



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Nela Rychterová

Testování ekvivalence a noninferiority

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Jaromír Antoch, CSc.

Konzultant diplomové práce: Christian Genest, PhD, FRSC

Studijní program: Matematika

Studijní obor: PMSE

Praha 2019

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

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

V dne

Podpis autora

Ráda bych poděkovala panu profesoru Jaromíru Antochovi za podnětné připomínky, velkou vstřícnost a čas, který mi během vedení této práce věnoval. Také bych ráda poděkovala profesoru Christianovi Genestovi, který mi umožnil získat studijní zkušenosti v zahraničí, za jeho cenné rady, ochotu a čas, který mi věnoval. Dále děkuji své rodině a přátelům za velkou podporu během studia.

Název práce: Testování ekvivalence a noninferiority

Autor: Nela Rychterová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: prof. RNDr. Jaromír Antoch, CSc.,
Katedra pravděpodobnosti a matematické statistiky

Konzultant diplomové práce: Christian Genest, PhD, FRSC,
Katedra matematiky a statistiky

Abstrakt: Tato diplomová práce se zabývá tématy, které jsou vztaženy k úloze, zda zákazníci poznají rozdíl mezi výrobky. Nejprve je zde probrána teorie testů ekvivalence a noninferiority, což jsou vhodné nástroje v úlohách, kdy je cílem ukázat, že dané výrobky jsou ekvivalentní, nebo že nový výrobek není podstatně horší než stávající. Dále je v práci popsán Thurstonův přístup, jak měřit a porovnávat působení podnětů na lidské smysly. Následně jsou, s využitím předchozích kapitol, rozebrány tři běžné nepovinné normy, které řeší úlohy sensorické analýzy typu, zda zákazníci poznají rozdíl mezi výrobky. Jedná se o zkoušku duo-trio, trojúhelníkovou a párovou porovnávací zkoušku. Je zde podrobně vysvětlen jejich statistický podklad a zkoušky jsou porovnány na základě síly testů. K normám je uveden i alternativní přístup založený na Thurstonově myšlence. V práci je navíc přiblížen Saatyho způsob odhadování vektoru priorit, téma užitečné, když vybíráme nejlepší z n objektů nebo objekty chceme seřadit či porovnat. Je představen i jiný způsob odhadování vektoru priorit, který je Saatyho přístupem inspirován.

Klíčová slova: Testování hypotéz, ekvivalence, noninferiorita, sensorika

Title: Testing equivalence and non inferiority

Author: Nela Rychterová

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Jaromír Antoch, CSc.,
Department of Probability and Mathematical Statistics

Consultant: Christian Genest, PhD, FRSC,
Department of Mathematics and Statistics

Abstract: This master thesis deals with topics related to the task whether customers are able to recognize a difference between products. First, testing of equivalence and non-inferiority is discussed in detail. It is an important tool when verifying that two products are equivalent or that a new product is not substantially worse than a current product. Afterwards, Thurstone's approach is introduced as a way to evaluate the impact of a stimulus on human senses. Subsequently, using the previous chapters, there is a detailed discussion dealing with three standards widely used in practice in the case when someone needs to apply sensory evaluation to verify whether customers are able to recognize a difference between products. In particular, these are duo-trio, triangle and paired comparison tests. There is a thorough explanation of their statistical base and the tests are compared according to their power. Furthermore, an approach based on the Thurstone's theory is introduced as an alternative to the standard methods. Moreover, this thesis introduces Saaty's approach to the estimation of a priority vector, which is a useful tool to compare, to order or to choose the best one from n objects. We also introduce another approach to estimation of a priority vector which is based on Saaty's idea.

Keywords: Hypotheses testing, equivalence, non inferiority, sensory evaluation

Značení a zkratky

\mathbb{N}	- přirozená čísla
\mathbb{N}_0	- $\mathbb{N} \cup 0$
\mathbb{Z}	- celá čísla
\mathbb{R}	- reálná čísla
\mathbb{R}_+ (\mathbb{R}_-)	- reálná kladná (záporná) čísla
\mathbb{R}^k	- k -rozměrný Euklidovský prostor
$\lfloor x \rfloor$	- dolní celá část x
$\mathbf{1}_{(B)}$, $\mathbf{1}(B)$	- indikátor množiny B
$\mathbf{0}_n$	- nulový vektor délky n
$\text{diag}(a_j)_{j=1,\dots,n}$	- diagonální matice $n \times n$ s prvky $a_j, j = 1, \dots, n$, na diagonále
$\text{Rank}(\mathbf{A})$	- hodnost matice \mathbf{A}
$\ \cdot\ $	- Euklidovská norma
u_α	- α -kvantil standardizovaného normálního rozdělení
Φ	- distribuční funkce náhodné veličiny se standardním normálním rozdělením
F_θ	- distribuční funkce, pokud je skutečná hodnota parametru θ
$P(A)$	- pravděpodobnost jevu A
$E X$	- střední hodnota náhodné veličiny X
$\text{var } X$	- rozptyl náhodné veličiny X
CLV	- centrální limitní věta

Úvod

Podniky v potravinářském průmyslu často řeší problém, zda zákazníci poznají, když částečně změní výrobu produktu. Někdy pozměňují výrobu za účelem, aby na trh uvedli nový produkt odlišný od již zažitého. V tomto případě podniky chtějí, aby zákazník produkty rozlišil. Jindy podniky výrobu pozměňují proto, že chtějí použít dostupnější (levnější, atd.) surovinu, ale nechtějí, aby zákazníci rozdíl mezi produkty poznali.

Dvě zmíněné úlohy týkající se pozměněné výroby produktů si jsou podobné, ale statistické hypotézy o nich musejí být formulovány odlišně. Hypotézy v „klasickém“ smyslu jsou formulovány tak, že nulová hypotéza vyjadřuje „rovnost“ produktů a alternativa jejich „odlišnost“. Tento přístup je vhodný, když podnik mění výrobu za účelem vytvoření nového produktu. V tomto případě se výrobce chce ujistit, že zákazníci rozdíl mezi produkty poznají. Zamítnutí nulové hypotézy vede k závěru, že produkty působí odlišně, ale při nezamítnutí nulové hypotézy nemůžeme říct nic o tom, zda jsou produkty shodné nebo ne.

Naopak pokud se výrobce chce ujistit, že dva produkty působí podobně a lze je použít zaměnitelně, „klasický“ přístup není vhodným nástrojem pro ukázání shody (ekvivalence) produktů. V takovém případě je nutné použít pozměněnou úlohu testování hypotéz, a to testy ekvivalence. V testech ekvivalence je nulová hypotéza stanovena jako „shoda“ a alternativa jako „neshoda“. Zamítnutí nulové hypotézy v tomto případě vede k závěru, že produkty jsou ekvivalentní. Testy ekvivalence jsou také velmi důležité v technickém odvětví nebo klinických zkouškách. S testováním ekvivalence a noninferiority se seznámíme v první kapitole.

Úloha, zda zákazníci poznají rozdíl mezi produkty, zahrnuje další problémy, se kterými je třeba se vypořádat. Zabýváme se porovnáváním a hodnocením produktů na základě lidských smyslů, což jsou problémy tzv. senzorické analýzy. Zde se potýkáme s problémem, jak kvantifikovat působení produktů (obecně podnětů) na lidské smysly a jak podněty porovnávat, což nemusí být vůbec přímočaré. Druhá kapitola je věnována Thurstonově myšlence, která se zabývá právě popisem, měřením a porovnáváním působení podnětů na člověka.

Třetí kapitola je věnována odhadu priorit n objektů, který je založený na Saatyho myšlence.

Dále existují nepovinné evropské normy, které podnikům radí, jak v úlohách senzorické analýzy postupovat. Normy mají statistický podklad, který ale často není nikde dobře vysvětlen. Ve čtvrté kapitole se podrobně zabýváme běžně užívanými normami, a to zkouškou duo-trio, trojúhelníkovou a párovou porovnávací zkouškou.

Obsah

Značení a zkratky	i
Úvod	iii
1 Testy ekvivalence a noninferiority	1
1.1 Stejně nejméně silnější testy ekvivalence a noninferiority	3
1.2 TOST procedura	13
2 Thurstonovská myšlenka porovnávání podnětů	15
3 Odhad vektoru priorit	21
3.1 Saatyho odhad vektoru priorit a další možnost odhadu	21
3.2 Simulační studie	25
4 Normy	29
4.1 Obecné informace k normám	29
4.2 Zkouška duo-trio	31
4.2.1 Provedení zkoušky duo-trio, pokyny pro vedoucího zkoušky	31
4.2.2 Vyhodnocení zkoušky duo-trio	32
4.2.3 Statistické vysvětlení zkoušky duo-trio	33
4.3 Trojúhelníková zkouška	36
4.3.1 Provedení trojúhelníkové zkoušky, pokyny pro vedoucího zkoušky	36
4.3.2 Vyhodnocení trojúhelníkové zkoušky	37
4.3.3 Statistické vysvětlení trojúhelníkové zkoušky	38
4.4 Párová porovnávací zkouška	40
4.4.1 Provedení párové porovnávací zkoušky, pokyny pro vedoucího zkoušky	40
4.4.2 Vyhodnocení párové porovnávací zkoušky	41
4.4.3 Statistické vysvětlení párové zkoušky	42
4.5 Porovnání zkoušek	49
4.6 Thurstonovský přístup	51
4.6.1 Thurstone–Ura model	51
4.6.2 Hypotézy a testování v Thurstone–Ura modelu	54
Závěr	59
Dodatek	61
Seznam použité literatury	63
Přílohy	65
A.1 Ukázka zdrojového kódu	65
A.2 Zdrojový kód k návrhu testu pro podobnost	68

1. Testy ekvivalence a noninferiority

V praxi často bývá třeba ukázat, že nový postup (výrobek, lék apod.) není podstatně horší než již zažitý postup (výrobek, lék apod.), nebo že jsou ekvivalentní, tj. „dostatečně“ podobné. Tato problematika vede na úlohu testování noninferiority a ekvivalence.

Uvedme příklady úloh, kde se testy ekvivalence a noninferiority uplatňují.

Příklad 1.1. Ze zkušenosti se ví, že při dané léčbě dojde k uzdravení 80 % pacientů. Při nové experimentální léčbě jsou použity méně toxické látky, které jsou šetrnější k organismu člověka. Otázkou je, zda není nová léčba významně horší než zažitá léčba vzhledem k procentu uzdravených pacientů. Pokud je nová léčba šetrnější, potom se možná lze spokojit se 79 % uzdravených pacientů.

Příklad 1.2. Při vyvíjení nové generické verze léku se zkoumá, zda je účinnost nového léku ekvivalentní s účinností zaběhlého léku. Zde by bylo ideální, kdyby účinnost nového léku byla totožná s účinností předchozího léku. Avšak tato striktní podmínka nemusí být ani reálná, ani potřebná. Stačí ukázat, že je účinnost nového léku dostatečně podobná předchozí verzi.

Příklad 1.3. V technické praxi lze některé hodnoty parametrů výrobního procesu považovat za ideální, např. tlak nebo teplotu. Často ale stačí mít hodnoty v určitém rozmezí kolem ideální hodnoty. V takovém případě je třeba kontrolovat, zda hodnoty zůstávají v daném rozmezí.

Testy ekvivalence a noninferiority jsou založeny na dvou hlavních myšlenkách:

- Ekvivalencí se nemyslí přesná shoda, ale „dostatečná“ podobnost, což jsme naznačili na příkladech výše.
- V testech ekvivalence je nulová hypotéza stanovena jako „neekvivalence“, alternativa je formulovaná jako ekvivalence. V testech noninferiority je nulová hypotéza stanovena jako nedosažení dané úrovně a alternativa jako dosažení dané úrovně. Formulace hypotéz u testů ekvivalence je „obráceně“ proti klasické formulaci, kde za nulové hypotézy platí požadovaná vlastnost, např. testy o předpokladech normality a homoskedasticity, Kolmogorovův Smirnovův test a test dobré shody. Pokud chybně formulujeme nulovou hypotézu jako ekvivalenci a alternativu jako neekvivalenci, nezamítnutí nulové hypotézy nelze brát jako „přijmutí“ ekvivalence. Testy ekvivalence jsou silnější nástroj pro „přijmutí“ shody než klasické testy.

Při testování ekvivalence a noninferiority se často zaměřujeme na reálný parametr θ , který vhodně vyjadřuje vlastnost zkoumaného pravděpodobnostního rozdělení.

U testů ekvivalence jsou hypotézy o parametru θ často formulovány oboustranně jako

$$H_0 : \theta \notin (\theta_0 - \epsilon_1, \theta_0 + \epsilon_2) \quad \text{proti} \quad H_1 : \theta \in (\theta_0 - \epsilon_1, \theta_0 + \epsilon_2), \quad (1.1)$$

kde θ_0 je daný parametr a ϵ_1, ϵ_2 dané kladné konstanty.

U testů noninferiority jsou hypotézy o parametru θ často formulovány jednostranně, a to

$$H_0 : \theta \leq \theta_0 - \epsilon \quad \text{proti} \quad H_1 : \theta > \theta_0 - \epsilon,$$

kde θ_0 je daný parametr a ϵ daná kladná konstanta.

Příklad 1.4. Porovnání testů klasických hypotéz a testů ekvivalence:

- Testování střední hodnoty μ normálního rozdělení:
Klasická formulace hypotéz:

$$H_0 : \mu = \mu_0 \quad \text{proti} \quad H_1 : \mu \neq \mu_0.$$

Formulace hypotéz při testování ekvivalence:

$$H_0 : \mu \notin (\mu_0 - \epsilon_1, \mu_0 + \epsilon_2) \quad \text{proti} \quad H_1 : \mu \in (\mu_0 - \epsilon_1, \mu_0 + \epsilon_2).$$

- Dvouvýběrový problém, testování středních hodnot μ_1 a μ_2 ve dvou populacích:

Klasická formulace hypotéz:

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{proti} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Formulace hypotéz při testování ekvivalence:

$$H_0 : \mu_1 - \mu_2 \notin (-\epsilon_1, \epsilon_2) \quad \text{proti} \quad H_1 : \mu_1 - \mu_2 \in (-\epsilon_1, \epsilon_2).$$

- Testování parametru p alternativního rozdělení:

Klasická formulace hypotéz:

$$H_0 : p = p_0 \quad \text{proti} \quad H_1 : p \neq p_0.$$

Formulace hypotéz při testování ekvivalence:

$$H_0 : p \notin (p_0 - \epsilon_1, p_0 + \epsilon_2) \quad \text{proti} \quad H_1 : p \in (p_0 - \epsilon_1, p_0 + \epsilon_2).$$

- Testování funkcí přežití S_1 a S_2 ve dvou populacích v případě cenzorovaných dat:

Klasická formulace hypotéz:

$$H_0 : \forall t \geq 0 : S_1(t) = S_2(t) \quad \text{proti} \quad H_1 : \exists t \geq 0 : S_1(t) \neq S_2(t).$$

Formulace hypotéz při testování ekvivalence:

$$H_0 : \sup_{t \geq 0} |S_1(t) - S_2(t)| \geq \delta \quad \text{proti} \quad H_1 : \sup_{t \geq 0} |S_1(t) - S_2(t)| < \delta,$$

kde $\delta > 0$.

Polem uplatnění testů ekvivalence a noninferiority jsou například klinické studie zahrnující aktivní kontrolu. V takovém experimentu je jedné skupině pacientů poskytnuta nejlepší dostupná léčba a druhé skupině nová léčba. Většinou se neočekává, že by nová léčba byla výrazně účinnější než původní léčba, ale může mít některé lepší vlastnosti, například může být šetrnější nebo levnější (příklad 1.1). V klinických studiích se obvykle uplatňují testy zkoumající parametr p alternativního rozdělení či testy o shodnosti funkcí přežití v případě cenzorovaných dat.

1.1 Stejněměrně nejsilnější testy ekvivalence a noninferiority

Vysvětlili jsme myšlenku testování ekvivalence a noninferiority a ukázali několik příkladů, jak formulovat hypotézy. Zaměříme se na testování hypotéz.

Důležitou a často užívanou třídou rodin pravděpodobnostních rozdělání jsou jednoparametrické rodiny exponenciálního typu.

Definice 1.1. *Nechť $\theta \in \mathbb{R}$. Nechť náhodný vektor \mathbf{X} má pravděpodobnostní rozdělení P_θ , které je absolutně spojitě vzhledem k σ -konečné míře μ a příslušná hustota $p_\theta = dP_\theta/d\mu$ má tvar*

$$p_\theta(\mathbf{x}) = C(\theta) \exp \{Q(\theta)T(\mathbf{x})\} h(\mathbf{x}),$$

kde $C(\theta) > 0$, $h(\mathbf{x}) \geq 0$. Potom říkáme, že $p_\theta(\mathbf{x})$ je jednoparametrická hustota exponenciálního typu.

Do takové třídy rozdělání patří například rodiny hustot:

- normálního rozdělání s neznámou střední hodnotou a známým rozptylem,
- normálního rozdělání se známou střední hodnotou a neznámým rozptylem,
- exponenciálního rozdělání s neznámou střední hodnotou,
- alternativního rozdělání s neznámým parametrem p ,
- binomického rozdělání se známým n a neznámým parametrem p ,
- Poissonova rozdělání s neznámou střední hodnotou. Atd.

Při testování hypotéz hledáme testy, které dodržují zvolenou hladinu a snažíme se nalézt testy s největší silou. Věta 1.1 uvádí jednostranné testy v jednoparametrických exponenciálních rodinách, které jsou stejněměrně nejsilnější, odkažme se na definici D.6 v dodatku.

Věta 1.1. *Nechť náhodný vektor \mathbf{X} má hustotu p_θ z jednoparametrické exponenciální rodiny hustot $\{p_\theta, \theta \in \Theta \subseteq \mathbb{R}\}$, $\theta_0 \in \Theta$ a funkce Q (z definice 1.1) je striktně monotónní. Potom pro statistický problém testování hypotéz*

$$H_0 : \theta \leq \theta_0 \quad \text{proti} \quad H_1 : \theta > \theta_0$$

(i) *existuje stejněměrně nejsilnější test s kritickou funkcí ϕ , která má tvar:*

$$\begin{array}{ll} a) \text{ je-li } Q \text{ rostoucí} & b) \text{ je-li } Q \text{ klesající} \\ \phi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) > C, \\ \gamma, & T(\mathbf{x}) = C, \\ 0, & T(\mathbf{x}) < C, \end{cases} & \phi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) < C, \\ \gamma, & T(\mathbf{x}) = C, \\ 0, & T(\mathbf{x}) > C, \end{cases} \end{array} \quad (1.2)$$

kde $C \in \mathbb{R}$ a $\gamma \in [0,1)$ splňují

$$E_{\theta_0} \phi(\mathbf{X}) = \alpha. \quad (1.3)$$

(ii) Test splňující (1.2) a (1.3) je stejnoměrně nejsilnějším testem hypotéz $H_0 : \theta \leq \theta$ proti $H_1 : \theta > \theta$ s hladinou α .

Důkaz. Viz Lehmann (1997, Corollary 2, str. 80).

Q.E.D.

