavenie medzi nulovou a alternatívnou hypotézou nie je symetrické. Ak zamietame  $H_0$ , potvrdzujeme platnosť  $H_1$ . Nezamietnutím nulovej hypotézy nemôžeme usúdiť, že  $H_0$  platí a  $H_1$  neplatí.

K zisteniu záveru používame štatistický test založený na vhodne zostavenej merateľnej funkcii z dát  $\mathbf{X}$  a  $\mathbf{Y}$ , ktorej hovoríme *testová štatistika* a rozhodovacom pravidle. Ak  $S(\mathbf{X}, \mathbf{Y}) \in W$  zamietame nulovú hypotézu v prospech alternatívy, ak  $S(\mathbf{X}, \mathbf{Y}) \notin W$ , nulovú hypotézu nezamietame. Množinu  $W \in \mathbb{R}$  nazývame *kritický obor*.

Nech  $P_0(S(\mathbf{X}, \mathbf{Y}) \in W)$  predstavuje podmienenú pravdepodobnosť, že zamietneme nulovú hypotézu za jej platnosti a  $P_1(S(\mathbf{X}, \mathbf{Y}) \in W)$  značí podmienenú pravdepodobnosť, že zamietnieme nulovú hypotézu za platnosti alternatívnej hypotézy.

V prípade, že zamietneme nulovú hypotézu napriek tomu, že je platná, hovoríme o *chybe prvého druhu*. Ak naopak neplatnú hypotézu nezamietneme, jedná sa o *chybu druhého druhu*. V zvyšných dvoch prípadoch test rozhodol správne. Keďže test je konštruovaný tak, že chyba prvého druhu je závažnejšia ako chyba druhého druhu, chceme kontrolovať pravdepodobnosť chyby prvého druhu. K tomu slúži hladina testu, ktorá je definovaná nasledovne:

**Definícia 1.** *Nech  $\alpha \in (0,1)$  je vopred zadané číslo. Ak platí*

$$\alpha = \sup_{H_0} P_0(S(\mathbf{X}, \mathbf{Y}) \in W),$$

*hovoríme, že hladina testu je presne  $\alpha$ . Ak platí*

$$\alpha = \sup_{H_0} \lim_{n,m \rightarrow \infty} P_0(S(\mathbf{X}, \mathbf{Y}) \in W)$$

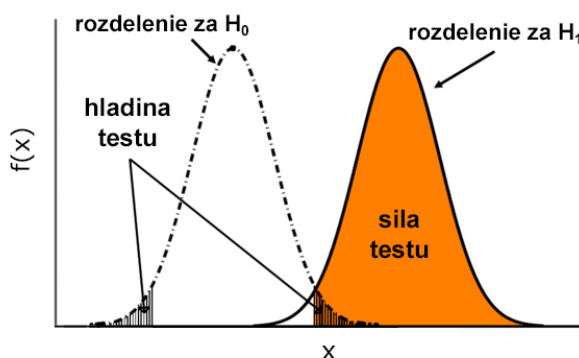
*pre  $\frac{n}{m} \rightarrow \lambda \in (0, \infty)$ , hovoríme, že test má hladinu  $\alpha$  asymptoticky.*

Hladina testu je teda najväčšia možná pravdepodobnosť chyby prvého druhu a volíme ju na začiatku testovania, zvyčajne sa volí  $\alpha = 0.05$ .

Pri dodržaní predpísanej hladiny testu sa snažíme minimalizovať pravdepodobnosť chyby druhého druhu. Doplnok pravdepodobnosti chyby druhého druhu sa nazýva sila testu a je definovaná nasledovne:

**Definícia 2.** Číslo  $\beta = P_1(S(\mathbf{X}, \mathbf{Y}) \in W)$  sa nazýva sila testu.

Sila testu vyjadruje pravdepodobnosť, že zamietneme neplatnú nulovú hypotézu. Test sa nazýva *konzistentný*, ak jeho sila konverguje k 1 pre rastúce rozsahy náhodných výberov, pričom oba rastú podobne rýchlo. Ďalšou vlastnosťou je *ne-strannosť*, ktorá požaduje, aby sila testu proti každej alternatíve bola väčšia alebo rovná ako  $\alpha$ .



Obr. 1.1: Grafické znázornenie sily a hladiny testu.

Na silu testu majú vplyv štyri faktory.

- **Hladina testu** pôsobí priamo úmerne, teda pre väčšiu hladinu testu je väčšia aj sila testu. V tomto prípade však musíme nájsť rozumný kompromis medzi silou a hladinou testu.
- **Alternatívna hypotéza**, respektíve vzdialenosť skutočnej hodnoty parametru od nulovej hypotézy. Čím väčší je tento rozdiel, tým väčšia je aj sila.
- **Variabilita dát**, ktorá je popísaná rozptylom alebo smerodatnou odchýlkou, má nepriamo úmerný vplyv. Ak majú dáta väčší rozptyl, sila testu je menšia a potrebujeme viac pozorovaní na presnejšie rozhodnutie o hypotéze.
- **Rozsah dát**, u ktorého platí, čím viac pozorovaní máme k dispozícii, tým väčšia je sila testu. Pri rozsahu dát sa zvyčajne kladie otázka, aké množstvo pozorovaní potrebujeme na to, aby test dosiahol vopred zadanú silu.

O zamietnutí alebo nezamietnutí nulovej hypotézy rozhodujeme najčastejšie pomocou p-hodnoty, ktorá tiež býva prezentovaná ako dosiahnutá hladina testu. Tá vyjadruje maximálnu možnú pravdepodobnosť, že za platnosti nulovej hypotézy napozorujeme dáta, ktoré sú s nulovou hypotézou vo väčšom alebo rovnakom rozpore ako aktuálne analyzované dáta. Nulovú hypotézu budeme zamietť pre p-hodnotu menšiu alebo rovnú ako  $\alpha$ .

Testovanie hypotéz môžeme využiť napríklad na porovnávanie distribučných funkcií dvoch rozdelení, ich stredných hodnôt alebo mediánov. Možnými testami sa budeme zaoberať v ďalších kapitolách.



## 2. Dvojvýberové testy pre úplné dáta

Majme nezávislý náhodný výber  $\mathbf{X} = (X_1, \dots, X_n)^T$  z rozdelenia s distribučnou funkciou  $F_X$  a na ňom nezávislý náhodný výber  $\mathbf{Y} = (Y_1, \dots, Y_m)^T$  z rozdelenia s distribučnou funkciou  $F_Y$ . Tieto rozdelenia patria do rovnakej rodiny rozdelení  $\mathcal{F}$ . Uvažujme ďalej funkcionál  $\tau : \mathcal{F} \rightarrow \mathbb{R}$ . Dvojvýberové testy porovnávajú  $\tau(F_X)$  a  $\tau(F_Y)$ . Testovaná hypotéza potom môže mať tvar  $H_0 : \tau(F_X) = \tau(F_Y)$  proti alternatíve  $H_1 : \tau(F_X) \neq \tau(F_Y)$ .

### 2.1 Presný t-test

Častými prípadmi testovaných parametrov sú stredné hodnoty  $\mu_X$  a  $\mu_Y$  rozdelení  $F_X$  a  $F_Y$ , teda  $\tau(F_X) = \mu_X$  a  $\tau(F_Y) = \mu_Y$ . Jedným z najpoužívanejších dvojvýberových testov na porovnanie stredných hodnôt je presný t-test. Ide o parametrický test, ktorý predpokladá, že  $X_i \sim \mathcal{N}(\mu_X, \sigma^2)$  a  $Y_j \sim \mathcal{N}(\mu_Y, \sigma^2)$ , kde  $\mu_X, \mu_Y \in \mathbb{R}, \sigma^2 > 0$  a  $i \in \{1, \dots, n\}, j \in \{1, \dots, m\}$ . Testujeme či sa stredné hodnoty líšia o vopred zadanú reálnu konštantu  $\delta$ . Zvolená nulová a alternatívna hypotéza majú tvar

$$H_0 : \mu_X = \mu_Y + \delta, \quad H_1 : \mu_X \neq \mu_Y + \delta. \quad (2.1)$$

Testová štatistika má tvar

$$S(\mathbf{X}, \mathbf{Y}) = \frac{\bar{X}_n - \bar{Y}_m - \delta}{\sqrt{S_{n,m}^2 \left(\frac{1}{n} + \frac{1}{m}\right)}},$$

kde  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\bar{Y}_m = \frac{1}{m} \sum_{i=1}^m Y_i$  a  $S_{n,m}^2$  je nestranný odhad spoločného rozptylu  $\sigma^2$ , pre ktorý platí

$$S_{n,m}^2 = \frac{n-1}{n+m-2} S_X^2 + \frac{m-1}{n+m-2} S_Y^2,$$

kde  $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  a  $S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2$ . O rozdelení testovej štatistiky hovorí nasledujúca veta.

**Veta 1.** *Nech  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  sú nezávislé náhodné výbery z normálnych rozdelení so strednými hodnotami  $\mu_X$  a  $\mu_Y$  a rovnakým rozptylom  $\sigma^2$ . Potom*

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{S_{n,m}^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{n+m-2}.$$

*Dôkaz.* Vid Omelka, M. (2019, Veta 6.2)

□

Za platnosti nulovej hypotézy v (2.1) má  $S(\mathbf{X}, \mathbf{Y})$  rozdelenie  $t_{n+m-2}$  a teda nulovú hypotézu zamietneme práve vtedy, keď

$$|S(\mathbf{X}, \mathbf{Y})| \geq t_{n+m-2} \left(1 - \frac{\alpha}{2}\right),$$

kde  $t_{n+m-2} \left(1 - \frac{\alpha}{2}\right)$  je  $\left(1 - \frac{\alpha}{2}\right)$  kvantil t-rozdelenia s  $n+m-2$  stupňami voľnosti.

