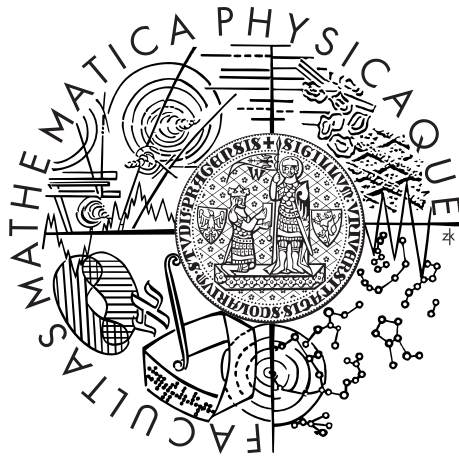


Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Bc. Lucie Pekařová

Metody pro analýzu změny od počáteční hodnoty ke konečné

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D.

Studijní program: Matematika

Studijní obor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2012

Chtěla bych poděkovat vedoucímu diplomové práce doc. Mgr. Michalu Kulichovi, Ph.D. za to, že se mi po celou dobu psaní této práce trpělivě věnoval a poskytoval mi cenné rady a připomínky. Dále bych chtěla poděkovat svému partnerovi za poskytnutí psychické podpory, která mi v průběhu práce velmi pomohla.

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

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

V Praze dne

Lucie Pekařová

Název práce: Metody pro analýzu změny od počáteční hodnoty ke konečné

Autor: Bc. Lucie Pekařová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V této práci se zabýváme odhadem efektu léčby v klinických randomizovaných studiích. Odhady efektu léčby konstruujeme na základě tří modelů. V první části práce se zabýváme chováním těchto odhadů za předpokladu, že efekt léčby nemusí být u každého pacienta stejný. Zjišťujeme, že všechny typy odhadů jsou konzistentní, a uvádíme jejich asymptotické rozdělení. Tyto odhady teoreticky porovnáváme na základě jejich asymptotických rozptylů. Teoretické závěry pak potvrzujeme simulační studií. V druhé části práce popisujeme případ, kdy měření počátečních a konečných hodnot pacientů je měřeno s chybou. Zde se zabýváme chováním dvou odhadů. Zjišťujeme, že oba odhady jsou konzistentní, a konstruujeme jejich asymptotické rozdělení. Tyto odhady také teoreticky porovnáváme.

Klíčová slova: randomizovaná studie, počáteční a konečná hodnota, odhad efektu léčby

Title: Methods for Analyzing Change From Baseline to Final Assessment

Author: Bc. Lucie Pekařová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Mgr. Michal Kulich, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this thesis, we analyze treatment effect estimate in randomized clinical studies. Treatment effect estimates are constructed on the basis of three models. The first part of this thesis is about the behaviour of these estimates when the treatment effects vary with patients. We find out that all types of estimates are consistent and we derived their asymptotic distribution. The estimates are compared by their asymptotic variances. The theoretical conclusions are confirmed by a simulation study. The second part describes the case where measurements of baseline and final values contain an error. Two estimates are analyzed. We find out that both estimates are consistent. We derive their asymptotic distribution and compare their variances.

Keywords: randomized study, baseline and final assessment, treatment effect estimate

Obsah

Úvod	1
1 Situace	2
2 Analýza nekonstantního efektu léčby	5
2.1 Model I a odhad efektu léčby $\hat{\delta}^I$ za jeho platnosti	5
2.2 Model II a odhad efektu léčby $\hat{\delta}^{II}$ za jeho platnosti	8
2.3 Model III a odhad efektu léčby $\hat{\delta}^{III}$ za jeho platnosti	11
3 Chování odhadů za nekonstantního efektu léčby	18
3.1 Chování odhadu $\hat{\delta}^I$ za platnosti modelu II	18
3.2 Chování odhadu $\hat{\delta}^{III}$ za platnosti modelu II	19
3.3 Chování odhadu $\hat{\delta}^I$ za platnosti modelu III	20
3.4 Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu III	21
3.5 Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu I	23
3.6 Chování odhadu $\hat{\delta}^{III}$ za platnosti modelu I	24
3.7 Porovnání chování odhadů	32
4 Simulační studie nekonstatního efektu léčby	35
4.1 Konečná hodnota pocházející z modelu II	36
4.2 Konečná hodnota pocházející z modelu III	37
4.3 Konečná hodnota pocházející z modelu I	40

5	Analýza efektu léčby při měření s chybou	49
5.1	Model I a odhad efektu léčby $\hat{\delta}^I$ za jeho platnosti	50
5.2	Model II a odhad efektu léčby $\hat{\delta}^{II}$ za jeho platnosti	52
6	Chování odhadů při měření s chybou	55
6.1	Chování odhadu $\hat{\delta}^I$ za platnosti modelu II	55
6.2	Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu I	56
	Závěr	58
	Seznam použité literatury	60
	Seznam použitých symbolů	61
	Přílohy	64

Úvod

Před zavedením nového léku na trh musí být tento lék otestován v klinických studiích. V těchto studiích se pacienti rozdělí na dvě skupiny. První skupině se podá lék a druhá dostane placebo. Většina klinických studií je randomizovaných, což znamená, že pacienti se do těchto skupin dělí náhodně. Pacientům je měřena veličina, na kterou má daný lék působit - například hmotnost či krevní tlak.

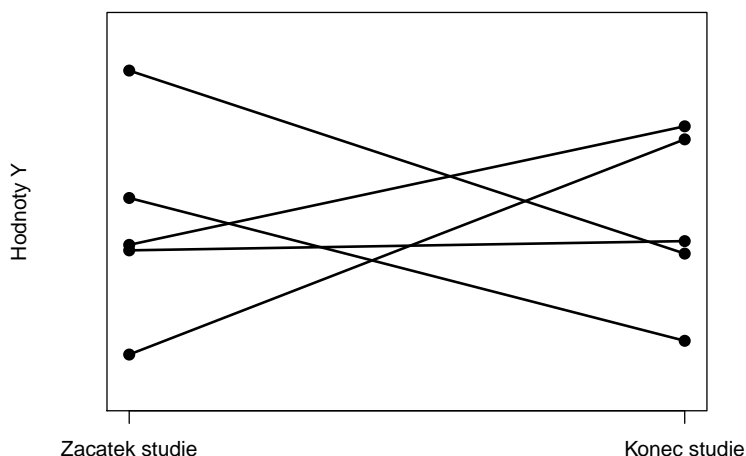
Po změření hodnot pacientů na začátku a konci studie nastupuje analýza, jejíž hlavním cílem je odhadnutí efektu léčby. Pro otestování vlivu léčby na danou měřenou veličinu lze zvolit jeden ze tří modelů. Mnoho matematiků tyto modely srovnávají za různých předpokladů a pokouší se nalézt optimální přístup. Názory na optimální přístup se však dodnes liší. Například Stephen Senn v článku [4] se zabývá analýzou efektu léčby za předpokladu, že střední hodnoty v experimentální a kontrolní skupině na začátku studie nemusí být stejné. V článku [5] se zase Yu-Kang Tu a Mark S. Gilthorpe zabývají porovnáním modelů za předpokladu, že zkoumanou veličinu měříme s chybou.

Donedávna však chyběla práce, která by analýzu efektu léčby uceleně shrnovala. Andrea Pacáková ve své diplomové práci [1] navržené přístupy vystavěla od základů a přehledně je porovnala za určitých předpokladů. Předpokládala, že se jedná o randomizovanou studii, každý pacient má stejný efekt léčby a zkoumanou veličinu měříme bez chyb. Tyto předpoklady jsou však velmi omezující.

V této práci výsledky z [1] zobecníme. Zaměříme se hlavně na situaci, kdy efekt léčby pro každého pacienta nemusí být stejný. Tuto situaci zkoumáme v kapitolách 2 a 3 teoreticky a simulačně v kapitole 4. Dále se zabýváme situací, kdy je veličina měřena s určitou chybou. Tuto situaci popisujeme v kapitolách 5 a 6. V obou situacích prozkoumáme odhady efektu léčby pomocí zmíněných modelů a následně tyto přístupy porovnáme mezi sebou.

1. Situace

Předpokládejme, že na začátku studie máme n pacientů, kterým měříme veličinu Y . Hodnoty této veličiny měříme pacientům na začátku studie a na konci. Počáteční, resp. konečnou hodnotu veličiny i -tého pacienta rozlišujeme 0, resp. 1 v dolním indexu Y (tj. Y_{0i} , resp. Y_{1i}). Na obrázku 1.1 vidíme hodnoty veličiny Y , které byly naměřeny pacientům na začátku a konci studie.



Obrázek 1.1: Počáteční a konečné hodnoty pěti pacientů.

Pacienty rozdělíme náhodně do dvou skupin - experimentální a kontrolní. Pacientům v experimentální skupině je podán lék a pacientům v kontrolní skupině je podáno placebo. Předpokládejme, že pacient je přiřazen do experimentální skupiny s pravděpodobností $\pi \in (0, 1)$. Přiřazení i -tého pacienta do experimentální skupiny popisujeme náhodnou veličinou $I_{[i \in E]}$, pro kterou platí

$$\mathcal{L}(I_{[i \in E]}) = \text{Alt}(\pi), \quad i = 1, \dots, n.$$

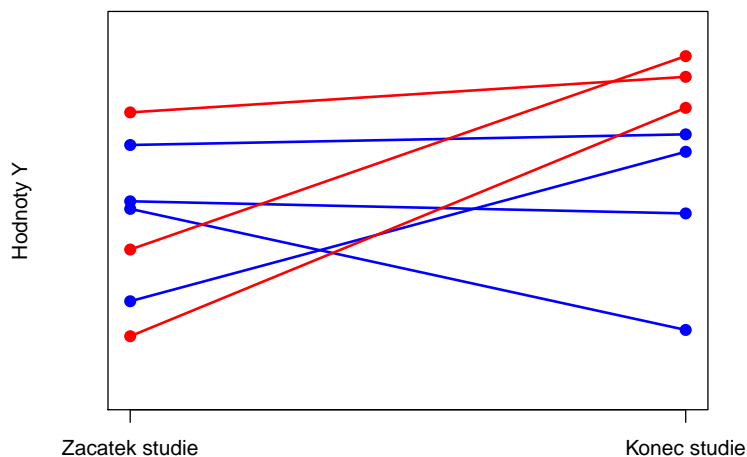
Předpokládejme, že veličiny $I_{[i \in E]}$, $i = 1, \dots, n$, jsou vzájemně nezávislé a zároveň nezávislé na počátečních hodnotách Y_{0i} . Rozdělení pacientů do experimentální a kontrolní skupiny a jejich hodnoty jsou znázorněny na obrázku 1.2.

Označme $n_E = \sum_{i=1}^n I_{[i \in E]}$ počet pacientů přiřazených do experimentální skupiny a $n_C = n - n_E$ počet pacientů v kontrolní skupině. Dále označme poměr počtu pacientů v experimentální skupině a kontrolní skupině

$$q_n = \frac{n_E}{n_C}.$$

Ze zákona velkých čísel platí

$$\frac{1}{n} \sum_{i=1}^n I_{[i \in E]} \xrightarrow{P} \pi, \quad n \rightarrow \infty,$$



Obrázek 1.2: Počáteční a konečné hodnoty pacientů - tyto hodnoty jsou v experimentální skupině značeny červeně, v kontrolní skupině modře.

proto

$$q_n \xrightarrow{P} \frac{\pi}{1-\pi}, \quad n \rightarrow \infty.$$

Označme $q = \frac{\pi}{1-\pi}$. Platí $q \in (0, \infty)$. Proto když $n_C \rightarrow \infty$, tak $n_E \rightarrow \infty$.

Označme písmenem E , resp. C v horním indexu počáteční a konečné hodnoty příslušnost i -tého pacienta v experimentální, resp. kontrolní skupině. Po přecházení získáváme dvojice pozorování

$$\begin{aligned} (Y_{0i}^E, Y_{1i}^E), & \quad i = 1, \dots, n_E, \\ (Y_{0i}^C, Y_{1i}^C), & \quad i = 1, \dots, n_C. \end{aligned} \tag{1.1}$$

Díky randomizaci máme

$$\mathcal{L}(Y_{0i}^E | i \in E) = \mathcal{L}(Y_{0i} | i \in C).$$

Označme

$$\mu_0 = \mathbb{E} Y_{0i}^E = \mathbb{E} Y_{0i}^C \quad \text{a} \quad \sigma_0^2 = \text{var} Y_{0i}^E = \text{var} Y_{0i}^C.$$

Na konci studie máme k dispozici hodnoty (1.1). Naším cílem je z těchto hodnot odhadnout efekt léčby, který představuje střední hodnotu rozdílu v konečných hodnotách mezi experimentální a kontrolní skupinou. Tento efekt léčby budeme značit δ . Pro modelování závislosti konečné hodnoty na počáteční uvedeme tři modely. První model (dále budeme značit jako model I) využívá toho, že se jedná o randomizovanou studii. Vyjadřuje konečnou hodnotu pacientů jako součet její střední hodnoty a náhodné chyby. Tento model platí obecně vždy a kromě

nezávislosti náhodných chyb nemusíme nic dalšího předpokládat. Určitou nevýhodou však představuje to, že tento model nebere v úvahu počáteční hodnoty pacientů. Druhý model (model II) se už počátečními hodnotami zabývá - vyjadřuje rozdíl mezi konečnou a počáteční hodnotou jako součet konstanty a náhodné chyby. Model II předpokládá, že tento rozdíl na počáteční hodnotě nezávisí, což přesně určuje tvar závislosti konečné hodnoty na počáteční. Poslední model (model III) je regresní model, který předpokládá obecnou lineární závislost konečné hodnoty na počáteční. Speciálním případem modelu III je model II, v kterém je závislost konečné hodnoty na počáteční se směrnici jedna.

Uvedené tři modely a jejich náležité odhady efektu léčby již byly prozkoumány a shrnuty v práci [1]. Výsledným doporučením této práce bylo vždy vzít jako odhad efektu léčby odhad založený na modelu III. Práce [1] má však omezující předpoklady. Předpokládal se zde konstantní efekt léčby, konstantní rozptyl v experimentální skupině a konstantní rozptyl v kontrolní skupině. Navíc se neuvažovala situace, kdy získané počáteční a konečné hodnoty pacientů jsou měřeny s chybou. V této práci výsledky [1] zobecníme a zjistíme, zda v jiné situaci není lepší jiný odhad efektu léčby. V kapitolách 2, 3 a 4 se budeme zabývat situací, kdy se experimentální skupina rozdělí na dvě podskupiny a v každé podskupině může mít léčba jiný efekt. V kapitolách 5 a 6 prozkoumáme situaci, kdy měření počáteční i konečné hodnoty je prováděno s chybou.

2. Analýza nekonstantního efektu léčby

V této kapitole se zaměříme na situaci, kdy podání léku nemá stejný účinek na každém pacientovi. Předpokládejme, že pacienti v experimentální skupině se rozdělí na dvě podskupiny podle náhodné veličiny $Z_i \in \{0, 1\}$, $i = 1, \dots, n_E$, a v každé podskupině může mít podaná léčba jiný efekt. V praxi nedokážeme napozorovat, do které podskupiny pacient patří. Podskupiny mohou představovat skupiny lidí, kteří mají jiným způsobem pracující enzymy. Může se stát, že efekt léčby v první podskupině může být zcela opačný než ve druhé podskupině. Je tedy otázkou, zda se celková léčba vůbec vyplatí, tj. zda je výsledný efekt léčby nenulový.

Předpokládejme, že veličiny Z_i jsou nezávislé náhodné veličiny s alternativním rozdělením s parametrem $p = \mathbf{P}(Z_i = 1) \in (0, 1)$. Dále předpokládejme, že identifikátor $\mathbb{1}_{[Z_i=1]}$ nezávisí na podskupině Z_i , počáteční hodnota Y_{0i}^E však na podskupině Z_i záviset může. Efekt léčby ve skupině $Z_i = 0$ označíme δ_0 , ve skupině $Z_i = 1$ označíme efekt léčby δ_1 . Odhadujeme celkový efekt léčby

$$\delta = (1 - p) \delta_0 + p \delta_1.$$

Pro naši analýzu máme k dispozici naměřené hodnoty na počátku a konci studie (1.1), neznáme však příslušnost pacienta v podskupinách $Z_i = 0$ a $Z_i = 1$ v experimentální skupině. Hodnoty na konci studie mohou být ovlivněny počáteční hodnotou, léčbou a podskupinou Z_i . Nevíme však jakým způsobem. Pro modelování těchto vlivů uvedeme již zmíněné tři modely. Na jejich základě odhadneme celkový efekt léčby. Zjistíme, jak se tyto odhady chovají, sestrojíme test hypotézy nulovosti efektu léčby a získáme asymptotické rozdělení daného odhadu.

2.1 Model I a odhad efektu léčby $\hat{\delta}^I$ za jeho platnosti

První model, kterým se budeme zabývat, budeme označovat jako model I. Tento model se nezabývá počátečními hodnotami Y_{0i} , vyjadřuje pouze konečné hodnoty Y_{1i}^C ve formě obecného zápisu náhodné veličiny a konečné hodnotě Y_{1i}^E navíc připisuje efekt léčby v závislosti na Z_i . Přesný zápis modelu I je

$$\begin{aligned} Y_{1i}^C &= \mu + \varepsilon_i^{I,C}, \\ Y_{1i}^E &= \mu + \delta_0 \mathbb{1}_{[Z_i=0]} + \delta_1 \mathbb{1}_{[Z_i=1]} + \varepsilon_i^{I,E} \end{aligned}$$

za podmínek

$$\begin{aligned} \mathbf{E} \varepsilon_i^{I,C} &= 0, & i &= 1, \dots, n_C, \\ \mathbf{E} \varepsilon_i^{I,E} &= 0, \quad \mathbf{E} \left(\varepsilon_i^{I,E} \mid Z_i \right) = 0, & i &= 1, \dots, n_E, \\ \mathbf{var} \varepsilon_i^{I,C} &= \sigma_{I,C}^2, & i &= 1, \dots, n_C, \\ \mathbf{var} \left(\varepsilon_i^{I,E} \mid Z_i \right) &= \sigma_{I,E,0}^2 \mathbf{1}_{[Z_i=0]} + \sigma_{I,E,1}^2 \mathbf{1}_{[Z_i=1]}, & i &= 1, \dots, n_E, \end{aligned}$$

a $\varepsilon_1^{I,C}, \dots, \varepsilon_{n_C}^{I,C}, \varepsilon_1^{I,E}, \dots, \varepsilon_{n_E}^{I,E}$ jsou nezávislé. Náhodné chyby $\varepsilon_i^{I,E}$ a $\varepsilon_i^{I,C}$ mohou záviset na počáteční hodnotě Y_{0i}^E, Y_{0i}^C .

Předpokládejme, že platí výše uvedený model I. My však máme k dispozici pouze hodnoty (1.1) a nevíme, zda pacient patří do podskupiny $Z_i = 0$ nebo $Z_i = 1$. Celkový efekt léčby tedy odhadneme na základě modelu I z práce [1], ve které se předpokládá, že efekt léčby je u každého pacienta stejný. Odhad efektu léčby na základě modelu I z práce [1] je

$$\hat{\delta}^I = \bar{Y}_1^E - \bar{Y}_1^C,$$

kde

$$\bar{Y}_1^E = \frac{1}{n_E} \sum_{i=1}^{n_E} Y_{1i}^E \quad \text{a} \quad \bar{Y}_1^C = \frac{1}{n_C} \sum_{i=1}^{n_C} Y_{1i}^C.$$

Dále se budeme zabývat vlastnostmi tohoto odhadu za platnosti modelu I, sestrojíme testovou statistiku pro testování hypotézy nulovosti efektu léčby a určíme asymptotické rozdělení tohoto odhadu.

Následující věta uvádí, k čemu odhad $\hat{\delta}^I$ konverguje v pravděpodobnosti.

Věta 2.1. *Za platnosti modelu I*

$$\hat{\delta}^I \xrightarrow{P} \delta, \quad n \rightarrow \infty,$$

kde $\delta = \delta_0 (1 - p) + \delta_1 p$.

Důkaz. Střední hodnota odhadu je

$$\mathbf{E} \hat{\delta}^I = \mathbf{E} \bar{Y}_1^E - \mathbf{E} \bar{Y}_1^C = \mu + \delta_0 (1 - p) + \delta_1 p - \mu = \delta_0 (1 - p) + \delta_1 p.$$

Rozptyl odhadu $\hat{\delta}^I$ se rovná

$$\mathbf{var} \hat{\delta}^I = \mathbf{var} \bar{Y}_1^C + \mathbf{var} \bar{Y}_1^E = \frac{1}{n_C^2} \sum_{i=1}^{n_C} \mathbf{var} Y_{1i}^C + \frac{1}{n_E^2} \sum_{i=1}^{n_E} \mathbf{var} Y_{1i}^E = \frac{\sigma_{I,C}^2}{n_C} + \frac{\mathbf{var} Y_{1i}^E}{n_E},$$

kde

$$\begin{aligned} \mathbf{var} Y_{1i}^E &= \mathbf{var} \mathbf{E} (Y_{1i}^E \mid Z_i) + \mathbf{E} \mathbf{var} (Y_{1i}^E \mid Z_i) \\ &= \mathbf{var} (\mu + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]}) + \mathbf{E} (\sigma_{I,E,0}^2 \mathbf{1}_{[Z_i=0]} + \sigma_{I,E,1}^2 \mathbf{1}_{[Z_i=1]}) \\ &= \mathbf{var} (\delta_0 + (\delta_1 - \delta_0) \mathbf{1}_{[Z_i=1]}) + p \sigma_{I,E,1}^2 + (1 - p) \sigma_{I,E,0}^2 \\ &= p (1 - p) (\delta_1 - \delta_0)^2 + p \sigma_{I,E,1}^2 + (1 - p) \sigma_{I,E,0}^2. \end{aligned}$$

Z předchozích rovností a ze slabého zákona velkých čísel plyne, že odhad $\hat{\delta}^I$ konverguje k efektu léčby $\delta = \delta_0 (1 - p) + \delta_1 p$ v pravděpodobnosti pro $n \rightarrow \infty$. \square

Testujme hypotézu $H_0 : \delta = 0$ proti alternativě $H_1 : \delta \neq 0$. K tomu využijeme následující větu.

Věta 2.2. *Nechť platí model I. Označme*

$$t_I = \frac{\bar{Y}_1^E - \bar{Y}_1^C}{\sqrt{\frac{S_{I,C}^2}{n_C} + \frac{S_{I,E}^2}{n_E}}},$$

kde

$$S_{I,C}^2 = \frac{1}{n_C - 1} \sum_{i=1}^{n_C} (Y_{1i}^C - \bar{Y}_1^C)^2 \quad \text{a} \quad S_{I,E}^2 = \frac{1}{n_E - 1} \sum_{i=1}^{n_E} (Y_{1i}^E - \bar{Y}_1^E)^2.$$

Je-li $\delta = 0$, pak platí

$$t_I \xrightarrow{d} \mathbf{N}(0, 1).$$

Důkaz. Z centrální limitní věty máme

$$\sqrt{n_C} (\bar{Y}_1^C - \mu) \xrightarrow{d} \mathbf{N}(0, \sigma_{I,C}^2), \quad n_C \rightarrow \infty,$$

a

$$\sqrt{n_E} (\bar{Y}_1^E - (\mu + \delta)) \xrightarrow{d} \mathbf{N}(0, \text{var } Y_{1i}^E), \quad n_E \rightarrow \infty,$$

z čehož plyne

$$\sqrt{n_C} (\bar{Y}_1^E - (\mu + \delta)) \xrightarrow{d} \mathbf{N}\left(0, \frac{1}{q} \text{var } Y_{1i}^E\right), \quad n_C \rightarrow \infty.$$

Z předchozích konvergencí a z nezávislosti \bar{Y}_1^E a \bar{Y}_1^C plyne za platnosti $\delta = 0$

$$\frac{\bar{Y}_1^E - \bar{Y}_1^C}{\sqrt{\frac{\sigma_{I,C}^2}{n_C} + \frac{\text{var } Y_{1i}^E}{q n_C}}} \xrightarrow{d} \mathbf{N}(0, 1).$$

Protože jsou $S_{I,C}^2$ a $S_{I,E}^2$ konzistentní odhady $\sigma_{I,C}^2$ a $\text{var } Y_{1i}^E$, platí za $\delta = 0$ dle Cramérovy-Sluckého věty

$$t_I = \frac{\bar{Y}_1^E - \bar{Y}_1^C}{\sqrt{\frac{S_{I,C}^2}{n_C} + \frac{S_{I,E}^2}{n_E}}} \xrightarrow{d} \mathbf{N}(0, 1).$$

□

Z předchozí věty jsme získali testovou statistiku t_I , která má za platnosti H_0 asymptoticky normované normální rozdělení. Hypotézu H_0 tedy zamítneme na hladině α , jestliže $|t_I| \geq u_{1-\frac{\alpha}{2}}$.