Poznámka. Pokud známe rozdělení $T(\mathbf{X})$ za platnosti parametru θ_0 , není těžké stejnoměrně nejsilnější test nalézt. Předpokládejme, že Q je rostoucí funkce. Konstanty C a γ jsou určeny rovností $E_{\theta_0} \phi(\mathbf{X}) = \alpha$, přičemž platí

$$E_{\theta_0} \phi(\mathbf{X}) = P_{\theta_0}\{T(\mathbf{X}) > C\} + \gamma P_{\theta_0}\{T(\mathbf{X}) = C\}.$$

Položme $C = \sup\{x : P_{\theta_0}\{T(\mathbf{X}) > x\} \leq \alpha\}$. Potom je

$$\gamma = \begin{cases} 0, & P_{\theta_0}\{T(\mathbf{X}) > C\} = \alpha, \\ \frac{\alpha - P_{\theta_0}\{T(\mathbf{X}) > C\}}{P_{\theta_0}\{T(\mathbf{X}) = C\}}, & P_{\theta_0}\{T(\mathbf{X}) > C\} < \alpha. \end{cases}$$

Dle věty 1.1 bodu (ii) se jedná o stejnoměrně nejsilnější test.

Následující věta uvádí stejnoměrně nejsilnější testy ekvivalence v oboustranném smyslu v jednoparametrických exponenciálních rodinách hustot.

Věta 1.2. *Nechť náhodný vektor \mathbf{X} má hustotu p_θ z jednoparametrické exponenciální rodiny hustot $\{p_\theta, \theta \in \Theta \subseteq \mathbb{R}\}$, $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ a funkce Q (z definice 1.1) je striktně monotónní. Potom pro statistický problém testování hypotéz*

$$H_0 : \theta \notin (\theta_1, \theta_2) \quad \text{proti} \quad H_1 : \theta \in (\theta_1, \theta_2)$$

(i) *existuje stejnoměrně nejsilnější test s kritickou funkcí ϕ , která má tvar:*

$$\phi(\mathbf{x}) = \begin{cases} 1, & T(\mathbf{x}) \in (C_1, C_2), \\ \gamma_i, & T(\mathbf{x}) = C_i, i = 1, 2, \\ 0, & T(\mathbf{x}) \notin [C_1, C_2], \end{cases} \quad (1.4)$$

kde $C_1, C_2 \in \mathbb{R}$ a $\gamma \in [0, 1]$ splňují rovnosti

$$E_{\theta_1} \phi(\mathbf{X}) = E_{\theta_2} \phi(\mathbf{X}) = \alpha. \quad (1.5)$$

(ii) *Test splňující (1.4) a (1.5) je stejnoměrně nejsilnějším testem hypotéz $H_0 : \theta \notin (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$ s hladinou α .*

Důkaz. Viz Lehmann (1997, Theorem 6, str. 101–103).

Q.E.D.

Ukažme stejnoměrně nejsilnější testy ekvivalence v další třídě rozdělení. Uplatňují se rodiny rozdělení, jejichž hustoty jsou, vzhledem k nějaké σ -konečné míře μ , striktně totálně pozitivní řádu tři, zkráceně STP₃. Uvedme definici 1.2.

Definice 1.2. Necht \mathcal{X} a Θ jsou lineárně uspořádané množiny a $\{p_\theta, \theta \in \Theta\}$ je rodina reálných funkcí s definičním oborem \mathcal{X} . Pro libovolné $n \in \mathbb{N}$ definujeme množiny

$$\mathcal{X}^{(n)} = \{(x_1, \dots, x_n); x_1 < x_2 < \dots < x_n, x_1, \dots, x_n \in \mathcal{X}\}$$

a

$$\Theta^{(n)} = \{(\theta_1, \dots, \theta_n); \theta_1 < \theta_2 < \dots < \theta_n, \theta_1, \dots, \theta_n \in \Theta\}.$$

Pro libovolné $(x_1, \dots, x_n) \in \mathcal{X}^{(n)}$ a $(\theta_1, \dots, \theta_n) \in \Theta^{(n)}$ označme

$$\Delta_n \begin{pmatrix} x_1, \dots, x_n \\ \theta_1, \dots, \theta_n \end{pmatrix} = \det \begin{pmatrix} p_{\theta_1}(x_1) & \dots & p_{\theta_1}(x_n) \\ \vdots & \ddots & \vdots \\ p_{\theta_n}(x_1) & \dots & p_{\theta_n}(x_n) \end{pmatrix}.$$

Necht $r \in \mathbb{N}$. Pokud pro každé $n = 1, \dots, r$ platí

$$\Delta_n \begin{pmatrix} x_1, \dots, x_n \\ \theta_1, \dots, \theta_n \end{pmatrix} > 0 \quad \forall ((x_1, \dots, x_n), (\theta_1, \dots, \theta_n)) \in \mathcal{X}^{(n)} \times \Theta^{(n)}, \quad (1.6)$$

potom rodinu $\{p_\theta, \theta \in \Theta\}$ nazveme striktně totálně pozitivní řádu r , zkráceně STP_r . Jestliže podmínka (1.6) platí pro každé $r \in \mathbb{N}$, rodinu $\{p_\theta, \theta \in \Theta\}$ nazveme striktně totálně pozitivní nekonečného řádu, zkráceně STP_∞ .

Poznámka. Definice 1.2 je uvedena v Wellek (2003, str. 249), původně v Karlin (1968, str. 11, 12)

Uvedme některé typy hustot, které jsou STP_∞ , a tím i STP_3 .

Lemma 1.3. Necht Θ je nedegenerovaný interval, \mathcal{X} je borelovská množina na \mathbb{R} a μ je σ -konečná míra. Necht náhodná veličina X má hustotu z rodiny $\{p_\theta, \theta \in \Theta\}$ vzhledem k μ , a to tvaru

$$p_\theta(x) = C(\theta) \exp(\theta x) h(x),$$

kde $C : \Theta \rightarrow \mathbb{R}_+$ a $h : \mathcal{X} \rightarrow \mathbb{R}_+$. Potom $\{p_\theta, \theta \in \Theta\}$ je STP_∞ .

Důkaz. Viz Wellek (2003, Lemma A.1.2, str. 249).

Q.E.D.

Poznámka.

- Tvrzení lemmatu 1.3 platí i pro $\Theta \subseteq \mathbb{R}$ borelovskou a hustoty tvaru

$$p_\theta(x) = C(\theta) \exp\{Q(\theta)x\} h(x),$$

kde $Q : \Theta \rightarrow \mathbb{R}$ je rostoucí funkce, $C : \Theta \rightarrow \mathbb{R}_+$ a $h : \mathcal{X} \rightarrow \mathbb{R}_+$.

- Lemma 1.3 ukazuje vztah mezi rodinami jednoparametrických hustot exponenciálního typu a rodinami STP_∞ . Z lemmatu 8 v Lehmann (1997, str. 58) plyne, že hustotu z definice 1.1 tvaru

$$p_\theta(\mathbf{x}) = C(\theta) \exp\{Q(\theta)T(\mathbf{x})\} h(\mathbf{x})$$

vzhledem k σ -konečné míře μ můžeme transformací $Y = T(\mathbf{X})$ převést na

$$p_\theta(y) = C(\theta) \exp\{Q(\theta) y\} \tilde{h}(y) \quad (1.7)$$

vzhledem k nějaké σ -konečné míře ν , kde $\tilde{h} : \mathcal{Y} \rightarrow \mathbb{R}_+$ a $\mathcal{Y} \subseteq \mathbb{R}$ je borelovská. Pokud je funkce Q rostoucí, je rodina hustot STP_∞ . Pokud Q není rostoucí, lze zkoumat transformovaný parametr $\delta = Q(\theta)$, v tomto případě je příslušná rodina hustot STP_∞ . Vidíme, že rodinu hustot, do které patří hustota $T(\mathbf{X})$, lze vzhledem k vhodné míře vyjádřit jako STP_∞ .

Lemma 1.4. *Pro libovolné $\nu \in \mathbb{N}$ a $\theta \in \mathbb{R}$ označme $p_\theta(x; \nu)$ hustotu necentrálního t -rozdělení s ν stupni volnosti a parametrem necentrality θ . Potom rodina hustot $\{p_\theta(x; \nu), \theta \in \mathbb{R}\}$ je STP_∞ .*

Důkaz. Viz Wellek (2003, Lemma A.1.3, str. 250).

Q.E.D.

Lemma 1.5. *Nechť $(\nu_1, \nu_2)^\top \in \mathbb{N}^2$ a V je náhodná veličina, která má F -rozdělení s ν_1 a ν_2 stupni volnosti. Pro libovolné $\rho \in \mathbb{R}_+$ označme $h_\rho(x; \nu_1, \nu_2)$ hustotu náhodné veličiny ρV . Potom rodina hustot $\{h_\rho(x; \nu_1, \nu_2), \rho > 0\}$ je STP_∞ .*

Důkaz. Viz Wellek (2003, Lemma A.1.4, str. 250).

Q.E.D.

Následující věta 1.6 udává podmínky pro existenci stejnoměrně nejsilnějších testů ve STP_3 rodinách hustot s jednorozměrným reálným parametrem.

Věta 1.6. *Nechť $\alpha \in (0,1)$, $\mathcal{X} \subseteq \mathbb{R}$ a $\Theta \subseteq \mathbb{R}$ jsou nedegenerované intervaly, $\mathcal{B}_\mathcal{X}$ borelovská σ -algebra na \mathcal{X} , μ je σ -konečná míra na $\mathcal{B}_\mathcal{X}$, jejíž nosič obsahuje alespoň dva různé body. Nechť náhodná veličina X je z rodiny hustot $\{p_\theta, \theta \in \Theta\}$ vzhledem k μ na \mathcal{X} , která je STP_3 . Dále předpokládejme, že funkce $(x, \theta) \mapsto p_\theta(x)$ je spojitá v obou argumentech. Nechť jsou dány body $\theta_1, \theta_2 \in \Theta$ splňující $\theta_1 < \theta_2$.*

Potom pro problém testování hypotézy $H_0 : \theta \in \Theta \setminus (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$

(i) existuje stejnoměrně nejsilnější test s hladinou významnosti α a kritickou funkcí $\phi : \mathcal{X} \rightarrow [0,1]$ tvaru:

$$\phi(x) = \begin{cases} 1, & x \in (C_1, C_2), \\ \gamma_i, & x = C_i, i = 1, 2, \\ 0, & x \in \mathcal{X} \setminus [C_1, C_2], \end{cases} \quad (1.8)$$

kde $C_1, C_2 \in \mathcal{X}, C_1 < C_2, \gamma_1, \gamma_2 \in [0,1]$ splňují rovnosti

$$E_{\theta_1} \phi(X) = E_{\theta_2} \phi(X) = \alpha. \quad (1.9)$$

(ii) Test splňující (1.8) a (1.9) je stejnoměrně nejsilnějším testem hypotézy $H_0 : \theta \in \Theta \setminus (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$ na hladině α .

Důkaz. Věta v Wellek (2003, Theorem A.1.5, str. 251) s odkazem na Lehmann (1997, kapitola 3.7) a Kallenberg a Janssen (1984, str. 54–58).

Q.E.D.

V poznámce pod lemmatem 1.3 jsme ukázali, že hustota veličiny $Y = T(\mathbf{X})$ z věty 1.2 vzhledem k vhodné σ -konečné míře patří do rodiny hustot STP_∞ . Při splnění podmínky spojitosti $(y, \theta) \mapsto p_\theta(y)$ v obou argumentech, kde $p_\theta(y)$ je hustota Y za platnosti θ , je vlastně věta 1.6 rozšířením věty 1.2.

Po přidání určitých předpokladů „symetrie“ ve větě 1.6 se zjednoduší tvar a výpočet konstant C_1 a C_2 . Uvedme větu 1.7.

Věta 1.7. *Nechť jsou splněny předpoklady věty 1.6 a náhodná veličina X má hustotu z rodiny $\{p_\theta, \theta \in \Theta\}$ vzhledem k σ -konečné míře μ . Navíc předpokládejme, že $\theta_1 = -\theta_2 = \epsilon$, kde $\epsilon > 0$, a že rozdělení X v případě, že skutečný parametr je ϵ , je stejné jako rozdělení $-X$ v případě, že skutečný parametr je $-\epsilon$. Potom kritická funkce stejnoměrně nejsilnějšího testu má tvar:*

$$\phi(x) = \begin{cases} 1, & |x| < C, \\ \gamma, & |x| = C, \\ 0, & |x| > C, \end{cases}$$

kde $C = \sup\{x \in [0, \infty) : P_\epsilon(|X| < x) \leq \alpha\}$ a

$$\gamma(x) = \begin{cases} \frac{\alpha - P_\epsilon(|X| < C)}{P_\epsilon(|X| = C)}, & P_\epsilon(|X| = C) > 0, \\ 0, & P_\epsilon(|X| = C) = 0. \end{cases}$$

Důkaz. Viz Wellek (2003, Lemma A.1.6, str. 252).

Q.E.D.

Poznámka. Ve větách 1.1, 1.2, 1.6 a 1.7 vystupují znáhodněné testy (odkažme se na definici D.4 v dodatku). Pokud nechceme znáhodněné testy, v praxi se problém řeší tak, že se naleznou meze C_1, C_2 , resp. C , a konstanty γ_1, γ_2 , resp. γ , jako v uvedených větách, a položíme $\gamma_1 = \gamma_2 = 0$, resp. $\gamma = 0$. To znamená, že na hranicích C_1 a C_1 , resp. C nulovou hypotézu nezamítáme. Takový test, pokud původní γ_1, γ_2 , resp. γ byla nenulová, je konzervativní, tzn. má menší hladinu než stanovené α . Případně se nalezne jiná nerandomizovaná verze zmíněných testů.

Věty 1.2 a 1.6 sice říkají, jak stejnoměrně nejsilnější testy vypadají, ale není jasné, jak hodnoty konstant $C_1, C_2, \gamma_1, \gamma_2$ získat. Uvedme algoritmus 1, který v případě veličin se spojitým rozdělením nalézá přibližné hodnoty C_1 a C_2 a algoritmus 2, který v případě veličin s diskrétním rozdělením nalézá konstanty $C_1, C_2, \gamma_1, \gamma_2$. Věta 1.10 pod algoritmy dokazuje správnost algoritmů.

Algoritmus 1

1. Zvolme počáteční konstanty $C_{1,L}^0, C_{1,U}^0$ tak, že
 - $C_{1,L}^0 < C_{1,U}^0, C_{1,L}^0, C_{1,U}^0 \in \mathcal{X}$,
 - konstanta C_1 se nachází v intervalu $(C_{1,L}^0, C_{1,U}^0)$,
 - $1 - F_{\theta_1}(C_{1,U}^0) \geq \alpha$.

Zvolme počáteční $C_1^0 \in (C_{1,L}^0, C_{1,U}^0)$ a toleranční mez $\kappa > 0$. Položme $k = 0$.

2. Spočtěme $C_2^k = F_{\theta_1}^{-1}[\alpha + F_{\theta_1}(C_1^k)]$, kde F_θ (resp. F_θ^{-1}) je distribuční (resp. kvantilová) funkce veličiny X v případě, kdy je θ skutečná hodnota parametru.
3. Spočtěme $\alpha_2^k = F_{\theta_2}(C_2^k) - F_{\theta_2}(C_1^k)$.
4. Pokud je $|\alpha_2^k - \alpha| < \kappa$, skončíme algoritmus s hodnotami C_1^k a C_2^k . V opačném případě:
Je-li je $\alpha_2^k > \alpha$, položme

$$(C_{1,L}^{k+1}, C_{1,U}^{k+1}) = (C_{1,L}^k, C_1^k) \quad \text{a} \quad C_1^{k+1} = (C_{1,L}^{k+1} + C_{1,U}^{k+1})/2.$$

Je-li $\alpha_2^k < \alpha$, položme

$$(C_{1,L}^{k+1}, C_{1,U}^{k+1}) = (C_1^k, C_{1,U}^k) \quad \text{a} \quad C_1^{k+1} = (C_{1,L}^{k+1} + C_{1,U}^{k+1})/2.$$

Položme $k = k + 1$. Vraťme se na krok 2.

Čím je menší toleranční mez $\kappa > 0$, tím je aproximace hodnot C_1 a C_2 přesnější.

Algoritmus 2 je pro veličiny s diskretním rozdělením nabývající celočíselných hodnot.

Algoritmus 2

1. Zvolme počáteční celočíselnou hodnotu C_1^0 takovou, že $\mathbb{P}_{\theta_1}(X > C_1^0) > \alpha$ a pro skutečné C_1 platí $C_1 \leq C_1^0$. Položme $k = 0$.
2. Nalezněme největší celočíselné $C_2^k > C_1^k$ tak, že

$$\mathbb{P}_{\theta_1}(X \in [C_1^k, C_2^k]) \leq \alpha, \quad \mathbb{P}_{\theta_2}(X \in [C_1^k, C_2^k]) \leq \alpha.$$

3. Nalezněme γ_1^k, γ_2^k jako řešení soustavy rovnic

$$\begin{aligned} \mathbb{P}_{\theta_1}(X \in [C_1^k, C_2^k]) + \gamma_1^k \mathbb{P}_{\theta_1}(X = C_1^k) + \gamma_2^k \mathbb{P}_{\theta_1}(X = C_2^k) &= \alpha \\ \mathbb{P}_{\theta_2}(X \in [C_1^k, C_2^k]) + \gamma_1^k \mathbb{P}_{\theta_2}(X = C_1^k) + \gamma_2^k \mathbb{P}_{\theta_2}(X = C_2^k) &= \alpha. \end{aligned}$$

4. Pokud jsou $\gamma_1^k, \gamma_2^k \in [0, 1)$, skončíme algoritmus s hodnotami $C_1^k, C_2^k, \gamma_1^k, \gamma_2^k$. Jinak položme $C_1^{k+1} = C_1^k - 1$, $k = k + 1$ a vraťme se na krok 2.

Návrhy algoritmů 1 a 2 jsou inspirovány prací Wellek (2003, str. 39, 40).

Formulujme pomocná lemmata 1.8 a 1.9.

Lemma 1.8. *Nechť náhodná veličina X je z rodiny hustot $\{p_\theta(x), \theta \in \Theta \subseteq \mathbb{R}\}$ vzhledem k σ -konečné míře μ a borelovská množina $\mathcal{X} \subseteq \mathbb{R}$ obsahuje nosič rozdělení X . Předpokládejme, že $p_\theta(x)$ je kladná funkce pro každé $\theta \in \Theta$ a každé $x \in \mathcal{X}$. Nechť pro každou dvojici $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ je $p_{\theta_2}(x)/p_{\theta_1}(x)$ striktně rostoucí funkce x na \mathcal{X} . Nechť $\psi(x)$ je reálná funkce taková, že existuje $x_0 \in \mathcal{X}$ splňující*

$$\begin{aligned} \psi(x) &\leq 0, & x < x_0, \\ \psi(x) &\geq 0, & x > x_0, \end{aligned} \quad x \in \mathcal{X}.$$

Předpokládejme, že $P_\theta[\psi(X) \neq 0] > 0$ pro každé $\theta \in \Theta$. Nechť existuje θ_0 splňující $E_{\theta_0} \psi(X) = 0$.

Potom pro každé $\theta < \theta_0$ je $E_\theta \psi(X) < 0$ a pro každé $\theta > \theta_0$ je $E_\theta \psi(X) > 0$.