Ukážeme, že za platnosti nulovej hypotézy je hladina testu presne  $\alpha$ . Pre jednoduchosť zápisu využijeme označenie  $t = t_{n+m-2} \left(1 - \frac{\alpha}{2}\right)$  a distribučnú funkciu t-rozdelenia s  $n + m - 2$  stupňami voľnosti označíme  $T$ . Z definície 1 potom

$$\begin{aligned} P_0(|S(\mathbf{X}, \mathbf{Y})| \geq t) &= 1 - P_0(-t \leq S(\mathbf{X}, \mathbf{Y}) \leq t) = \\ &= 1 - T(t) + T(-t) = 2(1 - T(t)) = 2\left(1 - \left(1 - \frac{\alpha}{2}\right)\right) = \\ &= \alpha, \end{aligned}$$

kde sme v tretej rovnosti využili symetriu t-rozdelenia.

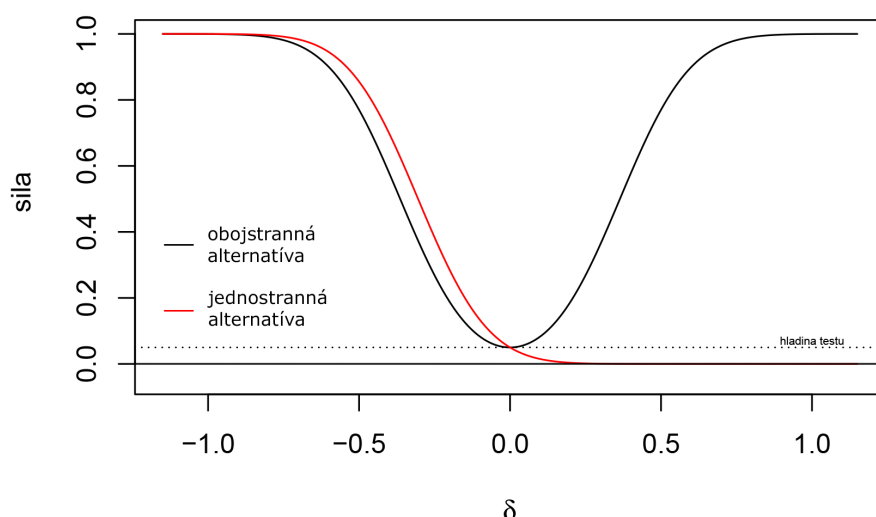
Presnú silu testu spočítame z definície 2. Uvažujme teda platnosť alternatívy  $\mu_X - \mu_Y = \delta_1 \neq \delta$ . Rovnako ako pri výpočte hladiny testu využijeme značenie  $t = t_{n+m-2} \left(1 - \frac{\alpha}{2}\right)$  a  $T$  označujúce distribučnú funkciu t - rozdelenia s  $n + m - 2$  stupňami voľnosti. Potom platí

$$\begin{aligned} \beta &= P_1(|S(\mathbf{X}, \mathbf{Y})| \geq t) = 1 - P_1(-t \leq S(\mathbf{X}, \mathbf{Y}) \leq t) \\ &= 1 - P\left(-t \leq \frac{\bar{X}_n - \bar{Y}_m - \delta_1}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq t\right) \\ &= 1 - P\left(-t \leq \frac{\bar{X}_n - \bar{Y}_m - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} - \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq t\right) \\ &= 1 - P\left(-t + \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq \frac{\bar{X}_n - \bar{Y}_m - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq t + \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right) \\ &= 1 - T\left(t + \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right) + T\left(-t + \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right). \end{aligned} \tag{2.2}$$

Pre jednostranný test, ktorý má nulovú hypotézu napríklad  $H_0 : \mu_X \leq \mu_Y + \delta$  proti alternatíve  $H_1 : \mu_X > \mu_Y + \delta$ , by sme nulovú hypotézu zamietali práve vtedy, keď  $S(\mathbf{X}, \mathbf{Y}) \geq t_{n+m-2}(1 - \alpha)$ , kde pravá strana predstavuje  $(1 - \alpha)$  kvantil t-rozdelenia s  $n + m - 2$  stupňami voľnosti, ktorý označíme  $t_1$ . Využijeme ďalej rovnaké značenie pre  $T$  ako vyššie a uvažujme  $\mu_X - \mu_Y = \delta_1 > \delta$ . Potom pre silu jednostranného testu platí podobnými úpravami ako v 2.2

$$\begin{aligned} \beta &= 1 - P\left(\frac{\bar{X}_n - \bar{Y}_m - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} - \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \leq t_1\right) \\ &= 1 - T\left(t_1 + \frac{\delta_1 - \delta}{\sqrt{S_{n,m}^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right). \end{aligned}$$

Na obrázku 2.1 vidíme, že pre jednostrannú alternatívnu hypotézu je presný t-test silnejší ako v prípade obojstrannej alternatívy.



Obr. 2.1: Porovnanie silofunkcií presného t-testu pre obojsstrannú a jednostrannú alternatívu. Náhodné výbery s rozsahmi  $n = 60$ ,  $m = 40$  pochádzajú z normálnych rozdelení s rovnakým rozptylom  $\sigma^2 = 1$ .

## 2.2 Asymptotický z-test

V prípade, že náhodné výbery nepochádzajú z normálnych rozdelení s rovnakým rozptylom môžeme na testovanie stredných hodnôt modifikáciu t-testu. Ten predpokladá, že nezávislé náhodné výbery pochádzajú z rozdelení, ktoré majú konečné a kladné rozptyly. Nulová a alternatívna hypotéza sú zvolené rovnako ako v (2.1). Testová štatistika má tvar

$$S(\mathbf{X}, \mathbf{Y}) = \frac{\bar{X}_n - \bar{Y}_m - \delta}{\sqrt{S_X^2/n + S_Y^2/m}}.$$

Môžeme si všimnúť, že testové štatistiky v t-teste a z-teste pracujú s inými odhadmi rozptylov. Ak sú rozsahy oboch náhodných výberov rovnaké, sú testové štatistiky t-testu a z-testu rovnaké.

**Veta 2.** *Nech  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  sú nezávislé náhodné výbery so strednými hodnotami  $\mu_X$  a  $\mu_Y$  a konečnými, kladnými rozptylmi  $\sigma_X^2, \sigma_Y^2$ . Potom*

$$\frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}} \xrightarrow{d} \mathcal{N}(0,1),$$

pre  $n, m \rightarrow \infty$ ,  $\frac{n}{m} \rightarrow \lambda \in (0, \infty)$ .

*Dôkaz.* Vid Omelka, M. (2019, Veta 6.3)

□

Z vety 2 vieme, že testová štatistika má za  $H_0$  asymptoticky normované normálne rozdelenie a z toho vyplýva, že nulovú hypotézu budeme zamietť práve vtedy, keď

$$|S(\mathbf{X}, \mathbf{Y})| \geq u_{1-\frac{\alpha}{2}},$$

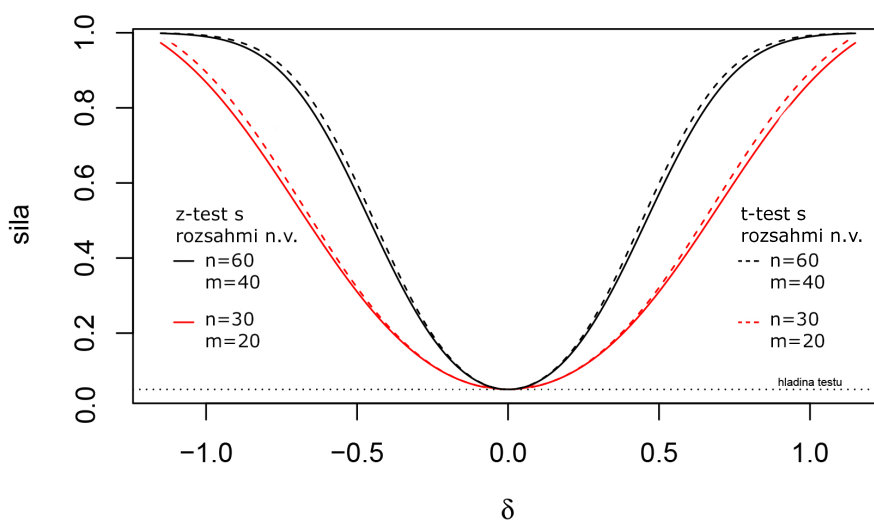
kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2})$  kvantil rozdelenia  $\mathcal{N}(0,1)$ .

Hladinu testu spočítame z definície 1 a využijeme značenia  $u = u_{1-\frac{\alpha}{2}}$ . Distribučnú funkciu  $\mathcal{N}(0,1)$  označíme  $\Phi$  a využijeme vlastnosť symetrie, teda že  $\Phi(-u) = 1 - \Phi(u)$ .

$$\begin{aligned}
 P_0(|S(\mathbf{X}, \mathbf{Y})| \geq u) &= 1 - P_0(-u \leq S(\mathbf{X}, \mathbf{Y}) \leq u) \\
 &\stackrel{as.}{=} 1 - \Phi(u) + \Phi(-u) = 2(1 - \Phi(u)) = 2\left(1 - \left(1 - \frac{\alpha}{2}\right)\right) \\
 &= \alpha.
 \end{aligned}
 \tag{2.3}$$

Pre výpočet sily testu uvažujme, že platí alternatíva  $\delta_1 = \mu_X - \mu_Y \neq \delta$ . Potom, keď  $\mu_X$  a  $\mu_Y$  sú skutočné parametre, tak rozdelenie testovej štatistiky  $S(\mathbf{X}, \mathbf{Y})$  je asymptoticky  $\mathcal{N}(\nu, 1)$ , kde  $\nu = \frac{\delta - \delta_1}{\sqrt{S_X^2/n + S_Y^2/m}}$ . Teda  $S(\mathbf{X}, \mathbf{Y}) - \nu$  má asymptoticky normované normálne rozdelenie a analogicky ako v (2.2) dostaneme

$$\beta \stackrel{as.}{=} 1 - \Phi(u - \nu) + \Phi(-u - \nu).$$



Obr. 2.2: Porovnanie silofunkcií asymptotického z-testu a presného t-testu pre dva prípady rôznych rozsahov náhodných výberov. Náhodné výbery pochádzajú z normálnych rozdelení s rovnakým rozptylom  $\sigma^2 = 1$ .