V další větě získáváme asymptotické rozdělení odhadu $\hat{\delta}^I$.

Věta 2.3. *Za platnosti modelu I je*

$$\sqrt{n} (\hat{\delta}^I - \delta) \xrightarrow{d} \mathbf{N}(0, V_{\hat{\delta}^I}),$$

kde

$$V_{\hat{\delta}^I} = (q + 1) \left(\sigma_{I,C}^2 + \frac{1}{q} \left(p(1-p)(\delta_0 - \delta_1)^2 + p\sigma_{I,E,1}^2 + (1-p)\sigma_{I,E,0}^2 \right) \right). \quad (2.1)$$

Důkaz. Jelikož

$$\begin{aligned}\text{var } \bar{Y}_1^C &= \frac{1}{n_C} \sigma_{I,C}^2, \\ \text{var } \bar{Y}_1^E &= \frac{1}{n_E} \left(p(1-p)(\delta_1 - \delta_0)^2 + p\sigma_{I,E,1}^2 + (1-p)\sigma_{I,E,0}^2 \right),\end{aligned}$$

použitím důkazu věty 2.2 a konvergence

$$\frac{n}{n_C} = 1 + q_n \rightarrow 1 + q,$$

získáváme tvrzení věty. □

Asymptotický rozptyl odhadu $\hat{\delta}^I$ z (2.1) se oproti asymptotickému rozptylu v práci [1] liší. Pokud bychom předpokládali, že rozptyly jsou v podskupinách $Z_i = 0$ a $Z_i = 1$ stejné, tj. $\sigma_{I,E,0}^2 = \sigma_{I,E,1}^2 = \sigma_{I,E}^2$, pak by asymptotický rozptyl z (2.1) byl větší o

$$\frac{q+1}{q} p(1-p)(\delta_1 - \delta_0)^2$$

než asymptotický rozptyl v práci [1].

V této podkapitole jsme dokázali, že odhad $\hat{\delta}^I$ je konzistentní odhad efektu léčby za platnosti modelu I a sestrojili jsme testovou statistiku hypotézy nulovosti efektu léčby. Tyto výsledky se shodují s prací [1]. Dále jsme zjistili asymptotické rozdělení odhadu $\hat{\delta}^I$. V speciálním případě konstantních rozptylů v experimentální skupině jako v práci [1] by asymptotický rozptyl z (2.1) byl větší než v práci [1].

2.2 Model II a odhad efektu léčby $\hat{\delta}^{II}$ za jeho platnosti

Dalším modelem, který popíšeme, je model II. Tento model předpokládá, že změna $Y_{1i}^C - Y_{0i}^C$ nezávisí na počáteční hodnotě Y_{0i}^C a popisuje střední hodnotu rozdílu konečné a počáteční hodnoty $Y_{1i}^C - Y_{0i}^C$ jako konstantu. V experimentální skupině se navíc připisuje podle podskupiny Z_i příslušný efekt léčby. Přesný tvar modelu II je

$$\begin{aligned}Y_{1i}^C - Y_{0i}^C &= \gamma + \varepsilon_i^{II,C}, \\ Y_{1i}^E - Y_{0i}^E &= \gamma + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{II,E}\end{aligned}$$

za podmínek

$$\begin{aligned}\mathbb{E} \varepsilon_i^{II,C} &= 0, & \mathbb{E} \left(\varepsilon_i^{II,C} \mid Y_{0i}^C \right) &= 0, & i &= 1, \dots, n_C, \\ \mathbb{E} \varepsilon_i^{II,E} &= 0, & \mathbb{E} \left(\varepsilon_i^{II,E} \mid Y_{0i}^E, Z_i \right) &= 0, & i &= 1, \dots, n_E, \\ \text{var } \varepsilon_i^{II,C} &= \sigma_{II,C}^2, & & & i &= 1, \dots, n_C, \\ \text{var} \left(\varepsilon_i^{II,E} \mid Z_i \right) &= \sigma_{II,E,0}^2 \mathbf{1}_{[Z_i=0]} + \sigma_{II,E,1}^2 \mathbf{1}_{[Z_i=1]}, & & & i &= 1, \dots, n_E,\end{aligned}$$

a $\varepsilon_1^{II,C}, \dots, \varepsilon_{n_C}^{II,C}, \varepsilon_1^{II,E}, \dots, \varepsilon_{n_E}^{II,E}$ jsou nezávislé. Náhodné chyby $\varepsilon_i^{II,C}, \varepsilon_i^{II,E}$ nezávisí na počáteční hodnotě podmíněně na podskupině Z_i . Tento model je podmodelem modelu I.

Pro odhadnutí celkového efektu léčby máme k dispozici hodnoty (1.1). Bohužel nevíme, zda pacient patří do podskupiny $Z_i = 0$ nebo $Z_i = 1$. Podobně jako v kapitole 2.1 využijeme odhad efektu léčby na základě modelu II z práce [1], ve které se předpokládá, že efekt léčby je u každého pacienta stejný. Odhad $\hat{\delta}^{II}$ je založen na rozdílu změn mezi konečnou a počáteční hodnotou v jednotlivých skupinách

$$\hat{\delta}^{II} = (\bar{Y}_1^E - \bar{Y}_0^E) - (\bar{Y}_1^C - \bar{Y}_0^C),$$

kde

$$\begin{aligned} \bar{Y}_0^E &= \frac{1}{n_E} \sum_{i=1}^{n_E} Y_{0i}^E, & \bar{Y}_0^C &= \frac{1}{n_C} \sum_{i=1}^{n_C} Y_{0i}^C, \\ \bar{Y}_1^E &= \frac{1}{n_E} \sum_{i=1}^{n_E} Y_{1i}^E & \text{a} & \bar{Y}_1^C = \frac{1}{n_C} \sum_{i=1}^{n_C} Y_{1i}^C. \end{aligned}$$

Dále se budeme zabývat vlastnostmi tohoto odhadu za platnosti modelu II, sestrojíme testovou statistiku pro testování hypotézy nulovosti efektu léčby a určíme asymptotické rozdělení tohoto odhadu.

Následující věta říká, k čemu odhad $\hat{\delta}^{II}$ konverguje v pravděpodobnosti.

Věta 2.4. *Za platnosti modelu II*

$$\hat{\delta}^{II} \xrightarrow{P} \delta, \quad n \rightarrow \infty,$$

kde $\delta = \delta_0(1-p) + \delta_1 p$.

Důkaz. Střední hodnota odhadu je

$$\mathbb{E} \hat{\delta}^{II} = \mathbb{E} (\bar{Y}_1^E - \bar{Y}_0^E) - \mathbb{E} (\bar{Y}_1^C - \bar{Y}_0^C) = \gamma + (1-p)\delta_0 + p\delta_1 - \gamma = (1-p)\delta_0 + p\delta_1.$$

Rozptyl odhadu $\hat{\delta}^{II}$ počítáme podobně jako v důkazu věty 2.1

$$\text{var} \hat{\delta}^{II} = \text{var} (\bar{Y}_1^C - \bar{Y}_0^C) + \text{var} (\bar{Y}_1^E - \bar{Y}_0^E) = \frac{\sigma_{II,C}^2}{n_C} + \frac{\text{var} (Y_{1i}^E - Y_{0i}^E)}{n_E},$$

kde

$$\text{var} (Y_{1i}^E - Y_{0i}^E) = p(1-p)(\delta_0 - \delta_1)^2 + p\sigma_{II,E,1}^2 + (1-p)\sigma_{II,E,0}^2.$$

Ze slabého zákona velkých čísel plyne, že odhad $\hat{\delta}^{II}$ konverguje k efektu léčby δ v pravděpodobnosti pro $n \rightarrow \infty$. \square

Testujme hypotézu $H_0 : \delta = 0$ proti alternativě $H_1 : \delta \neq 0$. Nejdříve sestrojíme testovou statistiku, o které pojednává následující věta.

Věta 2.5. *Nechť platí model II. Označme*

$$t_{II} = \frac{(\bar{Y}_1^E - \bar{Y}_0^E) - (\bar{Y}_1^C - \bar{Y}_0^C)}{\sqrt{\frac{S_{II,C}^2}{n_C} + \frac{S_{II,E}^2}{n_E}}},$$

kde

$$S_{II,C}^2 = \frac{1}{n_C - 1} \sum_{i=1}^{n_C} \left(Y_{1i}^C - Y_{0i}^C - (\bar{Y}_1^C - \bar{Y}_0^C) \right)^2,$$

$$S_{II,E}^2 = \frac{1}{n_E - 1} \sum_{i=1}^{n_E} \left(Y_{1i}^E - Y_{0i}^E - (\bar{Y}_1^E - \bar{Y}_0^E) \right)^2.$$

Je-li $\delta = 0$, pak platí

$$t_{II} \xrightarrow{d} \mathbf{N}(0, 1).$$

Důkaz. Dokáže se stejným způsobem jako ve větě 2.2. □

Za platnosti H_0 má tedy statistika t_{II} asymptoticky normované normální rozdělení. Nulovou hypotézu zamítáme na hladině α , pokud $|t_{II}| \geq u_{1-\frac{\alpha}{2}}$.

V další větě se zabýváme asymptotickým rozdělení odhadu $\hat{\delta}^{II}$.

Věta 2.6. *Za platnosti modelu II je*

$$\sqrt{n} \left(\hat{\delta}^{II} - \delta \right) \xrightarrow{d} \mathbf{N}(0, V_{\hat{\delta}^{II}}),$$

kde

$$V_{\hat{\delta}^{II}} = (q+1) \left(\sigma_{II,C}^2 + \frac{1}{q} \left(p(1-p)(\delta_0 - \delta_1)^2 + p\sigma_{II,E,1}^2 + (1-p)\sigma_{II,E,0}^2 \right) \right). \quad (2.2)$$

Důkaz. Dokáže se stejným způsobem jako ve větě 2.3. □

Asymptotický rozptyl odhadu $\hat{\delta}^{II}$ z (2.2) a asymptotický rozptyl tohoto odhadu v práci [1] se liší. Pokud bychom předpokládali, že rozptyly jsou v podskupinách $Z_i = 0$ a $Z_i = 1$ stejné, tj. $\sigma_{II,E,0}^2 = \sigma_{II,E,1}^2 = \sigma_{II,E}^2$, pak by asymptotický rozptyl z (2.2) byl větší o

$$\frac{q+1}{q} p(1-p)(\delta_1 - \delta_0)^2$$

než asymptotický rozptyl v práci [1].

V této podkapitole jsme dokázali, že odhad $\hat{\delta}^{II}$ je konzistentní odhad efektu léčby za platnosti modelu II a sestrojili jsme testovou statistiku hypotézy nulovosti efektu léčby. Tyto výsledky se shodují s prací [1]. Dále jsme zjistili asymptotické rozdělení odhadu $\hat{\delta}^{II}$. V speciálním případě konstantních rozptylů v experimentální skupině jako v práci [1] by asymptotický rozptyl z (2.2) byl větší než v práci [1].

2.3 Model III a odhad efektu léčby $\hat{\delta}^{III}$ za jeho platnosti

Model III vyjadřuje lineární závislost konečné hodnoty Y_{1i}^C na počáteční hodnotě Y_{0i}^C . Pro experimentální skupinu je přičten efekt léčby dle podskupiny Z_i . Platí-li tento model, jsou náhodné chyby podmíněně na Z_i nezávislé na počáteční hodnotě Y_{0i} . Model III má tvar

$$\begin{aligned} Y_{1i}^C &= \alpha + \beta Y_{0i}^C + \varepsilon_i^{III,C}, \\ Y_{1i}^E &= \alpha + \beta Y_{0i}^E + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \end{aligned}$$

za podmíněk

$$\begin{aligned} \mathbb{E} \varepsilon_i^{III,C} &= 0, \quad \mathbb{E} \left(\varepsilon_i^{III,C} \mid Y_{0i}^C \right) = 0, & i &= 1, \dots, n_C, \\ \mathbb{E} \varepsilon_i^{III,E} &= 0, \quad \mathbb{E} \left(\varepsilon_i^{III,E} \mid Y_{0i}^E, Z_i \right) = 0, & i &= 1, \dots, n_E, \\ \text{var} \varepsilon_i^{III,C} &= \sigma_{III,C}^2, & i &= 1, \dots, n_C, \\ \text{var} \left(\varepsilon_i^{III,E} \mid Z_i \right) &= \sigma_{III,E,0}^2 \mathbf{1}_{[Z_i=0]} + \sigma_{III,E,1}^2 \mathbf{1}_{[Z_i=1]}, & i &= 1, \dots, n_E, \end{aligned}$$

a $\varepsilon_1^{III,C}, \dots, \varepsilon_{n_C}^{III,C}, \varepsilon_1^{III,E}, \dots, \varepsilon_{n_E}^{III,E}$ jsou nezávislé. Model III je podmodelem modelu I a model II je podmodelem modelu III.

Pro odhadnutí celkového efektu léčby máme k dispozici pouze hodnoty (1.1). Nevíme, zda pacient patří do podskupiny $Z_i = 0$ nebo $Z_i = 1$. Model III si proto nejdříve přepíšeme do tvaru regresního modelu, ve kterém počáteční hodnota a identifikátor přiřazení pacienta do experimentální skupiny jsou jeho regresory. Tento tvar modelu byl využit k přepsání modelu III v práci [1], ve kterém se předpokládá, že efekt léčby je u každého pacienta stejný. Zde však regresní model může mít jiné parametry než model z práce [1]. Naším cílem je odhadnout parametry našeho regresního modelu a zjistit, k čemu tyto odhady konvergují.

Model III přepíšeme do tvaru

$$Y_{1i} = \alpha^* + \beta^* Y_{0i} + \delta^* \mathbf{1}_{[i \in E]} + \varepsilon_i^{III} \quad (2.3)$$

za podmíněk

$$\begin{aligned} \mathbb{E} \varepsilon_i^{III} &= 0, \quad \mathbb{E} \left(\varepsilon_i^{III} \mid Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i \right) = 0 \\ \text{var} \left(\varepsilon_i^{III} \mid \mathbf{1}_{[i \in E]}, Z_i \right) &= \sigma_{III,C}^2 + \left(\sigma_{III,E,0}^2 \mathbf{1}_{[Z_i=0]} + \sigma_{III,E,1}^2 \mathbf{1}_{[Z_i=1]} - \sigma_{III,C}^2 \right) \mathbf{1}_{[i \in E]} \end{aligned} \quad (2.4)$$

a nezávislosti chyb ε_i^{III} , $i = 1, \dots, n$. Jedná se o regresní model

$$\mathbb{E} (Y_i \mid \mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}^*, \quad (2.5)$$

kde

$$\mathbf{X} = \begin{pmatrix} 1 & Y_{01} & \mathbf{1}_{[1 \in E]} \\ \vdots & \vdots & \vdots \\ 1 & Y_{0n} & \mathbf{1}_{[n \in E]} \end{pmatrix} \quad \text{a} \quad \boldsymbol{\beta}^* = \begin{pmatrix} \alpha^* \\ \beta^* \\ \delta^* \end{pmatrix}$$

a \mathbf{X}_i označuje i -tý řádek matice \mathbf{X} . Naším cílem je odhadnout parametry $\boldsymbol{\beta}^*$ tohoto modelu metodou nejmenších čtverců a zjistit jejich asymptotické rozdělení.

Odhadneme parametry modelu (2.5) metodou nejmenších čtverců. Platí

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} \\ \hat{\delta}^{III} \end{pmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (2.6)$$

kde

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n} \end{pmatrix}.$$

Pro výpočet odhadů je potřeba zinvertovat matici $\mathbf{X}^T \mathbf{X}$ - k tomu lze například použít větu A10 z knihy [2].

Tvrzení 2.7. *Nechť $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}$ je symetrická pozitivně definitní bloková matice taková, že bloky \mathbf{A} a \mathbf{D} jsou čtvercové. Pak je $\mathbf{Q} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^T$ regulární a pozitivně definitní a platí*

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{Q}^{-1} & -\mathbf{Q}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{B}^T\mathbf{Q}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{B}^T\mathbf{Q}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{pmatrix}.$$

Důkaz tohoto tvrzení lze nalézt v knize [2]. Přesný vzorec pro odhad $\hat{\delta}^{III}$ zde uvádět nebudeme.

Podívejme se na asymptotické rozdělení odhadů regresních parametrů v modelu (2.5), ve kterém jsou porušeny předpoklady o shodě rozptylů. K tomu využijeme následující tvrzení, jehož důkaz lze nalézt v [1].

Tvrzení 2.8. *Nechť (\mathbf{X}_i, Y_i) , $i = 1, \dots, n$, jsou nezávislé stejně rozdělené náhodné vektory. Nechť platí*

$$\mathbb{E}(Y_i | \mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}^*$$

a nechť mají \mathbf{X}_i konečné čtvrté momenty a Y_i konečné rozptyly. Označme

$$\mathbf{U}_i(\boldsymbol{\beta}) = \mathbf{X}_i(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}).$$

Nechť $\hat{\boldsymbol{\beta}}$ je odhad parametru $\boldsymbol{\beta}^$ takový, který řeší rovnici*

$$\sum_{i=1}^n \mathbf{U}_i(\boldsymbol{\beta}^*) = 0$$

Potom je tento odhad konzistentní a platí pro něj

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{d} \mathbf{N}(0, \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}),$$

kde

$$\mathbf{D} = \mathbb{E} \mathbf{X}_i \mathbf{X}_i^T, \quad \mathbf{V} = \text{var} \mathbf{U}_i(\boldsymbol{\beta}^*).$$

Díky větě 2.8 můžeme zjistit chování odhadu $\hat{\delta}^{III}$. Toto chování je popsáno v následující větě. V důkazu této věty značíme pravděpodobnost přiřazení pacienta do podskupiny $Z_i = 1$ podmíněnou jeho počáteční hodnotou jako

$$p_Z(Y_{0i}) = \mathbf{P}(Z_i = 1 | Y_{0i})$$

a platí pro ni

$$\mathbf{E} p_Z(Y_{0i}) = \mathbf{E} \left[\mathbf{E}[Z_i | Y_{0i}] \right] = \mathbf{E} Z_i = p.$$

Toto značení budeme dále používat i v dalších kapitolách.

Věta 2.9. *Mějme model (2.3) s podmínkami (2.4) a předpokládejme, že pro tento model platí podmínky tvrzení 2.8. Mějme odhad $\hat{\beta}$ z (2.6) získaný metodou nejmenších čtverců. Pak $\hat{\delta}^{III}$ je konzistentní odhad δ .*

Důkaz. V modelu (2.3) platí

$$\mathbf{E}(Y_{1i} | \mathbf{X}_i) = \alpha + \beta Y_{0i} + \left(\delta_0 + (\delta_1 - \delta_0) p_Z(Y_{0i}) \right) \mathbf{1}_{[i \in E]} = \mathbf{X}_i^T \boldsymbol{\beta}^*.$$

Abychom věděli, k čemu konverguje odhad $\hat{\beta}$, musíme zjistit, čemu se rovná $\boldsymbol{\beta}^*$. Parametr $\boldsymbol{\beta}^*$ zjistíme z rovnice

$$\mathbf{E} \mathbf{U}_i(\boldsymbol{\beta}^*) = 0, \quad (2.7)$$

kde

$$\mathbf{U}_i(\alpha^*, \beta^*, \delta^*) = \begin{pmatrix} 1 \\ Y_{0i} \\ \mathbf{1}_{[i \in E]} \end{pmatrix} (Y_{1i} - \alpha^* - \beta^* Y_{0i} - \delta^* \mathbf{1}_{[i \in E]}). \quad (2.8)$$

Rovnici (2.7) rozepíšeme do soustavy rovnic

$$\begin{aligned} \mathbf{E} Y_{0i} &= \mathbf{E} (\alpha^* + \beta^* Y_{0i} + \delta^* \mathbf{1}_{[i \in E]}) \\ \mathbf{E} (Y_{0i} Y_{1i}) &= \mathbf{E} (\alpha^* Y_{0i} + \beta^* Y_{0i}^2 + \delta^* Y_{0i} \mathbf{1}_{[i \in E]}) \\ \mathbf{E} (Y_{1i} \mathbf{1}_{[i \in E]}) &= \mathbf{E} (\alpha^* \mathbf{1}_{[i \in E]} + \beta^* Y_{0i} \mathbf{1}_{[i \in E]} + \delta^* \mathbf{1}_{[i \in E]}) \end{aligned} \quad (2.9)$$

K vyřešení (2.9) provedeme nejdříve pomocné výpočty

$$\begin{aligned} \mathbf{E} (Y_{0i} Y_{1i}) &= \alpha \mu_0 + \beta \mathbf{E} Y_{0i}^2 + \mathbf{E} [(\delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]}) \mathbf{1}_{[i \in E]} Y_{0i} + \varepsilon_i^{III} Y_{0i}] \\ &= \alpha \mu_0 + \beta \mathbf{E} Y_{0i}^2 + \mathbf{E} [Y_{0i} \mathbf{E} [(\delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]}) \mathbf{1}_{[i \in E]} | Y_{0i}]] + \\ &\quad + \mathbf{E} [Y_{0i} \mathbf{E} [\varepsilon_i^{III} | Y_{0i}, Z_i, \mathbf{1}_{[i \in E]}]] \\ &= \alpha \mu_0 + \beta \mathbf{E} Y_{0i}^2 + \mathbf{E} \left[\left(\delta_0 + (\delta_1 - \delta_0) p_Z(Y_{0i}) \right) \pi Y_{0i} \right] \\ &= \alpha \mu_0 + \beta \mathbf{E} Y_{0i}^2 + \pi \left(\mu_0 \delta_0 + (\delta_1 - \delta_0) \mathbf{E} [Y_{0i} p_Z(Y_{0i})] \right) \\ &= \alpha \mu_0 + \beta \mathbf{E} Y_{0i}^2 + \pi \mu_0 \tilde{\delta}, \end{aligned}$$

kde

$$\tilde{\delta} = \delta_0 + (\delta_1 - \delta_0) \frac{1}{\mu_0} \mathbf{E} [Y_{0i} p_Z (Y_{0i})],$$

a

$$\begin{aligned} \mathbf{E} (Y_{1i} \mathbf{1}_{[i \in E]}) &= \alpha \pi + \beta \mu_0 \pi + \mathbf{E} [(\delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]}) \mathbf{1}_{[i \in E]}] + \\ &\quad + \mathbf{E} [\mathbf{1}_{[i \in E]} \mathbf{E} [\varepsilon_i^{III} \mid Y_{0i}, Z_i, \mathbf{1}_{[i \in E]}]] \\ &= \alpha \pi + \beta \mu_0 \pi + \delta \pi. \end{aligned}$$

Z (2.9) a z předchozích pomocných výpočtů získáváme soustavu rovnic

$$\begin{aligned} \alpha^* + \mu_0 \beta^* + \pi \delta^* &= \alpha + \mu_0 \beta + \pi \delta \\ \mu_0 \alpha^* + \mathbf{E} Y_{0i}^2 \beta^* + \mu_0 \pi \delta^* &= \mu_0 \alpha + \mathbf{E} Y_{0i}^2 \beta + \mu_0 \pi \tilde{\delta} \\ \pi \alpha^* + \mu_0 \pi \beta^* + \pi \delta^* &= \pi \alpha + \mu_0 \pi \beta + \pi \delta \end{aligned}$$

z níž postupně dostáváme

$$\begin{aligned} \delta^* &= \delta, \\ \beta^* &= \beta - \frac{\mu_0}{\sigma_0^2} \pi (\delta_0 - \tilde{\delta}), \\ \alpha^* &= \alpha + \frac{\mu_0^2}{\sigma_0^2} \pi (\delta_0 - \tilde{\delta}), \end{aligned}$$

z čehož podle věty 2.8 vyplývá, že $\hat{\delta}^{III}$ je konzistentní odhad δ . \square

Dále se podíváme na asymptotické rozdělení odhadu $\hat{\delta}^{III}$, které je popsáno v následující větě.