Důkaz. Vezměme libovolné $\theta_2 > \theta_1$. Označme $p_{\theta_2}(x_0)/p_{\theta_1}(x_0) = C_{\theta_1, \theta_2} > 0$. Ukážeme, že

$$E_{\theta_2} \psi(X) > C_{\theta_1, \theta_2} E_{\theta_1} \psi(X). \quad (1.10)$$

$$\begin{aligned} E_{\theta_2} \psi(X) - C_{\theta_1, \theta_2} E_{\theta_1} \psi(X) &= \int_{\mathcal{X}} (\psi p_{\theta_2} - C_{\theta_1, \theta_2} \psi p_{\theta_1}) d\mu \\ &= \int_{\mathcal{X}} \left(\psi \frac{p_{\theta_2}}{p_{\theta_1}} p_{\theta_1} - C_{\theta_1, \theta_2} \psi p_{\theta_1} \right) d\mu = \int_{\mathcal{X}} \psi p_{\theta_1} \left(\frac{p_{\theta_2}}{p_{\theta_1}} - C_{\theta_1, \theta_2} \right) d\mu. \end{aligned}$$

Funkce $\psi p_{\theta_1} (p_{\theta_2}/p_{\theta_1} - C_{\theta_1, \theta_2})$ je na $\mathcal{X} \cap \{(-\infty, x_0) \cup (x_0, \infty)\}$ kladná, v x_0 má hodnotu 0. Z předpokladů $E_{\theta_0} \psi(X) = 0$ a $P_\theta[\psi(X) \neq 0] > 0$ pro každé $\theta \in \Theta$ plyne, že míra μ není koncentrovaná jen na $\{x_0\}$, a tedy

$$E_{\theta_2} \psi(X) - C_{\theta_1, \theta_2} E_{\theta_1} \psi(X) > 0.$$

První nerovnost v tvrzení získáme volbou θ_0 za θ_2 v (1.10), druhou nerovnost v tvrzení získáme volbou θ_0 za θ_1 v (1.10).

Q.E.D.

Lemma 1.9. Nechť náhodná veličina X z rodiny hustot $\{p_\theta(x), \theta \in \Theta \subseteq \mathbb{R}\}$ splňuje podmínky lemmatu 1.8. Nechť $\theta_1, \theta_2 \in \Theta$. Nechť ϕ a ϕ^* jsou dvě funkce tvaru (1.8).

(i) Nechť $P_\theta[\phi(X) \neq \phi^*(X)] > 0$ pro každé $\theta \in \Theta$. Pokud $E_{\theta_1} \phi(x) = E_{\theta_1} \phi^*(x)$, potom $E_{\theta_2} \phi(x) < E_{\theta_2} \phi^*(x)$ pro $\theta_1 < \theta_2$, resp. $E_{\theta_2} \phi(x) > E_{\theta_2} \phi^*(x)$ pro $\theta_1 > \theta_2$.

(ii) Jestliže $\theta_1 \neq \theta_2$ a platí $E_{\theta_1} \phi(x) = E_{\theta_1} \phi^*(x)$ a $E_{\theta_2} \phi(x) = E_{\theta_2} \phi^*(x)$, potom $\phi = \phi^*$ s pravděpodobnostní jedna.

Důkaz. Položme $\psi(x) = \phi^*(x) - \phi(x)$. V případě (i) funkce ψ splňuje podmínky lemmatu (1.8). Platí $E_{\theta_1} \psi(X) = 0$, a tedy $E_{\theta_2} \psi(X) < 0$ pro $\theta_2 < \theta_1$, resp. $E_{\theta_2} \psi(X) > 0$ pro $\theta_2 > \theta_1$. Čímž je dokázána (i). Tvrzení (ii) plyne z (i).

Q.E.D.

Věta 1.10. Nechť náhodný vektor \mathbf{X} a funkce T jsou z věty 1.2, resp. X je náhodná veličina z věty 1.6.

(i) Nechť $Y = T(\mathbf{X})$ má spojitě rozdělení a vzhledem k nějaké σ -konečné míře μ hustotu tvaru $C(\theta)g(y) \exp[Q(\theta)y]$, kde Q je rostoucí na Θ , resp. náhodná veličina X má spojitě rozdělení. Potom algoritmus 1 končí po konečně mnoha krocích a limitně vede ke stejnoměrně nejsilnějšímu testu hypotéz $H_0 : \theta \in \Theta \setminus (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$ s hladinou α .

(ii) Nechť $Y = T(\mathbf{X})$, resp. X , je diskrétní náhodná veličina nabývající pouze celočíselných hodnot. Potom algoritmus 2 končí po konečně mnoha krocích a nalézá stejnoměrně nejsilnější test hypotéz $H_0 : \theta \in \Theta \setminus (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$ s hladinou α .

Důkaz. Náhodná veličina X , resp. $Y = T(\mathbf{X})$, splňuje podmínky lemmatu 1.8. Podle věty 1.2, resp. 1.6, víme, že test tvaru (1.4), resp. (1.8), s konstantami C_1, C_2, γ_1 a γ_2 splňující (1.5), resp. (1.9), existuje a je stejnoměrně nejsilnější. Bez újmy na obecnosti dále mluvíme pouze o veličině X (místo o X a Y).

(i) Konstanty γ_1, γ_2 jsou rovny nule, protože X má spojitě rozdělení. Označme $\phi_k(x) = \mathbf{1}[x \in (C_1^k, C_2^k)]$, což je kritická funkce v k -tém cyklu algoritmu 1. Vždy máme $\mathbf{E}_{\theta_1} \phi_k(X) = \alpha$, a pokud $\mathbf{E}_{\theta_2} \phi_k(X) < \alpha$ (resp. $\mathbf{E}_{\theta_2} \phi_k(X) > \alpha$), z lemmatu 1.9 bodu (i) vyplývá, že interval (C_1, C_2) přísluší stejnoměrně nejsilnějšímu testu leží napravo od (C_1^k, C_2^k) , tj. $C_1^k < C_1$ a $C_2^k \leq C_2$ (resp. leží nalevo od (C_1^k, C_2^k) , tj. $C_1 < C_1^k$ a $C_2 \leq C_2^k$). Algoritmus 1 definuje (C_1^{k+1}, C_2^{k+1}) napravo (resp. nalevo) od (C_1^k, C_2^k) . Pro každé k platí, že $C_1, C_1^k \in (C_{1,L}^k, C_{1,U}^k)$ a $C_{1,U}^k - C_{1,L}^k \leq (C_{1,U}^0 - C_{1,L}^0)/2^{k-1}$. Odtud plyne, že hodnota C_1^k je nejvýše $(C_{1,U}^0 - C_{1,L}^0)/2^{k-1}$ vzdálená od nějakého C_1 , ke kterému lze nalézt C_2 splňující (1.8) a (1.9). Tedy platí $C_1^k \rightarrow C_1$ pro $k \rightarrow \infty$. Ze spojitosti $\alpha_2(C) = F_{\theta_2}[g(C)] - F_{\theta_2}(C)$, kde $g(C) = F_{\theta_1}^{-1}[\alpha + F_{\theta_1}(C)]$, platí $\alpha_2^k \rightarrow \alpha$ pro $k \rightarrow \infty$. Tím je (i) dokázáno.

(ii) Platnost tvrzení pro diskrétní celočíselnou náhodnou veličinu je zřejmá.

Tím je věta 1.10 dokázána.

Q.E.D.

Poznámka. Poukažme na nevýhodu algoritmu 1. Výpočty $C_2^k = F_{\theta_1}^{-1}[\alpha + F_{\theta_1}(C_1^k)]$ a $\alpha_2^k = F_{\theta_2}(C_2^k) - F_{\theta_2}(C_1^k)$ v algoritmu 1 jsou zásadní. Správné fungování algoritmu v praxi je velmi ovlivněno správností těchto výpočtů. Pokud jsou (teoretické) hodnoty $F_{\theta_1}^{-1}[\alpha + F_{\theta_1}(C_1^k)]$ a $F_{\theta_1}^{-1}[\alpha + F_{\theta_1}(C_1^k) \pm \epsilon]$ velmi odlišné při malé změně ϵ , algoritmus nemusí fungovat spolehlivě.

Příklad 1.5. Mějme náhodný výběr X_1, \dots, X_n z exponenciálního rozdělení s neznámou střední hodnotou γ , tj. $X_i \sim f_\gamma(x) = (1/\gamma) \exp(-x/\gamma)$. Chceme nalézt stejnoměrně nejsilnější test hypotézy $H_0 : \gamma \notin (\gamma_1, \gamma_2)$ proti $H_1 : \gamma \in (\gamma_1, \gamma_2)$, kde $0 < \gamma_1 < \gamma_2$.

Sdružená hustota náhodného výběru

$$\prod_{i=1}^n f_\gamma(X_i) = \frac{1}{\gamma^n} \exp\left(-\frac{\sum_{i=1}^n X_i}{\gamma}\right)$$

je z jednoparametrické exponenciální rodiny hustot s $T(\mathbf{X}) = \sum_{i=1}^n X_i$. Veličina $Y = T(\mathbf{X})$ má gama rozdělení s parametry n a γ , příslušná hustota je

$$g_\gamma(y; n) = \frac{1}{\gamma^n (n-1)!} y^{n-1} \exp(-y/\gamma).$$

Funkce $Q(\gamma) = -1/\gamma$ je rostoucí na $(0, \infty)$. Rodina $\{g_\gamma(y; n), \gamma \in (0, \infty)\}$ je dle lemmatu 1.3 a poznámky pod ním STP_∞ . Funkce $g_\gamma(y; n)$ je spojitá v γ i y .

Jsou splněny podmínky věty 1.6. Pro stejnoměrně nejsilnější test hledáme konstanty C_1, C_2 splňující $\mathbf{E}_{\gamma_1} \phi(\mathbf{X}) = \mathbf{E}_{\gamma_2} \phi(\mathbf{X}) = \alpha$. Dle věty 1.10 můžeme pro nalezení C_1, C_2 použít algoritmus 1.

Příklad 1.6. Mějme X_1, \dots, X_n náhodný výběr z alternativního rozdělení s neznámým parametrem $p \in (0, 1)$. Chceme testovat hypotézu

$$H_0 : p \notin (p_1, p_2) \quad \text{proti} \quad H_1 : p \in (p_1, p_2),$$

kde $0 < p_1 < p_2 < 1$. Sdružená hustota náhodného výběru je

$$f_p(\mathbf{x}) = p^{\sum_i x_i} (1-p)^{n-\sum_i x_i} = (1-p)^n \exp \left[\sum_{i=1}^n x_i \log \left(\frac{p}{1-p} \right) \right].$$

Hustota $f_p(\mathbf{x})$ je exponenciálního typu s $T(\mathbf{X}) = \sum_{i=1}^n X_i$. Jsou splněny předpoklady věty 1.2, lze nalézt stejnoměrně nejsilnější test, který má tvar (1.4). Konstanty $C_1, C_2, \gamma_1, \gamma_2$ závisí na rozdělení $T(\mathbf{X})$. Veličina $Y = T(\mathbf{X})$ má binomické rozdělení s parametry p a n . Pro nalezení konstant $C_1, C_2, \gamma_1, \gamma_2$ lze použít algoritmus 2.

Příklad 1.7. Užití věty 1.7 ilustrujme na konstrukci stejnoměrně nejsilnějšího testu pro problém s normálně rozdělenými pozorováními. Necht X_1, \dots, X_n , kde $X_i \sim \mathcal{N}(\mu, \sigma_0^2)$, je náhodný výběr, $\mu \in \mathbb{R}$ je neznámý parametr a $\sigma_0^2 \in (0, \infty)$ je známé. Nalezneme pro

$$H_0 : \mu \notin (\mu_0 - \epsilon_1, \mu_0 + \epsilon_2) \quad \text{proti} \quad H_1 : \mu \in (\mu_0 - \epsilon_1, \mu_0 + \epsilon_2)$$

stejnoměrně nejsilnější test.

Nejprve transformujme X_i na

$$\widetilde{X}_i = \frac{X_i - \mu_0 - (\epsilon_2 - \epsilon_1)/2}{\sigma_0}.$$

Platí $\mathbf{E}_{\mu_0 - \epsilon_1} \widetilde{X}_i = -(\epsilon_1 + \epsilon_2)/(2\sigma_0)$, $\mathbf{E}_{\mu_0 + \epsilon_2} \widetilde{X}_i = (\epsilon_1 + \epsilon_2)/(2\sigma_0)$ a $\text{var}_{\mu} \widetilde{X}_i = 1$ pro libovolné μ .

Označme $\tilde{\epsilon} = (\epsilon_1 + \epsilon_2)/(2\sigma_0)$ a $\tilde{\mu} = \mathbf{E}_{\mu} \widetilde{X}_1 = [\mu - \mu_0 - (\epsilon_2 - \epsilon_1)/2]/\sigma_0$. Původní problém je ekvivalentní s

$$H_0 : \tilde{\mu} \notin (-\tilde{\epsilon}, \tilde{\epsilon}) \quad \text{proti} \quad H_1 : \tilde{\mu} \in (-\tilde{\epsilon}, \tilde{\epsilon}),$$

protože

$$\tilde{\mu} < -\tilde{\epsilon} \Leftrightarrow \frac{\mu - \mu_0 - (\epsilon_2 - \epsilon_1)/2}{\sigma_0} \leq -\frac{\epsilon_1 + \epsilon_2}{2\sigma_0} \Leftrightarrow \mu \leq \mu_0 - \epsilon_1$$

a analogicky $\tilde{\mu} \geq \tilde{\epsilon} \Leftrightarrow \mu \geq \mu_0 + \epsilon_2$.

Označme $\overline{\widetilde{X}} = 1/n \sum_{i=1}^n \widetilde{X}_i$. Veličina $X^* = \sqrt{n} \overline{\widetilde{X}}$ má normální rozdělení se střední hodnotou $\mu^* = \sqrt{n} \tilde{\mu}$ a rozptylem jedna. Položme $\epsilon^* = \sqrt{n} \tilde{\epsilon}$.

Původní problém je ekvivalentní s

$$H_0 : \mu^* \notin (-\epsilon^*, \epsilon^*) \quad \text{proti} \quad H_1 : \mu^* \in (-\epsilon^*, \epsilon^*). \quad (1.11)$$

Označme $p_{\mu^*}^*$ hustotu rozdělení X^* za platnosti μ^* . Z lemmatu 1.3 vyplývá, že rodina $\{p_{\mu^*}^*, \mu^* \in \mathbb{R}\}$ je STP_{∞} , a tím i STP_3 . Podmínky věty 1.7 jsou splněny. Nalezneme stejnoměrně nejsilnější test s hladinou α . Z absolutní spojitosti X^* vyplývá, že $\gamma = 0$, konstanta C splňuje $\mathbf{P}_{\epsilon^*}(|X^*| \leq C) = \alpha$. Pokud je skutečná hodnota parametru ϵ^* , veličina $(X^*)^2$ má necentrální χ^2 rozdělení s jedním stupněm volnosti a parametrem necentrality $(\epsilon^*)^2$. Označme α -kvantil tohoto rozdělení $\chi_{1,(\epsilon^*)^2}^2(\alpha)$. Potom

$$\alpha = \mathbf{P}_{\epsilon^*}\{|X^*| \leq C\} = \mathbf{P}_{\epsilon^*}\{(X^*)^2 \leq C^2\} = \mathbf{P}_{\epsilon^*}\{(X^*)^2 \leq \chi_{1,(\epsilon^*)^2}^2(\alpha)\}.$$

Odsud vyplývá, že hledaná konstanta C je rovna $\sqrt{\chi_{1,(\epsilon^*)^2}^2(\alpha)}$.

Stejněměrně nejsilnější test ekvivalence je dán pravidlem, že zamítáme nulovou hypotézu v (1.11), když $|X^*| < \sqrt{\chi_{1,(\epsilon^*)^2}^2(\alpha)}$, resp. v řeči původního výběru a testu, když

$$\left| \sum_{i=1}^n X_i - n\mu_0 - n(\epsilon_2 - \epsilon_1)/2 \right| < \sigma_0 \sqrt{n \chi_{1,(\epsilon^*)^2}^2(\alpha)}.$$

Odvození v příkladu 1.7 je převzato z knihy Wellek (2003, str. 45, 46).

Příklad 1.8. Oproti příkladu 1.7 je v praxi třeba řešit neznámý rozptyl σ^2 . Ukažme stejněměrně nejsilnější test v případě normálně rozdělených dat, pokud je parametrem zájmu $\gamma = \mu/\sigma$ (střední hodnota podělená směrodatnou odchylkou). Mějme náhodný výběr X_1, \dots, X_n z $\mathcal{N}(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}$ a $\sigma^2 \in (0, \infty)$ jsou neznámé. Uvažujme hypotézy

$$H_0 : \gamma \notin (\gamma_1, \gamma_2) \quad \text{proti} \quad H_1 : \gamma \in (\gamma_1, \gamma_2).$$

Označme \bar{X}_n výběrový průměr a S_n^2 výběrový rozptyl. Statistika $T_n = \sqrt{n} \bar{X}_n / S_n$ má necentrální t -rozdělení s $n-1$ stupni volnosti a parametrem necentrality $\sqrt{n} \gamma$. Označme $\sqrt{n} \gamma = \gamma^*$ a příslušnou hustotu p_{γ^*} . Podle lemmatu 1.4 je rodina hustot $\{p_{\gamma^*}, \gamma^* \in \mathbb{R}\}$ STP $_{\infty}$, a tím i STP $_3$. Označme $\gamma_1^* = \sqrt{n} \gamma_1$ a $\gamma_2^* = \sqrt{n} \gamma_2$. Původní problém je ekvivalentní s

$$H_0 : \gamma^* \notin (\gamma_1^*, \gamma_2^*) \quad \text{proti} \quad H_1 : \gamma^* \in (\gamma_1^*, \gamma_2^*).$$

Předpoklady věty 1.6 jsou splněny, tudíž stejněměrně nejsilnější test existuje. Meze C_1, C_2 lze nalézt algoritmem 1.

V symetrickém případě $\mu_1 = -\epsilon$ a $\mu_2 = \epsilon$ jsou splněny také předpoklady věty 1.7. V tomto zjednodušeném případě hledáme konstantu C pro stejněměrně nejsilnější test podobně jako v příkladu 1.7 s využitím faktu, že T_n^2 má necentrální F -rozdělení s 1 a $n-1$ stupni volnosti a parametrem necentrality $(\epsilon^*)^2$.

V knize Wellek (2003) nalezneme řadu dalších testů hypotéz, ke kterým existuje, a je odvozen, stejněměrně nejsilnější test ekvivalence. Například:

- Testování rozptylu v případě normálně rozdělených dat. Mějme náhodný výběr X_1, \dots, X_n z $\sim \mathcal{N}(\mu, \sigma^2)$ s neznámými parametry $\mu \in \mathbb{R}$ a $\sigma^2 \in (0, \infty)$. Parametr zájmu je σ . Příslušné hypotézy jsou $H_0 : \sigma \notin (\sigma_1, \sigma_2)$ proti $H_1 : \sigma \in (\sigma_1, \sigma_2)$.
- Párový problém. Mějme náhodný výběr $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ a předpokládejme, že $D_i = X_i - Y_i \sim \mathcal{N}(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}$ a $\sigma^2 \in (0, \infty)$ jsou neznámé. Parametr zájmu je $\theta = \mu/\sigma$. Příslušné hypotézy jsou $H_0 : \theta \notin (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$, kde $\theta_1 < \theta_2$.
- Dvouvýběrový problém. Mějme dva nezávislé náhodné výběry X_1, \dots, X_n z $\mathcal{N}(\xi, \sigma^2)$ a Y_1, \dots, Y_n z $\mathcal{N}(\mu, \sigma^2)$ s neznámými parametry $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$ a $\sigma^2 \in (0, \infty)$. Parametr zájmu je $\theta = (\xi - \mu)/\sigma$ a hypotézy jsou $H_0 : \theta \notin (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$, kde $\theta_1 < \theta_2$.

V knize Welles (2003) jsou popsány i stejnoměrně nejsilnější nestranné testy ekvivalence (odkážme se na definici D.8 v dodatku. Například:

- Test pro poměr šancí $\theta = [p_1(1 - p_2)]/[p_2(1 - p_1)]$, kde $X \sim \text{Bi}(n, p_1)$ a $Y \sim \text{Bi}(m, p_2)$ jsou nezávislé. Příslušné hypotézy jsou $H_0 : \theta \notin (\theta_1, \theta_2)$ proti $H_1 : \theta \in (\theta_1, \theta_2)$, kde $0 < \theta_1 < \theta_2$.