Ako vidíme na obrázku 2.2, pre náhodné výbery pochádzajúce z normálnych rozdelení s rovnakým rozptylom je presný t-test silnejší ako asymptotický z-test. V prípade, že sú rozsahy rovnaké, testové štatistiky t-testu a z-testu sa zhodujú.

## 2.3 Wilcoxonov test

Na testovanie posunutia dvoch rozdelení sa využíva Wilcoxonov test, ktorý predpokladá, že  $F_X$  a  $F_Y$  sú spojité a  $\exists \delta \in \mathbb{R} : F_X(x) = F_Y(x - \delta) \forall x \in \mathbb{R}$ . Nulová a alternatívna hypotéza potom majú tvar

$$H_0 : \delta = 0, \quad H_1 : \delta \neq 0. \quad (2.4)$$

V prípade existencie stredných hodnôt platí za nulovej hypotézy ich rovnosť. Ďalej si môžeme všimnúť, že nulová hypotéza v (2.4) sa dá ekvivalentne formulovať ako

$$H_0 : F_X(x) = F_Y(x) \forall x \in \mathbb{R} \quad (2.5)$$

Testová štatistika je založená na poradí náhodných veličín  $X_i$  v spojenom výbere  $X_1, \dots, X_n, Y_1, \dots, Y_m$ . Poradím rozumieme číslo  $R_i \in \{1, \dots, n+m\}$ , pre ktoré platí  $X_i = X_{(R_i)}$ , kde  $X_{(R_i)}$  je  $R_i$ -tá najmenšia hodnota medzi pozorovaniami  $X_1, \dots, X_n, Y_1, \dots, Y_m$ . Testová štatistika je tvaru

$$S(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^n R_i,$$

kde  $R_i, i \in \{1, \dots, n\}$  sú poradia náhodných veličín  $X_i$  v združenom výbere  $X_1, \dots, X_n, Y_1, \dots, Y_m$ .

**Veta 3.** Ak sú distribučné funkcie oboch náhodných výberov spojité a platí nulová hypotéza, tak

1.

$$E S(\mathbf{X}, \mathbf{Y}) = \frac{n(n+m+1)}{2}, \quad \text{var}(S(\mathbf{X}, \mathbf{Y})) = \frac{mn(n+m+1)}{12},$$

2.

$$\frac{S(\mathbf{X}, \mathbf{Y}) - \mu_S}{\sqrt{\sigma_S^2}} \xrightarrow{d} \mathcal{N}(0,1) \quad (2.6)$$

pre  $n, m \rightarrow \infty$ , kde  $\mu_S = E S(\mathbf{X}, \mathbf{Y})$  a  $\sigma_S^2 = \text{var}(S(\mathbf{X}, \mathbf{Y}))$ .

*Dôkaz.* Omelka, M. (2019, str. 104)

□

Vo vete 3 predpokladáme spojitosť distribučných funkcií, aby sa v združenom výbere skoro isto nevyskytovali zhody. V prípade, že by náhodné výbery pochádzali z diskretných rozdelení, mohli by sa v nich vyskytovať zhody, preto by bolo nutné pracovať s korektúrou testovej štatistiky, ktorú možno nájsť v Omelka, M. (2019). Z vety 3 poznáme asymptotické rozdelenie testovej štatistiky za nulovej hypotézy a teda

$$H_0 \text{ zamietneme} \iff \frac{|S(\mathbf{X}, \mathbf{Y}) - \mu_S|}{\sqrt{\sigma_S^2}} \geq u_{1-\frac{\alpha}{2}},$$

kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2})$  kvantil rozdelenia  $\mathcal{N}(0,1)$ .

Rovnako ako v prípade predchádzajúcich testov sa dá ukázať, že za platnosti modelu a nulovej hypotézy test dodržiava hladinu  $\alpha$  asymptoticky. Pri výpočte sily využijeme, že testová štatistika sa dá prepísať ako

$$\begin{aligned}
S(\mathbf{X}, \mathbf{Y}) &= \sum_{i=1}^n R_i = \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}[X_i \geq X_j] + \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}[X_i \geq Y_j] \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}[X_{(i)} \geq X_{(j)}] + \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}[X_i \geq Y_j] \\
&= \sum_{i=1}^n i + \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}[X_i \geq Y_j] \\
&= \frac{n(n+1)}{2} + \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}[X_i \geq Y_j].
\end{aligned}$$

Vo všeobecnosti potom pre rozptyl  $\text{var}(S(\mathbf{X}, \mathbf{Y}))$  platí

$$\begin{aligned}
\sigma^2 &= \text{var} \left( \frac{n(n+1)}{2} + \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}[X_i \geq Y_j] \right) \\
&= nm \text{var}(\mathbb{1}[X_i \geq Y_j]) + nm(m-1) \text{cov}(\mathbb{1}[X_i \geq Y_{j_1}]; \mathbb{1}[X_i \geq Y_{j_2}]) \\
&\quad + nm(n-1) \text{cov}(\mathbb{1}[X_{i_1} \geq Y_j]; \mathbb{1}[X_{i_2} \geq Y_j]) \\
&= nmp_1(1-p_1) + nm(m-1)(p_2 - p_1^2) + nm(n-1)(p_3 - p_1^2),
\end{aligned}$$

kde využívame značenia

$$\begin{aligned}
p_1 &= P(X_i \geq Y_j), \\
p_2 &= P(X_i \geq Y_{j_1} \wedge X_i \geq Y_{j_2}), \\
p_3 &= P(X_{i_1} \geq Y_j \wedge X_{i_2} \geq Y_j).
\end{aligned}$$

Nech  $\mathbf{x}=(x_1, \dots, x_n)^T$  a  $\mathbf{y}=(y_1, \dots, y_m)^T$  sú pozorované realizácie náhodných výberov  $\mathbf{X}$  a  $\mathbf{Y}$ . Potom  $p_1$ ,  $p_2$  a  $p_3$  môžeme odhadnúť pomocou

$$\begin{aligned}
\widehat{p}_1 &= \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m P(x_i \geq y_j), \\
\widehat{p}_2 &= \frac{1}{nm(m-1)} \sum_{i=1}^n \sum_{j_1 \neq j_2} P(x_i \geq y_{j_1} \wedge x_i \geq y_{j_2}), \\
\widehat{p}_3 &= \frac{1}{nm(n-1)} \sum_{j=1}^m \sum_{i_1 \neq i_2} P(x_{i_1} \geq y_j \wedge x_{i_2} \geq y_j).
\end{aligned}$$

Pre strednú hodnotu  $S(\mathbf{X}, \mathbf{Y})$  vo všeobecnosti platí

$$\mu = \frac{n(n+1)}{2} + nmp_1.$$

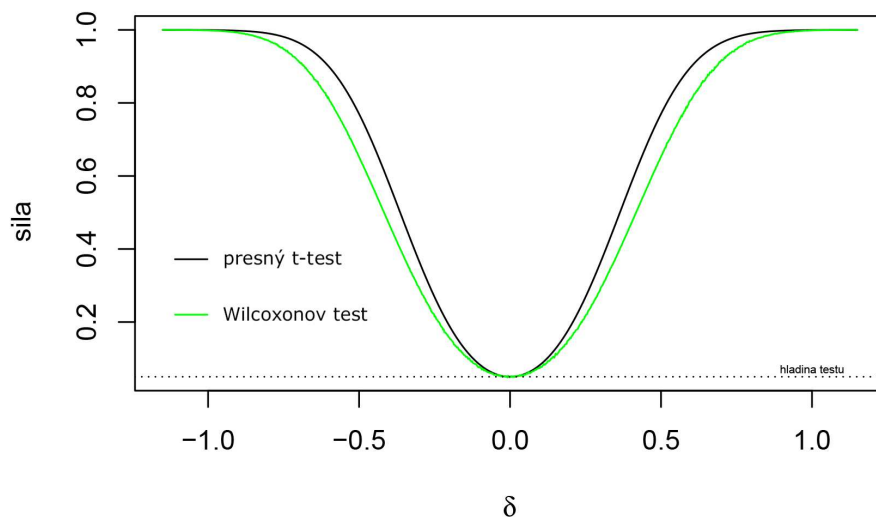
Ako uvádza Chow a kol. (2008), za platnosti alternatívnej hypotézy platí, že  $p_1 \neq \frac{1}{2}$  a testová štatistika  $S(\mathbf{X}, \mathbf{Y})$  môže byť aproximovaná náhodnou veličinou s normálnym rozdelením so strednou hodnotou  $\widehat{\mu}_1 = \frac{n(n+1)}{2} + nm\widehat{p}_1$  a rozptylom

$\widehat{\sigma}_1^2 = nm\widehat{p}_1(1 - \widehat{p}_1) + nm(m - 1)(\widehat{p}_2 - \widehat{p}_1^2) + nm(n - 1)(\widehat{p}_3 - \widehat{p}_1^2)$ . Potom teda pre silu Wilcoxonovho testu platí

$$\begin{aligned} \beta &= 1 - P_1 \left( -u \leq \frac{S(\mathbf{X}, \mathbf{Y}) - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_1^2}} \frac{\sqrt{\widehat{\sigma}_1^2}}{\sqrt{\widehat{\sigma}_S^2}} - \frac{\mu_S - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_S^2}} \leq u \right) \\ &= 1 - P_1 \left( \left( -u + \frac{\mu_S - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_S^2}} \right) \frac{\sqrt{\widehat{\sigma}_S^2}}{\sqrt{\widehat{\sigma}_1^2}} \leq \frac{S(\mathbf{X}, \mathbf{Y}) - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_1^2}} \leq \left( u + \frac{\mu_S - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_S^2}} \right) \frac{\sqrt{\widehat{\sigma}_S^2}}{\sqrt{\widehat{\sigma}_1^2}} \right) \\ &\stackrel{as.}{=} 1 - \Phi \left( \left( u + \frac{\mu_S - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_S^2}} \right) \frac{\sqrt{\widehat{\sigma}_S^2}}{\sqrt{\widehat{\sigma}_1^2}} \right) + \Phi \left( \left( -u + \frac{\mu_S - \widehat{\mu}_1}{\sqrt{\widehat{\sigma}_S^2}} \right) \frac{\sqrt{\widehat{\sigma}_S^2}}{\sqrt{\widehat{\sigma}_1^2}} \right), \end{aligned}$$

kde  $u = u_{1-\frac{\alpha}{2}}$ .