Věta 2.10. *Mějme model (2.3) s podmínkami (2.4) a předpokládejme, že pro tento model platí podmínky věty 2.8. Mějme odhad $\hat{\beta}$ parametrů tohoto modelu získaný metodou nejmenších čtverců. Pak platí*

$$\sqrt{n} (\hat{\delta}^{III} - \delta) \xrightarrow{d} \mathbf{N} (0, V_{\hat{\delta}^{III}}),$$

kde

$$\begin{aligned} V_{\hat{\delta}^{III}} &= (q+1) \left\{ \sigma_{III,C}^2 + \frac{1}{q} \left\{ (\delta_1 - \delta_0)^2 \mathbf{E} [p_Z (Y_{0i}) (1 - p_Z (Y_{0i}))] + \right. \right. \\ &\quad \left. \left. + p \sigma_{III,E,1}^2 + (1-p) \sigma_{III,E,0}^2 \right\} \right\} \end{aligned} \quad (2.10)$$

Důkaz. K důkazu této věty použijeme větu 2.8. Pro vyjádření $V_{\hat{\delta}^{III}}$ potřebujeme vypočítat inverzní matici \mathbf{D}^{-1} , matici \mathbf{V} a součin matic $\mathbf{D}^{-1} \mathbf{V} \mathbf{D}^{-1}$. Z tohoto součinu nám stačí znát třetí prvek třetího řádku, čemuž se rovná $V_{\hat{\delta}^{III}}$.

Ke spočítání $V_{\delta III}$ stačí z inverzní matice \mathbf{D}^{-1} pouze znát její poslední řádek a poslední sloupec. Potřebné prvky této matice jsou spočteny v práci [1] a jsou to

$$\mathbf{D}^{-1} = \begin{pmatrix} \cdot & \cdot & -\frac{1}{1-\pi} \\ \cdot & \cdot & 0 \\ -\frac{1}{1-\pi} & 0 & \frac{1}{\pi} + \frac{1}{1-\pi} \end{pmatrix}. \quad (2.11)$$

Dále spočítáme matici \mathbf{V} . Díky struktuře matice \mathbf{D}^{-1} stačí z matice \mathbf{V} vypočítat pouze prvky V_{11} , V_{13} a V_{33} . Matici \mathbf{V} rozepíšeme

$$\begin{aligned} \mathbf{V} &= \text{var } \mathbf{U}_i(\boldsymbol{\beta}^*) = \text{var } \mathbf{E}[\mathbf{U}_i(\boldsymbol{\beta}^*) \mid \mathbf{X}_i] + \mathbf{E} \text{var}[\mathbf{U}_i(\boldsymbol{\beta}^*) \mid \mathbf{X}_i] \\ &= \mathbf{E} \text{var}[\mathbf{U}_i(\boldsymbol{\beta}^*) \mid \mathbf{X}_i]. \end{aligned}$$

Z požadovaných prvků matice \mathbf{V} nejdříve spočítáme příslušné prvky podmíněné varianční matice $\text{var}[\mathbf{U}_i(\boldsymbol{\beta}^*) \mid \mathbf{X}_i]$ a poté aplikujeme střední hodnotu. Prvky vektoru $\mathbf{U}_i(\alpha^*, \beta^*, \delta^*)$ budeme značit číslem v závorce v dolním indexu. Platí

$$\begin{aligned} \text{var}[\mathbf{U}_{i(1)}(\boldsymbol{\beta}^*) \mid Y_{0i}, \mathbf{I}_{[i \in E]}] &= \text{var}\left(\mathbf{E}[Y_{1i} \mid Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i] \mid Y_{0i}, \mathbf{I}_{[i \in E]}\right) + \\ &+ \mathbf{E}\left[\text{var}(Y_{1i} \mid Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]}\right], \end{aligned} \quad (2.12)$$

kde

$$\begin{aligned} &\text{var}\left(\mathbf{E}[Y_{1i} \mid Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i] \mid Y_{0i}, \mathbf{I}_{[i \in E]}\right) = \\ &= \text{var}\left(\alpha + \beta Y_{0i} + (\delta_0 \mathbf{I}_{[Z_i=0]} + \delta_1 \mathbf{I}_{[Z_i=1]}) \mathbf{I}_{[i \in E]} \mid Y_{0i}, \mathbf{I}_{[i \in E]}\right) \\ &= \text{var}\left((\delta_1 - \delta_0) \mathbf{I}_{[Z_i=1]} \mid Y_{0i}\right) \mathbf{I}_{[i \in E]} = (\delta_1 - \delta_0)^2 p_Z(Y_{0i}) \left(1 - p_Z(Y_{0i})\right) \mathbf{I}_{[i \in E]} \end{aligned} \quad (2.13)$$

a

$$\begin{aligned} &\mathbf{E}\left[\text{var}(Y_{1i} \mid Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]}\right] = \\ &= \mathbf{E}\left[\sigma_{III,C}^2 + (\sigma_{III,E,0}^2 \mathbf{I}_{[Z_i=0]} + \sigma_{III,E,1}^2 \mathbf{I}_{[Z_i=1]} - \sigma_{III,C}^2) \mathbf{I}_{[i \in E]} \mid Y_{0i}, \mathbf{I}_{[i \in E]}\right] \\ &= \sigma_{III,C}^2 + \left(\sigma_{III,E,0}^2 + (\sigma_{III,E,1}^2 - \sigma_{III,E,0}^2) p_Z(Y_{0i}) - \sigma_{III,C}^2\right) \mathbf{I}_{[i \in E]}. \end{aligned} \quad (2.14)$$

Z rovnic (2.12), (2.13) a (2.14) dostáváme první prvek matice \mathbf{V}

$$\begin{aligned} V_{11} &= \mathbf{E} \text{var}[\mathbf{U}_{i(1)}(\boldsymbol{\beta}^*) \mid Y_{0i}, \mathbf{I}_{[i \in E]}] \\ &= (\delta_1 - \delta_0)^2 \mathbf{E}\left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i}))\right] \pi + \\ &+ (1 - \pi) \sigma_{III,C}^2 + \left(\sigma_{III,E,0}^2 + (\sigma_{III,E,1}^2 - \sigma_{III,E,0}^2) p\right) \pi. \end{aligned}$$

Pomocí (2.12) spočítáme další prvek V_{33} matice \mathbf{V}

$$\begin{aligned} \text{var}[\mathbf{U}_{i(3)}(\boldsymbol{\beta}^*) \mid Y_{0i}, \mathbf{I}_{[i \in E]}] &= \mathbf{I}_{[i \in E]} \text{var}[\mathbf{U}_{i(1)}(\boldsymbol{\beta}^*) \mid Y_{0i}, \mathbf{I}_{[i \in E]}] \\ &= (\delta_1 - \delta_0)^2 p_Z(Y_{0i}) \left(1 - p_Z(Y_{0i})\right) \mathbf{I}_{[i \in E]} + \\ &+ \left(\sigma_{III,E,0}^2 + (\sigma_{III,E,1}^2 - \sigma_{III,E,0}^2) p_Z(Y_{0i}) - \sigma_{III,C}^2\right) \mathbf{I}_{[i \in E]}, \end{aligned}$$

z čehož dostáváme

$$\begin{aligned} V_{33} &= \mathbf{E} \operatorname{var} [\mathbf{U}_{i(3)}(\boldsymbol{\beta}^*) \mid Y_{0i}, \mathbf{I}_{[i \in E]}] \\ &= \left\{ (\delta_1 - \delta_0)^2 \mathbf{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right] + \sigma_{III,E,0}^2 + (\sigma_{III,E,1}^2 - \sigma_{III,E,0}^2) p \right\} \pi. \end{aligned} \quad (2.15)$$

Nakonec spočítáme prvek V_{13}

$$\operatorname{cov} [\mathbf{U}_{i(1)}(\boldsymbol{\beta}^*), \mathbf{U}_{i(3)}(\boldsymbol{\beta}^*)] = \mathbf{I}_{[i \in E]} \operatorname{var} [\mathbf{U}_{i(1)}(\boldsymbol{\beta}^*) \mid Y_{0i}, \mathbf{I}_{[i \in E]}],$$

odkud plyne

$$V_{13} = V_{33}.$$

Z předchozích rovnic dostáváme matici \mathbf{V} , jejíž prvky jsou vyjádřeny pomocí V_{33} z (2.15)

$$\mathbf{V} = \begin{pmatrix} V_{33} + (1 - \pi) \sigma_{III,C}^2 & \cdot & V_{33} \\ \cdot & \cdot & \cdot \\ V_{33} & \cdot & V_{33} \end{pmatrix}.$$

Spočítejme první a třetí prvek třetího řádku matice $\mathbf{D}^{-1}\mathbf{V}$

$$\begin{aligned} [\mathbf{D}^{-1}\mathbf{V}]_{[3,1]} &= -\frac{1}{1 - \pi} (V_{33} + (1 - \pi) \sigma_{III,C}^2) + \left(\frac{1}{\pi} + \frac{1}{1 - \pi} \right) V_{33} \\ &= \frac{1}{\pi} - \sigma_{III,C}^2, \\ [\mathbf{D}^{-1}\mathbf{V}]_{[3,3]} &= -\frac{1}{1 - \pi} V_{33} + \left(\frac{1}{\pi} + \frac{1}{1 - \pi} \right) V_{33} = \frac{1}{\pi} V_{33}. \end{aligned}$$

Dostáváme asymptotický rozptyl odhadu $\hat{\delta}^{III}$

$$\begin{aligned} V_{\hat{\delta}^{III}} &= [\mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}]_{[3,3]} = -\frac{1}{1 - \pi} \left(\frac{1}{\pi} - \sigma_{III,C}^2 \right) + \left(\frac{1}{\pi} + \frac{1}{1 - \pi} \right) \frac{1}{\pi} V_{33} \\ &= \frac{1}{\pi} \left\{ (\delta_1 - \delta_0)^2 \mathbf{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right] + \sigma_{III,E,0}^2 + \right. \\ &\quad \left. + (\sigma_{III,E,1}^2 - \sigma_{III,E,0}^2) p \right\} + \frac{1}{1 - \pi} \sigma_{III,C}^2 \\ &= (q + 1) \left\{ \sigma_{III,C}^2 + \frac{1}{q} \left\{ (\delta_1 - \delta_0)^2 \mathbf{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right] + \right. \right. \\ &\quad \left. \left. + p \sigma_{III,E,1}^2 + (1 - p) \sigma_{III,E,0}^2 \right\} \right\} \end{aligned}$$

Použitím tvrzení 2.8 dostáváme platnost

$$\sqrt{n} (\hat{\delta}^{III} - \delta) \xrightarrow{d} \mathbf{N}(0, V_{\hat{\delta}^{III}}).$$

□

Asymptotický rozptyl odhadu $\hat{\delta}^{III}$ z (2.10) a asymptotický rozptyl tohoto odhadu v práci [1] se liší. Pokud bychom předpokládali, že rozptyly jsou v podskupinách $Z_i = 0$ a $Z_i = 1$ stejné, tj. $\sigma_{III,E,0}^2 = \sigma_{III,E,1}^2 = \sigma_{III,E}^2$, pak by asymptotický rozptyl z (2.10) byl větší o

$$\frac{q+1}{q} (\delta_1 - \delta_0)^2 \mathbf{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right]$$

než asymptotický rozptyl v práci [1].

Dále budeme testovat hypotézu $H_0 : \delta = 0$ proti alternativě $H_1 : \delta \neq 0$. V následující větě sestrojíme testovou statistiku pro tento test.

Věta 2.11. *Nechť platí předpoklady věty 2.10. Označme*

$$t_{III} = \frac{\sqrt{n}\hat{\delta}^{III}}{\hat{V}_{\hat{\delta}^{III}}},$$

kde

$$\hat{V}_{\hat{\delta}^{III}} = \left[\hat{\mathbf{D}}^{-1} \hat{\mathbf{V}} \hat{\mathbf{D}}^{-1} \right]_{[3,3]}.$$

a

$$\begin{aligned} \hat{\mathbf{D}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T, \\ \hat{\mathbf{V}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{U}_i(\boldsymbol{\beta}^*) \mathbf{U}_i(\boldsymbol{\beta}^*)^T. \end{aligned}$$

Je-li $\delta = 0$, pak platí

$$t_{III} \xrightarrow{d} \mathbf{N}(0, 1).$$

Důkaz. Plyne z konzistence odhadu rozptylu, zákona velkých čísel a Cramérový-Sluckého věty. \square

Asymptotické rozdělení testové statistiky t_{III} za H_0 je tedy normované normální. Hypotézu H_0 zamítneme na hladině α , pokud $|t_{III}| \geq u_{1-\frac{\alpha}{2}}$.

V této podkapitole jsme dokázali, že odhad $\hat{\delta}^{III}$ je konzistentní odhad efektu léčby za platnosti modelu III a sestrojili jsme testovou statistiku hypotézy nulovosti efektu léčby. Tyto výsledky se shodují s prací [1]. Dále jsme zjistili asymptotické rozdělení odhadu $\hat{\delta}^{III}$. V speciálním případě konstantních rozptylů v experimentální skupině jako v práci [1] by asymptotický rozptyl z (2.10) byl větší než v práci [1].

3. Chování odhadů za nekonstantního efektu léčby

Tato kapitola navazuje na předchozí kapitolu, ve které jsme uvedli modely I, II a III a chování odhadů efektu léčby $\hat{\delta}^I$, $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ za platnosti příslušných modelů. Ve skutečnosti však nevíme, ze kterého modelu data vznikla a jaký odhad efektu léčby bychom měli použít. V následujících podkapitolách postupně prozkoumáme, jaké vlastnosti má daný odhad za platnosti jednoho ze dvou zbývajících modelů. Podíváme se, zda je daný odhad za platnosti tohoto modelu konzistentní, a porovnáme asymptotické rozptyly tohoto odhadu za platnosti uvažovaného modelu a za platnosti modelu, na jehož základě odhad vznikl.

3.1 Chování odhadu $\hat{\delta}^I$ za platnosti modelu II

Platí-li model II, pak platí i model I. Odhad $\hat{\delta}^I$ je tedy za platnosti modelu II nestranný a konzistentní. Zajímá nás asymptotické rozdělení tohoto odhadu a jeho porovnání s odhadem $\hat{\delta}^{II}$ za platnosti modelu II. O tom pojednává následující věta.

Věta 3.1. *Za platnosti modelu II pro odhad $\hat{\delta}^I$ platí*

$$\sqrt{n} \left(\hat{\delta}^I - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, W_{\hat{\delta}^I}^{II} \right),$$

kde

$$W_{\hat{\delta}^I}^{II} = V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} \sigma_0^2.$$

Důkaz. Rozptyl $W_{\hat{\delta}^I}^{II}$ je roven asymptotickému rozptylu $V_{\hat{\delta}^I}$ z (2.1). Tento rozptyl vyjádříme pomocí parametrů z modelu II

$$W_{\hat{\delta}^I}^{II} = (q+1) \left(\text{var } Y_{1i}^C + \frac{1}{q} \text{var } Y_{1i}^E \right),$$

kde

$$\text{var } Y_{1i}^C = \text{var} \left(Y_{0i}^C + \varepsilon_i^{II,C} \right) = \text{var } Y_{0i}^C + \text{var } \varepsilon_i^{II,C} = \sigma_0^2 + \sigma_{II,C}^2$$

a

$$\begin{aligned} \text{var } Y_{1i}^E &= \text{var} \left(Y_{0i}^E + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{II,E} \right) \\ &= \text{var} \left(\mathbf{E} \left[Y_{0i}^E + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{II,E} \mid Z_i \right] \right) + \\ &\quad + \mathbf{E} \left[\text{var} \left(Y_{0i}^E + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{II,E} \mid Z_i \right) \right] \\ &= \text{var} \left(\mu_0 + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} \right) + \mathbf{E} \left[\sigma_0^2 + \sigma_{II,E,0}^2 \mathbf{1}_{[Z_i=0]} + \sigma_{II,E,1}^2 \mathbf{1}_{[Z_i=1]} \right] \\ &= (\delta_0 - \delta_1)^2 p(1-p) + \sigma_0^2 + (1-p) \sigma_{II,E,0}^2 + p \sigma_{II,E,1}^2. \end{aligned}$$

Celkem získáváme

$$\begin{aligned} W_{\hat{\delta}^I}^{II} &= (q+1) \left\{ \sigma_0^2 + \sigma_{II,C}^2 + \right. \\ &\quad \left. + \frac{1}{q} \left((\delta_0 - \delta_1)^2 p(1-p) + \sigma_0^2 + (1-p) \sigma_{II,E,0}^2 + p \sigma_{II,E,1}^2 \right) \right\} \\ &= V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} \sigma_0^2. \end{aligned}$$

□

Odhad $\hat{\delta}^I$ je sice nestranný a konzistentní odhad δ stejně jako odhad $\hat{\delta}^{II}$, ale za platnosti modelu II má odhad $\hat{\delta}^I$ vždy větší rozptyl než odhad $\hat{\delta}^{II}$. Tento výsledek se shoduje s výsledkem z práce [1].

3.2 Chování odhadu $\hat{\delta}^{III}$ za platnosti modelu II

Platí-li model II, pak platí model III s parametrem $\beta = 1$. Odhad $\hat{\delta}^{III}$ je tedy za platnosti modelu III nestranný a konzistentní. Chceme určit asymptotické rozdělení tohoto odhadu za platnosti modelu II a zda je za platnosti modelu II lepší použít odhad $\hat{\delta}^{II}$ či odhad $\hat{\delta}^{III}$. V následující větě zjistíme asymptotické rozdělení odhadu $\hat{\delta}^{III}$ za platnosti modelu II.

Věta 3.2. *Asymptotické rozdělení odhadu $\hat{\delta}^{III}$ za platnosti modelu II je*

$$\sqrt{n} \left(\hat{\delta}^{III} - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, W_{\hat{\delta}^{III}}^{II} \right),$$

kde

$$W_{\hat{\delta}^{III}}^{II} = V_{\hat{\delta}^I} - \frac{q+1}{q} (\delta_0 - \delta_1)^2 \text{var } p_Z(Y_{0i})$$

Důkaz. Jelikož za platnosti modelu II platí model III s parametrem $\beta = 1$, tak rozptyly náhodných chyb z modelů II a III se shodují

$$\sigma_{II,C}^2 = \sigma_{III,C}^2, \quad \sigma_{II,E,1}^2 = \sigma_{III,E,1}^2 \quad \text{a} \quad \sigma_{II,E,0}^2 = \sigma_{III,E,0}^2.$$

Rozptyl $W_{\hat{\delta}^{III}}^{II}$ je roven asymptotickému rozptylu $V_{\hat{\delta}^I}$ z (2.10). Použitím shodnosti rozptylů náhodných chyb z modelů II a III dostáváme

$$\begin{aligned} W_{\hat{\delta}^{III}}^{II} &= (q+1) \left\{ \sigma_{II,C}^2 + \frac{1}{q} \left\{ (\delta_1 - \delta_0)^2 \mathbb{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right] + \right. \right. \\ &\quad \left. \left. + p \sigma_{II,E,1}^2 + (1-p) \sigma_{II,E,0}^2 \right\} \right\} \\ &= V_{\hat{\delta}^I} + \frac{q+1}{q} (\delta_0 - \delta_1)^2 \left(\mathbb{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right] - p(1-p) \right). \end{aligned}$$

Jelikož

$$\begin{aligned} \mathbb{E}\left[p_Z(Y_{0i})(1-p_Z(Y_{0i}))\right] - p(1-p) &= p - \mathbb{E}p_Z^2(Y_{0i}) - (p-p^2) \\ &= -\text{var } p_Z(Y_{0i}), \end{aligned}$$

získáváme znění věty. □

Z předchozí věty plyne, že za platnosti modelu II má odhad $\hat{\delta}^{III}$ menší asymptotický rozptyl než odhad $\hat{\delta}^{II}$, pokud je funkce $p_Z(Y_{0i})$ nekonstantní. Pokud je tato funkce konstantní, pak jsou asymptotické rozptyly odhadů stejné. Tento výsledek se liší od práce [1], ve které se předpokládá konstantní efekt léčby. V práci [1] mají odhady $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ za platnosti modelu II vždy stejný asymptotický rozptyl.

3.3 Chování odhadu $\hat{\delta}^I$ za platnosti modelu III

Platí-li model III, pak platí i model I. Díky tomu je odhad $\hat{\delta}^I$ za platnosti modelu III nestranný a konzistentní. Zajímá nás asymptotické rozdělení tohoto odhadu a jeho porovnání s odhadem $\hat{\delta}^{III}$ za platnosti modelu I. O asymptotickém rozdělení odhadu $\hat{\delta}^I$ pojednává následující věta.