Také je zde popsána řada dalších testů ekvivalence a noninferiority. Uvedme pouze několik příkladů:

- Testy o parametrech ve více výběrových modelech s normálními výběry.
- Test ekvivalence pro funkce přežití.
- Neparametrické metody v modelech pro párové, dvouvýběrové a více výběrové problémy, kde se srovnávají různé charakteristiky rozdělení.

1.2 TOST procedura

Probrali jsme stejnoměrně nejsilnější testy ekvivalence a noninferiority v jednoparametrických třídách exponenciálního typu a rodinách hustot STP_3 . Existenci a tvar těchto testů jsme uvedli v několika větách. Nevýhodou těchto testů je, že jejich konstrukce nemusí být jednoduchá.

Uvedme proto další testy ekvivalence, jejichž tvar se v praxi dobře nalézají. Jejich nevýhodou je, že mohou být značně konzervativní. Tyto testy jsou založeny na dvou jednostranných testech, proto se těmto testům říká TOST (z angl. two one-sided tests) procedura. TOST proceduru pro konkrétní případ testování uvedl Schuirman (1987).

Věta 1.11. *Nechť $\alpha \in (0, 1)$. Mějme hypotézy*

$$H_0 : \theta \notin (\theta_1, \theta_2) \quad \text{proti} \quad H_1 : \theta \in (\theta_1, \theta_2), \quad (1.12)$$

kde $\theta_1 < \theta_2$. Uvažujme hypotézy

$$H_{0_1} : \theta \leq \theta_1 \quad \text{proti} \quad H_{1_1} : \theta > \theta_1,$$

$$H_{0_2} : \theta \geq \theta_2 \quad \text{proti} \quad H_{1_2} : \theta < \theta_2.$$

Stanovme test hypotéz (1.12) pravidlem, že H_0 zamítáme, když obě hypotézy H_{0_1} a H_{0_2} zamítneme, každou na hladině α . Potom daný test má hladinu nejvyšší α .

Důkaz. Pokud skutečný parametr θ splňuje $\theta \leq \theta_1$, potom platí

$$P_\theta(\text{zamítneme } H_0) \leq P_\theta(\text{zamítneme } H_{0_1}) = \alpha.$$

Pro skutečný parametr $\theta \geq \theta_2$ se $P_\theta(\text{zamítneme } H_0)$ ukáže analogicky.

Q.E.D.

Myšlenku TOST procedury v testech ekvivalence a noninferiority používá například Pardo (2014).

Welles (2003) popisuje testy, které odpovídají TOST proceduře a jejichž konstrukce je založena na intervalech spolehlivosti. Konstrukci testů shrňme ve větě 1.12.

Věta 1.12. Necht $\alpha \in (0,1)$. Předpokládejme, že náhodný vektor \mathbf{X} má hustotu z rodiny hustot $\{p_\theta, \theta \in \Theta \subseteq \mathbb{R}\}$, $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ a o parametru θ jsou sestaveny hypotézy

$$H_0 : \theta \notin (\theta_1, \theta_2) \quad \text{proti} \quad H_1 : \theta \in (\theta_1, \theta_2).$$

Označme $\underline{\theta}(\mathbf{X}, \alpha)$, respektive $\bar{\theta}(\mathbf{X}, \alpha)$, dolní, resp. horní, mez jednostranného $100(1 - \alpha)\%$ intervalu spolehlivosti pro θ , tzn.

$$P_\theta [(\underline{\theta}(\mathbf{X}, \alpha), \infty) \ni \theta] = 1 - \alpha, \quad P_\theta [(-\infty, \bar{\theta}(\mathbf{X}, \alpha)) \ni \theta] = 1 - \alpha.$$

Test daný pravidlem, že H_0 zamítáme, když

$$(\underline{\theta}(\mathbf{X}, \alpha), \bar{\theta}(\mathbf{X}, \alpha)) \subset (\theta_1, \theta_2), \quad \text{tj. když} \quad \theta_1 \leq \underline{\theta}(\mathbf{X}, \alpha) \ \& \ \theta_2 \geq \bar{\theta}(\mathbf{X}, \alpha), \quad (1.13)$$

má hladinu nejvýše α .

Důkaz. Procedura (1.13) odpovídá provedení dvou jednostranných testů s hladinou významnosti α :

- $H_{0_1} : \theta \leq \theta_1$ proti $H_{1_1} : \theta > \theta_1$, kdy H_{0_1} zamítáme, když $\theta_1 \leq \underline{\theta}(\mathbf{X}, \alpha)$.
- $H_{0_2} : \theta \geq \theta_2$ proti $H_{1_2} : \theta < \theta_2$, kdy H_{0_2} zamítáme, když $\theta_2 \geq \bar{\theta}(\mathbf{X}, \alpha)$.

Potom platnost věty 1.12 vyplývá z věty 1.11.

Q.E.D.

Příklad 1.9. Mějme dva nezávislé náhodné výběry X_1, \dots, X_{n_1} a Y_1, \dots, Y_{n_2} , kde $n_1, n_2 \in \mathbb{N}$. Předpokládejme, že $\text{var } X_1, \text{var } Y_1 \in (0, \infty)$ a označme $E X_1 = \mu_X$ a $E Y_1 = \mu_Y$. Označme $\delta = \mu_X - \mu_Y$. Chceme testovat ekvivalenci středních hodnot. Hypotézy jsou sestaveny jako

$$H_0 : |\delta| \geq \epsilon \quad \text{proti} \quad H_0 : |\delta| < \epsilon,$$

kde $\epsilon > 0$. Je známo, že

$$Z_{n_1, n_2} = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - \delta}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

když $n_1 + n_2 = n \rightarrow \infty$, $n_1/n_2 \rightarrow q \in (0,1)$ a δ je skutečný parametr.

Na základě asymptotické normality Z_{n_1, n_2} vezměme

$$\underline{\theta}(\mathbf{X}, \alpha) = \bar{X}_{n_1} - \bar{Y}_{n_2} - u_{1-\alpha} \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}, \quad \bar{\theta}(\mathbf{X}, \alpha) = \bar{X}_{n_1} - \bar{Y}_{n_2} + u_{1-\alpha} \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}.$$

Označme

$$A = \bar{X}_{n_1} - \bar{Y}_{n_2}, \quad B = u_{1-\alpha} \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}.$$

H_0 zamítáme, když

$$\begin{aligned} -\epsilon \leq A - B &\Leftrightarrow A \geq -\epsilon + B &\Leftrightarrow |\bar{X}_{n_1} - \bar{Y}_{n_2}| \leq \epsilon - u_{1-\alpha} \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}} \\ \epsilon \geq A + B &\Leftrightarrow A \leq \epsilon - B \end{aligned}$$

2. Thurstonovská myšlenka porovnávání podnětů

V různých úlohách je třeba rozhodnout, zda daný produkt působí lépe na lidské smysly, např. je voňavější, chutnější, sladší než produkt jiný. Jiným úkolem může být seřadit několik produktů od nejlepšího po nejhorší na základě smyslového vnímání. Abychom takové úlohy mohli řešit, je nutné se vypořádat s tím, jak popsat či měřit smyslové vnímání. Dále o produktech či věcech, kterými působíme na smysly, mluvíme jako o podnětech.

Příklad 2.1. Příklady konkrétních úloh:

- Porovnání chutí několika druhů mandarinek a seřazení od nejchutnější po nejméně chutné.
- Porovnání různých stylů písma a seřazení od nejúhlednějšího po nejméně úhledné.
- Určení škály bolesti.

Thurstone (1927a) navrhl způsob, jak popsat a měřit působení podnětů na smysly posuzovatele. Uvažoval, že v organismu je proces, jímž organismus reaguje na podněty, rozlišuje je a porovnává, ale ne nutně vždy se stejným závěrem. Následně navrhl pravděpodobnostní model pro problematiku působení podnětů na posuzovatele.

Vezměme jediného posuzovatele a sérii podnětů, které chceme porovnávat, například již zmíněné mandarinky. Teoretická úvaha je taková, že když podnět zapůsobí na posuzovatele, tj. ochutná mandarinku, posuzovatel přiřadí podnětu číselnou hodnotu. Uvažujme, že při opakovaném působení tímž podnětem posuzovatel může pokaždé soudit jinak. Předpokládejme, že čím intenzivnější zkoumanou vlastnost (např. sladkost, vůni) vyvolá podnět u posuzovatele, tím vyšší hodnocení posuzovatel podnětu přiřadí, tj. dá mu vyšší číselnou hodnotu. Dále předpokládejme, že takové hodnocení podnětu lze popsat náhodnou veličinou s normálním rozdělením, označme ji R . Mějme sérii K podnětů a náhodnou veličinu popisující posouzení i -tého podnětu, $i = 1, 2, \dots, K$, označme R_i . Předpokládejme, že $R_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, kde $\sigma_i^2 \in (0, \infty)$, $i = 1, \dots, K$. Pripusťme, že veličiny R_1, \dots, R_K mohou být závislé. V tomto případě doplníme předpoklad, že sdružené rozdělení náhodného vektoru $(R_1, \dots, R_K)^\top$ je mnohorozměrné normální $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

O předpokladu závislosti R_1, \dots, R_K krátce pojednává Thurstone (1927b). V praxi si lze závislost představit tak, že úsudek o jednom podnětu ovlivňuje úsudek o druhém podnětu. Uvedme příklady, jak lze na korelovanost nahlížet. Například, když je posuzovatel v dobré náladě, při hodnocení může dát jednomu i druhému rukopisu vyšší skóre, při špatné náladě nižší skóre (kladná korelovanost). Když posuzovatel ohodnocuje úhlednost rukopisů a jedno písmo působí rozházeně a nečitelně, posuzovatel může druhé písmo hodnotit výrazně lépe, než kdyby hodnotil každé písmo samostatně (záporná korelovanost). Nezávislost lze

v praxi interpretovat tak, že úsudek o jednom podnětu neovlivňuje úsudek o druhém podnětu.

K porovnání podnětů jsou vhodné hodnoty parametrů $\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2$ a korelační koeficienty. Střední hodnoty μ_1, \dots, μ_K vyjadřují pořadí podnětů vzhledem ke zkoumané vlastnosti. Mohou být od sebe různě vzdálené, což vyjadřuje, že některé podněty jsou si „vjemově podobnější“ než jiné. Menší hodnota rozptylu σ_i^2 vyjadřuje, že hodnocení i -tého podnětu jsou podobná, kdežto vysoká hodnota rozptylu značí určitou rozporuplnost podnětu.

Pro porovnání podnětů je třeba odhadnout střední hodnoty, rozptyly a případné korelace. Jak a jaká data získat a jaké odhady použít? Často přirozenější otázkou pro posuzovatele, než jakou hodnotu přiřadit podnětu, je vystavit posuzovatele dvěma podnětům a zeptat se, který by hodnotil výše (je voňavější, chutnější, kyselejší atd.) bez možnosti odpovědi, že jsou shodné. Dohodněme se, že podnět s vyšším hodnocením budeme nazývat lepší. Klíčovou úlohu pro odhady parametrů na základě Thurstone (1927a,b) hrají párová porovnávání.

Mějme dva různé podněty A a B . Z předpokladu normality hodnocení podnětů vyplývá, že

$$R_A - R_B \sim \mathcal{N}(\mu_A - \mu_B, \sigma_A^2 + \sigma_B^2 - 2\rho_{A,B}\sigma_A\sigma_B),$$

kde $\rho_{A,B}$ je korelační koeficient veličin R_A a R_B . Označme $R_{A,B} = R_A - R_B$, $\mu_{A,B} = \mu_A - \mu_B$, $\sigma_{A,B}^2 = \sigma_A^2 + \sigma_B^2 - 2\rho_{A,B}\sigma_A\sigma_B$.

Thurstone (1927a) předpokládá, že kladná realizace $R_{A,B}$ vede k odpovědi posuzovatele „ R_A je lepší než R_B “ a záporná realizace $R_{A,B}$ vede k odpovědi „ R_B je lepší než R_A “. Představujeme si, že posuzovatel ohodnotí oba podněty A a B , tj. realizuje se $(R_A, R_B)^\top$, a jako lepší podnět zvolí ten s vyšší hodnotou realizace. Pravděpodobnost, že posuzovatel uvede, že podnět A je lepší než podnět B , je

$$\mathbb{P}(R_{A,B} > 0) = \mathbb{P}\left(\frac{R_{A,B} - \mu_{A,B}}{\sqrt{\sigma_{A,B}^2}} > -\frac{\mu_{A,B}}{\sqrt{\sigma_{A,B}^2}}\right) = \Phi\left(\frac{\mu_{A,B}}{\sqrt{\sigma_{A,B}^2}}\right), \quad (2.1)$$

kde Φ je distribuční funkce standardizovaného normálního rozdělení. Z (2.1) vyplývá

$$\mu_{A,B} = \Phi^{-1}[\mathbb{P}(R_{A,B} > 0)]\sqrt{\sigma_{A,B}^2},$$

tedy

$$\mu_A - \mu_B = \Phi^{-1}[\mathbb{P}(R_A > R_B)]\sqrt{\sigma_A^2 + \sigma_B^2 - 2\rho_{A,B}\sigma_A\sigma_B}. \quad (2.2)$$

Jednoduché odvození rovnice (2.2) je inspirováno prací Tsukida a Gupta (2011).

Označme $\hat{\mathbb{P}}(R_A > R_B)$ odhad $\mathbb{P}(R_A > R_B)$. Odhad dosadíme do rovnice (2.2), získáme tak

$$\mu_A - \mu_B = \Phi^{-1}[\hat{\mathbb{P}}(R_A > R_B)]\sqrt{\sigma_A^2 + \sigma_B^2 - 2\rho_{A,B}\sigma_A\sigma_B}. \quad (2.3)$$

Rovnici (2.3) Thurstone (1927a,b) nazývá *zákon srovnávacího úsudku*. Z (2.3) se odhadují μ_i, σ_i^2 a korelační koeficienty. Teoretická rovnice (2.2) a její odvození není v pracích Thurstone (1927a,b) uvedeno, ale zřejmě odsud vychází rovnice (2.3). Proto si dovolme také o rovnici (2.2) přemýšlet jako o *teoretickém zákonu srovnávacího úsudku*.

Krátce se zabývejme odhadem $P(R_A > R_B)$ v praxi. Poznamenejme, že veličina $\mathbf{1}(R_A > R_B)$ má alternativní rozdělení s parametrem $p = P(R_A > R_B)$. V případě nezávislých stejně rozdělených pozorování z rozdělení $\mathbf{1}(R_A > R_B)$ je nestranným a konzistentním odhadem $P(R_A > R_B)$ výběrový průměr. Thurstone (1927b) nastiňuje dva případy (a) a (b) získání dat a odhadu $P(R_A > R_B)$.

- (a) Data jsou získána tak, že jediného posuzovatele opakovaně vystavujeme podnětům A a B , posuzovatel hodnotí, který podnět je lepší bez možnosti volby, že jsou stejné. Odhadem $P(R_A > R_B)$ je relativní četnost odpovědi „podnět A je lepší než podnět B .“
- (b) Druhým případem je skupina posuzovatelů. Zde je nutné zajistit, že skupina posuzovatelů je homogenní, tj. lze předpokládat, že jejich hodnocení podnětů jsou stejně rozdělená. Každý posuzovatel hodnotí právě jednou každý pár podnětů. Odhadem $P(R_A > R_B)$ je relativní četnost odpovědi „podnět A je lepší než podnět B .“
- (c) Další možností získání dat, kterou Thurstone (1927b) neuvádí, jsou opakovaná hodnocení více posuzovatelů, tj. každý posuzovatel opakovaně hodnotí každý pár podnětů.

Ve všech třech případech (a)–(c) je nutné pečlivě navrhnout experimenty, avšak návrhy experimentů se v této práci nebudeme zabývat.

Uvažujme K podnětů, které chceme porovnat. Abychom mohli podněty porovnat, potřebujeme odhadnout parametry $\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2$ a korelace $\rho_{i,j}, i, j \in \{1, \dots, K\}, i < j$, což je celkově $K + K + K(K - 1)/2$ parametrů. Zabývejme se odhadováním parametrů.

Pro každou dvojici různých podnětů $\{i, j\}, i, j \in 1, \dots, n, i \neq j$ existují dvě uspořádané dvojice (i, j) a (j, i) . Rovnice (2.2) příslušná uspořádané dvojici (i, j) má tvar

$$\mu_i - \mu_j = \Phi^{-1}[P(R_i > R_j)]\sqrt{\sigma_i^2 + \sigma_j^2 - 2\rho_{i,j}\sigma_i\sigma_j}$$

a je až na opačné znaménko shodná s rovnicí (2.2) příslušné uspořádané dvojici podnětů (j, i) . Proto ke každé dvojici podnětů $\{i, j\}$ má smysl uvažovat pouze jednu rovnici typu (2.2). Dále budeme pro jednoduchost zápisu bez újmy na obecnosti uvažovat rovnici (2.2) příslušnou uspořádané dvojici $(\min(i, j), \max(i, j))$.

Při počtu podnětů K máme $K(K - 1)/2$ různých dvojic podnětů. Tedy pro $K(K - 1)/2$ dvojic podnětů máme soustavu $K(K - 1)/2$ teoretických rovnic typu (2.2). Odhadneme-li pravděpodobnosti $P(R_i > R_j), i, j \in \{1, \dots, K\}, i < j$, máme soustavu $K(K - 1)/2$ rovnic typu (2.3) s neznámými parametry $\mu_1, \dots, \mu_K, \sigma_1^2, \dots, \sigma_K^2$ a $\rho_{i,j}, i, j \in \{1, \dots, K\}, i < j$. Thurstone (1927a) navrhl získat odhady parametrů středních hodnot, rozptylů a korelací ze soustavy rovnic typu (2.3). Zde vyvstávají problémy, se kterými se musíme vypořádat.

Prvním problémem je, že pro každých $K \geq 2$ podnětů máme $2K + K(K - 1)/2$ parametrů, které chceme odhadnout, což je více než $K(K - 1)/2$ rovnic typu (2.3). Thurstone (1927b) se proto omezuje na případy, kdy pro každé $i, j \in \{1, \dots, K\}, i < j$, lze uvažovat $\rho_{i,j} = \rho$. V tomto případě máme $2K + 1$ parametrů a pro $K \geq 5$ máme více rovnic typu (2.3) než neznámých parametrů. Omezení

$\rho_{i,j} = \rho$, tj. korelační koeficienty mezi všemi podněty jsou stejné, je velmi striktní a značně nerealistické hlavně pro větší počet podnětů, a otevírá se zde prostor pro návrh méně striktního omezení pro korelační koeficienty. Například, když podněty seřadíme vzestupně podle středních hodnot a označíme $1, \dots, K$, můžeme zkusit uvažovat korelační koeficienty $\rho_1, \dots, \rho_{K-1}$, kde i -té a j -té posouzení podnětů má korelaci $\rho_{|i-j|}$.

Dalším problémem je, že kvůli záměně $\widehat{P}(R_i > R_j)$ za $P(R_i > R_j)$ přesné řešení soustavy rovnic (2.3) nemusí existovat, což nás navádí na nějaké přibližné řešení.