Na obrázku nižšie porovnáme silofunkcie t-testu a Wilcoxonovho testu pre prípad, že náhodné výbery pochádzajú z normálnych rozdelení s rovnakým rozptylom a líšia sa strednými hodnotami. Distribučné funkcie sú posunuté o  $\delta \in \mathbb{R}$ .

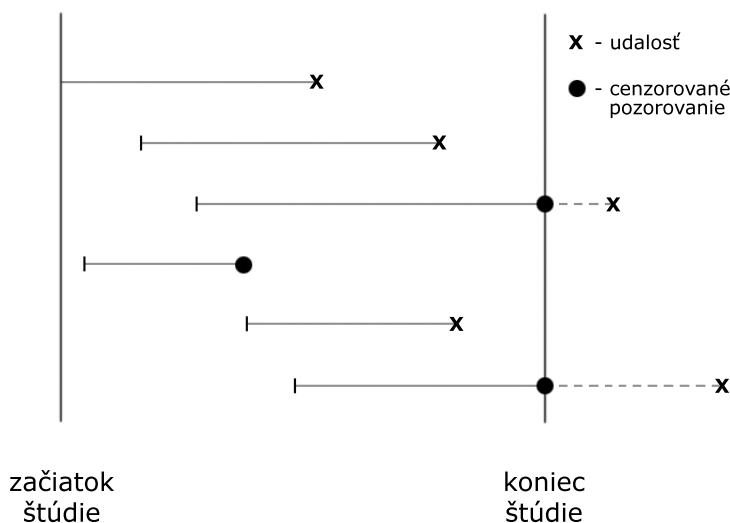


Obr. 2.3: Porovnanie silofunkcií presného t-testu a Wilcoxonovho testu na výberoch z normálnych rozdelení s rovnakým rozptylom. Pre rozsahy  $\mathbf{X}$  a  $\mathbf{Y}$  platí  $n = 60$ ,  $m = 40$ .

Obrázok 2.3 ukazuje, že ak náhodné výbery pochádzajú z normálnych rozdelení s rovnakým rozptylom, presný t-test dosahuje väčšiu silu ako Wilcoxonov test. Je to z toho dôvodu, že predpoklady Wilcoxonovho testu nie sú až tak obmedzujúce ako je to v prípade t-testu.

### 3. Dvojvýberové testy pre cenzorované dáta

Doteraz sme pracovali s testami, ktoré môžeme použiť v prípade, že sú všetky hodnoty pozorovaní skutočne známe. V niektorých prípadoch však môže z rôznych dôvodov dôjsť ku strate informácie a skutočnú hodnotu nepoznáme. V takom prípade hovoríme o cenzorovaných dátach, ktorých využitie je pri spracovaní dát žiadúce. S cenzorovanými dátami sa môžeme stretnúť napríklad v medicíne, kde môžeme sledovať čas, po ktorom sa rozvinie choroba. Ak k tejto udalosti nedôjde, poznáme iba čas, za aký sa choroba nerozvinula a ide o cenzorované pozorovanie.



Obr. 3.1: Príklad sprava cenzorovaných dát.

Majme nezáporné náhodné veličiny  $T$  a  $C$ , ktorým hovoríme *skutočná hodnota* a *cenzorovaná hodnota*. Rozdelenie náhodnej veličiny  $T$  môže byť popísané funkciou  $S(t) = P(T > t)$ , ktorá sa označuje ako *funkcia prežitia*. Ak máme  $F$ , distribučnú funkciu náhodnej veličiny  $T$ , tak platí, že  $S(t) = 1 - F(t)$ . Uvažujme ďalej, že  $\mathbf{X} = ((T_{X1}, C_{X1}), \dots, (T_{Xn}, C_{Xn}))$  a  $\mathbf{Y} = ((T_{Y1}, C_{Y1}), \dots, (T_{Ym}, C_{Ym}))$  sú nezávislé náhodné výbery s rozsahmi  $n$  a  $m$ . Nech navyše platí, že  $T_{X1}, \dots, T_{Xn}$  a  $T_{Y1}, \dots, T_{Ym}$  sú rovnako rozdelené nezáporné náhodné veličiny s funkciami prežitia  $S_X$  a  $S_Y$ . Tieto funkcie chceme pomocou zvolených dvojvýberových testov pre cenzorované dáta porovnať.

Pri analýze dát však nemáme k dispozícii  $T$  a  $C$ , ale iba  $X = \min(T, C)$  a *indikátor cenzorovania*  $\delta = \mathbf{1}[T \leq C]$ . Pozorované dáta potom majú tvar

$$\mathbf{X} = ((X_1, \delta_{X1}), \dots, (X_n, \delta_{Xn}))^T, \quad \mathbf{Y} = ((Y_1, \delta_{Y1}), \dots, (Y_m, \delta_{Ym}))^T. \quad (3.1)$$



### 3.1 Coxov-Mantelov test

Jedným z najpoužívanějších dvojvýberových testov pre cenzorované dáta je Coxov-Mantelov test, často tiež nazývaný logrank test. Test porovnáva funkcie prežitia, nulová a alternatívna hypotéza sú v našom prípade zvolené nasledovne:

$$H_0 : S_X(t) = S_Y(t) \quad \forall t \geq 0, \quad H_1 : \exists t \geq 0 : S_X(t) \neq S_Y(t). \quad (3.2)$$

Keďže platí  $S_X = 1 - F_X$  a  $S_Y = 1 - F_Y$ , zvolené nulové hypotézy v (3.2) a (2.5) sú totožné. Majme spojený výber  $\mathbf{Z}$  tvorený všetkými  $X_i$  a  $Y_i$ , pre ktoré sú  $\delta_{X_i}$  a  $\delta_{Y_i}$  rovné 1. Ide teda o všetky necenzorované pozorovania. Niektoré hodnoty v  $\mathbf{Z}$  sa môžu opakovať, preto označíme všetky rôzne hodnoty  $t_1, \dots, t_k$ . Ich usporiadaním od najmenšieho po najväčší získame usporiadaný výber  $t_{(1)} < \dots < t_{(k)}$ . Ďalej definujeme  $d_{t_i}$  ako početnosť  $t_i$ . Platí teda, že  $\sum_{i=1}^k d_{t_i} = r + s$ , kde  $r$  je počet necenzorovaných pozorovaní v  $\mathbf{X}$ , teda  $\delta_{X_i} = 1$ , práve  $r$ -krát a  $s$  je počet necenzorovaných pozorovaní v  $\mathbf{Y}$ . Ďalej  $R(t)$  je definované ako množina pozorovaní z oboch náhodných výberov, ktoré majú čas prežitia alebo cenzorovaný čas väčší alebo rovný ako  $t$ .  $R(t)$  obsahuje  $n_t$  pozorovaní z  $\mathbf{X}$  a  $m_t$  pozorovaní z  $\mathbf{Y}$  a počet pozorovaní v  $R(t_{(i)})$  je  $r_{(i)} = n_{t_{(i)}} + m_{t_{(i)}}$ . Ďalej definujeme  $A_{(i)} = \frac{m_{t_{(i)}}}{r_{(i)}}$  a

$$U = s - \sum_{i=1}^k d_{t_i} A_{(i)},$$

$$I = \sum_{i=1}^k \frac{d_{t_i} (r_{(i)} - d_{t_i})}{r_{(i)} - 1} A_{(i)} (1 - A_{(i)}).$$

Test je potom definovaný testovou štatistikou

$$S(\mathbf{X}, \mathbf{Y}) = \frac{U}{\sqrt{I}},$$

ktorá má za platnosti nulovej hypotézy asymptoticky normované normálne rozdelenie, ako uviedli Lee a Wang (2003). Na základe rozdelenia testovej štatistiky potom

$$H_0 \text{ zamietneme} \iff |S(\mathbf{X}, \mathbf{Y})| \geq u_{1-\frac{\alpha}{2}},$$

kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2})$ -tý kvantil rozdelenia  $\mathcal{N}(0,1)$ .

Test dosahuje hladinu  $\alpha$  asymptoticky a jej výpočet je identický ako v (2.3). Odvodenie presnej sily Coxovho-Mantelovho testu je príliš zložitá a preto budeme v štvrtej kapitole pracovať s empirickou silou odhadnutou pomocou simulácií.