Věta 3.3. *Za platnosti modelu III pro odhad $\hat{\delta}^I$ platí*

$$\sqrt{n}(\hat{\delta}^I - \delta) \xrightarrow{d} \mathbf{N}(0, W_{\hat{\delta}^I}^{III}),$$

kde

$$\begin{aligned} W_{\hat{\delta}^I}^{III} &= V_{\hat{\delta}^{III}} + \frac{(q+1)^2}{q} \beta^2 \sigma_0^2 + \frac{q+1}{q} \left\{ 2\beta(\delta_1 - \delta_0) \text{cov}(Y_{0i}^E, p_Z(Y_{0i}^E)) + \right. \\ &\quad \left. + (\delta_1 - \delta_0)^2 \text{var } p_Z(Y_{0i}^E) \right\}. \end{aligned}$$

Důkaz. Rozptyl $W_{\hat{\delta}^I}^{III}$ je roven asymptotickému rozptylu $V_{\hat{\delta}^I}$ z (2.1). Tento rozptyl vyjádříme pomocí parametrů z modelu III

$$W_{\hat{\delta}^I}^{III} = (q+1) \left(\text{var } Y_{1i}^C + \frac{1}{q} \text{var } Y_{1i}^E \right), \quad (3.1)$$

kde

$$\text{var } Y_{1i}^C = \text{var} \left(\beta Y_{0i}^C + \varepsilon_i^{III,C} \right) = \beta^2 \sigma_0^2 + \sigma_{III,C}^2$$

a rozptyl $\text{var } Y_{1i}^E$ rozepíšeme na

$$\begin{aligned} \text{var } Y_{1i}^E &= \text{var} \left(\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \right) \\ &= \text{var} \left(\mathbb{E} \left[\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \mid Y_{0i}^E, Z_i \right] \right) + \\ &\quad + \mathbb{E} \left[\text{var} \left(\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \mid Y_{0i}^E, Z_i \right) \right]. \end{aligned}$$

Z předchozí rovnosti postupně spočítáme

$$\begin{aligned}
\text{var}\left(\mathbb{E}\left[\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbb{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \mid Y_{0i}^E, Z_i\right]\right) &= \text{var}\left(\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbb{1}_{[Z_i=1]}\right) \\
&= \text{var}\left(\mathbb{E}\left[\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbb{1}_{[Z_i=1]} \mid Y_{0i}^E\right]\right) + \mathbb{E}\left[\text{var}\left(\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbb{1}_{[Z_i=1]} \mid Y_{0i}^E\right)\right] \\
&= \text{var}\left(\beta Y_{0i}^E + (\delta_1 - \delta_0) p_Z(Y_{0i}^E)\right) + \mathbb{E}\left[(\delta_1 - \delta_0)^2 p_Z(Y_{0i}^E) (1 - p_Z(Y_{0i}^E))\right] \\
&= \beta^2 \sigma_0^2 + (\delta_1 - \delta_0)^2 \text{var } p_Z(Y_{0i}^E) + 2\beta(\delta_1 - \delta_0) \text{cov}\left(Y_{0i}^E, p_Z(Y_{0i}^E)\right) + \\
&\quad + (\delta_1 - \delta_0)^2 \mathbb{E}\left[p_Z(Y_{0i}^E) (1 - p_Z(Y_{0i}^E))\right] \\
&= \beta^2 \sigma_0^2 + (\delta_1 - \delta_0)^2 \left\{ \mathbb{E}\left[p_Z(Y_{0i}^E) (1 - p_Z(Y_{0i}^E))\right] + \text{var } p_Z(Y_{0i}^E) \right\} + \\
&\quad + 2\beta(\delta_1 - \delta_0) \text{cov}\left(Y_{0i}^E, p_Z(Y_{0i}^E)\right)
\end{aligned}$$

a

$$\begin{aligned}
\mathbb{E}\left[\text{var}\left(\beta Y_{0i}^E + (\delta_1 - \delta_0) \mathbb{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \mid Y_{0i}^E, Z_i\right)\right] &= \\
&= \mathbb{E}\left[\sigma_{III,E,0}^2 \mathbb{1}_{[Z_i=0]} + \sigma_{III,E,1}^2 \mathbb{1}_{[Z_i=1]}\right] = (1-p) \sigma_{III,E,0}^2 + p \sigma_{III,E,1}^2.
\end{aligned}$$

Dosazením vypočítaných hodnot z předchozích rovnic do (3.1) získáváme

$$\begin{aligned}
W_{\hat{\delta}^I}^{III} &= (q+1) \left\{ \beta^2 \sigma_0^2 + \sigma_{III,C}^2 + \frac{1}{q} \left\{ \beta^2 \sigma_0^2 + 2\beta(\delta_1 - \delta_0) \text{cov}\left(Y_{0i}^E, p_Z(Y_{0i}^E)\right) + \right. \right. \\
&\quad \left. \left. + (\delta_1 - \delta_0)^2 \left\{ \mathbb{E}\left[p_Z(Y_{0i}^E) (1 - p_Z(Y_{0i}^E))\right] + \text{var } p_Z(Y_{0i}^E) \right\} + \right. \right. \\
&\quad \left. \left. + (1-p) \sigma_{III,E,0}^2 + p \sigma_{III,E,1}^2 \right\} \right\} \\
&= V_{\hat{\delta}^{III}} + \frac{(q+1)^2}{q} \beta^2 \sigma_0^2 + \frac{q+1}{q} \left\{ 2\beta(\delta_1 - \delta_0) \text{cov}\left(Y_{0i}^E, p_Z(Y_{0i}^E)\right) + \right. \\
&\quad \left. + (\delta_1 - \delta_0)^2 \text{var } p_Z(Y_{0i}^E) \right\}.
\end{aligned}$$

□

Díky vzorci z předchozí věty se domníváme, že za platnosti modelu III může mít odhad $\hat{\delta}^I$ větší i menší rozptyl než odhad $\hat{\delta}^{III}$. Pokud by funkce $p_Z(Y_{0i})$ byla konstantní, získali bychom stejný výsledek jako v práci [1], ve které se předpokládá konstantní efekt léčby. V práci [1] má odhad $\hat{\delta}^I$ v případě $\beta = 0$ stejný asymptotický rozptyl jako $\hat{\delta}^{III}$, jinak je asymptotický rozptyl $\hat{\delta}^I$ větší.

3.4 Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu III

Platí-li model III, pak obecně neplatí model II. Zajímá nás asymptotické rozdělení odhadu $\hat{\delta}^{II}$ za platnosti modelu III a jeho porovnání s odhadem $\hat{\delta}^{III}$ za platnosti modelu III. O tom pojednává následující věta.

Věta 3.4. *Asymptotické rozdělení odhadu $\hat{\delta}^{II}$ za platnosti modelu III je*

$$\sqrt{n} \left(\hat{\delta}^{II} - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, W_{\hat{\delta}^{II}}^{III} \right),$$

kde

$$\begin{aligned} W_{\hat{\delta}^{II}}^{III} &= V_{\hat{\delta}^{III}} + \frac{(q+1)^2}{q} (\beta-1)^2 \sigma_0^2 + \\ &+ \frac{q+1}{q} \left\{ 2(\beta-1)(\delta_1 - \delta_0) \operatorname{cov} \left(Y_{0i}^E, p_Z \left(Y_{0i}^E \right) \right) + \right. \\ &\left. + (\delta_1 - \delta_0)^2 \operatorname{var} p_Z \left(Y_{0i}^E \right) \right\}. \end{aligned}$$

Důkaz. Rozptyl $W_{\hat{\delta}^{III}}^{III}$ je roven

$$W_{\hat{\delta}^{III}}^{III} = (q+1) \left(\operatorname{var} \left(Y_{1i}^C - Y_{0i}^C \right) + \frac{1}{q} \operatorname{var} \left(Y_{1i}^E - Y_{0i}^E \right) \right).$$

Při počítání rozptylů v předchozí rovnici můžeme využít výpočty v důkazu věty 3.3, kde místo β dosazujeme $\beta - 1$. Počítejme

$$\operatorname{var} \left(Y_{1i}^C - Y_{0i}^C \right) = \operatorname{var} \left(\alpha + (\beta-1) Y_{0i}^C + \varepsilon_i^{III,C} \right) = (\beta-1)^2 \sigma_0^2 + \sigma_{III,C}^2$$

a

$$\begin{aligned} \operatorname{var} \left(Y_{1i}^E - Y_{0i}^E \right) &= \operatorname{var} \left(\alpha + (\beta-1) Y_{0i}^E + \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} + \varepsilon_i^{III,E} \right) \\ &= (\beta-1)^2 \sigma_0^2 + 2(\beta-1)(\delta_1 - \delta_0) \operatorname{cov} \left(Y_{0i}^E, p_Z \left(Y_{0i}^E \right) \right) + \\ &+ (\delta_1 - \delta_0)^2 \left\{ \mathbf{E} \left[p_Z \left(Y_{0i}^E \right) \left(1 - p_Z \left(Y_{0i}^E \right) \right) \right] + \operatorname{var} p_Z \left(Y_{0i}^E \right) \right\} + \\ &+ (1-p) \sigma_{III,E,0}^2 + p \sigma_{III,E,1}^2. \end{aligned}$$

Celkem získáváme

$$\begin{aligned} W_{\hat{\delta}^{III}}^{III} &= (q+1) \left\{ (\beta-1)^2 \sigma_0^2 + \sigma_{III,C}^2 + \frac{1}{q} \left\{ (\beta-1)^2 \sigma_0^2 + \right. \right. \\ &+ 2(\beta-1)(\delta_1 - \delta_0) \operatorname{cov} \left(Y_{0i}^E, p_Z \left(Y_{0i}^E \right) \right) + \\ &+ (\delta_1 - \delta_0)^2 \left\{ \mathbf{E} \left[p_Z \left(Y_{0i}^E \right) \left(1 - p_Z \left(Y_{0i}^E \right) \right) \right] + \operatorname{var} p_Z \left(Y_{0i}^E \right) \right\} + \\ &\left. \left. + (1-p) \sigma_{III,E,0}^2 + p \sigma_{III,E,1}^2 \right\} \right\} \end{aligned}$$

a po přepsání tohoto rozptylu pomocí $V_{\hat{\delta}^{III}}$ získáváme tvrzení věty. \square

Díky vzorci z předchozí věty se domníváme, že za platnosti modelu III může mít odhad $\hat{\delta}^{II}$ větší i menší rozptyl než odhad $\hat{\delta}^{III}$. Pokud by $p_Z(Y_{0i}) = p$, získali bychom stejný výsledek jako v práci [1], ve které se předpokládá konstantní efekt léčby. V práci [1] má odhad $\hat{\delta}^{II}$ v případě $\beta = 1$ stejný asymptotický rozptyl jako odhad $\hat{\delta}^{III}$, jinak je asymptotický rozptyl odhadu $\hat{\delta}^{II}$ větší.

3.5 Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu I

Platí-li model I, pak obecně neplatí model II. Chceme zjistit asymptotické rozdělení odhadu $\hat{\delta}^{II}$ za platnosti modelu I a porovnat tento odhad s odhadem $\hat{\delta}^I$. Asymptotické rozdělení odhadu $\hat{\delta}^{II}$ za platnosti modelu I získáváme v následující větě.

Věta 3.5. *Asymptotický rozptyl odhadu $\hat{\delta}^{II}$ za platnosti modelu I je*

$$\sqrt{n} \left(\hat{\delta}^{II} - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, W_{\hat{\delta}^{II}}^I \right),$$

kde

$$W_{\hat{\delta}^{II}}^I = V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} \sigma_0^2 - 2(q+1) \left(\text{cov} (Y_{0i}^C, Y_{1i}^C) + \frac{1}{q} \text{cov} (Y_{0i}^E, Y_{1i}^E) \right).$$

Důkaz. Asymptotický rozptyl $W_{\hat{\delta}^{II}}^I$ je roven

$$W_{\hat{\delta}^{II}}^I = (q+1) \left(\text{var} (Y_{1i}^C - Y_{0i}^C) + \frac{1}{q} \text{var} (Y_{1i}^E - Y_{0i}^E) \right),$$

kde

$$\text{var} (Y_{1i}^C - Y_{0i}^C) = \sigma_{I,C}^2 + \sigma_0^2 - 2 \text{cov} (Y_{0i}^C, Y_{1i}^C)$$

a použitím již spočítaného rozptylu $\text{var} Y_{1i}^E$ z důkazu věty 2.1 získáváme

$$\begin{aligned} \text{var} (Y_{1i}^E - Y_{0i}^E) &= (\delta_0 - \delta_1)^2 p(1-p) + p \sigma_{I,E,1}^2 + (1-p) \sigma_{I,E,0}^2 + \\ &+ \sigma_0^2 - 2 \text{cov} (Y_{0i}^E, Y_{1i}^E). \end{aligned}$$

Z předchozích rovnic celkem obdržíme

$$\begin{aligned} W_{\hat{\delta}^{II}}^I &= (q+1) \left(\text{var} (Y_{1i}^C - Y_{0i}^C) + \frac{1}{q} \text{var} (Y_{1i}^E - Y_{0i}^E) \right) \\ &= (q+1) \left\{ \sigma_{I,C}^2 + \sigma_0^2 - 2 \text{cov} (Y_{0i}^C, Y_{1i}^C) + \frac{1}{q} \left\{ (\delta_0 - \delta_1)^2 p(1-p) + \right. \right. \\ &\quad \left. \left. + p \sigma_{I,E,1}^2 + (1-p) \sigma_{I,E,0}^2 + \sigma_0^2 - 2 \text{cov} (Y_{0i}^E, Y_{1i}^E) \right\} \right\} \\ &= V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} \sigma_0^2 - 2(q+1) \left(\text{cov} (Y_{0i}^C, Y_{1i}^C) + \frac{1}{q} \text{cov} (Y_{0i}^E, Y_{1i}^E) \right). \end{aligned}$$

□

Odhad $\hat{\delta}^{II}$ má větší asymptotický rozptyl než $\hat{\delta}^I$ za platnosti modelu I, jestliže

$$\frac{\text{cov} (Y_{0i}^E, Y_{1i}^E)}{\sigma_0^2} < \frac{1}{2} \quad \text{a} \quad \frac{\text{cov} (Y_{0i}^C, Y_{1i}^C)}{\sigma_0^2} < \frac{1}{2}.$$

Tento výsledek se shoduje s výsledkem v práci [1], kde se předpokládá konstantní efekt léčby.

3.6 Chování odhadu $\hat{\delta}^{III}$ za platnosti modelu I

Platí-li model I, pak obecně neplatí model III. Abychom prozkoumali chování odhadu $\hat{\delta}^{III}$, vyjádříme model I jiným způsobem. Podíváme se, čemu se v tomto případě rovnají parametry modelu I. Pak zjistíme, zda je odhad $\hat{\delta}^{III}$ za platnosti modelu I konzistentní, a podíváme se na jeho asymptotické rozdělení.

Pro vyjádření modelu I v jiném tvaru rozepíšeme konečné hodnoty modelu I jako

$$\begin{aligned} Y_{1i} &= \mathbf{E}(Y_{1i} | Y_{0i}, i \in C) + \left(Y_{1i} - \mathbf{E}(Y_{1i} | Y_{0i}, i \in C) \right), & i \in C, \\ Y_{1i} &= \mathbf{E}(Y_{1i} | Y_{0i}, Z_i, i \in E) + \left(Y_{1i} - \mathbf{E}(Y_{1i} | Y_{0i}, Z_i, i \in E) \right), & i \in E. \end{aligned}$$

Dále označíme

$$\begin{aligned} h_C(Y_{0i}) &= \mathbf{E}(Y_{1i} | Y_{0i}, i \in C), \\ h_E(Y_{0i}, Z_i) &= \mathbf{E}(Y_{1i} | Y_{0i}, Z_i, i \in E), \end{aligned}$$

kde

$$h_E(Y_{0i}, Z_i) = h_{E,0}(Y_{0i}) \mathbf{1}_{[Z_i=0]} + h_{E,1}(Y_{0i}) \mathbf{1}_{[Z_i=1]},$$

a

$$\eta_i = Y_{1i} - \mathbf{E}(Y_{1i} | Y_{0i}, Z_i, \mathbf{1}_{[i \in E]}).$$

Celkem lze model I vyjádřit

$$Y_{1i} = h_C(Y_{0i}) \mathbf{1}_{[i \in C]} + h_E(Y_{0i}, Z_i) \mathbf{1}_{[i \in E]} + \eta_i, \quad i = 1, \dots, n, \quad (3.2)$$

přičemž pro η_i platí

$$\mathbf{E} \eta_i = \mathbf{E} [Y_{1i} - \mathbf{E}(Y_{1i} | Y_{0i}, Z_i, \mathbf{1}_{[i \in E]})] = 0$$

a

$$\mathbf{E} [\eta_i | Y_{0i}, Z_i, \mathbf{1}_{[i \in E]}] = \mathbf{E} [Y_{1i} | Y_{0i}, Z_i, \mathbf{1}_{[i \in E]}] - \mathbf{E} [Y_{1i} | Y_{0i}, Z_i, \mathbf{1}_{[i \in E]}] = 0.$$

Funkce h_C a h_E z vyjádření modelu I v (3.2) vyjadřují, jakým způsobem závisí konečná hodnota na počáteční. Tato závislost může být jiná v kontrolní skupině a v podskupinách experimentální skupiny.

V další větě uvádíme, čemu se rovnají parametry modelu I, vyjádříme-li ho způsobem uvedeným v (3.2).

Věta 3.6. *Uvažujme model I a jeho přepis (3.2). Pak platí*

$$\begin{aligned} \mu &= \mathbf{E} h_C(Y_{0i}), \\ \delta_0 &= \mathbf{E} h_E(Y_{0i}, 0) - \mathbf{E} h_C(Y_{0i}), \\ \delta_1 &= \mathbf{E} h_E(Y_{0i}, 1) - \mathbf{E} h_C(Y_{0i}), \\ \sigma_{I,C}^2 &= \mathbf{E} h_C^2(Y_{0i}) + \mathbf{E} (\eta_i^2 | i \in C) - \mu^2, \\ \sigma_{I,E,0}^2 &= \mathbf{E} h_E^2(Y_{0i}, 0) + \mathbf{E} [\eta_i^2 | i \in E, Z_i = 0] - (\mu - \delta_0)^2, \\ \sigma_{I,E,1}^2 &= \mathbf{E} h_E^2(Y_{0i}, 1) + \mathbf{E} [\eta_i^2 | i \in E, Z_i = 1] - (\mu - \delta_1)^2. \end{aligned}$$

Důkaz. Pro parametr μ zřejmě platí

$$\mu = \mathbf{E} h_C (Y_{0i}).$$

Dále určíme, čemu se rovnají parametry δ_0 a δ_1 . Platí

$$\begin{aligned} \mathbf{E} h_E (Y_{0i}, 0) &= \mathbf{E} h_{E,0} (Y_{0i}) = \mu + \delta_0, \\ \mathbf{E} h_E (Y_{0i}, 1) &= \mathbf{E} h_{E,1} (Y_{0i}) = \mu + \delta_1, \end{aligned}$$

z čehož plyne vyjádření parametrů

$$\begin{aligned} \delta_0 &= \mathbf{E} h_E (Y_{0i}, 0) - \mathbf{E} h_C (Y_{0i}), \\ \delta_1 &= \mathbf{E} h_E (Y_{0i}, 1) - \mathbf{E} h_C (Y_{0i}). \end{aligned}$$

Vyjádříme další parametr

$$\begin{aligned} \sigma_{I,C}^2 &= \text{var } \varepsilon_i^{I,C} = \text{var} (Y_{1i} - \mu \mid i \in C) = \text{var} (h_C (Y_{0i}) + \eta_i \mid i \in C) \\ &= \mathbf{E} [(h_C (Y_{0i}) + \eta_i)^2 \mid i \in C] - \left(\mathbf{E} [h_C (Y_{0i}) + \eta_i] \right)^2, \end{aligned}$$

kde

$$\begin{aligned} \mathbf{E} [(h_C (Y_{0i}) + \eta_i)^2 \mid i \in C] &= \mathbf{E} [\mathbf{E} [(h_C (Y_{0i}) + \eta_i)^2 \mid Y_{0i}, i \in C] \mid i \in C] \\ &= \mathbf{E} [h_C^2 (Y_{0i}) + 2 h_C (Y_{0i}) \mathbf{E} [\eta_i \mid Y_{0i}, i \in C] + \mathbf{E} (\eta_i^2 \mid Y_{0i}, i \in C) \mid i \in C] \\ &= \mathbf{E} h_C^2 (Y_{0i}) + \mathbf{E} (\eta_i^2 \mid i \in C) \end{aligned}$$

a

$$\mathbf{E} [h_C (Y_{0i}) + \eta_i] = \mu.$$

Celkem získáváme

$$\sigma_{I,C}^2 = \mathbf{E} h_C^2 (Y_{0i}) + \mathbf{E} (\eta_i^2 \mid i \in C) - \mu^2.$$

Podobným způsobem vyjádříme následující parametry

$$\begin{aligned} \sigma_{I,E,0}^2 &= \text{var} (\varepsilon_i^{I,E} \mid i \in E, Z_i = 0) \\ &= \text{var} (Y_{1i} - \mu - \delta_0 \mathbf{1}_{[Z_i=0]} + \delta_1 \mathbf{1}_{[Z_i=1]} \mid i \in E, Z_i = 0) \\ &= \text{var} (h_E (Y_{0i}, Z_i) + \eta_i \mid i \in E, Z_i = 0) \\ &= \mathbf{E} [(h_E (Y_{0i}, Z_i) + \eta_i)^2 \mid i \in E, Z_i = 0] - \\ &\quad - \left(\mathbf{E} [h_E (Y_{0i}, Z_i) + \eta_i \mid i \in E, Z_i = 0] \right)^2 \\ &= \mathbf{E} h_E^2 (Y_{0i}, 0) + \mathbf{E} [\eta_i^2 \mid i \in E, Z_i = 0] - (\mu - \delta_0)^2 \end{aligned}$$

a

$$\sigma_{I,E,1}^2 = \mathbf{E} h_E^2 (Y_{0i}, 1) + \mathbf{E} [\eta_i^2 \mid i \in E, Z_i = 1] - (\mu - \delta_1)^2,$$

čímž jsme větu dokázali. □

Dále chceme zjistit, jaké vlastnosti má odhad $\hat{\delta}^{III}$, platí-li model I. K tomu nám poslouží dvě tvrzení a dvě definice z [3]. Tvrzení se zabývají odhadem parametrů metodou maximální věrohodnosti, jestliže platí jiné rozdělení, než ze kterého tvoříme věrohodnostní funkci.

Definice 3.7. *Nechť X_1, \dots, X_n je náhodný výběr z rozdělení se spojitou distribuční funkcí G na měřitelném euklidovském prostoru Ω . Nechť g je hustota příslušná distribuční funkci G . Nechť Θ je kompaktní množina parametrů. Nechť $F_{\theta}(x)$ je rodina distribučních funkcí, jejichž hustoty $f_{\theta}(x)$ jsou měřitelné v x pro každé $\theta \in \Theta$ a spojitě v θ pro všechna $x \in \Omega$. Označme $\mathbf{X} = (X_1, \dots, X_n)^T$. Pak funkci*

$$L_n(\mathbf{X}, \theta) = \frac{1}{n} \sum_{i=1}^n \log f_{\theta}(X_i)$$

nazýváme kvazi-logaritmicou věrohodnostní funkci a odhad

$$\hat{\theta}_n^* = \operatorname{argmax}(L_n(\mathbf{X}, \theta), \theta \in \Theta)$$

kvazi-maximálně věrohodným odhadem parametru θ .

Tvrzení 3.8. *Za podmínek definice 3.7 kvazi-maximálně věrohodný odhad $\hat{\theta}_n^*$ existuje a je měřitelný pro všechna $n \in \mathbb{N}$.*

Definice 3.9. *Nechť jsou P a Q dvě pravděpodobnostní rozdělení s hustotami p a q . Pak*

$$K(p, q) = E_P \log \frac{p(x)}{q(x)}$$

nazýváme Kullbackovou-Leiblerovou vzdáleností rozdělení Q od P .

Tvrzení 3.10. *Nechť platí předpoklady definice 3.7. Předpokládejme dále, že*

(i) existuje střední hodnota $E \log g(X_n)$, $n \in \mathbb{N}$,

(ii) pro všechna $\theta \in \Theta$ platí

$$|\log f_{\theta}(x)| \leq m(x),$$

kde m je taková funkce, pro kterou $\int m(x) dG(x)$ existuje,

(iii) $K(g, f_{\theta})$ má jediné minimum v bodě θ^ .*

Pak platí

$$\hat{\theta}_n^* \xrightarrow{s.j.} \theta^*, \quad n \rightarrow \infty,$$

pro skoro všechny posloupnosti X_n .

V definici 3.7 a v předchozí definici zastupuje model I rozdělení G , protože se jedná o model, jehož platnost zde předpokládáme. Označme F_{θ} rozdělení náhodné veličiny Y_{1i}^* , kde Y_{1i}^* je odezvou v normálním regresním modelu

$$Y_{1i}^* = A + B Y_{0i} + \Delta|_{[i \in E]} + \varepsilon_i^*, \quad i = 1, \dots, n, \quad (3.3)$$

a označme parametry tohoto modelu $\boldsymbol{\theta} = (A, B, \Delta)^T$. Tento model je v podstatě model III z (2.5) doplněný o předpoklad normality a shodnosti rozptylů. V případě normálního regresního modelu je maximálně věrohodný odhad jeho parametrů roven odhadu metodou nejmenších čtverců. Odhad $\hat{\boldsymbol{\beta}}$ parametrů z modelu (2.5) získaný metodou nejmenších čtverců je tedy shodný s odhadem metodou maximální věrohodnosti z tohoto modelu, doplníme-li předpoklady modelu o normalitu a shodnost rozptylů. Naším cílem je zjistit, k čemu odhad $\hat{\boldsymbol{\beta}}$ za platnosti modelu I konverguje.