Zákon srovnávacího úsudku v obecném případě zachycuje rovnice (2.3). Thurstone (1927b) se kvůli velkému počtu neznámých parametrů omezuje na model, kdy lze uvažovat všechny korelační koeficienty rovny konstantě. Nyní se podívejme na jednodušší varianty Thurstonova modelu. Přidejme předpoklad nezávislosti mezi hodnoceními podnětů R_1, \dots, R_K . Rovnice (2.3) se zjednoduší na

$$\mu_A - \mu_B = \Phi^{-1}[\widehat{P}(R_A > R_B)]\sqrt{\sigma_A^2 + \sigma_B^2}$$

a počet neznámých parametrů se sníží na $2K$. Pro $K \geq 5$ je počet neznámých parametrů větší nebo roven počtu rovnic.

Nejjednodušší případ, který Thurstone (1927b) uvádí, je model s nezávislými hodnoceními a stejnými rozptyly $\sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$. Zde je $K + 1$ parametrů. Odhadněme parametry v tomto nejjednodušším případě. Rovnice (2.3) má tvar

$$\mu_A - \mu_B = \Phi^{-1}[\widehat{P}(R_A > R_B)]\sqrt{2}\sigma. \quad (2.4)$$

Uvažujeme-li rovnici (2.4) pro každý pár podnětů, získáváme přeurčenou soustavu rovnic. Označme $z_{i,j} = \Phi^{-1}[\widehat{P}(R_i > R_j)]$, $i < j$. Přeurčenou soustavu rovnic (2.4) lze zapsat ve tvaru

$$\begin{array}{c} \text{K} \\ \left\{ \begin{array}{c} \left(\begin{array}{cccccc} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \ddots & \ddots & 0 \\ 1 & 0 & 0 & \cdots & \cdots & 0 & -1 \end{array} \right) \\ \left(\begin{array}{cccccc} 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \ddots & \ddots & & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \ddots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 & -1 \end{array} \right) \\ \left(\begin{array}{cccccc} \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{array} \right) \end{array} \right. \end{array} \left(\begin{array}{c} \mu_1/\sigma \\ \mu_2/\sigma \\ \vdots \\ \mu_K/\sigma \end{array} \right) = \sqrt{2} \left(\begin{array}{c} z_{1,2} \\ z_{1,3} \\ \vdots \\ z_{K-1,K} \end{array} \right), \quad (2.5)$$

což je soustava lineárních rovnic $\mathbf{Ax} = \mathbf{b}$. Matice \mathbf{A} nemá plnou sloupcovou hodnotu, má hodnotu $K - 1$, což lze vidět z toho, že první blok matice \mathbf{A} je singulární matice, ale po vynechání prvního sloupce je první blok regulární čtvercová matice.

Označme $\mu_i/\sigma = \gamma_i$. Pro řešení soustavy $\mathbf{Ax} = \mathbf{b}$ zvolme metodu nejmenších čtverců. Nalezneme $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_K)^\top$ jako

$$\hat{\gamma} = \arg \min_{\gamma} \sum_{i=1}^{K-1} \sum_{j=i+1}^K [(\gamma_i - \gamma_j) - \sqrt{2} z_{i,j}]^2.$$

Funkci

$$\sum_{i=1}^{K-1} \sum_{j=i+1}^K [(\gamma_i - \gamma_j) - \sqrt{2} z_{i,j}]^2. \quad (2.6)$$

označme $f(\gamma)$.

Řešení $\hat{\gamma}$ není jednoznačné, protože matice \mathbf{A} nemá plnou sloupcovou hodnot. Zkusme zjistit, v jakých bodech funkce $f(\gamma)$ nabývá minima, tedy jaké jsou možné odhady metodou nejmenších čtverců. Označme $\sqrt{2} z_{i,j} = z_{i,j}^*$, $i < j$. Podívejme se na parciální derivace funkce $f(\gamma)$.

$$\frac{\partial f(\gamma)}{\partial \gamma_i} = \sum_{\{j; i < j\}} 2(\gamma_i - \gamma_j - z_{i,j}^*) - \sum_{\{j; j < i\}} 2(\gamma_j - \gamma_i - z_{j,i}^*).$$

Označme $z_{j,i}^* = -z_{i,j}^*$, $i < j$ a zmiňme, že $z_{j,i}^* = \sqrt{2} \Phi^{-1}[1 - \hat{\mathbf{P}}(R_i > R_j)]$, $i < j$. Položme derivaci rovnou nule a již použijeme značení pro odhady.

$$\begin{aligned} \sum_{\{j; i < j\}} 2(\hat{\gamma}_i - \hat{\gamma}_j - z_{i,j}^*) - \sum_{\{j; i > j\}} 2(\hat{\gamma}_j - \hat{\gamma}_i + z_{i,j}^*) &= 0 \\ (K-1)\hat{\gamma}_i - \sum_{\{j; i < j\}} (\hat{\gamma}_j + z_{i,j}^*) - \sum_{\{j; i > j\}} (\hat{\gamma}_j + z_{i,j}^*) &= 0 \\ K \hat{\gamma}_i - \sum_{j=1}^K \hat{\gamma}_j - \sum_{\{j; i \neq j\}} z_{i,j}^* &= 0 \\ \frac{1}{K} \sum_{j=1}^K \hat{\gamma}_j + \frac{1}{K} \sum_{\{j; i \neq j\}} z_{i,j}^* &= \hat{\gamma}_i \end{aligned}$$

Položme $\bar{\gamma}_K = K^{-1} \sum_{j=1}^K \hat{\gamma}_j$ a $G_i = K^{-1} \sum_{\{j; i \neq j\}} z_{i,j}^*$ pro $i = 1, \dots, K$. Získáváme soustavu rovnic

$$\begin{aligned} \hat{\gamma}_1 &= \bar{\gamma}_K + G_1 \\ &\vdots \\ \hat{\gamma}_K &= \bar{\gamma}_K + G_K \end{aligned} \quad (2.7)$$

$$\sum_{j=1}^K \hat{\gamma}_j = K \bar{\gamma}_K$$

Všimněme si, že $\sum_{j=1}^K G_j = 0$, protože v součtu $\sum_{j=1}^K G_j$ se každý člen $z_{i,j}^*$, $i, j \in \{1, \dots, n\}$, $i \neq j$, objeví právě jednou, přičemž platí $z_{i,j}^* + z_{j,i}^* = 0$. Tato soustava má nekonečně mnoho řešení. Je proto třeba přidat nějakou podmínku, abychom získali jednoznačné řešení. Vyberme jeden z podnětů, řekněme $s \in \{1, \dots, K\}$ a zvolme pevně hodnotu $\hat{\mu}_s$ (například 0). Zvolme pevně $\hat{\sigma}^2$ (například 1). Odhady středních hodnot a rozptylů zbylých podnětů budou vztaheny k této počáteční volbě. Máme $\hat{\gamma}_s = \hat{\mu}_s/\hat{\sigma}$. Z (2.7) vyjádříme $\bar{\gamma}_K = \hat{\gamma}_s - G_s$ a $\hat{\gamma}_i = \hat{\gamma}_s - G_s + G_i$ pro $i \neq s$. Získáváme $(\hat{\gamma}_1, \dots, \hat{\gamma}_K)^\top$. Z rovností $\hat{\mu}_i = \hat{\gamma}_i \hat{\sigma}$ dostaneme $(\hat{\mu}_1, \dots, \hat{\mu}_K, \hat{\sigma})$.

Pokud bychom například chtěli, aby nejmenší odhad střední hodnoty byl roven nule, stačí všechny odhady $\hat{\mu}_i$ posunout o vhodnou konstantu c tak, aby

$$\min_i \{(\hat{\mu}_i + c)\} = 0.$$

Stále to jsou odhady získané metodou nejmenších čtverců, neboť v minimalizované funkci $f(\boldsymbol{\gamma})$ z (2.6) vystupují pouze rozdíly γ_i a γ_j .

V modelech, kde nepředpokládáme rovnost rozptylů $\sigma_1^2, \dots, \sigma_K^2$, a třeba ještě uvažujeme korelace mezi hodnoceními, je získání odhadů složitější. Navíc musíme uvažovat omezení $\sigma_1^2, \dots, \sigma_K^2 \in (0, \infty)$ a $\rho \in (-1, 1)$.

Shrnutí

L.L.Thurstone ve svých pracích Thurstone (1927a,b) uvádí přístup, jak měřit smyslové vnímání. Předpokládá, že když podnět zapůsobí na posuzovatele, posuzovatel mu přiřadí numerickou hodnotu, a že tento proces posouzení (vnímání) podnětu se řídí normálním rozdělením. Příslušné střední hodnoty, rozptyly a korelační koeficienty dávají možnost porovnat vnímání různých podnětů. Pořadí středních hodnot lze chápat jako pořadí podnětů a vzdálenosti mezi středními hodnotami zachycují, že některé podněty působí podobněji než jiné. Rozptyly vypovídají o variabilitě vnímání. Na základě párového porovnávání podnětů Thurstone (1927a) uvádí rovnici (2.3), kterou nazývá *zákon srovnávacího úsudku*. Rovnice (2.3) vychází z teoretické rovnice (2.2), které říkáme *teoretický zákon srovnávacího úsudku*. Takové rovnice existují pro každý pár podnětů, z čehož vzniká soustava rovnic, která umožňuje odhadovat příslušné parametry středních hodnot, rozptylů a korelačních koeficientů.

3. Odhad vektoru priorit

Lidé se často potýkají s rozhodováním mezi několika možnostmi. Rozhodují se o běžných věcech, jako je například výběr z n nabízených druhů sýrů, typů triček či jiných potravin nebo oblečení a mnoho dalšího. Rozhodují se také o věcech s dalekosáhlejšími důsledky, jako je například výběr vysoké školy či zaměstnání.

Jednu z možností, která může pomoci s výběrem mezi n objekty z nějakého souboru, popsal Saaty (1977). Předpokládejme, že pro soubor s $n \in \mathbb{N}$ objekty, mezi kterými se rozhodujeme, existuje neznámý vektor $\mathbf{w} = (w_1, \dots, w_n)^\top$, který nazýváme vektorem priorit. Prvek w_i , $i \in \{1, \dots, n\}$, vyjadřuje důležitost prvku i mezi objekty v souboru. Předpokládejme, že $w_1 > 0, \dots, w_n > 0$, a navíc $\sum_{i=1}^n w_i = 1$. Čím má prvek (váha) w_i vyšší hodnotu, tím je důležitější a říkáme, že prvek i je (w_i/w_j) krát důležitější než prvek j . Pokud neznámý vektor priorit odhadneme, můžeme se o výběru ze souboru n objektů rozhodnout na základě odhadnutých vah $\hat{w}_1, \dots, \hat{w}_n$.

Odhadování vektoru priorit je základní úloha, která se uplatňuje i při řešení složitějších problémů. Například když rodiče volí základní školu pro své dítě, zaměřují se na kritéria jako jsou úroveň školy, zaměření školy, přístup k dětem apod. Nebo když na důležitou příležitost vybíráme jeden z n druhů kávy a záleží nám na vůni, chuti a ceně. Tyto problémy patří mezi úlohy rozhodování s více kritérii. Saaty (1977) navrhl metodu sloužící k porovnávání možností s více kritérii. Tato metoda je založena právě na odhadu vektoru priorit a Saaty (1977) ji nazval analytický hierarchický proces. Ačkoliv analytický hierarchický proces je velmi zajímavý, v této práci se jím zabývat nebudeme a zůstaneme u základní úlohy odhadování vektoru priorit.

Odhad vektoru priorit je důležitá úloha, která má široké uplatnění. V této kapitole popíšeme odhad vektoru priorit podle Saaty (1977), nový návrh odhadu vektoru priorit také založený na Saatyho myšlence a dva zmíněné odhady porovnáme v simulační studii.

3.1 Saatyho odhad vektoru priorit a další možnost odhadu

Mějme $n \geq 3$ objektů, které chceme porovnat a předpokládejme, že existuje příslušný vektor priorit $\mathbf{w} = (w_1, \dots, w_n)^\top$, $w_1 > 0, \dots, w_n > 0$ a $\sum_{i=1}^n w_i = 1$. Saaty (1977) popsal odhad vektoru priorit $\mathbf{w} = (w_1, \dots, w_n)^\top$, který je založený na párových porovnáváních. Saaty (1977) reprezentuje párová porovnávání maticí $n \times n$ tvaru

$$\mathbf{R} = \mathbf{w} (\mathbf{w}^{-1})^\top = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix}, \quad (3.1)$$

kde značíme $\mathbf{w}^{-1} = (w_1^{-1}, \dots, w_n^{-1})$. Prvky w_i/w_j , $i, j \in \{1, \dots, n\}$ matice \mathbf{R} vyjadřují, kolikrát je prvek i důležitější (resp. preferovanější, lepší, větší, těžší, chutnější atd.) než prvek j .

Poukažme na některé vlastnosti matice \mathbf{R} , ze kterých budeme v odhadování vektoru priorit vycházet. Matice $\mathbf{R} = \mathbf{w}(\mathbf{w}^{-1})^\top$ má hodnotu jedna. Jediné nenulové vlastní číslo matice \mathbf{R} je $\lambda_{\max} = n$, což plyne z faktu, že součet vlastních čísel matice je roven její stopě. Dále je ihned vidět, že $\mathbf{R}\mathbf{w} = n\mathbf{w}$ a také, že $\mathbf{R}^\top \mathbf{w}^{-1} = n\mathbf{w}^{-1}$.

Nyní se zaměříme na získávání dat pro odhad vektoru priorit. Pozorovateli pokládáme otázku: „Kolikrát je objekt i důležitější než objekt j ?“ Několik poznámek k praktickému provedení uvádíme na stránce 25. Celkově máme $n(n-1)/2$ různých dvojic objektů. Pro $i \neq j$ vybereme uspořádanou dvojici (i, j) (volbu pořadí $\{i, j\}$ v uspořádané dvojici zde neřešíme). V každé uspořádané dvojici objektů se ptáme: Kolikrát je objekt i důležitější než objekt j ? Odpověď označme jako r_{ij} a definujme $r_{ji} = 1/r_{ij}$ a $r_{i,i} = 1$ pro $i = j$. Hodnoty $r_{ij}, i, j \in \{1, \dots, n\}$ jsou pro nás odhady w_i/w_j . Definujme matici dat $(r_{i,j})_{i,j \in \{1, \dots, n\}}$ a označme ji \mathbf{R} . Dále o matici \mathbf{R} tvořené daty budeme mluvit jako o matici \mathbf{R} s chybami. Symbolem \mathbf{R} značíme i teoretickou matici (3.1). Aby nedošlo k záměně, o teoretické matici (3.1) budeme mluvit jako o matici \mathbf{R} bez chyb.

Matice $\mathbf{R} = (r_{i,j})_{i,j \in \{1, \dots, n\}}$ s chybami má všechny prvky kladné. Z Perronovy–Frobeniovy věty vyplývá, že existuje reálné kladné vlastní číslo λ_{\max} matice \mathbf{R} , které splňuje $|\lambda_{\max}| \geq |\lambda|$ pro libovolné vlastní číslo λ matice \mathbf{R} . Číslo λ_{\max} se nazývá Perronovo–Frobeniovo vlastní číslo. Také platí, že vlastní vektor příslušný λ_{\max} je reálný a má komponenty se stejným znaménkem. Saaty (1977) navrhl odhad \mathbf{w} jako řešení rovnice

$$\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}.$$

Za podmínek $\hat{\mathbf{w}} \geq 0$ a $\sum_{i=1}^n \hat{w}_i = 1$ je řešení $\hat{\mathbf{w}}$ jednoznačné.

Pokud místo $\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}$ řešíme $\mathbf{R}^\top \hat{\mathbf{v}} = \lambda_{\max}\hat{\mathbf{v}}$ (matice \mathbf{R} a \mathbf{R}^\top mají stejná vlastní čísla), obecně neplatí $\hat{\mathbf{v}}^{-1} \propto \hat{\mathbf{w}}$. Dokonce ani pořadí prvků $\{\hat{w}_1, \dots, \hat{w}_n\}$ a $\{1/\hat{v}_1, \dots, 1/\hat{v}_n\}$ se nemusí shodovat, na tuto „nesymetričnost“ poukazuje například Johnson a kol. (1979).

Příklad 3.1. (Ilustrativní) Mějme pozorovanou matici

$$\mathbf{R} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 1/2 \\ 1/3 & 1/2 & 1 \end{bmatrix}.$$

Zde máme $\lambda_{\max} \doteq 3.560$, normované vektory $\hat{\mathbf{w}} \doteq (0.469, 0.393, 0.138)^\top$ a $\hat{\mathbf{v}}^{-1} \doteq (0.373, 0.356, 0.271)^\top$. Vidíme, že $\hat{\mathbf{w}}$ není úměrný $\hat{\mathbf{v}}^{-1}$.

Probrali jsme Saatyho odhad vektoru priorit. Nyní uveďme jiný způsob odhadu. Zahrňme do teoretické představy o lidském chování myšlenku, že hodnocení objektu i proti objektu j není to samé jako hodnocení j proti i . Předpokládejme, že matice \mathbf{R} bez chyb má tvar

$$\mathbf{R} = \mathbf{u}\mathbf{v}^\top = \begin{bmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & u_2v_2 & \cdots & u_2v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nv_1 & u_nv_2 & \cdots & u_nv_n \end{bmatrix},$$

kde $\mathbf{u} = (u_1, \dots, u_n)^\top$ a $\mathbf{v} = (v_1, \dots, v_n)^\top$. Předpokládejme, že pro každé $i \in \{1, \dots, n\}$ je $u_i > 0$ a $v_i > 0$. Hodnota u_i vyjadřuje preferenci objektu i a v_i vyjadřuje „nepreferenci“ objektu i .

Matice $\mathbf{R} = \mathbf{w}(\mathbf{w}^{-1})^\top$ je speciálním případem $\mathbf{R} = \mathbf{u}\mathbf{v}^\top$, kde $u_i = \ell w_i$ a $v_i = 1/(\ell w_i)$ pro každé $i \in \{1, \dots, n\}$ a nějakou konstantu $\ell > 0$.

Zabývejme se odhady $u_i, v_i, i \in \{1, \dots, n\}$. Uvedme odhad založený na singulárním rozkladu matic a metodě nejmenších čtverců. Předpokládejme nyní, že matice \mathbf{R} je bez chyb, tedy $\mathbf{R} = \mathbf{u}\mathbf{v}^\top$. Dyadický rozvoj matice \mathbf{R} je $\mathbf{R} = \sigma\mathbf{p}\mathbf{q}^\top$, tedy $\mathbf{u} \propto \mathbf{p}$ a $\mathbf{v} \propto \mathbf{q}$.

Nyní uvažujme, že matice \mathbf{R} obsahuje chyby, tedy že matice $\mathbf{R} = (r_{ij})_{i,j=1,\dots,n}$ je tvořena daty stejným způsobem jako výše. Matici \mathbf{R} lze zapsat ve tvaru

$$\mathbf{R} = \sum_{i=1}^r \sigma_i \mathbf{p}_i \mathbf{q}_i^\top, \quad (3.2)$$

kde $1 \leq r \leq n$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$, $\{\mathbf{p}_1, \dots, \mathbf{p}_r\}$ a $\{\mathbf{q}_1, \dots, \mathbf{q}_r\}$ jsou dvě množiny reálných ortonormálních vektorů.