### 3.2 Logrank test

Ďalším často používaným testom na porovnanie funkcií prežitia je logrank test, ktorý porovnáva počty necenzorovaných pozorovaní s ich očakávaným počtom. V literatúre možno túto verziu nájsť aj pod menom  $\chi^2$  test. Nulová a alternatívna hypotéza sú zvolené rovnako ako v (3.2). K definovaniu testovej štatistiky majme rovnako ako v prípade Coxovho-Mantelovho testu časy  $t_1, \dots, t_k$ . Ďalej definujeme očakávané počty udalostí v skupinách v čase  $t$  ako

$$e_{Xt} = \frac{n_t}{n_t + m_t} d_t \quad \text{a} \quad e_{Yt} = \frac{m_t}{n_t + m_t} d_t,$$

kde  $n_t = \sum_{i=1}^n \mathbb{1}[X_i \geq t]$ ,  $m_t = \sum_{i=1}^m \mathbb{1}[Y_i \geq t]$  a  $d_t$  je početnosť  $t$ . Potom

$$\xi_X = \sum_{t \in t_1, \dots, t_k} e_{Xt} \quad a \quad \xi_Y = \sum_{t \in t_1, \dots, t_k} e_{Yt}$$

sú očakávané počty udalostí v náhodných výberoch  $\mathbf{X}$  a  $\mathbf{Y}$ . Testová štatistika má potom tvar

$$S(\mathbf{X}, \mathbf{Y}) = \frac{(r - \xi_X)^2}{\xi_X} - \frac{(s - \xi_Y)^2}{\xi_Y},$$

kde  $r, s$  sú počty necenzorovaných časov v  $\mathbf{X}$  a  $\mathbf{Y}$ . Podľa Lee a Wang (2003) má  $S(\mathbf{X}, \mathbf{Y})$  asymptoticky  $\chi^2$  rozdelenie s jedným stupňom voľnosti, z čoho potom nulovú hypotézu zamietneme práve vtedy, keď

$$S(\mathbf{X}, \mathbf{Y}) \geq \chi_1^2(1 - \alpha),$$

kde  $\chi_1^2(1 - \alpha)$  predstavuje  $(1 - \alpha)$ -kvantil rozdelenia  $\chi_1^2$ .

Označme ďalej  $F$  distribučnú funkciu  $\chi_1^2$  rozdelenia. Potom pre hladinu logrank testu platí podľa definície 1

$$P_0(S(\mathbf{X}, \mathbf{Y}) \geq \chi_1^2(1 - \alpha)) \stackrel{as.}{=} 1 - F(\chi_1^2(1 - \alpha)) = \alpha.$$

Vyjadrenie presnej sily testu je nad rámec bakalárskej práce, v ďalšej časti práce preto budeme využívať silu odhadnutú pomocou simulácií.

### 3.3 Gehanov test

Pre prípad, kedy testované dáta obsahujú aj cenzorované sa používa viacero rôznych modifikácií dvojjvýberového Wilcoxonovho testu. V práci sa budeme venovať Gehanovej modifikácii, ktorú uvedieme pre rovnako zvolenú nulovú hypotézu ako v (3.2).

Definujme si náhodnú veličinu  $U_{ij}$ , ktorá porovnáva pozorovania oboch náhodných výberov. V prípade, že žiadne dáta nie sú cenzorované, má  $U_{ij}$  tvar

$$U_{ij} = \begin{cases} +1 & \text{ak } X_i > Y_j \\ -1 & \text{ak } X_i < Y_j \\ 0 & \text{ak } X_i = Y_j \end{cases}$$

a test založený na testovej štatistike  $\sum_{i=1}^n \sum_{j=1}^m U_{ij}$  je ekvivalentný s Wilcoxonovým testom, čo uvádza Gehan (1965). Zovšeobecnenie náhodnej veličiny  $U_{ij}$  pre cenzorované dáta má tvar

$$U_{ij} = \begin{cases} +1 & \text{ak } (X_{i,1}) > (Y_{j,1}) \text{ alebo } (X_{i,0}) \geq (Y_{j,1}) \\ -1 & \text{ak } (X_{i,1}) < (Y_{j,1}) \text{ alebo } (X_{i,1}) \leq (Y_{j,0}) \\ 0 & \text{inak.} \end{cases}$$

Spojením  $\mathbf{X}$  a  $\mathbf{Y}$  do jedného výberu, značeného ako  $((v_1, \delta_1), \dots, (v_{n+m}, \delta_{n+m}))$ , môžeme funkciu prepísať ako

$$U_{ij} = \begin{cases} +1 & \text{ak } (v_i, 1) > (v_j, 1) \text{ alebo } (v_i, 0) \geq (v_j, 1) \\ -1 & \text{ak } (v_i, 1) < (v_j, 1) \text{ alebo } (v_i, 1) \leq (v_j, 0) \\ 0 & \text{inak} \end{cases}$$

a označíme  $U_i = \sum_{j=1}^{n+m} U_{ij}$ . Potom za platnosti nulovej hypotézy platí podľa Lee a Wang (2003), že  $W = \sum_{i=1}^n U_i$  má asymptoticky normálne rozdelenie s nulovou strednou hodnotou a rozptylom

$$\text{var}(W) = \frac{nm \sum_{i=1}^{n+m} U_i^2}{(n+m)(n+m-1)}.$$

Testová štatistika tvaru

$$S(\mathbf{X}, \mathbf{Y}) = \frac{W}{\sqrt{\text{var}(W)}}$$

má potom asymptoticky rozdelenie  $\mathcal{N}(0,1)$  a nulovú hypotézu budeme zamietat práve vtedy, keď

$$|S(\mathbf{X}, \mathbf{Y})| \geq u_{1-\frac{\alpha}{2}},$$

kde  $u_{1-\frac{\alpha}{2}}$  je  $(1 - \frac{\alpha}{2})$ -tý kvantil rozdelenia  $\mathcal{N}(0,1)$  a rovnako ako v (2.3) možno ukázať, že hladina testu je asymptoticky  $\alpha$ .

Keďže ide o neparametrický test, odvodenie presnej sily je nad rámec tejto práce a využijeme empirickú silu získanú pomocou simulácií.

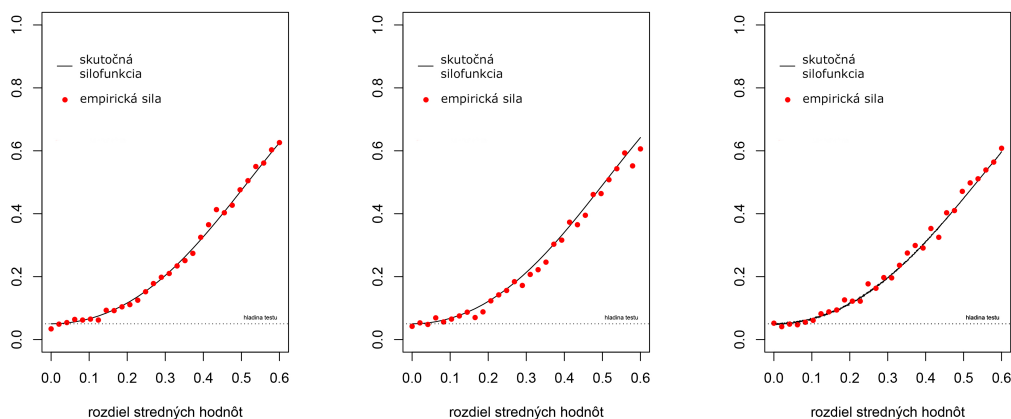
## 4. Simulácie

Keďže pre testy s cenzorovanými dátami nevieme odvodiť skutočnú silu testu, využijeme na jej odhadnutie empirickú silu, ktorú získame pomocou simulačnej štúdie v štatistickom softwari R. Pomocou simulácií tiež porovnáme silu testov za rôznych podmienok.

### 4.1 Porovnanie skutočných a empirických silofunkcií

Skutočné silofunkcie presného t-testu, z-testu a Wilcoxonovho testu najskôr porovnáme s empirickými silofunkciami pre rôzne rozsahy náhodných výberov. V každom z tisíc opakovaní vygenerujeme dáta zo zadaných rozdelení a prevedieme na nich zvolený test. Výbery najskôr generujeme tak, aby platila nulová hypotéza. Potom hladinu testu môžeme odhadnúť ako pomer zamietnutých nulových hypotéz a počet opakovaní. Vo zvyšných prípadoch generujeme dáta tak, aby platila alternatívna hypotéza. V takom prípade tento pomer vyjadruje silu testu.

Pre každé opakovanie boli vygenerované dáta z normálnych rozdelení s rovnakým rozptylom a rozdielom stredných hodnôt uvedených na horizontálnej osi. Rozsahy náhodných výberov  $\mathbf{X}$  a  $\mathbf{Y}$  boli zvolené  $n = m = 30$ , respektíve  $n = m = 70$  v druhom prípade.



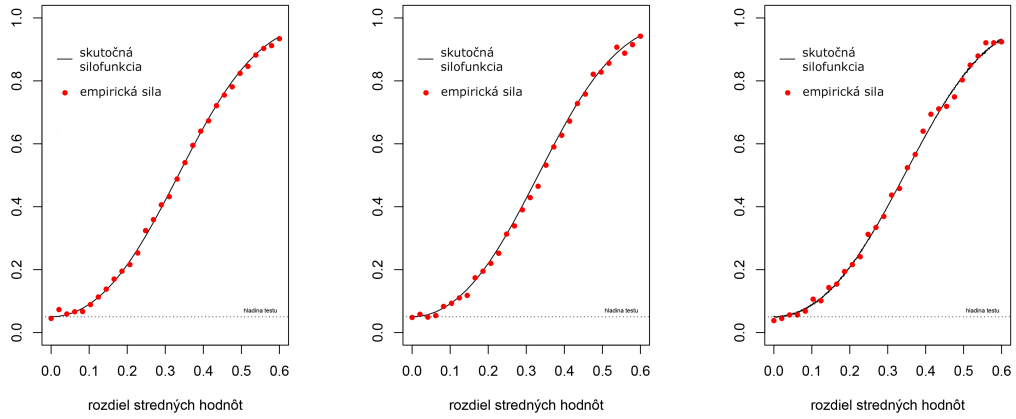
(a) Presný t-test.