Věta 3.11. *Nechť platí předpoklady věty 3.10. Označme $F_{\boldsymbol{\theta}}$ rozdělení Y_{1i}^* v normálním regresním modelu (3.3). Předpokládejme, že existují derivace $\frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{\theta}}(x)$ pro každé $i = 1, 2, 3$ a můžeme zaměnit derivaci a integrál*

$$\frac{\partial}{\partial \theta_i} \mathbf{E}_{model I} [\log f_{\boldsymbol{\theta}}(X)] = \mathbf{E}_{model I} \left[\frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{\theta}}(X) \right].$$

Pak za platnosti modelu I platí

$$\begin{aligned} \hat{\alpha} &\xrightarrow{P} \mu - \frac{\mu_0}{\sigma_0^2} \left\{ (1 - \pi) \text{cov}(Y_{0i}, h_C(Y_{0i})) + \pi \text{cov}(Y_{0i}, h_E(Y_{0i}, Z_i)) \right\}, \\ \hat{\beta} &\xrightarrow{P} \frac{1}{\sigma_0^2} \left\{ (1 - \pi) \text{cov}(Y_{0i}, h_C(Y_{0i})) + \pi \text{cov}(Y_{0i}, h_E(Y_{0i}, Z_i)) \right\}, \\ \hat{\delta}^{III} &\xrightarrow{P} \delta. \end{aligned}$$

Důkaz. Odhad $\hat{\boldsymbol{\beta}}$ je kvazi-maximálně věrohodným odhadem $\boldsymbol{\theta} = (A, B, \Delta)^T$ z definice 3.7. Dle tvrzení 3.10 chceme nalézt $\boldsymbol{\theta}^*$, které minimalizuje

$$K(g, f_{\boldsymbol{\theta}}) = \mathbf{E}_{model I} \left[\log \frac{g(X)}{f_{\boldsymbol{\theta}}(X)} \right] = \mathbf{E}_{model I} [\log g(X)] - \mathbf{E}_{model I} [\log f_{\boldsymbol{\theta}}(X)]$$

Parametr $\boldsymbol{\theta}^*$ tedy získáme vyřešením rovnic

$$\mathbf{E}_{model I} \left[\frac{\partial}{\partial \theta_i} \log f_{\boldsymbol{\theta}}(X) \right], \quad i = 1, 2, 3.$$

V tomto případě je předchozí soustava shodná s vyřešením

$$\mathbf{E} \mathbf{U}_i(A, B, \Delta) = 0 \tag{3.4}$$

za platnosti modelu I, kde \mathbf{U}_i je definováno v (2.8).

Abychom zjistili, k čemu odhad $\hat{\boldsymbol{\beta}}$ konverguje, je třeba vypočítat střední hodnotu prvků vektoru $\mathbf{U}_i(A, B, \Delta)$ za platnosti modelu I. Spočítejme střední hodnotu prvního prvku

$$\mathbf{E} [\mathbf{U}_{i(1)}(A, B, \Delta)] = \mathbf{E} Y_{1i} - A - B \mu_0 - \Delta \pi,$$

kde

$$\begin{aligned} \mathbf{E} Y_{1i} &= \mathbf{E} [h_C(Y_{0i}) \mathbf{1}_{\{i \in C\}} + h_E(Y_{0i}, Z_i) \mathbf{1}_{\{i \in E\}} + \eta_i] \\ &= \mu(1 - \pi) + \pi \mathbf{E} \left[\mathbf{E} [h_E(Y_{0i}, Z_i) \mid Z_i] \right] \\ &= \mu(1 - \pi) + \pi \mathbf{E} [(\mu + \delta_0) \mathbf{1}_{\{Z_i=0\}} + (\mu + \delta_1) \mathbf{1}_{\{Z_i=1\}}] \\ &= \mu(1 - \pi) + \pi \left((\mu + \delta_0)(1 - p) + (\mu + \delta_1)p \right) \\ &= \mu + \pi \left((1 - p)\delta_0 + p\delta_1 \right) = \mu + \delta \pi \end{aligned}$$

Střední hodnota druhého prvku je

$$\mathbf{E} [\mathbf{U}_{i(2)} (A, B, \Delta)] = \mathbf{E} [Y_{0i} Y_{1i}] - A \mu_0 - B \mathbf{E} Y_{0i}^2 - \Delta \pi \mu_0,$$

kde

$$\begin{aligned} \mathbf{E} [Y_{0i} Y_{1i}] &= \mathbf{E} [Y_{0i} h_C (Y_{0i}) \mathbf{I}_{[i \in C]}] + \mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i) \mathbf{I}_{[i \in E]}] + \mathbf{E} [Y_{0i} \eta_i] \\ &= (1 - \pi) \mathbf{E} [Y_{0i} h_C (Y_{0i})] + \pi \mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i)] + \\ &\quad + \mathbf{E} [Y_{0i} \mathbf{E} [\eta_i | Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i]] \\ &= (1 - \pi) \mathbf{E} [Y_{0i} h_C (Y_{0i})] + \pi \mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i)]. \end{aligned}$$

Střední hodnota třetího prvku je

$$\mathbf{E} [\mathbf{U}_{i(3)} (A, B, \Delta)] = \mathbf{E} [Y_{1i} \mathbf{I}_{[i \in E]}] - A \pi - B \mu_0 \pi - \Delta \pi,$$

kde

$$\begin{aligned} \mathbf{E} [Y_{1i} \mathbf{I}_{[i \in E]}] &= \mathbf{E} [h_E (Y_{0i}, Z_i) \mathbf{I}_{[i \in E]}] + \mathbf{E} [\eta_i \mathbf{I}_{[i \in E]}] \\ &= \pi \mathbf{E} [\mathbf{E} [h_E (Y_{0i}, Z_i) | Z_i]] + \mathbf{E} [\mathbf{I}_{[i \in E]} \mathbf{E} [\eta_i | Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i]] \\ &= \pi \mathbf{E} [(\mu + \delta_0) \mathbf{I}_{[Z_i=0]} + (\mu + \delta_1) \mathbf{I}_{[Z_i=1]}] \\ &= \pi ((\mu + \delta_0) (1 - p) + (\mu + \delta_1) p) \\ &= \pi \mu + \pi \delta \end{aligned}$$

Díky předchozím rovnicím, můžeme rovnici (3.4) přepsat do soustavy rovnic

$$\begin{aligned} A + \mu_0 B + \pi \Delta &= \mu + \pi \delta \\ \mu_0 A + \mathbf{E} Y_{0i}^2 B + \mu_0 \pi \Delta &= (1 - \pi) \mathbf{E} [Y_{0i} h_C (Y_{0i})] + \pi \mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i)] \\ \pi A + \mu_0 \pi B + \pi \Delta &= \pi \mu + \pi \delta \end{aligned} \tag{3.5}$$

Z první a třetí rovnice získáváme

$$\Delta = \delta.$$

Vynásobením μ_0 první rovnice a odečtením druhé v (3.5) získáváme

$$(\mu_0^2 - \mathbf{E} Y_{0i}^2) B = \mu \mu_0 + \mu_0 \pi \delta - (1 - \pi) \mathbf{E} [Y_{0i} h_C (Y_{0i})] - \pi \mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i)],$$

z čehož dostáváme

$$\begin{aligned} B &= -\frac{1}{\sigma_0^2} \left\{ \mu \mu_0 + \mu_0 \pi \delta - (1 - \pi) \mathbf{E} [Y_{0i} h_C (Y_{0i})] - \pi \mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i)] \right\} \\ &= \frac{1}{\sigma_0^2} \left\{ (1 - \pi) \left(\mathbf{E} [Y_{0i} h_C (Y_{0i})] - \mu \mu_0 \right) + \pi \left(\mathbf{E} [Y_{0i} h_E (Y_{0i}, Z_i)] - \mu \mu_0 - \mu_0 \delta \right) \right\} \\ &= \frac{1}{\sigma_0^2} \left\{ (1 - \pi) \text{cov}(Y_{0i}, h_C (Y_{0i})) + \pi \text{cov}(Y_{0i}, h_E (Y_{0i}, Z_i)) \right\} \end{aligned}$$

Dosazením vypočítaných hodnot Δ a B do první rovnice v (3.5) dostáváme poslední parametr

$$A = \mu - \frac{\mu_0}{\sigma_0^2} \left\{ (1 - \pi) \text{cov}(Y_{0i}, h_C(Y_{0i})) + \pi \text{cov}(Y_{0i}, h_E(Y_{0i}, Z_i)) \right\}$$

Z věty 3.10 a z vypočítaných rovností pro A , B a Δ plyne tvrzení této věty. \square

V další větě se budeme zabývat asymptotickým rozdělení odhadu $\hat{\delta}^{III}$, platí-li model I.

Věta 3.12. *Nechť platí předpoklady věty 3.11. Označme*

$$\begin{aligned} \psi(Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i) &= \text{var}(\eta_i | Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i), \\ K(Y_{0i}) &= \pi \mathbf{E} \left[(h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i}))^2 p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right], \\ L(Y_{0i}, \mathbf{1}_{[i \in E]}) &= \left\{ h_{E,0}(Y_{0i}) (1 - p_Z(Y_{0i})) + h_{E,1}(Y_{0i}) p_Z(Y_{0i}) \right\} \mathbf{1}_{[i \in E]} - \\ &\quad - \alpha - \beta Y_{0i} - \delta \mathbf{1}_{[i \in E]}. \end{aligned}$$

Pak asymptotické rozdělení odhadu $\hat{\delta}^{III}$ za platnosti modelu I je

$$\sqrt{n} (\hat{\delta}^{III} - \delta) \xrightarrow{d} \mathbf{N}(0, W_{\hat{\delta}^{III}}^I),$$

kde

$$\begin{aligned} W_{\hat{\delta}^{III}}^I &= \frac{1}{(1 - \pi)^2} \mathbf{E} \psi(Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i) + \frac{1 - 2\pi}{\pi^2 (1 - \pi)^2} \mathbf{E} \left[\psi(Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i) \mathbf{1}_{[i \in E]} \right] + \\ &\quad + \frac{1}{\pi^2} K(Y_{0i}) + \frac{1}{(1 - \pi)^2} \text{var} \left(h_C(Y_{0i}) \mathbf{1}_{[i \in C]} + L(Y_{0i}, \mathbf{1}_{[i \in E]}) \right) - \\ &\quad - \frac{2}{\pi (1 - \pi)^2} \text{cov} \left(L(Y_{0i}, \mathbf{1}_{[i \in E]}), L(Y_{0i}, \mathbf{1}_{[i \in E]}) \mathbf{1}_{[i \in E]} \right) + \\ &\quad + \frac{1}{\pi^2 (1 - \pi)^2} \text{var} \left(L(Y_{0i}, \mathbf{1}_{[i \in E]}) \mathbf{1}_{[i \in E]} \right). \end{aligned}$$

Důkaz. Nechť platí model I. K výpočtu asymptotického rozptylu $W_{\hat{\delta}^{III}}^I$ použijeme větu 2.8. Rozptyl $W_{\hat{\delta}^{III}}^I$ spočítáme jako

$$W_{\hat{\delta}^{III}}^I = [\mathbf{D}^{-1} \mathbf{V} \mathbf{D}^{-1}]_{[3,3]},$$

kde inverzní matice \mathbf{D}^{-1} je uvedena v (2.11) a díky jejímu tvaru stačí pro počítání $W_{\hat{\delta}^{III}}^I$ z matice \mathbf{V} znát pouze prvky V_{11} , V_{13} a V_{33} .

Spočítejme žádané prvky matice \mathbf{V} . Tuto matici rozepíšeme

$$\mathbf{V} = \mathbf{E} \left[\text{var}(\mathbf{U}_i(\alpha, \beta, \delta) | Y_{0i}, \mathbf{1}_{[i \in E]}) \right] + \text{var} \left(\mathbf{E}[\mathbf{U}_i(\alpha, \beta, \delta) | Y_{0i}, \mathbf{1}_{[i \in E]}] \right) \quad (3.6)$$

Vyjádříme žádané prvky z prvního sčítance matice \mathbf{V} . U každého prvku nejdříve vypočítáme podmíněný rozptyl či kovarianci a na výsledek aplikujeme střední hodnotu. Počítejme z matice $\text{var}(\mathbf{U}_i(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]})$ první prvek

$$\begin{aligned}
\text{var}(\mathbf{U}_{i(1)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}) &= \text{var}(Y_{1i} \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \\
&= \text{var}(h_C(Y_{0i}) \mathbf{I}_{[i \in C]} + h_E(Y_{0i}, Z_i) \mathbf{I}_{[i \in E]} + \eta_i \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \\
&= \mathbb{E} \left[\text{var}(h_E(Y_{0i}, Z_i) \mathbf{I}_{[i \in E]} + \eta_i \mid Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]} \right] + \\
&\quad + \text{var} \left(\mathbb{E} \left[h_E(Y_{0i}, Z_i) \mathbf{I}_{[i \in E]} + \eta_i \mid Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i \right] \mid Y_{0i}, \mathbf{I}_{[i \in E]} \right) \\
&= \mathbb{E} \left[\psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]} \right] + \text{var}(h_E(Y_{0i}, Z_i) \mathbf{I}_{[i \in E]} \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \\
&= \mathbb{E} \left[\psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]} \right] + \\
&\quad + \mathbf{I}_{[i \in E]} \text{var}(h_{E,0}(Y_{0i}) + (h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i})) \mathbf{I}_{[Z_i=1]} \mid Y_{0i}) \\
&= \mathbb{E} \left[\psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]} \right] + \\
&\quad + \mathbf{I}_{[i \in E]} (h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i}))^2 p_Z(Y_{0i}) (1 - p_Z(Y_{0i}))
\end{aligned}$$

a aplikujme střední hodnotu

$$\begin{aligned}
\mathbb{E} \left[\text{var}(\mathbf{U}_{i(1)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \right] &= \\
&= \mathbb{E} \psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) + \pi \mathbb{E} \left[(h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i}))^2 p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right].
\end{aligned}$$

Spočítejme druhý prvek prvního sčítance matice \mathbf{V} . Nejdříve vypočítáme

$$\begin{aligned}
\text{var}(\mathbf{U}_{i(3)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}) &= \mathbf{I}_{[i \in E]} \text{var}(Y_{1i} \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \\
&= \mathbb{E} \left[\mathbf{I}_{[i \in E]} \psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \mid Y_{0i}, \mathbf{I}_{[i \in E]} \right] + \\
&\quad + \mathbf{I}_{[i \in E]} (h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i}))^2 p_Z(Y_{0i}) (1 - p_Z(Y_{0i}))
\end{aligned}$$

a pak aplikujme střední hodnotu

$$\begin{aligned}
\mathbb{E} \left[\text{var}(\mathbf{U}_{i(3)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \right] &= \mathbb{E} \left[\mathbf{I}_{[i \in E]} \psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \right] + \\
&\quad + \pi \mathbb{E} \left[(h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i}))^2 p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right].
\end{aligned}$$

Podívejme se na poslední prvek prvního sčítance matice \mathbf{V} . Jelikož platí

$$\text{cov}(\mathbf{U}_{i(1)}(\alpha, \beta, \delta), \mathbf{U}_{i(3)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}) = \text{var}(\mathbf{U}_{i(3)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}),$$

tak

$$\begin{aligned}
\mathbb{E} \left[\text{cov}(\mathbf{U}_{i(1)}(\alpha, \beta, \delta), \mathbf{U}_{i(3)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{I}_{[i \in E]}) \right] &= \mathbb{E} \left[\mathbf{I}_{[i \in E]} \psi(Y_{0i}, \mathbf{I}_{[i \in E]}, Z_i) \right] + \\
&\quad + \pi \mathbb{E} \left[(h_{E,1}(Y_{0i}) - h_{E,0}(Y_{0i}))^2 p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right].
\end{aligned}$$

Spočítejme podmíněnou střední hodnotu z druhého sčítance matice \mathbf{V} z (3.6).
Počítejme

$$\mathbb{E}[\mathbf{U}_{i(1)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{1}_{[i \in E]}] = \mathbb{E}[Y_{1i} \mid Y_{0i}, \mathbf{1}_{[i \in E]}] - \alpha - \beta Y_{0i} - \delta \mathbf{1}_{[i \in E]},$$

kde

$$\begin{aligned} \mathbb{E}[Y_{1i} \mid Y_{0i}, \mathbf{1}_{[i \in E]}] &= \mathbb{E}\left[h_C(Y_{0i}) \mathbf{1}_{[i \in C]} + h_E(Y_{0i}, Z_i) \mathbf{1}_{[i \in E]} + \eta_i \mid Y_{0i}, \mathbf{1}_{[i \in E]}\right] \\ &= h_C(Y_{0i}) \mathbf{1}_{[i \in C]} + \mathbf{1}_{[i \in E]} \mathbb{E}\left[h_{E,0}(Y_{0i}) \mathbf{1}_{[Z_i=0]} + h_{E,1}(Y_{0i}) \mathbf{1}_{[Z_i=1]} \mid Y_{0i}\right] \\ &= h_C(Y_{0i}) \mathbf{1}_{[i \in C]} + \left\{h_{E,0}(Y_{0i}) (1 - p_Z(Y_{0i})) + h_{E,1}(Y_{0i}) p_Z(Y_{0i})\right\} \mathbf{1}_{[i \in E]}, \end{aligned}$$

a podmíněnou střední hodnotu

$$\begin{aligned} \mathbb{E}[\mathbf{U}_{i(3)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{1}_{[i \in E]}] &= \mathbf{1}_{[i \in E]} \mathbb{E}[\mathbf{U}_{i(1)}(\alpha, \beta, \delta) \mid Y_{0i}, \mathbf{1}_{[i \in E]}] \\ &= \left\{h_{E,0}(Y_{0i}) (1 - p_Z(Y_{0i})) + h_{E,1}(Y_{0i}) p_Z(Y_{0i})\right\} \mathbf{1}_{[i \in E]} - \\ &\quad - \alpha \mathbf{1}_{[i \in E]} - \beta Y_{0i} \mathbf{1}_{[i \in E]} - \delta \mathbf{1}_{[i \in E]}. \end{aligned}$$

Z předchozích rovnic získáváme žádané prvky matice \mathbf{V} . Použijeme značení ve znění věty. Prvky matice \mathbf{V} jsou

$$\begin{aligned} V_{11} &= \mathbb{E} \psi(Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i) + K(Y_{0i}) + \text{var}\left(h_C(Y_{0i}) \mathbf{1}_{[i \in C]} + L(Y_{0i}, \mathbf{1}_{[i \in E]})\right), \\ V_{33} &= \mathbb{E}\left[\psi(Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i) \mathbf{1}_{[i \in E]}\right] + K(Y_{0i}) + \text{var}\left(L(Y_{0i}, \mathbf{1}_{[i \in E]}) \mathbf{1}_{[i \in E]}\right), \\ V_{13} &= \mathbb{E}\left[\psi(Y_{0i}, \mathbf{1}_{[i \in E]}, Z_i) \mathbf{1}_{[i \in E]}\right] + K(Y_{0i}) + \\ &\quad + \text{cov}\left(L(Y_{0i}, \mathbf{1}_{[i \in E]}), L(Y_{0i}, \mathbf{1}_{[i \in E]}) \mathbf{1}_{[i \in E]}\right). \end{aligned}$$

Prvky matice \mathbf{V} máme již spočítané. Můžeme spočítat asymptotický rozptyl $W_{\hat{\delta}_{III}}^I$. Třetí řádek matice $\mathbf{D}^{-1}\mathbf{V}$ je

$$\begin{aligned} [\mathbf{D}^{-1}\mathbf{V}]_{[3,1]} &= -\frac{1}{1-\pi}V_{11} + \frac{1}{\pi(1-\pi)}V_{13}, \\ [\mathbf{D}^{-1}\mathbf{V}]_{[3,3]} &= -\frac{1}{1-\pi}V_{13} + \frac{1}{\pi(1-\pi)}V_{33}. \end{aligned}$$

Rozptyl $W_{\hat{\delta}_{III}}^I$ se rovná

$$\begin{aligned} W_{\hat{\delta}_{III}}^I &= [\mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}]_{[3,3]} \\ &= -\frac{1}{1-\pi} \left(-\frac{1}{1-\pi}V_{11} + \frac{1}{\pi(1-\pi)}V_{13}\right) \\ &\quad + \frac{1}{\pi(1-\pi)} \left(-\frac{1}{1-\pi}V_{13} + \frac{1}{\pi(1-\pi)}V_{33}\right) \\ &= \frac{1}{(1-\pi)^2}V_{11} - \frac{2}{\pi(1-\pi)^2}V_{13} + \frac{1}{\pi^2(1-\pi)^2}V_{33} \end{aligned}$$

a po dosazení prvků V_{11} , V_{13} a V_{33} získáváme vzorec pro $W_{\hat{\delta}_{III}}^I$ dle znění věty. \square

V této podkapitole jsme zjistili, že odhad $\hat{\delta}^{III}$ je za platnosti modelu I konzistentní a určili jsme jeho asymptotické rozdělení. Díky vzorci z předchozí věty se domníváme, že při vhodné volbě parametrů může být asymptotický rozptyl odhadu $\hat{\delta}^{III}$ větší i menší než asymptotický rozptyl odhadu $\hat{\delta}^I$.

3.7 Porovnání chování odhadů

V této podkapitole shrneme výsledky získané v předchozích kapitolách. Jelikož odhady $\hat{\delta}^I$, $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ jsou za platnosti všech modelů nestranné a konzistentní, budeme tyto odhady porovnávat podle asymptotického rozptylu. V každém ze tří modelů vyjádříme asymptotické rozptyly odhadů pomocí asymptotického rozptylu odhadu založeného na základě zkoumaného modelu a následně je porovnáme.

Chování odhadů za platnosti modelu II

Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu II jsme prozkoumali v kapitole 2.2. Asymptotický rozptyl tohoto odhadu je

$$V_{\hat{\delta}^{II}} = (q+1) \left(\sigma_{II,C}^2 + \frac{1}{q} \left(p(1-p)(\delta_0 - \delta_1)^2 + p\sigma_{II,E,1}^2 + (1-p)\sigma_{II,E,0}^2 \right) \right).$$

Dále jsme v kapitolách 3.1 a 3.2 získali vyjádření asymptotických rozptylů odhadů $\hat{\delta}^I$ a $\hat{\delta}^{III}$ pomocí asymptotického rozptylu $V_{\hat{\delta}^{II}}$

$$\begin{aligned} W_{\hat{\delta}^I}^{II} &= V_{\hat{\delta}^{II}} + \frac{(q+1)^2}{q} \sigma_0^2, \\ W_{\hat{\delta}^{III}}^{II} &= V_{\hat{\delta}^{II}} - \frac{q+1}{q} (\delta_0 - \delta_1)^2 \text{var } p_Z(Y_{0i}). \end{aligned}$$

Vidíme, že asymptotický rozptyl odhadu $\hat{\delta}^I$ je vždy větší než asymptotický rozptyl odhadu $\hat{\delta}^{II}$ a asymptotický rozptyl odhadu $\hat{\delta}^{III}$ je vždy menší nebo rovný asymptotickému rozptylu odhadu $\hat{\delta}^{II}$. Celkem získáváme

$$W_{\hat{\delta}^I}^{II} > V_{\hat{\delta}^{II}} \geq W_{\hat{\delta}^{III}}^{II},$$

přičemž rovnost napravo nastává, jestliže je funkce $p_Z(Y_{0i})$ konstantní.