Z Eckartovy–Youngovy–Mirského věty plyne, že matice $\sigma_1 \mathbf{p}_1 \mathbf{q}_1^\top$ je nejlepší aproximací matice \mathbf{R} hodnosti jedna. Platí

$$\arg \min_{\mathbf{B}; \text{rank}(\mathbf{B})=1} \|\mathbf{R} - \mathbf{B}\|_{\mathbb{F}}^2 = \arg \min_{\mathbf{B}; \text{rank}(\mathbf{B})=1} \sum_i \sum_j (r_{ij} - b_{ij})^2 = \sigma_1 \mathbf{p}_1 \mathbf{q}_1^\top,$$

kde $\|\cdot\|_{\mathbb{F}}$ je Frobeniova norma. Uvědomme si, že podmínka $\text{rank}(\mathbf{B}) = 1$ je ekvivalentní podmínce $\mathbf{B} = \mathbf{x}\mathbf{y}^\top$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Jinak řečeno, matice $\sigma_1 \mathbf{p}_1 \mathbf{q}_1^\top$ je řešením úlohy nejmenších čtverců

$$\min_{\mathbf{B}=\mathbf{x}\mathbf{y}^\top; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n} \sum_i \sum_j (r_{ij} - x_i y_j)^2. \quad (3.3)$$

Vektor \mathbf{p}_1 je vlastní vektor příslušný Perronovu–Frobeniovu vlastnímu číslu matice $\mathbf{R}\mathbf{R}^\top$ a \mathbf{q}_1 je vlastní vektor příslušný Perronovu–Frobeniovu vlastnímu číslu matice $\mathbf{R}^\top\mathbf{R}$. Z Perronovy–Frobeniovy věty plyne, že složky vektoru \mathbf{p}_1 , resp. \mathbf{q}_1 , mají stejné znaménko. V dyadickém rozvoji matice \mathbf{R} (3.2) mají vektory \mathbf{p}_1 a \mathbf{q}_1 buď oba kladné složky nebo oba záporné složky. Kdyby složky vektorů \mathbf{p}_1 a \mathbf{q}_1 měly opačná znaménka, matice $\sigma_1 \mathbf{p}_1 \mathbf{q}_1^\top$ by měla všechny prvky záporné a nebyla by řešením (3.3). Například v bodě $-\sigma_1 \mathbf{p}_1 \mathbf{q}_1^\top$ by minimalizovaná funkce (3.3) by měla menší hodnotu, neboť $r_{i,j} > 0$ pro všechna i, j .

Bez újmy na obecnosti předpokládejme, že \mathbf{p}_1 a \mathbf{q}_1 mají všechny složky kladné. Vezměme odhady $\hat{\mathbf{u}} \propto \mathbf{p}_1$ a $\hat{\mathbf{v}} \propto \mathbf{q}_1$. Vektory \mathbf{p}_1 a \mathbf{q}_1 splňují $\sum_{i=1}^n p_{1,i}^2 = \sum_{i=1}^n q_{1,i}^2 = 1$, ale spíše požadujeme, aby $\sum_{i=1}^n \hat{u}_i = \sum_{i=1}^n \hat{v}_i = 1$.

Věnujme se singulárnímu rozkladu matice \mathbf{R}^\top . Předpokládejme nejprve, že matice $\mathbf{R}^\top = \mathbf{v}\mathbf{u}^\top$ neobsahuje chyby. Dyadický rozvoj matice \mathbf{R}^\top je $\mathbf{R}^\top = \sigma\mathbf{q}\mathbf{p}^\top$ a platí $\mathbf{u} \propto \mathbf{p}$ a $\mathbf{v} \propto \mathbf{q}$. Nyní uvažujme matici \mathbf{R}^\top s chybami. Dyadický rozvoj matice \mathbf{R}^\top je

$$\mathbf{R}^\top = \sum_{i=1}^r \sigma_i \mathbf{q}_i \mathbf{p}_i^\top.$$

Zde je matice $\sigma_1 \mathbf{q}_1 \mathbf{p}_1^\top$ nejlepší aproximací matice \mathbf{R}^\top hodnosti jedna a zároveň řešením úlohy nejmenších čtverců

$$\min_{\mathbf{B}=\mathbf{x}\mathbf{y}^\top; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n} \sum_i \sum_j (r_{ji} - x_i y_j)^2.$$

Odhadujeme \mathbf{u}, \mathbf{v} opět jako $\hat{\mathbf{u}} \propto \mathbf{p}_1$ a $\hat{\mathbf{v}} \propto \mathbf{q}_1$.

Předpokládejme teď, že matice \mathbf{R} bez chyb má speciální tvar $\mathbf{R} = \mathbf{w}(\mathbf{w}^{-1})^\top$. Na základě úvah výše zvolme odhad $\hat{w}_i \propto \sqrt{\hat{u}_i/\hat{v}_i}$, $i \in \{1, \dots, n\}$. Smysluplnost tohoto odhadu vychází z úvahy, že v modelu $\mathbf{R} = \mathbf{u}\mathbf{v}^\top = \mathbf{w}(\mathbf{w}^{-1})^\top$ je $\mathbf{u} = \ell\mathbf{w}$ a $\mathbf{v} = 1/(\ell\mathbf{w})$, a tedy $u_i/v_i = w_i^2$.

Příklad 3.2. Porovnejme řešení $\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}$ s $\hat{w}_i = \sqrt{\hat{u}_i/\hat{v}_i}$ na příkladu, který je převzatý z Saaty (1977), příklad 1. Bylo vybráno šest měst, a to Káhira, Tokio, Chicaga, San Francisco, Londýn a Montreal. Posuzovatel odhadoval, kolikrát je město i proti městu j vzdáleno od Philadelphie. V matici (3.4) jsou (v uvedeném pořadí měst) odhady pozorovatele.

$$R = \begin{pmatrix} 1 & 1/3 & 8 & 3 & 3 & 7 \\ 3 & 1 & 9 & 3 & 3 & 9 \\ 1/8 & 1/9 & 1 & 1/6 & 1/5 & 2 \\ 1/3 & 1/3 & 6 & 1 & 1/3 & 6 \\ 1/3 & 1/3 & 5 & 3 & 1 & 6 \\ 1/7 & 1/9 & 1/2 & 1/6 & 1/6 & 1 \end{pmatrix}. \quad (3.4)$$

V tabulce 3.1 jsme pro porovnání uvedli skutečné relativní vzdálenosti, Saatyho odhad $\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}$ a odhady $\hat{w}_i = \sqrt{\hat{u}_i/\hat{v}_i}$. Chyby odhadů jsou měřeny v L_2 -normě a v daném pořadí jsou 0.046 a 0.066. Vzhledem k dané míře chyby je Saatyho přístup lepší.

město	skutečnost	$\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}$	$\sqrt{\hat{u}_i/\hat{v}_i}$
Káhira	0.278	0.262	0.228
Tokio	0.361	0.398	0.395
Chicaga	0.032	0.033	0.043
San Francisco	0.132	0.116	0.143
Londýn	0.177	0.164	0.160
Montreal	0.019	0.027	0.031
chyba odhadů		0.046	0.066

Pozn.: Chyba odhadů je měřena v L_2 -normě.

Tabulka 3.1: Porovnání skutečných relativních vzdáleností měst od Philadelphie a daných odhadů v příkladu 3.2.

Příklad 3.3. Ukažme ještě jeden příklad ze Saaty (1977), příklad 4. Úkolem pozorovatele bylo porovnat pět předmětů podle váhy. Nejprve pozorovatel zvedal jeden předmět po druhém, aby získal představu o rozsahu hmotností. Potom pozorovatel porovnával předměty v párech tak, že předměty postupně zvedal v pravé ruce. V matici (3.5) jsou odhady posuzovatele, kolikrát je jeden předmět těžší než druhý:

$$R = \begin{pmatrix} 1 & 1/5 & 1/3 & 1/4 & 4 \\ 5 & 1 & 2 & 2 & 8 \\ 3 & 1/2 & 1 & 1/2 & 4 \\ 4 & 1/2 & 2 & 1 & 7 \\ 1/4 & 1/8 & 1/4 & 1/7 & 1 \end{pmatrix}. \quad (3.5)$$

věc	skutečnost	$\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}$	$\sqrt{\hat{u}_i/\hat{v}_i}$
rádio	0.10	0.088	0.104
psací stroj	0.39	0.405	0.399
velký příruč. kufřík	0.20	0.183	0.183
projektor	0.27	0.286	0.273
malý příruč. kufřík	0.04	0.038	0.041
chyba odhadů		0.030	0.020

Pozn.: Chyba odhadů je měřena v L_2 -normě.

Tabulka 3.2: Porovnání skutečných relativních vah vybraných předmětů a daných odhadů v příkladu 3.3.

V tabulce 3.2 jsme pro porovnání uvedli skutečné relativní váhy předmětů, Saatyho odhad $\mathbf{R}\hat{\mathbf{w}} = \lambda_{\max}\hat{\mathbf{w}}$ a odhady $\hat{w}_i = \sqrt{\hat{u}_i/\hat{v}_i}$. Chyby odhadů jsou měřeny v L_2 -normě a v daném pořadí jsou 0.030 a 0.020. V tomto případě je nový odhad lepší.

Několik poznámek k porovnávání v praxi, které jsou převzaté z Saaty (1977):

- V praxi nelze pozorovatele vystavit velkému počtu otázek najednou, získané odpovědi by nemusely být hodnotné. Saaty (1977) uvádí, že rozumný počet porovnávaných objektů n je menší než deset.
- Otázka „Kolikrát je objekt i preferovanější než objekt j ?“ může působit trochu uměle. Saaty (1977) navrhl stupnici porovnávání podnětů i proti j s hodnotami $\{1/9, 1/8, \dots, 1/2, 1, 2, 3, \dots, 9\}$:

1: stejná preference objektů i a j ,

3: trochu vyšší preference objektu i ,

5: silná preference objektu i ,

7: velmi silná preference objektu i ,

9: naprostá preference objektu i ,

hodnoty 2, 4, 6, 8 jsou mezistupně. Hodnoty $1/2, \dots, 1/9$ se použijí analogicky, pokud je objekt j preferovanější než objekt i .

3.2 Simulační studie

Abychom zjistili, jak nový odhad vystihuje realitu, provedli jsme simulační studii. Popíšme jeden běh simulace. Zvolili jsme počet objektů n a vektor priorit w . Vygenerovali jsme matici $(w_i/w_j)_{i,j=1,\dots,n}$. Pro každou dvojici (i,j) , $i < j$ jsme nezávisle vygenerovali $\epsilon_{ij} \sim \mathcal{LN}(0,1)$, kde $\mathcal{LN}(\mu, \sigma^2)$ značí logaritmicko-normální rozdělení s parametry (μ, σ^2) *. Sestavili jsme matici $\mathbf{R} = (r_{ij})_{i,j=1,\dots,n}$, kde

$$r_{ij} = \begin{cases} (w_i/w_j)\epsilon_{ij}, & i < j, \\ 1, & i = j, \\ (w_i/w_j)(1/\epsilon_{ji}), & i > j. \end{cases}$$

*Používáme parametrizaci takovou, že pro náhodnou veličinu $X \sim \mathcal{LN}(\mu, \sigma^2)$ platí $\log(X) \sim \mathcal{N}(\mu, \sigma^2)$.

Z datové matice \mathbf{R} jsme odhadli vektor w novým (tj. $\hat{w}_i \propto \sqrt{\hat{u}_i/\hat{v}_i}$) a Saatyho způsobem. U obou odhadů jsme spočetli chybu měřenou v L_2 -normě, tj. $\sqrt{\sum_{i=1}^n (\hat{w}_i - w_i)^2}$ a rozdíl chyb odhadů.

Pro simulační studii jsme vybrali počty objektů $n = 3, 5, 10$. Zvolili jsme různé vektory priorit $w = (w_1, \dots, w_n)$, a to:

- 1 Všechny váhy stejné: $w_1 = \dots = w_n$.
- 2 Jedna váha $7\times$ větší než ostatní váhy: $w_j \propto 1, j \in \{1, \dots, n-1\}, w_n \propto 7$.
- 3 Přibližně polovina vah $5\times$ větší než zbylé váhy:
 $w_i \propto 1, i \in \{1, \dots, \lfloor (n+1)/2 \rfloor\}$ a $w_i \propto 5, i \in \{\lfloor (n+1)/2 \rfloor + 1, \dots, \lfloor n/2 \rfloor\}$.
- 4 Jedna váha $7\times$ menší než ostatní váhy: $w_j \propto 7, j \in \{1, \dots, n-1\}, w_n \propto 1$.
- 5 Jedna váha menší a jedna váha větší než zbylé váhy:
 $w_1, \dots, w_{n-2} \propto 5, w_{n-1} \propto 1, w_n \propto 9$.
- 6 Váhy s hodnotami rovnoměrně rozloženými mezi 1–9:
 $w_i \propto 1 + (i-1)8/(n-1), i = 1, \dots, n$.

Pro každou variantu 1–6 vektoru priorit w jsme provedli 10^4 běhů simulace. V simulacích, když jsme měli vektory priorit ve tvarech 2–6, jsme v každém běhu simulace vektor priorit w sestavili z náhodně permutovaných vah (w_1, \dots, w_n) . To znamená, že jsme v k -tém běhu $k \in \{1, \dots, 10^4\}$ uvažovali vektor priorit $w^k = (w_{1_k}, \dots, w_{n_k})$, kde $(1_k, \dots, n_k)$ je náhodná permutace $(1, \dots, n)$.

Výstupy simulací jsme shrnuli v tabulce 3.3. V tabulce vidíme průměrné chyby odhadů z 10^4 běhů simulací, které jsou měřeny v L_2 -normě a v závorce jsou doplněny odhady směrodatných odchylek. Ty jsou počítány jako odmocniny z výběrových rozptylů chyb v L_2 -normě. Saatyho odhad průměrně vystihuje skutečný vektor priorit lépe než nový odhad, a to pro každou kombinaci n a w . Zdá se, že Saatyho přístup nejlépe vystihuje skutečnost pro $n = 10$, potom pro $n = 5$ a následně pro $n = 3$. Nový přístup nejlépe vystihuje skutečnost pro $n = 10$, dále nelze říct, zda je nový přístup lepší pro hodnotu $n = 5$ nebo $n = 3$. Když se zaměříme na typy vektorů priorit 1–6, Saatyho i nový přístup nejlépe vystihují realitu v případě 2, tedy když má jedna váha $7\times$ větší hodnotu než zbylé váhy. V tomto případě si jsou průměrné chyby odhadů nejpodobnější. Zdá se, že velikosti chyb obou přístupů korespondují pro stejné kombinace n a w , tím myslíme, že pro stejné kombinace n a w mají oba přístupy větší chyby a pro stejné kombinace n a w menší chyby.

Na závěr této kapitoly řekněme, že odhad vektoru priorit, který T. L. Saaty popsal v práci Saaty (1977), je velmi používaná metoda, která má uplatnění v nejrůznějších oblastech, kde se vyskytují úlohy výběru z více možností, například v senzorické analýze, kterou se v této diplomové práci zabýváme. Je to široké téma, které jsme „natukli“ a ponechali jako velký prostor pro další možnost zkoumání.

		typ zvoleného vektoru priorit \mathbf{w}					
		1		2		3	
n	odhad	průměrné chyby odhadů [L_2 -norma (sm. od.)]					
n = 3	nový	0.261	(0.124)	0.144	(0.104)	0.174	(0.115)
	Saaty	0.225	(0.110)	0.144	(0.105)	0.169	(0.113)
	rozdíl	0.036	(0.014)	0.001	(-0.001)	0.005	(0.002)
n = 5	nový	0.220	(0.087)	0.147	(0.074)	0.200	(0.105)
	Saaty	0.173	(0.066)	0.137	(0.076)	0.168	(0.086)
	rozdíl	0.047	(0.021)	0.009	(-0.003)	0.032	(0.020)
n = 10	nový	0.136	(0.051)	0.133	(0.059)	0.157	(0.067)
	Saaty	0.107	(0.031)	0.107	(0.048)	0.118	(0.043)
	rozdíl	0.028	(0.020)	0.026	(0.011)	0.039	(0.023)
		typ zvoleného vektoru priorit \mathbf{w}					
		4		5		6	
n	odhad	průměrné chyby odhadů [L_2 -norma (sm. od.)]					
n = 3	nový	0.224	(0.137)	0.223	(0.136)	0.214	(0.136)
	Saaty	0.207	(0.127)	0.206	(0.125)	0.199	(0.127)
	rozdíl	0.017	(0.010)	0.016	(0.010)	0.015	(0.009)
n = 5	nový	0.220	(0.091)	0.217	(0.089)	0.211	(0.095)
	Saaty	0.177	(0.075)	0.175	(0.074)	0.173	(0.079)
	rozdíl	0.043	(0.016)	0.042	(0.015)	0.039	(0.017)
n = 10	nový	0.156	(0.052)	0.151	(0.052)	0.154	(0.060)
	Saaty	0.110	(0.033)	0.110	(0.033)	0.114	(0.039)
	rozdíl	0.046	(0.018)	0.041	(0.019)	0.041	(0.021)

Tabulka 3.3: Porovnání nového odhadu vektoru priorit a Saatyho odhadu. V tabulce jsou uvedeny průměrné chyby odhadů (měřeny v L_2 -normě) z 10^4 simulačních běhů a v závorce jsou doplněny odhady směrodatných odchylek. V tabulce jsou také uvedeny rozdíly průměrných chyb a směrodatných odchylek.

4. Normy

V praxi se používají normy, které popisují postupy pro řešení otázek senzorické analýzy. Abychom přiblížili, co se senzorickou analýzou myslí, například citujme Vítová (2011):

„Senzorická analýza je definována jako analytická metoda, při níž se tzv. organoleptické vlastnosti potravin stanoví výhradně lidskými smysly. Její význam spočívá v tom, že postihuje takové kvalitativní ukazatele, které není možno, alespoň ne úplně, charakterizovat přístrojovou technikou.“

Poznámka. Organoleptické vlastnosti jsou vlastnosti vnímatelné smysly. Například mezi organoleptické vlastnosti vody patří barva, chuť, pach a teplota.

V této kapitole rozebereme vybrané normy, které byly vytvořeny pro potřeby senzorické analýzy, konkrétně pro zjištění existence vnímatelného rozdílu nebo podobnosti mezi dvěma různými výrobky. Zaměříme se přitom na běžně používané metody zpracované v normách:

- zkouška duo-trio,
- trojúhelníková zkouška,
- párová porovnávací zkouška.

Nejprve v sekci 4.1 uvedeme základní informace k normám. Ze sekce 4.1 lze pokračovat do libovolné sekce 4.2, 4.3, 4.4, kde jsou rozebrány normy pro jednotlivé zkoušky. V sekcích jednotlivých norem nejprve stručně popíšeme provedení zkoušky a vyhodnocení zkoušky uvedené v normách, potom statistický podklad. Modely v normách jsou si dost podobné, ale pro ucelené vysvětlení každé normy nebudeme shrnovat stejné či podobné části do jednoho textu příslušícímu více normám (kromě některých částí párové zkoušky, které jsou naprosto shodné se zkouškou duo-trio).

4.1 Obecné informace k normám

Zkoušky duo-trio a trojúhelníková se používají pro dva problémy:

- (a) Určení, zda mezi výrobky existuje vnímatelný rozdíl.

Příklad: Výrobce chce zkusit novou recepturu sušenek, ale zavádět další výrobek má smysl pouze tehdy, pokud konzumenti poznají rozdíl od již zaběhlého výrobku.

- (b) Určení, zda je vnímatelný rozdíl mezi výrobky dostatečně malý.

Příklad: Výrobce by rád změnil droždí v receptu na chléb, který je velmi oblíbený u zákazníků. Chce se ujistit, že rozdíl chutí mezi zaběhlým a novým chlebem je minimální.

Párová porovnávací zkouška se používá pro dva podobné problémy. Liší se v tom, že předmětem zájmu není celkový rozdíl mezi výrobky, ale zkoumá se rozdíl v intenzitě určité senzorické vlastnosti výrobků. Zkoumané problémy jsou:

- (a) Určení, zda mezi intenzitami zkoumané vlastnosti výrobků existuje vnímatelný rozdíl.