(b) Asymptotický z-test.

(c) Wilcoxonov test.

Obr. 4.1: Porovnanie skutočnej a empirickej silofunkcie testov pre rozsahy náhodných výberov  $n = m = 30$ .

Porovnaním obrázkov 4.1 a 4.2 vidíme, že pre väčšie rozsahy náhodných výberov je rozdiel medzi skutočnou a empirickou silofunkciou menší. Zároveň sa ukazuje, že rozdiel medzi odhadnutou a skutočnou silou je v prípade presného t-testu o niečo menší ako v prípade zvyšných dvoch testov.



(a) Presný t-test.

(b) Asymptotický z-test.

(c) Wilcoxonov test.

Obr. 4.2: Porovnanie skutočnej a empirickej silofunkcie testov pre rozsahy náhodných výberov  $n = m = 70$ .

## 4.2 Porovnanie sily testov pre úplné dáta

V nasledujúcej podkapitole porovnáme empirické silofunkcie pre úplné náhodné výbery. V každom z tisíc opakovaní vygenerujeme dáta predstavujúce  $\mathbf{X}$  a  $\mathbf{Y}$  s rozsahmi  $n, m$  a na týchto dátach prevedieme testy uvedené v práci na hladine  $\alpha = 0.05$ . Testované hypotézy sú rovnakého tvaru ako boli uvedené pre jednotlivé testy. Najskôr uvažujme  $\mathbf{X}$  z  $Exp(1)$ , a  $\mathbf{Y}$  pochádzajúci z  $Exp(\lambda_Y)$ . V druhom prípade budú náhodné výbery pochádzať z lognormálnych rozdelení,  $X_i \sim LN(0,1)$  pre  $i = 1, \dots, n$  a  $Y_j \sim LN(\mu_Y, 1)$  pre  $j = 1, \dots, m$ . Platí, že ak má náhodná veličina  $V$  lognormálne rozdelenie s parametrami  $\mu$  a  $\sigma^2$ , tak  $\log(V)$  má normálne rozdelenie so strednou hodnotou  $\mu$  a rozptylom  $\sigma^2$ . Po použití tejto transformácie teda budú splnené predpoklady presného t-testu. Na záver uvažujme Weibullovo rozdelenie s hustotou

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

kde  $k > 0$  je parameter tvaru a  $\lambda > 0$  je parameter škálovania. Najskôr budeme porovnávať  $\mathbf{X}$  z  $Weibull(1.1, 100)$  a  $\mathbf{Y}$  pochádzajúci z  $Weibull(k_Y, \lambda_Y)$ . V druhom prípade potom  $X_i \sim Weibull(0.95, 100)$  pre  $i = 1, \dots, n$ .

Výsledky získané simuláciami nájdeme v tabuľkách 4.1, 4.2 a 4.3 a na obrázkoch nižšie. Po transformácii dát z lognormálnych rozdelení je pre akékoľvek rozsahy náhodných výberov najsilnejší presný t-test, ktorý tiež dodržiava predpísanú hladinu, ako vidíme v tabuľke 4.2. V prípade exponenciálnych rozdelení je najsilnejším Coxov-Mantelov test, ktorý ale v niektorých prípadoch nedodržiava hladinu testu, čo vidíme v tabuľke 4.1. Veľmi podobné výsledky ako v prípade exponenciálnych rozdelení sme dostali aj v prípade, že náhodné výbery pochádzajú z Weibullových rozdelení s rovnakým parametrom tvaru. V práci sú preto uvedené výsledky iba pre rôzne  $k_Y$  a  $k_X$ . Na obrázku 4.5a vidíme, že ak je parameter tvaru menší v rozdelení, kde sa tiež mení parameter škálovania, najsilnejším testom je Coxov-Mantelov. V opačnom prípade to je Gehanov test. Získané poznatky platia vo všeobecnosti pre akékoľvek  $k_Y > k_X$ , respektíve  $k_Y < k_X$ .

				hladina		sila testu	
$\lambda_X = 1$		$\lambda_Y$		1.000	1.222	1.555	2.000
test	n	m					
Wilcoxonov				0.050	0.149	0.459	0.825
Coxov-Mantelov				0.051	0.214	0.599	0.919
logrank test	50	50		0.048	0.174	0.557	0.897
z-test				0.050	0.172	0.567	0.906
t-test				0.050	0.173	0.567	0.909
Wilcoxonov				0.055	0.217	0.683	0.963
Coxov-Mantelov				0.057	0.296	0.812	0.993
logrank test	70	100		0.050	0.250	0.784	0.991
z-test				0.053	0.236	0.775	0.987
t-test				0.054	0.273	0.802	0.99
Wilcoxonov				0.055	0.095	0.273	0.629
Coxov-Mantelov				0.050	0.149	0.422	0.788
logrank test	40	25		0.048	0.119	0.378	0.742
z-test				0.044	0.140	0.409	0.786
t-test				0.043	0.096	0.319	0.694

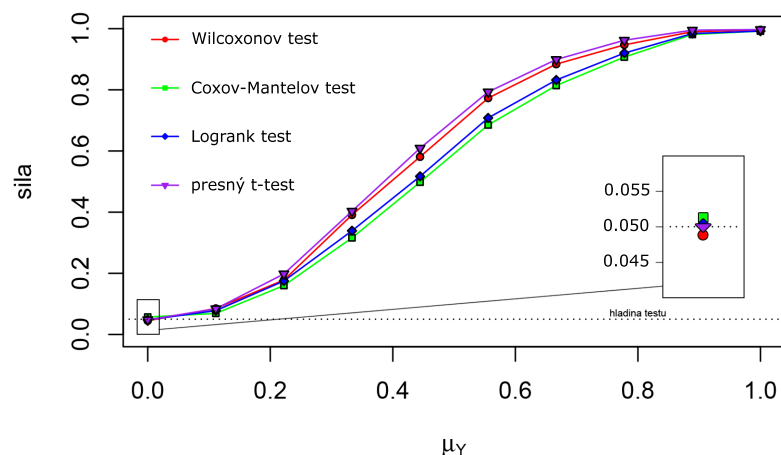
Tabulka 4.1: Porovnanie sily testov pre úplné dáta pochádzajúce z exponenciálnych rozdelení.

				hladina		sila testu	
$\mu_X = 0$		$\mu_Y$		0.000	0.222	0.555	1.000
test	n	m					
Wilcoxonov				0.049	0.159	0.384	0.903
Coxov-Mantelov				0.051	0.128	0.303	0.827
logrank test	50	50		0.050	0.145	0.328	0.850
z-test				0.050	0.165	0.398	0.915
t-test				0.050	0.165	0.398	0.915
Wilcoxonov				0.045	0.283	0.532	0.983
Coxov-Mantelov				0.051	0.237	0.453	0.965
logrank test	70	100		0.042	0.260	0.487	0.973
z-test				0.047	0.297	0.550	0.984
t-test				0.050	0.301	0.552	0.985
Wilcoxonov				0.044	0.133	0.239	0.711
Coxov-Mantelov				0.055	0.102	0.197	0.609
logrank test	40	25		0.045	0.108	0.201	0.627
z-test				0.046	0.140	0.235	0.742
t-test				0.044	0.142	0.248	0.743

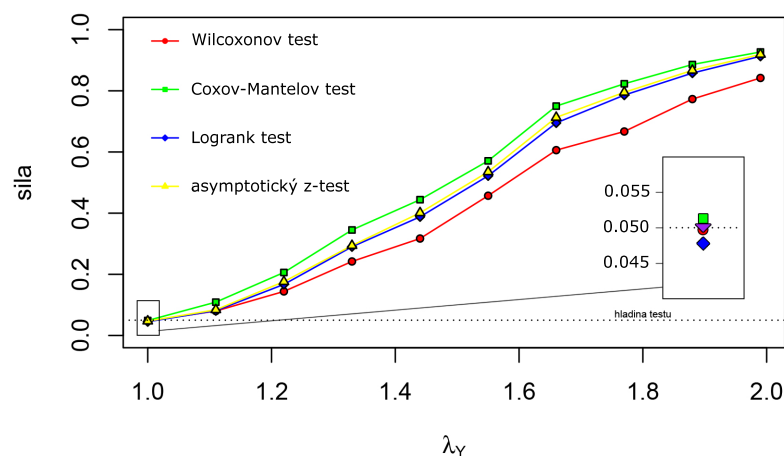
Tabulka 4.2: Porovnanie sily testov pre úplné dáta, kde  $X_1, \dots, X_n$  pochádzajú z  $LN(0,1)$  a  $Y_1, \dots, Y_m$  z  $LN(\mu_Y, 1)$ .