Chování odhadů za platnosti modelu III

Chování odhadu $\hat{\delta}^{III}$ za platnosti modelu III jsme prozkoumali v kapitole 2.3. Asymptotický rozptyl tohoto odhadu je

$$\begin{aligned} V_{\hat{\delta}^{III}} &= (q+1) \left\{ \sigma_{III,C}^2 + \frac{1}{q} \left\{ (\delta_1 - \delta_0)^2 \mathbb{E} \left[p_Z(Y_{0i}) (1 - p_Z(Y_{0i})) \right] \right. \right. \\ &\quad \left. \left. + p\sigma_{III,E,1}^2 + (1-p)\sigma_{III,E,0}^2 \right\} \right\}. \end{aligned}$$

Dále jsme v kapitolách 3.3 a 3.4 získali vyjádření asymptotických rozptylů odhadů $\hat{\delta}^I$ a $\hat{\delta}^{II}$ pomocí asymptotického rozptylu $V_{\hat{\delta}^{III}}$

$$W_{\hat{\delta}^I}^{III} = V_{\hat{\delta}^{III}} + \frac{(q+1)^2}{q} \beta^2 \sigma_0^2 + \frac{q+1}{q} \left\{ 2\beta (\delta_1 - \delta_0) \text{cov}(Y_{0i}, p_Z(Y_{0i})) + (\delta_1 - \delta_0)^2 \text{var } p_Z(Y_{0i}) \right\},$$

$$W_{\hat{\delta}^{II}}^{III} = V_{\hat{\delta}^{III}} + \frac{(q+1)^2}{q} (\beta - 1)^2 \sigma_0^2 + \frac{q+1}{q} \left\{ 2(\beta - 1) (\delta_1 - \delta_0) \text{cov}(Y_{0i}, p_Z(Y_{0i})) + (\delta_1 - \delta_0)^2 \text{var } p_Z(Y_{0i}) \right\}.$$

Předpokládejme, že $p_Z(Y_{0i}) = p$. Pak asymptotický rozptyl $W_{\hat{\delta}^I}^{III}$ je větší než $V_{\hat{\delta}^{III}}$, pokud $\beta \neq 0$. Podobně asymptotický rozptyl $W_{\hat{\delta}^{II}}^{III}$ je větší než $V_{\hat{\delta}^{III}}$, pokud $\beta \neq 1$. Dále asymptotický rozptyl $W_{\hat{\delta}^I}^{III}$ je větší než $W_{\hat{\delta}^{II}}^{III}$, jestliže

$$\frac{(q+1)^2}{q} \beta^2 \sigma_0^2 > \frac{(q+1)^2}{q} (\beta - 1)^2 \sigma_0^2,$$

což po úpravě dává $\beta > \frac{1}{2}$.

Celkem získáváme následující. Platí-li rovnost $p_Z(Y_{0i}) = p$ a $\beta > \frac{1}{2}$, pak asymptotické rozptyly odhadů za platnosti modelu II splňují

$$W_{\hat{\delta}^I}^{III} > W_{\hat{\delta}^{II}}^{III} \geq V_{\hat{\delta}^{III}}.$$

Platí-li rovnost $p_Z(Y_{0i}) = p$ a $\beta < \frac{1}{2}$, pak asymptotické rozptyly odhadů za platnosti modelu II splňují

$$W_{\hat{\delta}^{II}}^{III} > W_{\hat{\delta}^I}^{III} \geq V_{\hat{\delta}^{III}}.$$

Chování odhadů za platnosti modelu I

Chování odhadu $\hat{\delta}^I$ za platnosti modelu I jsme prozkoumali v kapitole 2.1. Asymptotický rozptyl tohoto odhadu je

$$V_{\hat{\delta}^I} = (q+1) \left(\sigma_{I,C}^2 + \frac{1}{q} \left(p(1-p) (\delta_0 - \delta_1)^2 + p \sigma_{I,E,1}^2 + (1-p) \sigma_{I,E,0}^2 \right) \right).$$

Dále jsme v kapitolách 3.5 a 3.6 získali vyjádření asymptotických rozptylů odhadů $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ pomocí asymptotického rozptylu $V_{\hat{\delta}^I}$

$$W_{\hat{\delta}^{II}}^I = V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} \sigma_0^2 - 2(q+1) \left(\text{cov}(Y_{0i}^C, Y_{1i}^C) + \frac{1}{q} \text{cov}(Y_{0i}^E, Y_{1i}^E) \right).$$

a vzorec pro $W_{\hat{\delta}^{III}}^I$ nalezneme ve větě 3.12, zde ho z důvodu jeho složitosti uvádět nebudeme.

Odhad $\hat{\delta}^{II}$ má větší asymptotický rozptyl než $\hat{\delta}^I$ za platnosti modelu I, jestliže

$$\frac{\text{cov}(Y_{0i}^E, Y_{1i}^E)}{\sigma_0^2} < \frac{1}{2} \quad \text{a} \quad \frac{\text{cov}(Y_{0i}^C, Y_{1i}^C)}{\sigma_0^2} < \frac{1}{2}.$$

Díky složitosti vzorce pro $W_{\hat{\delta}^{III}}^I$ nevíme, zda odhad $\hat{\delta}^{III}$ má větší či menší asymptotický rozptyl než odhad $\hat{\delta}^I$ nebo $\hat{\delta}^{II}$. Domníváme se, že při vhodné volbě parametrů mohou nastat všechny situace.

4. Simulační studie nekonstatního efektu léčby

V této části se budeme zabývat ověřováním spočítaných teoretických výsledků. V kapitolách 2 a 3 jsme předpokládali, že podání léku nemá stejný účinek na každém pacientovi. V kapitole 2 jsme uvedli tři modely a odhady efektu léčby na základě těchto modelů. V kapitole 3 jsme se zabývali porovnáním asymptotických rozptylů těchto odhadů za platnosti jednotlivých modelů. V této kapitole prověříme získané teoretické výsledky pomocí simulací. V práci [1] nalezneme podobné ověření za předpokladu, že podání léku má stejný účinek na každém pacientovi. Výsledkem této práce bylo doporučení vždy použít odhad efektu léčby $\hat{\delta}^{III}$.

Následné výpočty provádíme na základě 5000 simulací, přičemž v každé simulaci vygenerujeme počáteční a konečné hodnoty pacientů. Z těchto hodnot spočítáme průměrné odhady efektu léčby $\hat{\delta}^I$, $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$, empirické směrodatné odchylky, průměrné odhady asymptotických odchylek a pravděpodobnosti pokrytí. Tyto hodnoty budeme prezentovat pro každé zadání parametrů ve výsledných tabulkách. Výpočty provádíme v programu R.

Nejdříve získáme počáteční hodnoty pacientů. Vygenerujeme n počátečních hodnot veličiny Y ze zadaného rozdělení, čímž získáme hodnoty Y_{0i} , $i = 1, \dots, n$. Následně roztrídíme pacienty do experimentální a kontrolní skupiny. Pacienta přiřadíme do experimentální skupiny s předem zadanou pravděpodobností π . Pokud se pacient nachází v experimentální skupině, přiřadíme ho do podskupiny $Z_i = 1$ s pravděpodobností $p_Z(Y_{0i})$. Celkem získáme počáteční hodnoty pacientů Y_{0i}^C a Y_{0i}^E a roztrídění pacientů v experimentální skupině do podskupin.

Dále chceme vygenerovat konečné hodnoty Y_{1i} . V podkapitole 3.6 jsme zapisovali model I pomocí funkcí $h_C(Y_{0i})$ a $h_E(Y_{0i}, Z_i)$ a chyby η_i s podmíněným rozptylem $\psi(Y_{0i}, \mathbb{1}_{[i \in E]}, Z_i)$. Každý ze tří modelů se dá pomocí těchto funkcí zapsat. Výslednou konečnou hodnotu Y_{1i} spočítáme jako součet

$$h_C(Y_{0i}) \mathbb{1}_{[i \in C]} + h_E(Y_{0i}, Z_i) \mathbb{1}_{[i \in E]}$$

a náhodné chyby se zadaným rozdělením a s rozptylem $\psi(Y_{0i}, \mathbb{1}_{[i \in E]}, Z_i)$. Získáme tak konečné hodnoty pacientů Y_{1i} , které můžeme uspořádat jako hodnoty Y_{1i}^C a Y_{1i}^E .

V každé ze simulací spočítáme odhady $\hat{\delta}^I$, $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$. Vypočítané průměry těchto odhadů ze všech simulací uvádíme ve výsledné tabulce. Po skončení všech simulací spočítáme pro každý odhad jeho empirickou směrodatnou odchylku, kterou ve výsledné tabulce také uvádíme.

Dále v každé ze simulací spočítáme odhady jejich asymptotických rozptylů (viz kapitola 2). Odhad asymptotického rozptylu $\hat{\delta}^{III}$ počítáme dvojím způsobem.

První odhad je spočítaný na základě regresního modelu. Druhý odhad získáme pomocí sendvičové varianční matice $\hat{D}^{-1}\hat{V}\hat{D}^{-1}$, kde \hat{V} je empirická varianční matice vektoru $U_i(\hat{\alpha}, \hat{\beta}, \hat{\delta}^{III})$ a

$$\hat{D} = \frac{1}{n} \sum_{i=1}^n (1, Y_{0i}, I_{[i \in E]})^T (1, Y_{0i}, I_{[i \in E]}).$$

Výsledná tabulka udává průměry přes simulace všech těchto odhadů asymptotických rozptylů.

V každé ze simulací spočítáme pro všechny odhady 95%-ní intervaly spolehlivosti pro δ

$$\left(\hat{\delta} - u_{0.975}\sqrt{v}, \hat{\delta} + u_{0.975}\sqrt{v} \right),$$

kde $\hat{\delta}$ je některý z odhadů $\hat{\delta}^I$, $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ a v je jeho odhadnutý asymptotický rozptyl (pro $\hat{\delta}^{III}$ uvažujeme dva typy odhadů). V každé simulaci zjistíme, zda tento interval spolehlivosti pokrývá příslušný typ průměrného odhadu δ . Po skončení všech simulací spočítáme celkový podíl pokrytí, který uvádíme ve výsledné tabulce. Pravděpodobnost pokrytí δ tímto intervalem spolehlivosti je v tabulce vyšší než ve skutečnosti, jelikož vlastně počítáme pravděpodobnost pokrytí průměrného odhadu δ .

V následujících podkapitolách se zabýváme těmito simulacemi za předpokladu, že data pocházejí z jednoho z modelů. Pro každý model je nutno zvolit velké množství parametrů, z čehož vzniká ještě větší množství kombinací. Prováděli jsme simulace pro mnoho těchto kombinací. V každé kapitole uvádíme pouze vybrané kombinace parametrů, které ilustrují zajímavé situace.

4.1 Konečná hodnota pocházející z modelu II

Nejdříve se podíváme na simulace, ve kterých konečnou hodnotu modelujeme pomocí modelu II. Model II vyjádříme pomocí již zmíněných funkcí $h_C(Y_{0i})$, $h_E(Y_{0i}, Z_i)$ a náhodné chyby η_i s podmíněným rozptylem $\psi(Y_{0i}, I_{[i \in E]}, Z_i)$. V modelu II platí

- $h_C(x) = \gamma + x$,
- $h_E(x, z) = \gamma + x + \delta_0(1 - z) + \delta_1 z$,
- $\psi(x, y, z) = a + by + cyz$.

Závislost konečné hodnoty na počáteční je v tomto modelu stejná v kontrolní i experimentální skupině. V experimentální skupině připisujeme příslušné efekty léčby dle podskupiny.

V tabulce 4.1 volíme počáteční hodnotu z dvojité exponenciálního rozdělení se střední hodnotou $\mu_0 = 100$ a směrodatnou odchylkou $\sigma_0 = 10$. Pacient v experimentální skupině je přiřazen do podskupiny $Z_i = 1$ s konstantní pravděpodobností $p_Z(x) = 0.4$, efekt léčby v této podskupině je $\delta_1 = 2$ a v podskupině $Z_i = 0$ je $\delta_0 = 5$. Rozdělení náhodných chyb η_i volíme normální. Vidíme, že empirické odchylky odhadů i jejich průměrné odhady asymptotických odchylek vychází podobně u odhadů $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$. Odhad $\hat{\delta}^I$ má signifikantně větší směrodatnou odchylku větší než ostatní odhady. V tomto případě považujeme tedy odhady $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ za stejně dobré a odhad $\hat{\delta}^I$ za horší než ostatní.

V tabulce 4.2 volíme počáteční hodnotu z normálního rozdělení se střední hodnotou $\mu_0 = 100$ a směrodatnou odchylkou $\sigma_0 = 10$. Pacient v experimentální skupině je přiřazen do podskupiny $Z_i = 1$ s pravděpodobností

$$p_Z(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - 100),$$

efekt léčby v této podskupině je $\delta_0 = 5$ a v podskupině $Z_i = 1$ je $\delta_0 = 20$. Rozdělení náhodných chyb η_i volíme zesílené normální s parametrem šikmosti 5. Odhad $\hat{\delta}^{III}$ má v tomto případě signifikantně menší směrodatnou odchylku než odhady $\hat{\delta}^I$ a $\hat{\delta}^{II}$.

Výsledky v tabulkách 4.1 a 4.2 korespondují s výsledky v kapitole 3.7. V této kapitole se dozvídáme, že největší asymptotický rozptyl má za platnosti modelu II odhad $\hat{\delta}^I$. Asymptotický rozptyl odhadu $\hat{\delta}^{II}$ je větší než asymptotický rozptyl odhadu $\hat{\delta}^{III}$ v případě, že funkce $p_Z(x)$ není konstantní (viz tabulka 4.2). Je-li funkce $p_Z(x)$ konstantní (viz tabulka 4.1), pak mají odhady $\hat{\delta}^{II}$ a $\hat{\delta}^{III}$ stejný asymptotický rozptyl. Jako nejlepší odhad považujeme tedy odhad $\hat{\delta}^{III}$.

4.2 Konečná hodnota pocházející z modelu III

Dále se podíváme na simulace, ve kterých konečnou hodnotu modelujeme pomocí modelu III. Model III také vyjádříme pomocí zmíněných funkcí $h_C(Y_{0i})$, $h_E(Y_{0i}, Z_i)$ a náhodné chyby η_i s podmíněným rozptylem $\psi(Y_{0i}, \mathbb{1}_{[i \in E]}, Z_i)$. V modelu III platí

- $h_C(x) = \alpha + \beta x$,
- $h_E(x, z) = \alpha + \beta x + \delta_0(1 - z) + \delta_1 z$,
- $\psi(x, y, z) = a + b y + c y z$.

Závislost konečné hodnoty na počáteční je v tomto modelu stejná v kontrolní i experimentální skupině, pouze v experimentální skupině přisuzujeme příslušné efekty léčby dle podskupiny.

V tabulkách 4.3 a 4.4 se zabýváme případem, kdy funkce přiřazení pacienta v experimentální skupině do podskupiny $Z_i = 1$ je konstantní. Počáteční hodnota má

Volba parametrů		
$n = 850$	$\pi = 0.4$	$p_Z(x) = 0.4$
rozdělení Y_{0i} dvojitě exponenciální	$\mu_0 = 100$	$\sigma_0 = 10$
rozdělení η_i normální		

Funkce modelu II	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 10 + x$	$\hat{\delta}^I$	4.997	0.795	0.791	0.950
$h_E(x, z) = \begin{cases} 17 + x, & z = 0 \\ 12 + x, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	4.995	0.369	0.369	0.952
$\psi(x, y, z) = 20 + 5y + 5yz$	$\hat{\delta}^{III}$	4.995	0.369	sendvičově z modelu	0.952
				0.351	0.936

Tabulka 4.1: Výsledky simulace pro data pocházející z modelu II.

Volba parametrů		
$n = 850$ rozdělení Y_{0i} normální rozdělení η_i zesíklené normální s parametrem šikmosti 5	$\pi = 0.4$ $\mu_0 = 100$	$p_Z(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - 100)$ $\sigma_0 = 4$

Funkce modelu II	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 10 + x$	$\hat{\delta}^I$	12.490	0.652	0.663	0.955
$h_E(x, z) = \begin{cases} 15 + x, & z = 0 \\ 30 + x, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	12.492	0.489	0.494	0.954
$\psi(x, y, z) = 20 + 5y + 5yz$	$\hat{\delta}^{III}$	12.494	0.451	sendvičově z modelu	0.949 0.927

Tabulka 4.2: Výsledky simulace pro data pocházející z modelu II.

normální rozdělení a rozdělení náhodných chyb η_i je zešikmené normální. V tabulce 4.3 volíme směrnici $\beta = 0.7$. V tomto případě má nejmenší směrodatné odchyly odhad $\hat{\delta}^{III}$, mírně větší odchyly má odhad $\hat{\delta}^{II}$ a největší odchyly má odhad $\hat{\delta}^I$. Změníme-li směrnici na $\beta = -1$ (viz tabulka 4.4), pak se pořadí velikosti odchylek změní. Nejmenší odchyly má znova odhad $\hat{\delta}^{III}$, větší má pak odhad $\hat{\delta}^I$ a největší odchyly má odhad $\hat{\delta}^{II}$. Tento výsledek koresponduje s výsledkem z kapitoly 3.7. V případě, že $p_Z(x)$ je konstantní, má vždy nejmenší asymptotický rozptyl odhad $\hat{\delta}^{III}$. Pokud je směrnice $\beta < 0.5$, tak je asymptotický rozptyl odhadu $\hat{\delta}^{II}$ větší než odhadu $\hat{\delta}^I$.

V kapitolách 3.3 a 3.4 vyjadřujeme domněnku, že odhady $\hat{\delta}^I$ a $\hat{\delta}^{II}$ můžou mít větší i menší asymptotický rozptyl než odhad $\hat{\delta}^{III}$. Víme, že je-li funkce $p_Z(x)$ konstantní, je asymptotický rozptyl odhadu $\hat{\delta}^{III}$ vždy nejmenší (viz předchozí odstavec). V tabulkách 4.5 a 4.6 uvažujeme případ, kdy funkce $p_Z(x)$ není konstantní. V tabulce 4.5 vidíme, že odchyly odhadu $\hat{\delta}^{III}$ jsou větší než odchyly odhadu $\hat{\delta}^{II}$. V tabulce 4.6 je zase uvedena situace, kdy směrodatné odchyly odhadu $\hat{\delta}^{III}$ jsou větší než odchyly odhadu $\hat{\delta}^I$. Tyto případy jsou však ojedinělé. Z převážné části má odhad $\hat{\delta}^{III}$ nejmenší směrodatné odchyly.

Celkem tuto situaci můžeme tedy rozdělit na dva případy. V prvním případě, kdy funkce $p_Z(x)$ je konstantní, vychází vždy nejlépe odhad $\hat{\delta}^{III}$. Ve druhém případě, kdy funkce $p_Z(x)$ není konstantní, směrodatná odchyly odhadu $\hat{\delta}^{III}$ může být signifikantně větší i menší než u ostatních odhadů. Případy, kdy je odchyly odhadu $\hat{\delta}^{III}$ větší než u některého z odhadů, jsou však spíše ojedinělé. Jako nejlepší odhad tedy považujeme odhad $\hat{\delta}^{III}$.

4.3 Konečná hodnota pocházející z modelu I

Zabývejme se nyní případem, kdy konečná hodnota pochází z modelu I. Model I můžeme vyjádřit pomocí funkcí $h_C(Y_{0i})$, $h_E(Y_{0i}, Z_i)$ a náhodné chyby η_i s podmíněným rozptylem $\psi(Y_{0i}, \mathbb{1}_{\{i \in E\}}, Z_i)$. Tyto funkce však nemůžeme zapsat pro celý model jednotně, jelikož nevíme, jakým způsobem závisí konečná hodnota na počáteční. Dále proto zkoumáme různé tvary těchto funkcí.

Podívejme se na případ, kdy

- $h_C(x) = x$,
- $h_E(x, z) = \begin{cases} 120, & z = 0, \\ 80 + 25\sqrt[3]{x - 80}, & z = 1. \end{cases}$

Výsledky simulace v tomto případě nalezneme v tabulce 4.7. Vidíme, že empirická směrodatná odchyly i průměrný odhad asymptotické odchyly jsou největší pro odhad $\hat{\delta}^I$, menší pro odhad $\hat{\delta}^{II}$ a nejmenší pro odhad $\hat{\delta}^{III}$. Pro data vzniklá pomocí výše uvedených funkcí je tedy nejlepší odhad $\hat{\delta}^{III}$, což se dá vysvětlit tím, že závislost konečné hodnoty na počáteční se dá nejlépe proložit přímkou s nenulovou směrnici.

Volba parametrů		
$n = 850$	$\pi = 0.4$	$p_Z(x) = 0.4$
rozdělení Y_{0i} normální	$\mu_0 = 100$	$\sigma_0 = 10$
rozdělení η_i zesílené normální s parametrem šikmosti 5		

Funkce modelu III	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 110 + 0.7x$	$\hat{\delta}^I$	5.013	0.842	0.865	0.953
$h_E(x, z) = \begin{cases} 117 + 0.7x, & z = 0 \\ 112 + 0.7x, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	5.009	0.730	0.743	0.952
	$\hat{\delta}^{III}$	5.010	0.694	sendvičově z modelu	0.712 0.709
$\psi(x, y, z) = 20 + 5y + 5yz$					

Tabulka 4.3: Výsledky simulace pro data pocházející z modelu III.

Volba parametrů		
$n = 850$ rozdělení Y_{0i} normální rozdělení η_i zesílené normální s parametrem šikmosti 5	$\pi = 0.4$ $\mu_0 = 100$	$p_Z(x) = 0.4$ $\sigma_0 = 10$

Funkce modelu III	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 110 - x$	$\hat{\delta}^I$	4.995	0.984	0.999	0.956
$h_E(x, z) = \begin{cases} 117 - x, & z = 0 \\ 112 - x, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	5.001	1.554	1.571	0.952
$\psi(x, y, z) = 20 + 5y + 5yz$	$\hat{\delta}^{III}$	4.988	0.700	sendvičově z modelu	0.953 0.952

Tabulka 4.4: Výsledky simulace pro data pocházející z modelu III.

Volba parametrů		
$n = 850$	$\pi = 0.3$	$p_Z(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - 50)$
rozdělení Y_{0i} normální	$\mu_0 = 50$	$\sigma_0 = 10$
rozdělení η_i normální		

Funkce modelu III	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 110 + 0.1x$	$\hat{\delta}^I$	22.500	1.188	1.182	0.947
$h_E(x, z) = \begin{cases} 115 + 0.1x, & z = 0 \\ 150 + 0.1x, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	22.500	0.920	0.911	0.950
	$\hat{\delta}^{III}$	22.500	0.996	sendvičově z modelu	0.946
$\psi(x, y, z) = 5 + 5y + 20yz$				0.696	0.830

Tabulka 4.5: Výsledky simulace pro data pocházející z modelu III.

Volba parametrů		
$n = 850$	$\pi = 0.3$	$p_Z(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - 100)$
rozdělení Y_{0i} normální	$\mu_0 = 100$	$\sigma_0 = 10$
rozdělení η_i normální		

Funkce modelu III	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 110 - x$	$\hat{\delta}^I$	22.519	0.940	0.943	0.951
$h_E(x, z) = \begin{cases} 115 - x, & z = 0 \\ 150 - x, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	22.508	1.242	1.248	0.953
$\psi(x, y, z) = 20 + 5y + 5yz$	$\hat{\delta}^{III}$	22.527	1.030	sendvičově z modelu	0.948 0.838

Tabulka 4.6: Výsledky simulace pro data pocházející z modelu III.

Dále uvažujme situaci, kdy

- $h_C(x) = 80 + \left(\frac{x-80}{10}\right)^3$,
- $h_E(x, z) = \begin{cases} 120, & z = 0, \\ 80 + 10\sqrt[3]{x - 80}, & z = 1. \end{cases}$

Výsledky simulace v tomto případě uvádíme v tabulce 4.8. Vidíme, že empirická směrodatná odchylka i průměrný odhad asymptotické odchylky jsou největší pro odhad $\hat{\delta}^{II}$, menší pro odhad $\hat{\delta}^{III}$ a nejmenší pro odhad $\hat{\delta}^I$. Nejlepším odhadem v této situaci je tedy odhad $\hat{\delta}^I$, který se snaží proložit data přímkou s nulovou směrnici. Odhad $\hat{\delta}^{III}$ prokládá data přímkou se stejnou směrnici v kontrolní a experimentální skupině a díky kontrolní skupině se přímka proloží s nenulovou směrnici.