Příklad: Výrobce chce zvýšit množství soli dávané do pomazánky, aby pomazánka byla chutnější. Zajímá ho, zda spotřebitel pozná, že je nová pomazánka slanější než původní.

- (b) Určení, zda je vnímatelný rozdíl mezi intenzitami zkoumané vlastnosti výrobků dostatečně malý.

Příklad: Při změně výroby se má použít látka, která může způsobit nepříjemnou kyselejší chuť výrobku. Výrobce se chce ujistit, že nový výrobek nebude mít kyselejší chuť než původní.

Pokud se dále v textu budeme odkazovat na problém (a) nebo (b), budeme tím rozumět problémy vztažené jak ke zkoušce duo-trio a trojúhelníkové, tak k párové porovnávací zkoušce. Statistickému testu vztaženému k problému (a) říkáme *test pro rozdíl* a testu vztaženému k problému (b) říkáme *test pro podobnost*.

V normách jsou uvedeny jak pojmy související s provedením zkoušky, tak pojmy používané ve statistickém testování. Vysvětlení pojmů se v normách občas mírně liší ve formulaci. Použijme terminologii užitou v normě ČSN EN ISO 10399 (2010) pro zkoušku duo-trio, uveďme nejdůležitější pojmy a citujme jejich vysvětlení uvedené v normách (v uvozovkách a kurzivou) a u vybraných pojmů uveďme naši statistickou představu o nich (pod vysvětlením v normách):

- α -riziko: „pravděpodobnost závěru, že vnímatelný rozdíl existuje, když žádný neexistuje

POZNÁMKA Toto se také nazývá chyba prvního druhu, hladina významnosti nebo falešně pozitivní výsledek.“

Značení α se v normách používá ve dvou významech:

- Hladina významnosti testu pro rozdíl,
- $(1 - \alpha)$ jako síla testu pro podobnost.

- β -riziko: „pravděpodobnost závěru, že neexistuje vnímatelný rozdíl, když existuje

POZNÁMKA Toto se také nazývá chyba druhého druhu nebo falešně negativní výsledek.“

Značení β se v normách používá ve dvou významech:

- Hladina významnosti testu pro podobnost,
- $(1 - \beta)$ jako síla testu pro rozdíl.

- výrobek: „materiál, který se posuzuje“

- vzorek: „připravená část výrobku, která je předkládána a hodnocena během jednoho sezení zkoušky“

- rozdíl: „situace, ve které může být vzorek rozpoznán na základě sensorických vlastností

POZNÁMKA Podíl hodnocení, během kterých je detekován vnímatelný rozdíl sensorických vlastností mezi dvěma výrobky se označuje symbolem p_d .“

Číslu p_d rozumíme tak, že v populaci posuzovatelů je $p_d \times 100$ % posuzovatelů, kteří jsou schopni rozlišit výrobky a $(1 - p_d) \times 100$ % posuzovatelů, kteří nejsou schopni rozlišit výrobky.

- podobnost: „*situace, při které každý vnímatelný rozdíl mezi vzorky je tak malý, že výrobky mohou být použity zaměnitelně*“
- citlivost: „*obecný termín, který se používá, aby shrnul charakteristiky zkoušky* POZNÁMKA *Ve statistice je citlivost zkoušky definovaná hodnotami α , β a p_d .*“

Citlivostí zkoušky se v normách zpravidla myslí síla testu proti konkrétní alternativě.

Pojem p_d z poznámky u pojmu rozdíl je klíčový. Hodnota p_d závisí na odlišnosti výrobků a na panelu (populaci) posuzovatelů. Hodnotou p_d měříme vnímatelnou odlišnost (či podobnost) výrobků. Diskutujme o správnosti interpretace p_d .

Hodnotu p_d interpretujeme jako podíl posuzovatelů v populaci, kteří jsou vždy schopni rozlišit výrobky a $(1 - p_d)$ jako podíl posuzovatelů, kteří nejsou nikdy schopni rozlišit výrobky. Takový předpoklad spíše není reálný, jak zmiňuje Ennis (1993).

Jiný pohled na p_d by mohl být takový, že jedinec pozná rozdíl s pravděpodobností p_d , tudíž někdy rozdíl pozná, jindy nepozná. Taková možnost se také nezdá být reálná.

Ačkoliv ani jeden z přístupů zcela nevystihuje skutečnost, zvolme první možnost interpretace.

Zkoušky duo-trio, trojúhelníková a párová porovnávací představují tři různé experimenty. Princip všech tří zkoušek je ten, že posuzovatelé musí odpovědět na požadovanou otázku a nesmí přitom odpovědět, že neví. Takové zkoušce se říká zkouška s vynucenou odpovědí.

4.2 Zkouška duo-trio

Ve zkoušce duo-trio se posuzovateli předloží jeden referenční vzorek a následně se předloží další dva vzorky, o nichž posuzovatel ví, že jeden je shodný s referenčním a jeden je odlišný od referenčního, ale neví, který je který. Posuzovatel má za úkol říct, který vzorek je shodný s referenčním a který se od něj liší, a to bez možnosti odpovědět, že neví.

4.2.1 Provedení zkoušky duo-trio, pokyny pro vedoucího zkoušky

Shrňme, jak zkouška probíhá, detailnější popis najdeme v normě ČSN EN ISO 10399 (2010). Zvolí se hladina významnosti testu. V případě testu pro podobnost se zvolí p_{d_0} *. Vybere se počet posuzovatelů n pro zkoušku. Počet posuzovatelů

*Značení p_{d_0} jsme zvolili pro přehlednost, v normě se používá značení p_d .

je možno volit na základě požadované síly testu podle tabulky A.3 v ČSN EN ISO 10399 (2010). Bližší vysvětlení k tabulce A.3 je uvedeno v sekci 4.2.3. Pro zkoušku se vyberou posuzovatelé, kteří mají stejný stupeň kvalifikace.

Ve zkoušce duo-trio jsou uvedeny dvě techniky provedení, z nichž se jedna vybere. Citujme z normy ČSN EN ISO 10399 (2010), jakou techniku vybrat: „*Pokud posuzovatelé výrobek znají (např. kontrolní vzorek z výrobní linky), použije se stálá referenční technika. Pokud je jeden vzorek známější než ostatní, pak se použije vyvážená referenční technika.*“ Označme dva různé výrobky ve zkoušce A a B .

(i) *Stálá referenční technika:*

Uvažujme dvě pořadí vzorků výrobků A a B : (A -referenční, A,B) a (A -referenční, B,A). Posuzovatelé se rozdělí do dvojic. V každé dvojici se náhodně jednomu posuzovateli přiřadí pořadí (A -referenční, A,B) a druhému (A -referenční, B,A).

(ii) *Vyvážená referenční technika:*

Uvažujme čtyři pořadí vzorků (A -referenční, A,B), (A -referenční, B,A), (B -referenční, A,B), (B -referenční, B,A). Posuzovatelé se rozdělí do skupin po čtyřech. V každé skupině se každému posuzovateli náhodně přiřadí právě jedno ze čtyř pořadí vzorků.

Další postup je společný pro stálou i vyváženou referenční techniku. Každému posuzovateli se předloží vzorky zleva doprava v pořadí, které mu bylo přiřazeno. Posuzovatel postupuje v daném pořadí. Poté zaznamená, který vzorek je shodný s referenčním a který se od referenčního liší bez možnosti odpovědi, že neví. Posuzovatelé musí odpovídat nezávisle. Podrobnější instrukce k postupu, který je třeba dodržet, jsou uvedeny v normě ČSN EN ISO 10399 (2010).

Správné odpovědi posuzovatelů se sečtou a jejich počet se označí X .

4.2.2 Vyhodnocení zkoušky duo-trio

Test pro rozdíl

V normě ČSN EN ISO 10399 (2010) se v tabulce A.1 nalezne buňka příslušná zvolené hladině testu α a počtu posuzovatelů n . Číslo v příslušné buňce označme k . Pokud je počet posuzovatelů n vyšší, než je uvedeno v tabulce, číslo k se spočítá jako nejbližší celé číslo vyšší než

$$\frac{n}{2} + u_{1-\alpha} \sqrt{\frac{n}{4}},$$

kde $u_{1-\alpha}$ je $(1 - \alpha)$ -kvantil standardizovaného normálního rozdělení.

Vyhodnocení zkoušky:

- *Počet správných odpovědí je menší než k*

Výsledek zkoušky je ten, že jsme neprokázali vnímatelný rozdíl mezi výrobky na hladině významnosti α .

- *Počet správných odpovědí je alespoň k*

Výsledek zkoušky je ten, že jsme prokázali vnímatelný rozdíl mezi výrobky na hladině významnosti α .

Test pro podobnost

V normě ČSN EN ISO 10399 (2010) se v tabulce A.2 nalezne buňka příslušná zvolené hladině testu β , hodnotě p_{d_0} a počtu posuzovatelů n . Číslo v příslušné buňce označme ℓ . Pokud je počet posuzovatelů n vyšší, než je uvedeno v tabulce A.2, spočítá se číslo Al :

$$Al = \left(\frac{2X}{n} - 1 \right) + 2u_{1-\beta} \sqrt{\frac{nX - X^2}{n^3}},$$

kde X je počet správných odpovědí a $u_{1-\beta}$ je $(1 - \beta)$ -kvantil standardizovaného normálního rozdělení.

Vyhodnocení zkoušky:

- Počet správných odpovědí je nejvýše ℓ , resp. $Al < p_{d_0}$:

Výsledek zkoušky je ten, že jsme prokázali dostatečnou podobnost výrobků na hladině významnosti β .

- Počet správných odpovědí je více než ℓ , resp. $Al \geq p_{d_0}$:

Výsledek zkoušky je ten, že jsme neprokázali dostatečnou podobnost výrobků na hladině významnosti β .

4.2.3 Statistické vysvětlení zkoušky duo-trio

Předpokládá se, že posuzovatel odpoví správně s pravděpodobností p , kterou neznáme. V normách je vztah mezi p a p_d dán jako $p = p_d + (1 - p_d)/2$. Zřejmě se vychází z následující úvahy:

Označme „správná odpověď“ jako SO, „posuzovatel pozná rozdíl“ jako PPR a „posuzovatel nepozná rozdíl“ jako PNR. Věta o podmíněné pravděpodobnosti říká: $p = P(\text{SO}) = P(\text{SO}|\text{PPR}) P(\text{PPR}) + P(\text{SO}|\text{PNR}) P(\text{PNR})$.

Pravděpodobnost, že posuzovatel rozezná výrobky, je $P(\text{PPR}) = p_d$. Předpokládáme, že tento posuzovatel vždy odpoví správně, tedy $P(\text{SO}|\text{PPR}) = 1$. Pravděpodobnost, že posuzovatel nerozezná výrobky, je $P(\text{PNR}) = 1 - p_d$. Předpokládáme, že tento posuzovatel hádá a správně uhádne s pravděpodobností $P(\text{SO}|\text{PNR}) = 1/2$. Odvodili jsme vztah $p = p_d + (1 - p_d)/2$.

Definujme $\mathbf{1}(\text{SO})$ indikátor, že posuzovatel odpověděl správně. Veličina $\mathbf{1}(\text{SO})$ má alternativní rozdělení s parametrem $p = p_d + (1 - p_d)/2$.

Pro $i \in \{1, \dots, n\}$ označme $\mathbf{1}(\text{SO})_i$ indikátor, že i -tý posuzovatel odpověděl správně. Předpokládáme, že $\mathbf{1}(\text{SO})_1, \dots, \mathbf{1}(\text{SO})_n$ tvoří náhodný výběr rozsahu n z alternativního rozdělení s parametrem p . Počet správných odpovědí X z odpovědí n posuzovatelů, se tedy řídí binomickým rozdělením s parametry n a $p = p_d + (1 - p_d)/2$.

Statistické hypotézy pro problémy (a) a (b) jsou sestaveny o parametru p_d , jakožto míře rozdílnosti či podobnosti dvou výrobků (sekce 4.1).

Test pro rozdíl

Uvažujme hladinu testu α .

Norma ČSN EN ISO 10399 (2010) zřejmě vychází z hypotéz o p_d ve tvaru:

$$H_0 : p_d = 0 \quad \text{proti} \quad H_1 : 0 < p_d \leq 1. \quad (4.1)$$

Nulová hypotéza vyjadřuje, že mezi výrobky není vnímatelný rozdíl. Alternativa vyjadřuje, že mezi výrobky vnímatelný rozdíl je.

Test hypotéz (4.1) je ekvivalentní testu hypotéz

$$H_0 : p = 1/2 \quad \text{proti} \quad H_1 : 1/2 < p \leq 1. \quad (4.2)$$

Proti nulové hypotéze svědčí velké hodnoty X .

Pro testování (4.2) se v normě pro malé hodnoty n (přibližně $n < 100$) používá přesný jednostranný test pro binomické rozdělení. Nulovou hypotézu zamítáme, když $X \geq k$, kde

$$k = \min\{j; j \in \mathbb{N}, P(X \geq j | n, p = 1/2) \leq \alpha\}.$$

Hodnota $P(X \geq j | n, p = 1/2)$ je za platnosti H_0 rovna

$$\sum_{i=j}^n \binom{n}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{n-i} = \left(\frac{1}{2}\right)^n \sum_{i=j}^n \binom{n}{i}.$$

Hodnoty k pro vybraná n jsou uvedeny v normě ČSN EN ISO 10399 (2010) v tabulce A.1.

Pro testování (4.2) se v normě pro velké hodnoty n používá asymptotický test založený na CLV. Nulovou hypotézu zamítáme, když

$$\sqrt{n} \frac{X/n - 1/2}{\sqrt{(1/2)(1 - 1/2)}} > u_{1-\alpha}, \quad \text{tj. když} \quad X > \frac{n}{2} + u_{1-\alpha} \sqrt{\frac{n}{4}}.$$

Síla přesného testu hypotéz (4.2) proti alternativě p , kde $p > 1/2$ (neboli hodnota silofunkce v bodě p), je

$$\sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}.$$

Test pro podobnost

Uvažujme hladinu testu β (značení β použijme z důvodu shody s normou). Norma ČSN EN ISO 10399 (2010) zřejmě vychází z hypotéz o p_d ve tvaru:

$$H_0 : p_{d_0} \leq p_d \leq 1 \quad \text{proti} \quad H_1 : 0 \leq p_d < p_{d_0}, \quad (4.3)$$

kde vhodné p_{d_0} musíme zvolit. Hodnotu p_{d_0} volíme tak, abychom mohli říct, že výrobky si jsou dostatečně podobné (pro naše účely), pokud je rozliší méně než $100p_{d_0}\%$ posuzovatelů. Nulová hypotéza vyjadřuje, že výrobky si nejsou dostatečně podobné a alternativa vyjadřuje dostatečnou podobnost. Normy nabízí hodnoty $p_{d_0} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

Test hypotéz (4.3) je ekvivalentní testu hypotéz

$$H_0 : p_0 \leq p \leq 1 \quad \text{proti} \quad H_1 : 1/2 \leq p < p_0, \quad (4.4)$$

kde $p_0 = p_{d_0} + (1 - p_{d_0})/2$. Proti nulové hypotéze svědčí malé hodnoty X .

Pro testování (4.4) se v normě pro malé hodnoty n (přibližně $n < 100$) používá přesný jednostranný test pro binomické rozdělení. Nulovou hypotézu zamítáme, když $X \leq \ell$, kde

$$\ell = \max\{j; j \in \mathbb{N}, \mathbf{P}(X \leq j | n, p = p_0) \leq \beta\}.$$

Hodnota $\mathbf{P}(X \leq j | n, p = p_0)$ je rovna

$$\sum_{i=0}^j \binom{n}{i} p_0^i (1 - p_0)^{n-i}.$$

Hodnoty ℓ pro vybraná n , β a p_{d_0} jsou uvedeny v normě ČSN EN ISO 10399 (2010) v tabulce A.2.

Pro testování (4.4) se v normě pro velké hodnoty n používá asymptotický test založený na CLV. Nulovou hypotézu zamítáme, když

$$\sqrt{n} \frac{X/n - p_0}{\sqrt{(X/n)(1 - X/n)}} < -u_{1-\beta}. \quad (4.5)$$

Dosazením $p_0 = p_{d_0} + (1 - p_{d_0})/2$ do (4.5) a úpravou získáme tvar pro zamítnutí H_0 , který je uvedený v normě ČSN EN ISO 10399 (2010):

$$\left(\frac{2X}{n} - 1\right) + 2u_{1-\beta} \sqrt{\frac{nX - X^2}{n^3}} < p_{d_0}.$$

Síla přesného testu hypotéz (4.4) proti alternativě p , kde $p \in [1/2p_0]$, je

$$\sum_{i=0}^{\ell} \binom{n}{i} p^i (1 - p)^{n-i}.$$

Odvoďme (asymptotický) $100(1 - \alpha)\%$ interval spolehlivosti pro p_d , který je uvedený v příloze B.3 normy ČSN EN ISO 10399 (2010). Vycházíme z aproximace normálním rozdělením, tj. z CLV. Užijme značení zavedené v normě $\hat{p} = X/n$, $\hat{p}_d = 2\hat{p} - 1$.

$$\mathbf{P} \left[-u_{1-\frac{\alpha}{2}} < \sqrt{n} \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}} < u_{1-\frac{\alpha}{2}} \right] \rightarrow 1 - \alpha, n \rightarrow \infty,$$

pokud je p skutečná hodnota parametru.

$$\mathbf{P} \left[\hat{p} - u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1 - \hat{p})/n} < p < \hat{p} + u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1 - \hat{p})/n} \right] \rightarrow 1 - \alpha, n \rightarrow \infty.$$

Použijme vztah $p = 1/2 + 1/2p_d$, tj. $p_d = 2p - 1$.

$$\mathbf{P} \left[\hat{p}_d - 2u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1 - \hat{p})/n} < p_d < \hat{p}_d + 2u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1 - \hat{p})/n} \right] \rightarrow 1 - \alpha, n \rightarrow \infty,$$

tedy $100(1 - \alpha)\%$ interval je

$$\left(\hat{p}_d - 2u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1 - \hat{p})/n}, \hat{p}_d + 2u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1 - \hat{p})/n} \right).$$

Volba počtu posuzovatelů pro zkoušku, tabulka A.3

Při výběru počtu posuzovatelů n pro zkoušku duo-trio se norma ČSN EN ISO 10399 (2010) odkazuje na tabulku A.3. Tabulka A.3 je sestavena z buněk, které přísluší trojicím čísel (α, β, p_d) , kde $\alpha, \beta \in \{0.001, 0.01, 0.05, 0.1, 0.2\}$ a $p_d \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. V buňkách jsou uvedeny počty posuzovatelů n , které odpovídají následujícím pravidlům:

- **Test pro rozdíl**

Buňka v tabulce A.3 příslušná trojici (α, β, p_d) obsahuje nejmenší počet posuzovatelů n pro zkoušku takový, že přesný test hypotéz (4.2) se stanovenou hladinou α má proti alternativě $p = p_d + (1 - p_d)/2$ sílu alespoň $1 - \beta$.

- **Test pro podobnost**

Buňka v tabulce A.3 příslušná trojici (α, β, p_d) obsahuje nejmenší počet posuzovatelů n pro zkoušku takový, že přesný test hypotéz (4.4) se zvolenou hodnotou $p_0 = p_d + (1 - p_d)/2$ a stanovenou hladinou β má proti alternativě $p = 1/2$ sílu alespoň $1 - \alpha$.