		<b>hladina</b>		<b>sila testu</b>		
$k_X = 0.95$		$k_Y$	0.950		0.800	
$\lambda_X = 100$		$\lambda_Y$	100.000	87.778	69.444	45.000
<b>test</b>	<b>n</b>	<b>m</b>				
Wilcoxonov			0.047	0.114	0.329	0.904
Coxov-Mantelov			0.051	0.078	0.245	0.898
logrank test	50	50	0.049	0.059	0.218	0.871
z-test			0.048	0.057	0.201	0.864
t-test			0.048	0.058	0.202	0.865
Wilcoxonov			0.043	0.139	0.510	0.988
Coxov-Mantelov			0.050	0.059	0.386	0.992
logrank test	70	100	0.045	0.052	0.355	0.988
z-test			0.045	0.049	0.314	0.983
t-test			0.043	0.046	0.333	0.988
Wilcoxonov			0.048	0.090	0.233	0.712
Coxov-Mantelov			0.061	0.074	0.209	0.734
logrank test	40	25	0.048	0.061	0.180	0.680
z-test			0.050	0.058	0.181	0.694
t-test			0.047	0.055	0.139	0.600
$k_X = 1.1$		$k_Y$	1.100		1.250	
$\lambda_X = 100$		$\lambda_Y$	100.000	87.778	69.444	45.000
Wilcoxonov			0.045	0.085	0.371	0.96
Coxov-Mantelov			0.053	0.192	0.653	0.995
logrank test	50	50	0.051	0.152	0.596	0.992
z-test			0.050	0.157	0.590	0.994
t-test			0.050	0.158	0.596	0.994
Wilcoxonov			0.044	0.099	0.573	0.996
Coxov-Mantelov			0.055	0.283	0.867	1.000
logrank test	70	100	0.051	0.242	0.824	1.000
z-test			0.049	0.210	0.818	1.000
t-test			0.049	0.249	0.850	1.000
Wilcoxonov			0.052	0.055	0.225	0.841
Coxov-Mantelov			0.054	0.112	0.430	0.962
logrank test	40	25	0.051	0.091	0.371	0.945
z-test			0.050	0.108	0.414	0.961
t-test			0.049	0.063	0.316	0.932

Tabuľka 4.3: Porovnanie sily testov pre úplné dáta, kde  $X_i \sim Weibull(k_X, \lambda_X)$  pre  $i = 1, \dots, n$  a  $Y_j \sim Weibull(k_Y, \lambda_Y)$ ,  $j = 1, \dots, m$ .



Obr. 4.3: Porovnanie sily testov pre úplné dáta z lognormálnych rozdelení s rozsahmi  $n = m = 50$ .  $\mathbf{X}$  je z  $LN(0,1)$ ,  $\mathbf{Y}$  pochádza z  $LN(\mu_Y,1)$ .

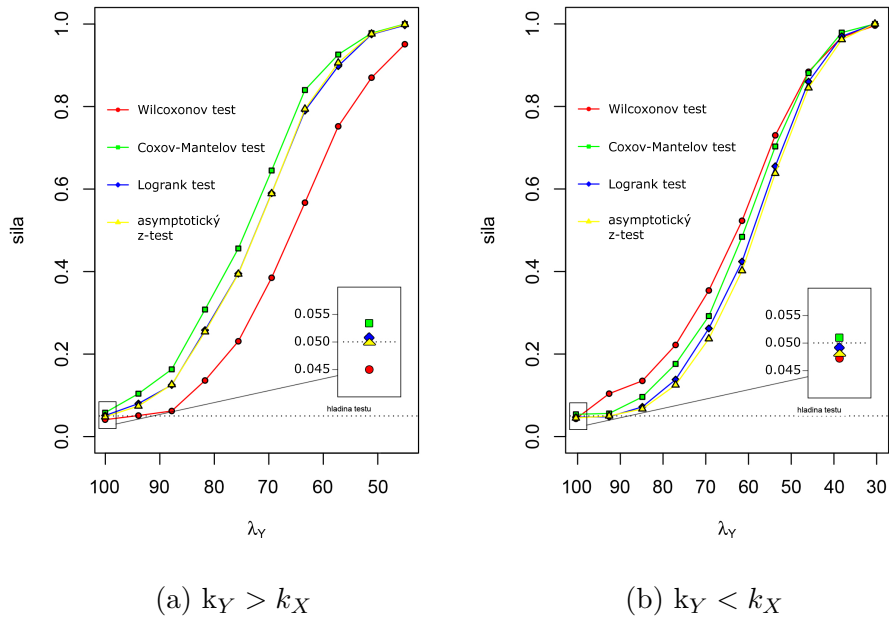


Obr. 4.4: Porovnanie sily testov pre úplné dáta z exponenciálnych  $Exp(1)$  a  $Exp(\lambda_Y)$  rozdelení s rozsahmi rovnými 50.

### 4.3 Porovnanie sily testov pre cenzorované dáta

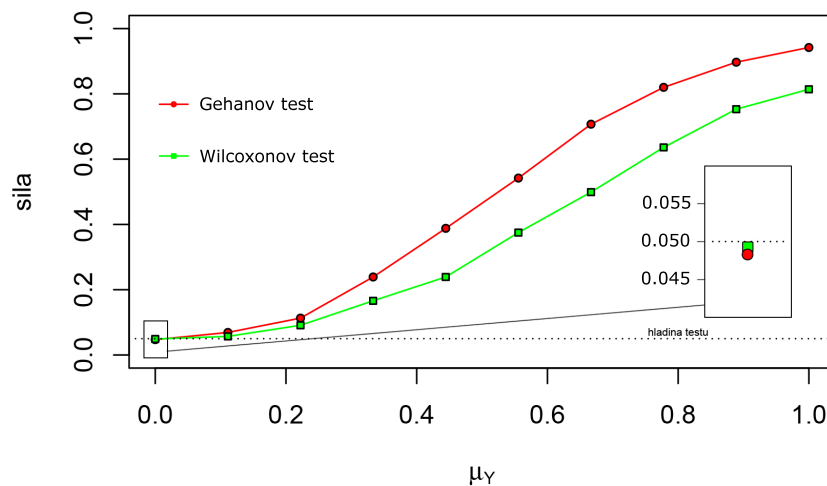
V ďalšej časti budeme uvažovať, že náhodné výbery obsahujú aj cenzorované dáta. Populácie najskôr postupne vygenerujeme z exponenciálnych, lognormálnych a Weibullových rozdelení rovnako ako to bolo uvedené na začiatku predchádzajúcej podkapitoli. Cenzorovanie budeme simulovať pomocou binomického rozdelenia, kde vygenerujeme opäť výbery, vyjadrujúce či je pozorovanie cenzorované alebo nie. Pravdepodobnosť úspechu v binomickom rozdelení predstavuje podiel cenzorovaných dát.





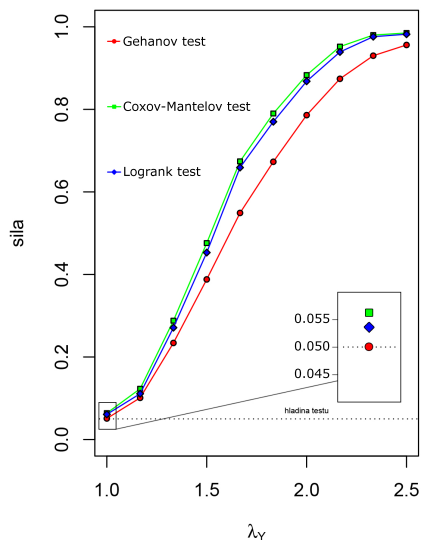
Obr. 4.5: Porovnanie sily testov pre úplné dáta z Weibulloвого rozdelenia. Rozsahy oboch náhodných výberov sú rovné 50.

Porovnáme Gehanov test s Wilcoxonovým na rovnako veľkých výberoch z lognormálnych rozdelení, pričom u Wilcoxonovho testu cenzorované dáta, ktoré tvoria 40 % výberu, vynecháme. Na obrázku 4.8 potom vidíme, že oba testy dodržia predpísanú hladinu, no Gehanov dosahuje väčšiu silu. Je to z toho dôvodu, že vynechaním cenzorovaných dát sa znížia rozsahy výberov a tiež strácame dodatočné informácie, ktoré cenzorované dáta poskytujú. V zvyšnej časti preto budeme porovnávať iba testy uvedené pre cenzorované dáta.

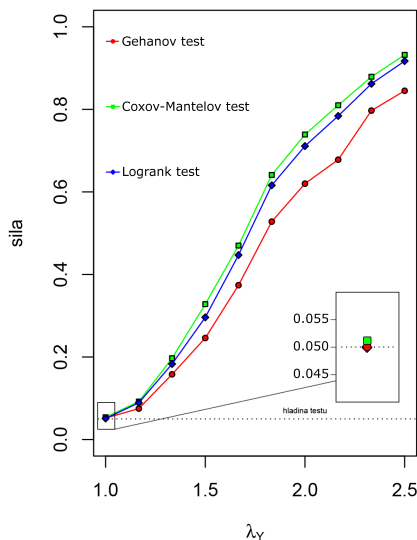


Obr. 4.6: Porovnanie sily Wilcoxonovho a Gehanovho testu pre dáta z lognormálnych rozdelení  $LN(0,1)$  a  $LN(\mu_Y,1)$  s rovnakými rozsahmi  $n = m = 50$ . Podiel cenzorovaných dát v oboch náhodných výberoch je 40 %.

Na obrázku 4.7 a v tabuľke 4.4 vidíme, že podiel cenzorovaných dát nemá na hladinu testu vplyv, no výrazne ovplyvňuje silu testu. Výsledky boli pre rôzne podiely cenzorovaných dát podobné. S narastajúcim množstvom cenzorovaných pozorovaní klesala sila testu, čo potvrdzuje porovnanie obrázkov 4.7a a 4.7b. Obrázok 4.8 a tabuľka 4.5 zase ukazujú, že ak náhodné výbery pochádzajú z lognormálnych rozdelení a líšia sa parametrami  $\mu_X$  a  $\mu_Y$ , sú ich distribučné funkcie posunuté a najsilnejší je Gehanov test. Výsledky simulácií pre prípad Weibullovoho rozdelenia nájdeme v tabuľke 4.6 a na obrázku 4.9. Vidíme, že ak je  $k_Y < k_X$ , najsilnejším testom Gehanov, v opačnom prípade je najsilnejší Coxov-Mantelov test.