V tabulce 4.9 jsou výsledky situace, kdy

- $h_C(x) = 80 + \left(\frac{x-80}{10}\right)^3$,
- $h_E(x, z) = \begin{cases} x, & z = 0, \\ 80 + \left(\frac{x-80}{10}\right)^3, & z = 1. \end{cases}$

Vidíme, že empirická směrodatná odchylka i průměrný odhad asymptotické odchylky jsou největší pro odhad $\hat{\delta}^I$, menší pro odhad $\hat{\delta}^{III}$ a nejmenší pro odhad $\hat{\delta}^{II}$. Za nejlepší odhad lze tedy v této situaci považovat odhad $\hat{\delta}^{II}$, který závislost konečné hodnoty na počáteční prokládá v experimentální i kontrolní skupině přímkou se směrnici 1.

Při dalších volbách funkcí a parametrů v simulaci se téměř vždy jevil nejlépe odhad $\hat{\delta}^{III}$. Ve většině případů má odhad $\hat{\delta}^{III}$ signifikantně menší směrodatnou odchylku než u ostatních odhadů. Proto i v této situaci doporučujeme vzít jako odhad efektu léčby odhad $\hat{\delta}^{III}$.

Volba parametrů		
$n = 850$	$\pi = 0.4$	$p_Z(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - 100)$
rozdělení Y_{0i} zešikmené normální s parametrem šikmosti 5	$\mu_0 = 100$	$\sigma_0 = 10$
rozdělení η_i dvojitě exponenciální		

Funkce modelu I	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = x$	$\hat{\delta}^I$	35.683	1.224	1.194	0.944
$h_E(x, z) = \begin{cases} 120, & z = 0 \\ 80 + 25\sqrt[3]{x - 80}, & z = 1 \\ 5 + (x - 100)^2, & y = 0 \end{cases}$	$\hat{\delta}^{II}$	35.715	0.794	0.771	0.940
$\psi(x, y, z) = \begin{cases} 5 + x - 100 , & y = 1, z = 0 \\ 5 + x - 100 , & y = 1, z = 1 \end{cases}$	$\hat{\delta}^{III}$	35.719	0.756	sendvičově z modelu	0.942 0.942

Tabulka 4.7: Výsledky simulace pro data pocházející z modelu I.

Volba parametrů		
$n = 850$	$\pi = 0.3$	$p_Z(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x - 100)$
rozdělení Y_{0i} normální	$\mu_0 = 100$	$\sigma_0 = 4$
rozdělení η_i dvojitě exponenciální		

Funkce modelu I	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 80 + \left(\frac{x-80}{10}\right)^3$	$\hat{\delta}^I$	25.148	0.529	0.526	0.947
$h_E(x, z) = \begin{cases} 120, & z = 0 \\ 80 + 10\sqrt[3]{x-80}, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	25.142	0.639	0.637	0.946
$\psi(x, y, z) = \begin{cases} 5 + (x-100)^2, & y = 0 \\ 5 + x-100 , & y = 1, z = 0 \\ 5 + (x-100)^2, & y = 1, z = 1 \end{cases}$	$\hat{\delta}^{III}$	25.146	0.583	sendvičově z modelu	0.945 0.896

Tabulka 4.8: Výsledky simulace pro data pocházející z modelu I.

Volba parametrů		
$n = 850$	$\pi = 0.3$	$p_Z(x) = 0.4$
rozdělení Y_{0i} dvojitě exponenciální	$\mu_0 = 100$	$\sigma_0 = 10$
rozdělení η_i zešikmené normální s parametrem šikmosti 5		

Funkce modelu I	Typ odhadu efektu léčby	Hodnota odhadu	Empirická směrodatná odchylka	Průměrný odhad asymptotické odchylky	Pravděpodobnost pokrytí δ
$h_C(x) = 80 + 25\sqrt[3]{x - 80}$	$\hat{\delta}^I$	-51.562	1.413	1.390	0.943
$h_E(x, z) = \begin{cases} x, & z = 0 \\ 80 + \left(\frac{x-80}{25}\right)^3, & z = 1 \end{cases}$	$\hat{\delta}^{II}$	-51.553	1.209	1.175	0.941
$\psi(x, y, z) = \begin{cases} 5 + x - 100 , & y = 0 \\ 5 + x - 100 , & y = 1, z = 0 \\ 5 + x - 100 , & y = 1, z = 1 \end{cases}$	$\hat{\delta}^{III}$	-51.572	1.263	sendvičově z modelu	0.940 0.939

Tabulka 4.9: Výsledky simulace pro data pocházející z modelu I.

5. Analýza efektu léčby při měření s chybou

V této kapitole se zabýváme situací, kdy veličina Y je naměřena s chybou. Veličinu Y tedy neznáme, místo ní pozorujeme hodnotu W . Pro ilustraci si představme, že pacientovi chceme změřit krevní tlak, který představuje veličinu Y . Krevní tlak však během dne podléhá značným výkyvům podle fyzické či psychické zátěže. Pacientovi je tak naměřena hodnota W , která může být vůči skutečné hodnotě odlišná právě kvůli fyzické nebo psychické zátěži.

Počáteční a konečné hodnoty naměřené veličiny W i -tého pacienta v experimentální či kontrolní skupině budeme dále značit stejným způsobem jako v kapitole 1 veličinu Y . Předpokládejme, že naměřené počáteční hodnoty W splňují

$$\begin{aligned} W_{0i}^C &= Y_{0i}^C + \xi_{0i}^C, & i = 1, \dots, n_C, \\ W_{0i}^E &= Y_{0i}^E + \xi_{0i}^E, & i = 1, \dots, n_E, \end{aligned}$$

kde $\xi_{01}^C, \dots, \xi_{0n_C}^C$ a $\xi_{01}^E, \dots, \xi_{0n_E}^E$ jsou nezávislé náhodné veličiny, které nezávisejí na Y_{0i}^C a Y_{0i}^E a pro které platí

$$E \xi_{0i}^C = E \xi_{0i}^E = 0$$

a

$$\text{var } \xi_{0i}^C = \text{var } \xi_{0i}^E = \tau_0^2.$$

Dále předpokládejme, že pro konečné hodnoty W platí

$$\begin{aligned} W_{1i}^C &= Y_{1i}^C + \xi_{1i}^C, & i = 1, \dots, n_C, \\ W_{1i}^E &= Y_{1i}^E + \xi_{1i}^E, & i = 1, \dots, n_E, \end{aligned}$$

kde $\xi_{11}^C, \dots, \xi_{1n_C}^C$ a $\xi_{11}^E, \dots, \xi_{1n_E}^E$ jsou nezávislé náhodné veličiny, které nezávisejí na počátečních či konečných hodnotách Y a pro které platí

$$E \xi_{1i}^C = 0, \quad \text{var } \xi_{1i}^C = \tau_{1,C}^2,$$

$$E \xi_{1i}^E = 0, \quad \text{var } \xi_{1i}^E = \tau_{1,E}^2,$$

a náhodné chyby ξ_{0i}^C a ξ_{1i}^C , resp. ξ_{0i}^E a ξ_{1i}^E nemusí být nezávislé.

K dispozici pro naši analýzu máme naměřené hodnoty W , nikoliv skutečné hodnoty Y . Naším cílem je odhadnout efekt léčby z naměřených hodnot

$$\begin{aligned} (W_{0i}^E, W_{1i}^E), & \quad i = 1, \dots, n_E, \\ (W_{0i}^C, W_{1i}^C), & \quad i = 1, \dots, n_C. \end{aligned} \tag{5.1}$$

Předpokládáme zde, že chyby měření veličiny Y nezávisí na hodnotě Y , což je velmi omezující. V praxi se často stává, že větší hodnotu měříme s větší chybou.

Závislost konečné hodnoty na počáteční budeme modelovat pomocí modelů I a II, které jsme představili v kapitole 1. Model III nebudeme rozebírat, protože překračuje rozsah této práce. Modely I a II jsou totožné s modely v práci [1]. V tomto případě však neznáme skutečné hodnoty Y a odhady efektu léčby sestrojíme z hodnot W . Dále sestrojíme test hypotézy nulovosti efektu léčby a získáme asymptotické rozdělení odhadu efektu léčby.

5.1 Model I a odhad efektu léčby $\hat{\delta}^I$ za jeho platnosti

Model I se nezabývá počátečními hodnotami Y_{0i} , vyjadřuje pouze konečné hodnoty Y_{1i}^C ve formě obecného zápisu náhodné veličiny a konečné hodnotě Y_{1i}^E . Model I je tvaru

$$\begin{aligned} Y_{1i}^C &= \mu + \varepsilon_i^{I,C}, \\ Y_{1i}^E &= \mu + \delta + \varepsilon_i^{I,E} \end{aligned}$$

za podmínek

$$\begin{aligned} \mathbb{E} \varepsilon_i^{I,C} &= 0, \quad \text{var} \varepsilon_i^{I,C} = \sigma_{I,C}^2, \quad i = 1, \dots, n_C, \\ \mathbb{E} \varepsilon_i^{I,E} &= 0, \quad \text{var} \varepsilon_i^{I,E} = \sigma_{I,E}^2, \quad i = 1, \dots, n_E, \end{aligned}$$

kde $\varepsilon_1^{I,C}, \dots, \varepsilon_{n_C}^{I,C}, \varepsilon_1^{I,E}, \dots, \varepsilon_{n_E}^{I,E}$ jsou nezávislé náhodné veličiny. Náhodné chyby $\varepsilon_i^{I,C}$ a $\varepsilon_i^{I,E}$ mohou záviset na počáteční hodnotě.

Dále budeme předpokládat, že platí model I. Na základě získaných dat (5.1) odhadneme efekt léčby δ a zjistíme, zda je tento odhad konzistentní. Dále sestrojíme testovou statistiku pro test nulovosti efektu léčby a podíváme se na asymptotické rozdělení odhadu efektu léčby.

Odhad v práci [1] na základě modelu I zde nemůžeme použít. V práci [1] se totiž předpokládá, že hodnoty měřené veličiny Y známe. My však máme k dispozici pouze naměřené hodnoty W z (5.1). Pro odhad efektu léčby δ proto použijeme odhad

$$\hat{\delta}^I = \bar{W}_1^E - \bar{W}_1^C,$$

kde

$$\bar{W}_1^E = \frac{1}{n_E} \sum_{i=1}^{n_E} W_{1i}^E \quad \text{a} \quad \bar{W}_1^C = \frac{1}{n_C} \sum_{i=1}^{n_C} W_{1i}^C.$$

Následující věta nám říká, že tento odhad je konzistentní.

Věta 5.1. *Za platnosti modelu I platí*

$$\hat{\delta}^I \xrightarrow{P} \delta, \quad n \rightarrow \infty.$$

Důkaz. Spočítejme střední hodnotu tohoto odhadu

$$\mathbb{E} \hat{\delta}^I = \frac{1}{n_E} \sum_{i=1}^{n_E} \mathbb{E} W_{1i}^E - \frac{1}{n_C} \sum_{i=1}^{n_C} \mathbb{E} W_{1i}^C = \mathbb{E} Y_{1i}^E - \mathbb{E} Y_{1i}^C = \mu + \delta - \mu = \delta$$

a jeho rozptyl

$$\begin{aligned} \text{var } \hat{\delta}^I &= \text{var } \bar{W}_1^C + \text{var } \bar{W}_1^E \\ &= \frac{1}{n_C^2} \sum_{i=1}^{n_C} \text{var} (Y_{1i}^C + \xi_i^C) + \frac{1}{n_E^2} \sum_{i=1}^{n_E} \text{var} (Y_{1i}^E + \xi_i^E) \\ &= \frac{1}{n_C} (\sigma_{I,C}^2 + \tau_{1,C}^2) + \frac{1}{n_E} (\sigma_{I,E}^2 + \tau_{1,E}^2). \end{aligned}$$

Dle slabého zákona velkých čísel a z předchozích rovností plyne, že $\hat{\delta}^I$ je konzistentní odhad δ . \square

Testujme hypotézu $H_0 : \delta = 0$ proti alternativě $H_1 : \delta \neq 0$. K tomu využijeme následující větu.

Věta 5.2. *Nechť platí model I. Označme*

$$t_I = \frac{\bar{W}_1^E - \bar{W}_1^C}{\sqrt{\frac{S_{I,C}^2}{n_C} + \frac{S_{I,E}^2}{n_E}}},$$

kde

$$\begin{aligned} S_{I,C}^2 &= \frac{1}{n_C - 1} \sum_{i=1}^{n_C} (W_{1i}^C - \bar{W}_1^C)^2, \\ S_{I,E}^2 &= \frac{1}{n_E - 1} \sum_{i=1}^{n_E} (W_{1i}^E - \bar{W}_1^E)^2. \end{aligned}$$

Je-li $\delta = 0$, pak platí

$$t_I \xrightarrow{d} \mathbf{N}(0, 1).$$

Důkaz. Dokáže se stejným způsobem jako důkaz věty 2.2. \square

Z předchozí věty jsme získali testovou statistiku t_I , která má za platnosti H_0 asymptoticky normované normální rozdělení. Nulovou hypotézu zamítáme na hladině α , jestliže $|t_I| \geq u_{1-\frac{\alpha}{2}}$.

V další větě získáváme asymptotické rozdělení odhadu $\hat{\delta}^I$.

Věta 5.3. *Za platnosti modelu I je*

$$\sqrt{n} (\hat{\delta}^I - \delta) \xrightarrow{d} \mathbf{N}(0, V_{\hat{\delta}^I}),$$

kde

$$V_{\hat{\delta}^I} = (q + 1) \left(\sigma_{I,C}^2 + \tau_{1,C}^2 + \frac{1}{q} (\sigma_{I,E}^2 + \tau_{1,E}^2) \right). \quad (5.2)$$

Důkaz. Důkaz je stejný jako ve větě 2.3. □

Porovnejme asymptotický rozptyl odhadu $\hat{\delta}^I$ z předchozí věty s asymptotickým rozptylem z práce [1], ve které se nepředpokládá, že veličiny Y jsou měřeny s chybou. Rozptyl odhadu $\hat{\delta}^I$ z (5.2) je o

$$(q + 1) \left(\tau_{1,C}^2 + \frac{1}{q} \tau_{1,E}^2 \right)$$

větší než rozptyl odhadu $\hat{\delta}^I$ z práce [1].

V této podkapitole jsme dokázali, že odhad $\hat{\delta}^I$ je konzistentní odhad efektu léčby za platnosti modelu I a sestrojili jsme testovou statistiku hypotézy nulovosti efektu léčby. Tyto výsledky se shodují s prací [1]. Dále jsme zjistili asymptotické rozdělení odhadu $\hat{\delta}^I$. Asymptotický rozptyl tohoto odhadu je vždy větší než asymptotický rozptyl z práce [1].

5.2 Model II a odhad efektu léčby $\hat{\delta}^{II}$ za jeho platnosti

Model II předpokládá, že rozdíl konečné a počáteční hodnoty nezávisí na počáteční hodnotě a popisuje střední hodnotu rozdílu konečné a počáteční hodnoty jako konstantu. Model II je tvaru

$$\begin{aligned} Y_{1i}^C - Y_{0i}^C &= \gamma + \varepsilon_i^{II,C}, \\ Y_{1i}^E - Y_{0i}^E &= \gamma + \delta + \varepsilon_i^{II,E} \end{aligned}$$

za podmínek

$$\begin{aligned} \mathbf{E} \varepsilon_i^{II,C} &= 0, & \mathbf{E} \left(\varepsilon_i^{II,C} \mid Y_{0i}^C \right) &= 0, & i &= 1, \dots, n_C, \\ \mathbf{E} \varepsilon_i^{II,E} &= 0, & \mathbf{E} \left(\varepsilon_i^{II,E} \mid Y_{0i}^E \right) &= 0, & i &= 1, \dots, n_E, \\ \text{var} \varepsilon_i^{II,C} &= \sigma_{II,C}^2, & & & i &= 1, \dots, n_C, \\ \text{var} \varepsilon_i^{II,E} &= \sigma_{II,E}^2, & & & i &= 1, \dots, n_E, \end{aligned}$$

kde $\varepsilon_1^{II,C}, \dots, \varepsilon_{n_C}^{II,C}, \varepsilon_1^{II,E}, \dots, \varepsilon_{n_E}^{II,E}$ jsou nezávislé náhodné veličiny, které nezávisí na počáteční hodnotě.

Naším cílem je na základě získaných dat (5.1) odhadnout efekt léčby δ a zjistit, zda je tento odhad konzistentní. Dále sestrojíme testovou statistiku pro test nulovosti efektu léčby a podíváme se na asymptotické rozdělení odhadu efektu léčby.

Odhad v práci [1] na základě modelu II zde nemůžeme použít. V práci [1] se totiž předpokládá, že hodnoty měřené veličiny Y známe. My však máme k dispozici

pouze naměřené hodnoty W z (5.1). Pro odhad efektu léčby δ proto použijeme odhad

$$\hat{\delta}^{II} = (\bar{W}_1^E - \bar{W}_0^E) - (\bar{W}_1^C - \bar{W}_0^C),$$

kde

$$\begin{aligned}\bar{W}_0^E &= \frac{1}{n_E} \sum_{i=1}^{n_E} W_{0i}^E, & \bar{W}_0^C &= \frac{1}{n_C} \sum_{i=1}^{n_C} W_{0i}^C, \\ \bar{W}_1^E &= \frac{1}{n_E} \sum_{i=1}^{n_E} W_{1i}^E & \text{a} & \bar{W}_1^C = \frac{1}{n_C} \sum_{i=1}^{n_C} W_{1i}^C.\end{aligned}$$

Následující věta nám říká, že tento odhad je konzistentní.

Věta 5.4. *Nechť platí model II. Pak*

$$\hat{\delta}^{II} \xrightarrow{P} \delta, \quad n \rightarrow \infty.$$

Důkaz. Spočítejme střední hodnotu odhadu

$$\begin{aligned}\mathbb{E} \hat{\delta}^{II} &= \frac{1}{n_E} \sum_{i=1}^{n_E} \mathbb{E} (W_{1i}^E - W_{0i}^E) - \frac{1}{n_C} \sum_{i=1}^{n_C} \mathbb{E} (W_{1i}^C - W_{0i}^C) \\ &= \mathbb{E} (Y_{1i}^E - Y_{0i}^E) - \mathbb{E} (Y_{1i}^C - Y_{0i}^E) = \gamma + \delta - \gamma = \delta.\end{aligned}$$

Spočítejme rozptyl odhadu

$$\begin{aligned}\text{var} \hat{\delta}^{II} &= \text{var} (\bar{W}_1^C - \bar{W}_0^C) + \text{var} (\bar{W}_1^E - \bar{W}_0^E) \\ &= \frac{1}{n_C^2} \sum_{i=1}^{n_C} \text{var} (Y_{1i}^C - Y_{0i}^C + \xi_{1i}^C - \xi_{0i}^C) + \frac{1}{n_E^2} \sum_{i=1}^{n_E} \text{var} (Y_{1i}^E - Y_{0i}^E + \xi_{1i}^E - \xi_{0i}^E) \\ &= \frac{1}{n_C} \left(\sigma_{II,C}^2 + \tau_{1,C}^2 + \tau_0^2 - 2 \text{cov} (\xi_{0i}^C, \xi_{1i}^C) \right) + \\ &\quad + \frac{1}{n_E} \left(\sigma_{II,E}^2 + \tau_{1,E}^2 + \tau_0^2 - 2 \text{cov} (\xi_{0i}^E, \xi_{1i}^E) \right).\end{aligned}$$

Z předchozích rovností a ze zákona velkých čísel plyne tvrzení věty. \square

Testujme hypotézu $H_0 : \delta = 0$ proti alternativě $H_1 : \delta \neq 0$. K tomu využijeme následující větu.

Věta 5.5. *Nechť platí model II. Označme*

$$t_{II} = \frac{(\bar{W}_1^E - \bar{W}_0^E) - (\bar{W}_1^C - \bar{W}_0^C)}{\sqrt{\frac{S_{II,C}^2}{n_C} + \frac{S_{II,E}^2}{n_E}}},$$

kde

$$\begin{aligned}S_{II,C}^2 &= \frac{1}{n_C - 1} \sum_{i=1}^{n_C} \left(W_{1i}^C - W_{0i}^C - (\bar{W}_1^C - \bar{W}_0^C) \right)^2, \\ S_{II,E}^2 &= \frac{1}{n_E - 1} \sum_{i=1}^{n_E} \left(W_{1i}^E - W_{0i}^E - (\bar{W}_1^E - \bar{W}_0^E) \right)^2.\end{aligned}$$

Jestliže $\delta = 0$, pak platí

$$t_{II} \xrightarrow{d} \mathbf{N}(0, 1).$$

Důkaz. Důkaz se provede stejným způsobem jako ve větě 2.2. \square

Z předchozí věty jsme získali testovou statistiku t_{II} , která má za platnosti H_0 asymptoticky normované normální rozdělení. Nulovou hypotézu zamítáme na hladině α , jestliže $|t_{II}| \geq u_{1-\frac{\alpha}{2}}$.

V další větě získáváme asymptotické rozdělení odhadu $\hat{\delta}^{II}$.

Věta 5.6. *Nechť platí model II. Pak pro odhad $\hat{\delta}^{II}$ platí*

$$\sqrt{n} \left(\hat{\delta}^{II} - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, V_{\hat{\delta}^{II}} \right),$$

kde

$$\begin{aligned} V_{\hat{\delta}^{II}} = & (q+1) \left\{ \sigma_{II,C}^2 + \tau_{1,C}^2 - 2 \operatorname{cov} \left(\xi_{0i}^C, \xi_{1i}^C \right) + \right. \\ & \left. + \frac{1}{q} \left(\sigma_{II,E}^2 + \tau_{1,E}^2 - 2 \operatorname{cov} \left(\xi_{0i}^E, \xi_{1i}^E \right) \right) \right\} + \frac{(q+1)^2}{q} \tau_0^2. \end{aligned} \quad (5.3)$$

Důkaz. Jelikož

$$\operatorname{var} \left(\bar{W}_1^C - \bar{W}_0^C \right) = \frac{1}{n_C} \left(\sigma_{II,C}^2 + \tau_{1,C}^2 + \tau_0^2 - 2 \operatorname{cov} \left(\xi_{0i}^C, \xi_{1i}^C \right) \right)$$

a

$$\operatorname{var} \left(\bar{W}_1^E - \bar{W}_0^E \right) = \frac{1}{n_E} \left(\sigma_{II,E}^2 + \tau_{1,E}^2 + \tau_0^2 - 2 \operatorname{cov} \left(\xi_{0i}^E, \xi_{1i}^E \right) \right),$$

získáváme podobně jako ve větě 2.3 rozptyl

$$\begin{aligned} V_{\hat{\delta}^{II}} = & (q+1) \left\{ \sigma_{II,C}^2 + \tau_{1,C}^2 + \tau_0^2 - 2 \operatorname{cov} \left(\xi_{0i}^C, \xi_{1i}^C \right) + \right. \\ & \left. + \frac{1}{q} \left(\sigma_{II,E}^2 + \tau_{1,E}^2 + \tau_0^2 - 2 \operatorname{cov} \left(\xi_{0i}^E, \xi_{1i}^E \right) \right) \right\}, \end{aligned}$$

který se rovná rozptylu ve znění věty. \square

Porovnejme asymptotický rozptyl odhadu $\hat{\delta}^{II}$ z předchozí věty s asymptotickým rozptylem z práce [1], ve které se nepředpokládá, že veličiny Y jsou měřeny s chybou. Rozptyl odhadu $\hat{\delta}^{II}$ z (5.3) je o

$$(q+1) \left(\tau_{1,C}^2 + \tau_0^2 - 2 \operatorname{cov} \left(\xi_{0i}^C, \xi_{1i}^C \right) + \frac{1}{q} \left(\tau_{1,C}^2 + \tau_0^2 - 2 \operatorname{cov} \left(\xi_{0i}^C, \xi_{1i}^C \right) \right) \right)$$

větší než rozptyl tohoto odhadu z práce [1].

V této podkapitole jsme dokázali, že odhad $\hat{\delta}^{II}$ je konzistentní odhad efektu léčby za platnosti modelu II a sestrojili jsme testovou statistiku hypotézy nulovosti efektu léčby. Tyto výsledky se shodují s prací [1]. Dále jsme zjistili asymptotické rozdělení odhadu $\hat{\delta}^{II}$. Asymptotický rozptyl tohoto odhadu je vždy větší než asymptotický rozptyl z práce [1].

6. Chování odhadů při měření s chybou

Tato kapitola navazuje na předchozí kapitolu, ve které jsme uvedli modely I a II a chování odhadů efektu léčby $\hat{\delta}^I$ a $\hat{\delta}^{II}$ za platnosti příslušných modelů. Nevíme však, ze kterého modelu data vznikla. V následujících podkapitolách postupně prozkoumáme, jaké vlastnosti má daný odhad za platnosti modelu, na jehož základě tento odhad nevznikl. Podíváme se, zda je daný odhad za platnosti tohoto modelu konzistentní, a porovnáme asymptotické rozptyly tohoto odhadu za platnosti uvažovaného modelu a za platnosti modelu, na jehož základě odhad vznikl. Dále porovnáme asymptotické rozptyly odhadů z této práce a z práce [1], ve které se nepředpokládá měření hodnot s chybou.

6.1 Chování odhadu $\hat{\delta}^I$ za platnosti modelu II

Platí-li model II, pak platí i model I. Odhad $\hat{\delta}^I$ je tedy za platnosti modelu II nestranný a konzistentní. Zajímá nás asymptotické rozdělení tohoto odhadu a jeho porovnání s odhadem $\hat{\delta}^{II}$ za platnosti modelu II. O tom pojednává následující věta.

Věta 6.1. *Za platnosti modelu II pro odhad $\hat{\delta}^I$ platí*

$$\sqrt{n} \left(\hat{\delta}^I - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, W_{\hat{\delta}^I}^{II} \right),$$

kde

$$W_{\hat{\delta}^I}^{II} = V_{\hat{\delta}^{II}} + \frac{(q+1)^2}{q} (\sigma_0^2 - \tau_0^2) + 2(q+1) \left(\text{cov}(\xi_{0i}^C, \xi_{1i}^C) + \frac{1}{q} \text{cov}(\xi_{0i}^E, \xi_{1i}^E) \right).$$

Důkaz. Asymptotický rozptyl $W_{\hat{\delta}^I}^{II}$ je roven asymptotickému rozptylu $V_{\hat{\delta}^I}$ z (5.2). Tento rozptyl vyjádříme pomocí parametrů z modelu II

$$W_{\hat{\delta}^I}^{II} = (q+1) \left(\text{var} W_{1i}^C + \frac{1}{q} \text{var} W_{1i}^E \right),$$

kde

$$\text{var} W_{1i}^C = \text{var} \left(Y_{0i}^C + \xi_{0i}^C + \varepsilon_i^{II,C} \right) = \sigma_0^2 + \sigma_{II,C}^2 + \tau_0^2,$$

$$\text{var} W_{1i}^E = \text{var} \left(Y_{0i}^E + \xi_{0i}^E + \varepsilon_i^{II,E} \right) = \sigma_0^2 + \sigma_{II,E}^2 + \tau_0^2.$$

Celkem získáváme

$$\begin{aligned} W_{\hat{\delta}^I}^{II} &= (q+1) \left(\sigma_0^2 + \sigma_{II,C}^2 + \tau_0^2 + \frac{1}{q} (\sigma_0^2 + \sigma_{II,E}^2 + \tau_0^2) \right) \\ &= V_{\hat{\delta}^{II}} + \frac{(q+1)^2}{q} (\sigma_0^2 - \tau_0^2) + 2(q+1) \left(\text{cov}(\xi_{0i}^C, \xi_{1i}^C) + \frac{1}{q} \text{cov}(\xi_{0i}^E, \xi_{1i}^E) \right). \end{aligned}$$

□

Odhad $\hat{\delta}^I$ má za platnosti modelu I větší rozptyl než odhad $\hat{\delta}^{II}$, jestliže

$$\tau_0^2 < 2 \min \{ \text{cov}(\xi_{0i}^C, \xi_{1i}^C), \text{cov}(\xi_{0i}^E, \xi_{1i}^E) \} + \sigma_0^2.$$

Pokud jsou tedy náhodné chyby na začátku a na konci měření v obou skupinách kladně korelovány a rozptyl počátečních hodnot Y je větší než rozptyl počátečních náhodných chyb, tak má odhad $\hat{\delta}^I$ větší asymptotický rozptyl než $\hat{\delta}^{II}$. Tento výsledek se liší od výsledku z práce [1], kde se neuvažuje měření hodnot s chybou. V práci [1] je asymptotický rozptyl odhadu $\hat{\delta}^I$ vždy větší než asymptotický rozptyl odhadu $\hat{\delta}^{II}$.

6.2 Chování odhadu $\hat{\delta}^{II}$ za platnosti modelu I

Platí-li model I, pak obecně neplatí model II. Naším cílem je získat asymptotické rozdělení odhadu $\hat{\delta}^{II}$ a jeho porovnání s odhadem $\hat{\delta}^I$ za platnosti modelu I. V následující větě zjistíme asymptotické rozdělení odhadu $\hat{\delta}^{II}$ za platnosti modelu I.

Věta 6.2. *Asymptotické rozdělení odhadu $\hat{\delta}^{II}$ za platnosti modelu I je*

$$\sqrt{n} \left(\hat{\delta}^{II} - \delta \right) \xrightarrow{d} \mathbf{N} \left(0, W_{\hat{\delta}^{II}}^I \right),$$

kde

$$W_{\hat{\delta}^{II}}^I = V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} (\sigma_0^2 + \tau_0^2) - 2(q+1) \left\{ \text{cov}(Y_{0i}^C, Y_{1i}^C) + \text{cov}(\xi_{0i}^C, \xi_{1i}^C) + \frac{1}{q} \left(\text{cov}(Y_{0i}^E, Y_{1i}^E) + \text{cov}(\xi_{0i}^E, \xi_{1i}^E) \right) \right\}.$$

Důkaz. Asymptotický rozptyl $W_{\hat{\delta}^{II}}^I$ se rovná

$$W_{\hat{\delta}^{II}}^I = (q+1) \left(\text{var}(W_{1i}^C - W_{0i}^C) + \frac{1}{q} \text{var}(W_{1i}^E - W_{0i}^E) \right),$$

kde rozptyl první rozptyl spočítáme jako

$$\begin{aligned} \text{var}(W_{1i}^C - W_{0i}^C) &= \text{var}(Y_{1i}^C + \xi_{1i}^C - Y_{0i}^C - \xi_{0i}^C) \\ &= \text{var}(Y_{1i}^C - Y_{0i}^C) + \text{var}(\xi_{1i}^C - \xi_{0i}^C) \\ &= \sigma_{I,C}^2 + \sigma_0^2 - 2 \text{cov}(Y_{0i}^C, Y_{1i}^C) + \tau_{1,C}^2 + \tau_{0,C}^2 - 2 \text{cov}(\xi_{0i}^C, \xi_{1i}^C) \end{aligned}$$

a stejným způsobem spočítáme druhý rozptyl

$$\text{var}(W_{1i}^E - W_{0i}^E) = \sigma_{I,E}^2 + \sigma_0^2 - 2 \text{cov}(Y_{0i}^E, Y_{1i}^E) + \tau_{1,E}^2 + \tau_{0,E}^2 - 2 \text{cov}(\xi_{0i}^E, \xi_{1i}^E).$$

Celkem získáváme

$$\begin{aligned}
W_{\hat{\delta}^{II}}^I &= (q+1) \left\{ \sigma_{I,C}^2 + \sigma_0^2 - 2 \operatorname{cov}(Y_{0i}^C, Y_{1i}^C) + \tau_{1,C}^2 + \tau_{0,C}^2 - 2 \operatorname{cov}(\xi_{0i}^C, \xi_{1i}^C) + \right. \\
&\quad \left. + \frac{1}{q} \left(\sigma_{I,E}^2 + \sigma_0^2 - 2 \operatorname{cov}(Y_{0i}^E, Y_{1i}^E) + \tau_{1,E}^2 + \tau_{0,E}^2 - 2 \operatorname{cov}(\xi_{0i}^E, \xi_{1i}^E) \right) \right\} \\
&= V_{\hat{\delta}^I} + \frac{(q+1)^2}{q} (\sigma_0^2 + \tau_0^2) - 2(q+1) \left\{ \operatorname{cov}(Y_{0i}^C, Y_{1i}^C) + \operatorname{cov}(\xi_{0i}^C, \xi_{1i}^C) + \right. \\
&\quad \left. + \frac{1}{q} \left(\operatorname{cov}(Y_{0i}^E, Y_{1i}^E) + \operatorname{cov}(\xi_{0i}^E, \xi_{1i}^E) \right) \right\}.
\end{aligned}$$

□

Odhad $\hat{\delta}^{II}$ má za platnosti modelu I větší rozptyl než odhad $\hat{\delta}^I$, jestliže

$$\tau_0^2 > 2 \max \left\{ \operatorname{cov}(Y_{0i}^C, Y_{1i}^C) + \operatorname{cov}(\xi_{0i}^C, \xi_{1i}^C), \operatorname{cov}(Y_{0i}^E, Y_{1i}^E) + \operatorname{cov}(\xi_{0i}^E, \xi_{1i}^E) \right\} - \sigma_0^2.$$

Je-li tedy součet rozptylu počátečních hodnot Y a rozptylu počátečních náhodných chyb dostatečně velký, tak odhad $\hat{\delta}^{II}$ má větší rozptyl než odhad $\hat{\delta}^I$.

Výsledný asymptotický rozptyl $W_{\hat{\delta}^{II}}^I$ se liší od asymptotického rozptylu z práce [1], kde se neuvažuje měření hodnot s chybou. Kdybychom předpokládali vzájemnou nezávislost chyb ξ_{0i}^C a ξ_{1i}^C , resp. ξ_{0i}^E a ξ_{1i}^E , pak asymptotický rozptyl $W_{\hat{\delta}^{II}}^I$ by byl o

$$\frac{(q+1)^2}{q} \tau_0^2$$

větší než asymptotický rozptyl v práci [1].

Závěr

V této práci jsme se zabývali klinickými randomizovanými studii. Pacienti byli náhodně rozděleni do dvou skupin, pacientům v experimentální skupině byl podán lék a pacientům v kontrolní skupině bylo podáno placebo. Pacientům se na začátku a konci studie změřila hodnota veličiny, kterou měl lék ovlivnit. Z těchto hodnot jsme odhadovali efekt léčby, který jsme definovali jako střední hodnotu rozdílu měřené veličiny v experimentální a kontrolní skupině. Tento odhad efektu léčby jsme zkoumali ve dvou situacích. V těchto situacích jsme efekt léčby odhadovali na základě tří modelů. První model (model I) se nezabývá počátečními hodnotami, vyjadřuje konečnou hodnotu jako součet konstanty a náhodné chyby. Ve druhém modelu (model II) předpokládáme, že rozdíl mezi konečnou a počáteční hodnotou na počáteční hodnotě nezávisí a vyjadřuje tento rozdíl jako součet konstanty a náhodné chyby. Poslední model (model III) je regresní model, který předpokládá obecnou lineární závislost konečné hodnoty na počáteční.

V první situaci jsme předpokládali, že lék nemusí mít stejný efekt na všechny pacienty. Tato situace byla hlavním objektem našeho zájmu. V kapitole 2 jsme sestrojili odhady efektu léčby na základě všech zmíněných modelů. Dále jsme v této kapitole sestrojili test hypotézy nulovosti efektu léčby a zjistili asymptotické rozdělení těchto odhadů za platnosti modelů, na jejichž základě odhady vznikly. V kapitole 3 jsme se pak zabývali asymptotickým rozdělením odhadů za platnosti modelů, na jejichž základě odhady nevznikly. Jelikož ve všech devíti případech (tři způsoby odhadu za platnosti tří různých modelů) jsou odhady nestranné a konzistentní, porovnávali jsme vhodnost použití odhadů podle jejich asymptotických rozptylů. Za platnosti modelu II vyšel vždy nejlépe odhad $\hat{\delta}^{III}$. Za platnosti modelu III byl nejlepším odhadem odhad $\hat{\delta}^{III}$ za předpokladu, že pravděpodobnost přiřazení pacienta do podskupin v experimentální skupině nezávisí na počáteční hodnotě pacienta. Za platnosti modelu I jsme nebyli teoreticky schopni asymptotické rozptyly odhadů porovnat díky jejich složitosti.

V kapitole 4 jsme ověřovali spočítané teoretické výsledky pomocí simulací. Potvrdili jsme, že nejlepším odhadem za platnosti modelu II je odhad $\hat{\delta}^{III}$. Za platnosti modelu I a III jsme uvedli situace, ve kterých měl odhad $\hat{\delta}^{III}$ větší i menší směrodatné odchyly než ostatní odhady. Situace, ve kterých měl odhad $\hat{\delta}^{III}$ nejmenší směrodatné odchyly ze všech odhadů, však převažovaly. Proto naším výsledným doporučením je použít odhad $\hat{\delta}^{III}$. Toto doporučení se shoduje s doporučením v práci [1], ve které se předpokládá konstantní efekt léčby.

V druhé situaci jsme předpokládali, že měření zkoumané veličiny je prováděno s chybou. Dále jsme předpokládali, že tyto chyby nezávisí na počáteční hodnotě. To je velmi omezující předpoklad, jelikož se v praxi často setkáváme s případem, kdy větší hodnota měřené veličiny je měřena s větší chybou. Odhady efektu léčby v této situaci jsme prozkoumali pouze pomocí modelů I a II. Modelem III jsme se nezabývali, jelikož významně překračuje rozsah práce. V kapitole 5 jsme uvedli odhady efektu léčby a zjistili asymptotické rozdělení těchto odhadů

za platnosti modelů, na jejichž základě odhady vznikly. Dále jsme sestrojili test hypotézy nulovosti efektu léčby. V kapitole 6 jsme získali asymptotické rozdělení odhadů za platnosti modelů, na jejichž základě odhady nevznikly. Odhady jsme znovu porovnávali pomocí jejich asymptotických rozptylů, jelikož ve všech čtyřech případech jsou odhady nestranné a konzistentní. Asymptotické rozptyly odhadů jsme porovnávali také v kapitole 6, kde jsme podrobně napsali podmínky pro vybrání jednoho z odhadů. Pokud mají počáteční chyby dostatečně velký rozptyl, je vhodné jako odhad efektu léčby použít odhad z modelu II.

Celkově jsme v této práci uceleně shrnuli analýzu efektu léčby za méně omezujících předpokladů než v práci [1]. Jedná se o situaci, kdy efekt léčby nemusí být u každého pacienta stejný, a o situaci, kdy je veličina měřena s chybou. Oba případy jsme teoreticky prozkoumali a první situaci simulačně podpořili. Věty v této práci, které jsou označeny jako „Věty“, jsme sami zformulovali a dokázali. Věty, které jsou označeny jako „Tvrzení“, jsou převzaty z jiných publikací. Další rozšíření práce vidíme v prozkoumání modelu III v druhé situaci a dále v uvažování situace, kdy chyby měření mohou záviset na měřené hodnotě.

Seznam použité literatury

- [1] PACÁKOVÁ, Andrea. *Analýza změny od počáteční hodnoty ke konečné*. Praha, 2010. Diplomová práce na Matematicko-Fyzikální fakultě Univerzity Karlovy na katedře Pravděpodobnosti a matematické statistiky.
- [2] ANDĚL, Jiří. *Základy matematické statistiky*. 2. vydání. Praha, 2007. ISBN 80-7378-001-1.
- [3] WHITE, Halbert. *Maximum likelihood estimation of misspecified models*. *Econometrica* **50**: 1-26.
- [4] SENN, Stephen. *Change from baseline and analysis of covariance revisited*. *Statistics in medicine* **25**: 4334-4344.
- [5] TU, Yu-Kang - GILTHORPE, Mark S. *Revisiting the relation between change and initial value: A review and evaluation*. *Statistics in medicine* **26**: 443-457.

Seznam použitých symbolů

n	počet pacientů
n_C	počet pacientů v kontrolní skupině
n_E	počet pacientů v experimentální skupině
Y	měřená veličina
Y_{0i}	hodnota veličiny Y měřená i -tému pacientovi na začátku studie
Y_{1i}	hodnota veličiny Y měřená i -tému pacientovi na konci studie
Y_{0i}^C, Y_{1i}^C	hodnota veličiny Y měřená i -tému pacientovi v kontrolní skupině na začátku/konci studie
Y_{0i}^E, Y_{1i}^E	hodnota veličiny Y měřená i -tému pacientovi v experimentální skupině na začátku/konci studie
μ_0	střední hodnota veličin Y_{0i}^C, Y_{0i}^E
σ_0^2	rozptyl veličin Y_{0i}^C, Y_{0i}^E
W	naměřená hodnota Y s chybou
W_{0i}	s chybou naměřená hodnota Y měřená i -tému pacientovi na začátku studie
W_{1i}	s chybou naměřená hodnota Y měřená i -tému pacientovi na konci studie
W_{0i}^C, W_{1i}^C	s chybou naměřená hodnota Y měřená i -tému pacientovi v kontrolní skupině na začátku/konci studie
W_{0i}^E, W_{1i}^E	s chybou naměřená hodnota Y měřená i -tému pacientovi v experimentální skupině na začátku/konci studie
ξ_{0i}^C, ξ_{1i}^C	chyba měření veličiny Y v kontrolní skupině na začátku/konci studie
ξ_{0i}^E, ξ_{1i}^E	chyba měření veličiny Y v experimentální skupině na začátku/konci studie
τ_0^2	rozptyl chyb měření ξ_{0i}^C a ξ_{0i}^E
$\tau_{1,C}^2$	rozptyl chyby měření ξ_{1i}^C
$\tau_{1,E}^2$	rozptyl chyby měření ξ_{1i}^E
$I_{[i \in E]}$	identifikátor přiřazení i -tého pacienta do experimentální skupiny
π	pravděpodobnost přiřazení pacienta do experimentální skupiny
q_n	podíl počtu pacientů v experimentální a kontrolní skupině
q	podíl pravděpodobností přiřazení pacienta do experimentální a kontrolní skupiny

Z_i	náhodná veličina přiřazující i -tého pacienta v experimentální skupině do jedné ze dvou podskupin
p	pravděpodobnost přiřazení pacienta v experimentální skupině do podskupiny $Z_i = 1$
$p_Z(Y_{0i})$	pravděpodobnost, že i -tý pacient bude přiřazen do skupiny $Z_i = 1$ za podmínky jeho počáteční hodnoty Y_{0i}
δ	efekt léčby
δ_0	efekt léčby v podskupině $Z_i = 0$
δ_1	efekt léčby v podskupině $Z_i = 1$
$\hat{\delta}^I$	odhad δ na základě modelu I
$\hat{\delta}^{II}$	odhad δ na základě modelu II
$\hat{\delta}^{III}$	odhad δ na základě modelu III
μ	parametr modelu I
γ	parametr modelu II
α, β	parametry modelu III
$\varepsilon_i^{I,C}$	náhodná chyba v kontrolní skupině v modelu I
$\varepsilon_i^{I,E}$	náhodná chyba v experimentální skupině v modelu I
$\varepsilon_i^{II,C}$	náhodná chyba v kontrolní skupině modelu II
$\varepsilon_i^{II,E}$	náhodná chyba v experimentální skupině modelu II
$\varepsilon_i^{III,C}$	náhodná chyba v kontrolní skupině modelu III
$\varepsilon_i^{III,E}$	náhodná chyba v experimentální skupině modelu III
$\sigma_{I,C}^2$	rozptyl chyby $\varepsilon_i^{I,C}$ v modelu I
$\sigma_{I,E,0}^2$	rozptyl chyby $\varepsilon_i^{I,E}$ v modelu I za podmínky, že i -tý pacient je v podskupině $Z_i = 0$
$\sigma_{I,E,1}^2$	rozptyl chyby $\varepsilon_i^{I,E}$ v modelu I za podmínky, že i -tý pacient je v podskupině $Z_i = 1$
$\sigma_{II,C}^2$	rozptyl chyby $\varepsilon_i^{II,C}$ v modelu II
$\sigma_{II,E,0}^2$	rozptyl chyby $\varepsilon_i^{II,E}$ v modelu II za podmínky, že i -tý pacient je v podskupině $Z_i = 0$
$\sigma_{II,E,1}^2$	rozptyl chyby $\varepsilon_i^{II,E}$ v modelu II za podmínky, že i -tý pacient je v podskupině $Z_i = 1$
$\sigma_{III,C}^2$	rozptyl chyby $\varepsilon_i^{III,C}$ v modelu III
$\sigma_{III,E,0}^2$	rozptyl chyby $\varepsilon_i^{III,E}$ v modelu III za podmínky, že i -tý pacient je v podskupině $Z_i = 0$
$\sigma_{III,E,1}^2$	rozptyl chyby $\varepsilon_i^{III,E}$ v modelu III za podmínky, že i -tý pacient je v podskupině $Z_i = 1$
$S_{I,C}^2$	odhad rozptylu naměřené Y_{1i}^C na základě modelu I
$S_{I,E}^2$	odhad rozptylu naměřené Y_{1i}^E na základě modelu I
$S_{II,C}^2$	odhad rozptylu naměřeného rozdílu $Y_{1i}^C - Y_{0i}^C$ na základě modelu II

$S_{II,E}^2$	odhad rozptylu naměřeného rozdílu $Y_{1i}^E - Y_{0i}^E$ na základě modelu II
$V_{\hat{\delta}^I}$	asymptotický rozptyl odhadu $\hat{\delta}^I$ za platnosti modelu I
$V_{\hat{\delta}^{II}}$	asymptotický rozptyl odhadu $\hat{\delta}^{II}$ za platnosti modelu II
$V_{\hat{\delta}^{III}}$	asymptotický rozptyl odhadu $\hat{\delta}^{III}$ za platnosti modelu III
$W_{\hat{\delta}^I}^{II}$	asymptotický rozptyl odhadu $\hat{\delta}^I$ za platnosti modelu II
$W_{\hat{\delta}^I}^{III}$	asymptotický rozptyl odhadu $\hat{\delta}^I$ za platnosti modelu III
$W_{\hat{\delta}^{II}}^I$	asymptotický rozptyl odhadu $\hat{\delta}^{II}$ za platnosti modelu I
$W_{\hat{\delta}^{II}}^{III}$	asymptotický rozptyl odhadu $\hat{\delta}^{II}$ za platnosti modelu III
$W_{\hat{\delta}^{III}}^I$	asymptotický rozptyl odhadu $\hat{\delta}^{III}$ za platnosti modelu I
$W_{\hat{\delta}^{III}}^{II}$	asymptotický rozptyl odhadu $\hat{\delta}^{III}$ za platnosti modelu II
$h_C(Y_{0i})$	střední hodnota konečné hodnoty Y_{1i} za podmínky počáteční hodnoty Y_{0i} v kontrolní skupině
$h_E(Y_{0i}, Z_i)$	střední hodnota konečné hodnoty Y_{1i} za podmínky počáteční hodnoty Y_{0i} v podskupině Z_i experimentální skupiny
$h_{E,0}(Y_{0i})$	střední hodnota konečné hodnoty Y_{1i} za podmínky počáteční hodnoty Y_{0i} v podskupině $Z_i = 0$ experimentální skupiny
$h_{E,1}(Y_{0i})$	střední hodnota konečné hodnoty Y_{1i} za podmínky počáteční hodnoty Y_{0i} v podskupině $Z_i = 1$ experimentální skupiny
η_i	náhodná chyba v modelu, v němž se konečná hodnota Y_{1i} vyjadřuje pomocí funkcí h_C a h_E (resp. $h_{E,0}$ a $h_{E,1}$)
$\psi(Y_{0i}, \mathbb{1}_{[i \in E]}, Z_i)$	rozptyl náhodné chyby η_i za podmínky Y_{0i} , $\mathbb{1}_{[i \in E]}$ a Z_i

Přílohy

1. CD s prací v pdf a zdrojové kódy simulační studie.