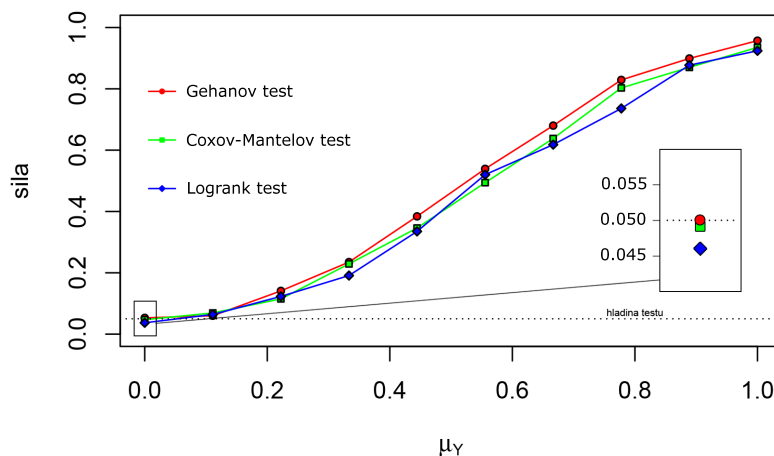


(a) Náhodné výbery s 10% podielom cenzorovaných dát.



(b) Náhodné výbery so 40% podielom cenzorovaných dát.

Obr. 4.7: Porovnanie sily testov pre cenzorované dáta z exponenciálnych rozdelení  $Exp(1)$  a  $Exp(\lambda_Y)$  s rozsahmi  $n = m = 50$



Obr. 4.8: Porovnanie sily testov pre cenzorované dáta z lognormálnych rozdelení s rovnakými rozsahmi a 40% podielom cenzorovaných dát.

podiel cenzor. dát	$\lambda_X = 0$ test			hladina		sila testu	
		n	m	1.000	1.222	1.555	2.000
10 %	Gehanov			0.050	0.122	0.415	0.778
	Coxov-Mantelov	50	50	0.056	0.169	0.520	0.883
	logrank test			0.054	0.156	0.50	0.872
	Gehanov			0.044	0.223	0.639	0.943
	Coxov-Mantelov	70	100	0.047	0.237	0.758	0.981
	logrank test			0.044	0.221	0.745	0.980
	Gehanov			0.052	0.100	0.276	0.585
	Coxov-Mantelov	40	25	0.053	0.097	0.340	0.694
	logrank test			0.046	0.112	0.362	0.684
40 %	Gehanov			0.050	0.101	0.301	0.635
	Coxov-Mantelov	50	50	0.051	0.121	0.394	0.752
	logrank test			0.050	0.108	0.372	0.735
	Gehanov			0.052	0.128	0.464	0.847
	Coxov-Mantelov	70	100	0.054	0.166	0.581	0.929
	logrank test			0.052	0.154	0.563	0.920
	Gehanov			0.047	0.066	0.196	0.428
	Coxov-Mantelov	40	25	0.051	0.077	0.246	0.553
	logrank test			0.045	0.071	0.230	0.520

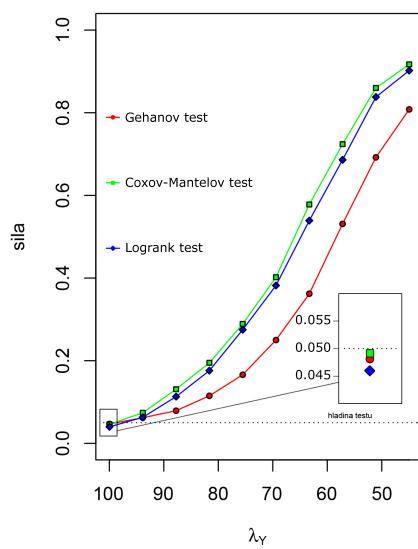
Tabuľka 4.4: Porovnanie sily testov pre dáta pochádzajúce z exponenciálnych rozdelení pre rôzny podiel cenzorovania.

$\mu_X = 0$ test			hladina		sila testu	
	n	m	0.000	0.222	0.555	1.000
Gehanov			0.050	0.107	0.522	0.953
Coxov-Mantelov	50	50	0.049	0.105	0.490	0.924
logrank test			0.046	0.098	0.474	0.917
Gehanov			0.051	0.160	0.765	0.999
Coxov-Mantelov	70	100	0.049	0.153	0.735	0.995
logrank test			0.048	0.146	0.724	0.995
Gehanov			0.052	0.090	0.360	0.823
Coxov-Mantelov	40	25	0.057	0.077	0.305	0.761
logrank test			0.049	0.068	0.280	0.731

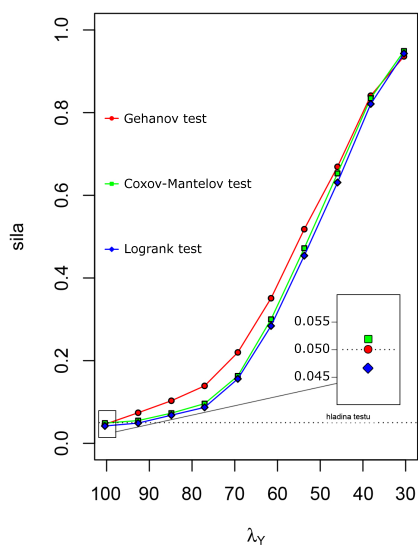
Tabuľka 4.5: Porovnanie sily testov pre dáta so 40 % podielom cenzorovaných dát, kde  $\mathbf{X}$  s rozsahom  $n$  pochádza z  $LN(0,1)$  a  $\mathbf{Y}$  rozsahu  $m$  z  $LN(\mu_Y,1)$ .

				hladina		sila testu	
$k_X = 0.95$		$k_Y$		0.950		0.800	
$\lambda_X = 100$		$\lambda_Y$		100.000	87.778	69.444	45.000
<b>test</b>	<b>n</b>	<b>m</b>					
Gehanov				0.050	0.084	0.229	0.710
Coxov-Mantelov	50	50		0.052	0.058	0.175	0.707
logrank test				0.047	0.055	0.167	0.697
Gehanov				0.048	0.099	0.326	0.861
Coxov-Mantelov	70	100		0.048	0.059	0.223	0.853
logrank test				0.047	0.055	0.217	0.845
Gehanov				0.051	0.076	0.155	0.457
Coxov-Mantelov	40	25		0.053	0.062	0.139	0.486
logrank test				0.044	0.056	0.130	0.468
$k_X = 1.1$		$k_Y$		1.100		1.250	
$\lambda_X = 100$		$\lambda_Y$		100.000	87.778	69.444	45.000
Gehanov				0.050	0.058	0.173	0.589
Coxov-Mantelov	50	50		0.054	0.092	0.266	0.762
logrank test				0.047	0.079	0.246	0.728
Gehanov				0.049	0.061	0.381	0.958
Coxov-Mantelov	70	100		0.057	0.140	0.624	0.992
logrank test				0.055	0.129	0.600	0.990
Gehanov				0.050	0.058	0.173	0.589
Coxov-Mantelov	40	25		0.054	0.092	0.266	0.762
logrank test				0.047	0.079	0.246	0.728

Tabuľka 4.6: Porovnanie sily testov pre dáta pochádzajúce z Weibullových rozdelení s parametrami tvaru  $k_X, k_Y$  a parametrami škálovania  $\lambda_X, \lambda_Y$ . Podiel cenzorovaných dát v oboch náhodných výberoch je 40 %.



(a)  $k_Y > k_X$



(b)  $k_Y < k_X$

Obr. 4.9: Porovnanie sily testov pre cenzorované dáta z Weibullovoho rozdelenia s rovnakými rozsahmi oboch náhodných výberov. Podiel cenzorovaných dát v oboch výberoch je približne 40 %.

# Záver

V práci sme predstavili niektoré základné dvojvýberové testy, odvodili ich skutočnú silu a porovnali ju pre rôzne rozsahy náhodných výberov či rôzne zvolenú nulovú a alternatívnu hypotézu. Skutočnú silu sme následne porovnali s empirickou, ktorú sme odhadli na základe simulácií.

Ďalej sme uviedli niektoré testy, ktoré je vhodné použiť ak analyzované dáta nie sú kompletne, ale obsahujú cenzorované pozorovania. Ukázali sme, že ak vynecháme cenzorované pozorovania a použijeme Wilcoxonov test, dosahuje menšiu silu ako Gehanov test, ktorý je modifikáciou Wilcoxonovho pre cenzorované dáta. Ďalej sme ukázali, že podiel cenzorovaných dát nemá vplyv na hladinu testu, no nepriamo úmerne ovplyvňuje silu testu. S rastúcim množstvom cenzorovaných pozorovaní bola sila testu menšia.

Uvedené testy sme tiež porovnali pre prípad úplných dát. Ak náhodné výbery pochádzajú z exponenciálnych rozdelení, najsilnejším testom bol Coxov-Mantelov. V prípade lognormálnych rozdelení sme použili transformáciu logaritmom a najsilnejším testom bol presný t-test. Ak dáta pochádzali z Weibullovoho rozdelenia s rovnakými parametrami tvaru, získané výsledky boli veľmi podobné ako v prípade exponenciálnych rozdelení. Ak bol parameter tvaru väčší v rozdelení, kde sa tiež menili parametre škálovania, najsilnejším bol opäť Coxov-Mantelov test, v opačnom prípade to bol Wilcoxonov test, respektíve Gehanov test v prípade cenzorovaných dát.

# Zoznam použitej literatúry

CHOW, S.-C., SHAO, J. a WANG, H. (2008). *Sample size calculations in clinical research*. Second Edition. Taylor Francis Group, LLC, Boca Raton. ISBN 978-1-58488-982-3.

GEHAN, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, **52**, 203–223.

LEE, E. T. a WANG, J. (2003). *Statistical methods for survival data analysis*. Third Edition. John Wiley Sons, Inc., Hoboken. ISBN 0-471-36997-7.

OMELKA, M. (2019). Poznámky k přednášce NMSA331 Matematická statistika. URL <https://www.karlin.mff.cuni.cz/~omelka/Soubory/nmsa331/ms1.pdf>. [Přístup 17-Júl-2020].