Nepovedlo se nám ukázat a nevíme, zda jsou výše daná pravidla pro test pro rozdíl a podobnost ekvivalentní pro každou trojici (α, β, p_d) , kde $\alpha, \beta, p_d \in (0, 1)$. Pouze jsme ověřili, že pro všechny kombinace hodnot α, β, p_d v tabulce A.3 ekvivalence platí.

Ve výpočetním prostředí R (R Core Team (2018)) jsme vytvořili funkce, které slouží k výpočtu tabulek A.1–A.3 a funkce k počítání hladiny a síly příslušných testů. K práci jsme přiložili CD, na kterém je R skript s naprogramovanými funkcemi. Ukázku zdrojového kódu lze nalézt v příloze A.1.

4.3 Trojúhelníková zkouška

U trojúhelníkové zkoušky se posuzovateli předloží tři vzorky, dva stejné a jeden odlišný, o nichž neví, který je který. Posuzovatel má za úkol říct, který vzorek se liší od dalších dvou bez možnosti odpovědi, že neví.

4.3.1 Provedení trojúhelníkové zkoušky, pokyny pro vedoucího zkoušky

Shrňme, jak zkouška probíhá, detailnější popis najdeme v normě ČSN EN ISO 4120 (2009). Zvolí se hladina významnosti testu. V případě testu pro podobnost se zvolí p_{d_0} [†]. Vybere se počet posuzovatelů n pro zkoušku. Počet posuzovatelů je možno volit na základě požadované síly testu podle tabulky A.3 v normě ČSN EN ISO 4120 (2009). Bližší vysvětlení k tabulce A.3 je uvedeno v sekci 4.3.3. Pro zkoušku se vyberou posuzovatelé, kteří mají stejný stupeň kvalifikace.

Uvažujme šest pořadí vzorků výrobků A a B : $AAB, ABA, BAA, ABB, BAB, BBA$. Posuzovatelé se rozdělí do skupin po šesti. V každé skupině se mezi posuzovatele náhodně rozdělí šest pořadí vzorků.

Každému posuzovateli se předloží vzorky zleva doprava v pořadí, jaké mu bylo přiřazeno. Posuzovatel postupuje v daném pořadí a zaznamená, které dva

[†]Značení p_{d_0} jsme zvolili pro přehlednost, v normě se používá značení p_d .

vzorky jsou shodné a který se liší, opět bez možnosti odpovědi, že neví. Odpovědi posuzovatelů musí být nezávislé. Podrobnější instrukce k postupu, který je třeba dodržet, jsou uvedeny v normě ČSN EN ISO 4120 (2009).

Nakonec se sečtou správné odpovědi posuzovatelů.

4.3.2 Vyhodnocení trojúhelníkové zkoušky

Test pro rozdíl

V normě ČSN EN ISO 4120 (2009) se v tabulce A.1 nalezne buňka příslušná zvolené hladině testu α a počtu posuzovatelů n . Číslo v příslušné buňce označme k . Pokud je počet posuzovatelů n vyšší, než je uvedeno v tabulce A.1, číslo k se spočítá jako nejbližší celé číslo vyšší než

$$\frac{n}{3} + u_{1-\alpha} \sqrt{\frac{2n}{9}},$$

kde $u_{1-\alpha}$ je $(1 - \alpha)$ -kvantil standardizovaného normálního rozdělení.

Vyhodnocení zkoušky:

- *Počet správných odpovědí je menší než k*

Výsledek zkoušky je ten, že jsme neprokázali vnímatelný rozdíl mezi výrobky na hladině významnosti α .

- *Počet správných odpovědí je alespoň k*

Výsledek zkoušky je ten, že jsme prokázali vnímatelný rozdíl mezi výrobky na hladině významnosti α .

Test pro podobnost

V normě ČSN EN ISO 4120 (2009) se v tabulce A.2 nalezne buňka příslušná zvolené hladině testu β , počtu posuzovatelů n a hodnotě p_{d_0} . Číslo v příslušné buňce označme ℓ . Pokud je počet posuzovatelů n vyšší, než je uvedeno v tabulce A.2, spočítá se číslo Al :

$$Al = \left(\frac{3X}{2n} - \frac{1}{2} \right) + \frac{3}{2} u_{1-\beta} \sqrt{\frac{nX - X^2}{n^3}},$$

kde X je počet správných odpovědí a $u_{1-\beta}$ je $(1 - \beta)$ -kvantil standardizovaného normálního rozdělení.

Vyhodnocení zkoušky:

- *Počet správných odpovědí je nejvýše ℓ , resp. $Al < p_{d_0}$:*

Výsledek zkoušky je ten, že jsme prokázali dostatečnou podobnost výrobků na hladině významnosti β .

- *Počet správných odpovědí je větší než ℓ , resp. $Al \geq p_{d_0}$:*

Výsledek zkoušky je ten, že jsme neprokázali dostatečnou podobnost výrobků na hladině významnosti β .

4.3.3 Statistické vysvětlení trojúhelníkové zkoušky

Předpokládá se, že posuzovatel odpoví správně s pravděpodobností p , kterou neznáme. V normách je vztah mezi p a p_d dán jako $p = p_d + (1 - p_d)/3$. Zřejmě se vychází z následující úvahy:

Označme „správnou odpověď“ jako SO, „posuzovatel pozná rozdíl“ jako PPR a „posuzovatel nepozná rozdíl“ jako PNR. Věta o podmíněné pravděpodobnosti říká: $p = P(\text{SO}) = P(\text{SO}|\text{PPR}) P(\text{PPR}) + P(\text{SO}|\text{PNR}) P(\text{PNR})$.

Pravděpodobnost, že posuzovatel rozezná výrobky, je $P(\text{PPR}) = p_d$. Předpokládáme, že tento posuzovatel vždy odpoví správně, tedy $P(\text{SO}|\text{PPR}) = 1$. Pravděpodobnost, že posuzovatel nerozezná výrobky, je $P(\text{PNR}) = 1 - p_d$. Předpokládáme, že tento posuzovatel hádá a správně uhádne s pravděpodobností $P(\text{SO}|\text{PNR}) = 1/3$. Odvodili jsme vztah $p = p_d + (1 - p_d)/3$.

Definujme $\mathbf{1}(\text{SO})$ indikátor, že posuzovatel odpověděl správně. Veličina $\mathbf{1}(\text{SO})$ má alternativní rozdělení s parametrem $p = p_d + (1 - p_d)/3$.

Pro $i \in \{1, \dots, n\}$ označme $\mathbf{1}(\text{SO})_i$ indikátory, že i -tý posuzovatel odpověděl správně. Předpokládáme, že $\mathbf{1}(\text{SO})_1, \dots, \mathbf{1}(\text{SO})_n$ tvoří náhodný výběr rozsahu n z alternativního rozdělení s parametrem p . Z tohoto předpokladu plyne, že $X =$ počet správných odpovědí n posuzovatelů se řídí binomickým rozdělením s parametry n a $p = p_d + (1 - p_d)/3$.

Statistické hypotézy pro problémy (a) a (b) jsou sestaveny o parametru p_d , jakožto míře rozdílnosti či podobnosti dvou výrobků (sekce 4.1).

Test pro rozdíl

Uvažujme hladinu testu α .

Norma ČSN EN ISO 4120 (2009) zřejmě vychází z hypotéz o p_d ve tvaru

$$H_0 : p_d = 0 \quad \text{proti} \quad H_1 : 0 < p_d \leq 1. \quad (4.6)$$

Nulová hypotéza vyjadřuje, že mezi výrobky není vnímatelný rozdíl. Alternativa vyjadřuje, že mezi výrobky vnímatelný rozdíl je.

Test hypotéz (4.6) je ekvivalentní testu hypotéz

$$H_0 : p = 1/3 \quad \text{proti} \quad H_1 : 1/3 < p \leq 1. \quad (4.7)$$

Proti nulové hypotéze svědčí velký počet správných odpovědí X .

Pro testování (4.7) se v normě pro malé hodnoty n (přibližně $n < 100$) používá přesný jednostranný test pro binomické rozdělení. Nulovou hypotézu zamítáme, když $X \geq k$, kde

$$k = \min\{j; j \in \mathbb{N}, P(X \geq j | n, p = 1/3) \leq \alpha\}.$$

Hodnoty k pro vybraná n jsou uvedeny v normě ČSN EN ISO 4120 (2009) v tabulce A.1.

Pro testování (4.7) se v normě pro velké hodnoty n používá asymptotický test založený na CLV. Nulovou hypotézu zamítáme, když

$$\sqrt{n} \frac{X/n - 1/3}{\sqrt{(1/3)(1 - 1/3)}} > u_{1-\alpha}, \quad \text{tj. když} \quad X > \frac{n}{3} + u_{1-\alpha} \sqrt{\frac{2n}{9}}.$$

Síla přesného testu hypotéz (4.7) proti alternativě p , kde $p > 1/3$ (neboli hodnota silofunkce v bodě p), je

$$\sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}.$$

Test pro podobnost

Uvažujme hladinu testu β (značení β použijeme z důvodu shody s normou).

Norma ČSN EN ISO 4120 (2009) zřejmě vychází z hypotéz o p_d ve tvaru

$$H_0 : p_{d_0} \leq p_d \leq 1 \quad \text{proti} \quad H_1 : 0 \leq p_d < p_{d_0}, \quad (4.8)$$

kde vhodné p_{d_0} musíme zvolit. Hodnotu p_{d_0} volíme tak, abychom mohli říct, že výrobky si jsou dostatečně podobné (pro naše účely), pokud je rozlíší méně než $p_{d_0} * 100$ % posuzovatelů. Nulová hypotéza vyjadřuje, že výrobky si nejsou dostatečně podobné a alternativa vyjadřuje dostatečnou podobnost. Normy nabízí hodnoty $p_{d_0} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

Test hypotéz (4.8) je ekvivalentní testu hypotéz

$$H_0 : p_0 \leq p \leq 1 \quad \text{proti} \quad H_1 : 1/3 \leq p < p_0, \quad (4.9)$$

kde $p_0 = p_{d_0} + (1 - p_{d_0})/3$.

Proti nulové hypotéze svědčí malé hodnoty X .

Pro testování (4.9) se v normě pro malé hodnoty n (přibližně $n < 100$) používá přesný jednostranný test pro binomické rozdělení. Nulovou hypotézu zamítáme, když $X \leq \ell$, kde

$$\ell = \max\{j; j \in \mathbb{N}, \mathbf{P}(X \leq j | n, p = p_0) \leq \beta\}.$$

Hodnoty ℓ pro vybraná n , β a p_{d_0} jsou uvedeny v normě ČSN EN ISO 4120 (2009) v tabulce A.2.

Pro testování (4.9) se v normě pro velké hodnoty n používá asymptotický test založený na CLV. Nulovou hypotézu zamítáme, když

$$\sqrt{n} \frac{X/n - p_0}{\sqrt{X/n(1 - X/n)}} < -u_{1-\beta}. \quad (4.10)$$

Dosazením $p_0 = p_{d_0} + (1 - p_{d_0})/3$ do (4.10) a úpravou získáme tvar pro zamítnutí H_0 , který je uvedený v normě ČSN EN ISO 4120 (2009):

$$\left(\frac{3X}{2n} - \frac{1}{2}\right) + \frac{3}{2}u_{1-\beta}\sqrt{\frac{nX - X^2}{n^3}} < p_{d_0}.$$

Síla přesného testu hypotéz (4.9) proti alternativě p , kde $p \in (1/3, p_0)$, je

$$\sum_{i=0}^{\ell} \binom{n}{i} p^i (1-p)^{n-i}.$$

Odvoďme (asymptotický) $100(1 - \alpha)\%$ interval spolehlivosti pro p_d uvedený v příloze B.3 normy ČSN EN ISO 4120 (2009). Vycházíme z aproximace normálním rozdělením, tj. z CLV. Označme $\hat{p} = X/n$, $\hat{p}_d = (3/2)\hat{p} - 1/2$.

$$\mathbf{P} \left[-u_{1-\frac{\alpha}{2}} < \sqrt{n} \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}} < u_{1-\alpha/2} \right] \rightarrow 1 - \alpha, \quad n \rightarrow \infty,$$

pokud p je skutečná hodnota parametru. Úpravou získáme

$$P \left[\hat{p} - u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n} \right] \rightarrow 1 - \alpha, n \rightarrow \infty.$$

Použijme vztah $p = 1/3 + 2/3p_d$, tj. $p_d = (3/2)p - 1/2$.

$$P \left[\hat{p}_d - \frac{3}{2} u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n} < p_d < \hat{p}_d + \frac{3}{2} u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n} \right] \rightarrow 1 - \alpha, n \rightarrow \infty.$$

Odvodili jsme (asymptotický) $100(1 - \alpha)\%$ interval spolehlivosti pro p_d

$$\left(\hat{p}_d - (3/2) u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p}_d + (3/2) u_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n} \right).$$

Volba počtu posuzovatelů pro zkoušku, tabulka A.3

Při výběru počtu posuzovatelů n pro trojúhelníkovou zkoušku se norma ČSN EN ISO 4120 (2009) odkazuje na tabulku A.3. Tabulka A.3 je sestavena z buněk, které přísluší trojicím (α, β, p_d) , kde $\alpha, \beta \in \{0.001, 0.01, 0.05, 0.1, 0.2\}$ a $p_d \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. V buňkách jsou uvedeny počty posuzovatelů n , které odpovídají následujícím pravidlům:

- **Test pro rozdíl**

Buňka v tabulce A.3 příslušná hodnotám α , β a $p_d > 0$ obsahuje nejmenší počet posuzovatelů n pro zkoušku takový, že přesný test hypotéz (4.7) se stanovenou hladinou α má proti alternativě $p = p_d + (1 - p_d)/3$ sílu alespoň $1 - \beta$.

- **Test pro podobnost**

Buňka v tabulce A.3 příslušná hodnotám α , β a p_d obsahuje nejmenší počet posuzovatelů n pro zkoušku takový, že přesný test hypotéz (4.9) se zvolenou hodnotou $p_0 = p_d + (1 - p_d)/3$ a stanovenou hladinou β má proti alternativě $p = 1/3$ sílu alespoň $1 - \alpha$.

Nepovedlo se nám ukázat a nevíme, zda jsou výše uvedená pravidla pro test pro rozdíl a podobnost ekvivalentní pro každou trojici (α, β, p_d) , kde $\alpha, \beta, p_d \in (0, 1)$. Pouze jsme ověřili, že pro všechny kombinace hodnot α, β, p_d v tabulce A.3 ekvivalence platí.

4.4 Párová porovnávací zkouška

Při párové porovnávací zkoušce se posuzovateli předloží dva vzorky. Posuzovatel má za úkol říct, který ze vzorků má intenzivnější zkoumanou vlastnost bez možnosti odpovědi, že neví.

4.4.1 Provedení párové porovnávací zkoušky, pokyny pro vedoucího zkoušky

Nejprve stručně shrňme, jak zkouška probíhá, detailnější popis najdeme v normě ČSN EN ISO 5495 (2009). Rozlišují se přitom dva případy, pro které se používá párová zkouška, a to:

- (i) *Když výrobce ví, který z výrobků A a B má intenzivnější vlastnost.*
Příklad: Výrobce zajímá, zda chuť sušenek bude sladší, pokud v receptu na sušenky zvýší množství cukru o 5 %. V tomto případě výrobce ví, který z výrobků má intenzivnější sladkost.
- (ii) *Když výrobce neví, který z výrobků A a B má intenzivnější vlastnost.*
Příklad: Výrobce si přeje určit, který ze dvou druhů paprik v pomazánce vyvolá pálivější chuť. V tomto případě výrobce neví, který z výrobků má intenzivnější pálivost.

Zvolí se hladina významnosti testu α pro rozdíl, β pro podobnost. V případě testu pro podobnost se zvolí p_{d_0} [‡]. Vybere se počet posuzovatelů n pro zkoušku. Počet posuzovatelů je možno volit na základě požadované síly testu podle tabulek A.4 vztažené k (i) a A.5 vztažené k (ii) v normě ČSN EN ISO 5495 (2009). Bližší vysvětlení k tabulkám A.4 a A.5 je uvedené níže v sekci 4.4.3. Pro zkoušku se vyberou posuzovatelé, kteří mají stejný stupeň kvalifikace.

Uvažujme možná pořadí vzorků výrobků A a B , a to AB a BA . Posuzovatelé se rozdělí do dvojic. V každé dvojici se náhodně jednomu posuzovateli přiřadí pořadí AB a druhému BA .

Každému posuzovateli se předloží vzorky zleva doprava v pořadí, které mu bylo přiřazeno. Posuzovatel postupuje v daném pořadí. Posuzovatel má za úkol říct, který ze vzorků má intenzivnější zkoumanou vlastnost bez možnosti odpovědět, že neví. Je nutné zajistit nezávislost hodnocení mezi posuzovateli.

- V případě (i) se správné odpovědi posuzovatelů sečtou.
- V případě (ii) se místo počtu „správných“ odpovědí definuje počet „shodných“ odpovědí. Označme a počet odpovědí „vzorek A má intenzivnější vlastnost než vzorek B “ a b počet odpovědí „vzorek B má intenzivnější vlastnost než vzorek A “. Počet shodných odpovědí definujeme jako maximum z čísel a a b .

4.4.2 Vyhodnocení párové porovnávací zkoušky

Test pro rozdíl

- *Případ (i), tj. výrobce ví, který z výrobků A a B má intenzivnější vlastnost:*
V normě ČSN EN ISO 5495 (2009) se v tabulce A.1 nalezne buňka příslušná zvolené hladině testu α a počtu posuzovatelů n . Číslo v příslušné buňce označme k . Pokud je počet posuzovatelů n větší, než je uvedeno v tabulce, číslo k se spočítá jako nejbližší celé číslo vyšší než

$$\frac{n+1}{2} + u_{1-\alpha} \sqrt{\frac{n}{4}},$$

kde $u_{1-\alpha}$ je $(1 - \alpha)$ -kvantil standardizovaného normálního rozdělení.

- *Případ (ii), tj. výrobce neví, který z výrobků A a B má intenzivnější vlastnost:*
V normě ČSN EN ISO 5495 (2009) se v tabulce A.2 nalezne buňka příslušná zvolené hladině testu α a počtu posuzovatelů n . Číslo v příslušné buňce

[‡]Značení p_{d_0} jsme zvolili pro přehlednost, v normě se používá značení p_d .

označme k . Pokud je počet posuzovatelů větší, než je uvedeno v tabulce, číslo k se spočítá jako nejbližší celé číslo vyšší než

$$\frac{n+1}{2} + u_{1-\alpha/2} \sqrt{\frac{n}{4}}.$$

Vyhodnocení zkoušky souhrnně pro oba případy (i) a (ii):

- *Počet správných, resp. shodných, odpovědí je menší než k :*

Výsledek zkoušky je ten, že jsme neprokázali vnímatelný rozdíl mezi intenzitami zkoumané vlastnosti výrobků na hladině významnosti α .

- *Počet správných, resp. shodných, odpovědí je alespoň k :*

Výsledek zkoušky je ten, že jsme prokázali vnímatelný rozdíl mezi intenzitami zkoumané vlastnosti výrobků na hladině významnosti α . V případě, kdy výrobce neví, který z výrobků A a B má intenzivnější vlastnost, rozhodneme, že výrobek s intenzivnější vlastností je ten, který více posuzovatelů volilo jako výrobek s intenzivnější vlastností.

Test pro podobnost

Souhrnně pro oba případy (i) a (ii):

V normě ČSN EN ISO 5495 (2009) se v tabulce A.3 nalezne buňka příslušná zvolené hladině testu β , počtu posuzovatelů n a hodnotě p_{d_0} . Číslo v příslušné buňce označme ℓ . Pokud je počet posuzovatelů n vyšší, než je uvedeno v tabulce A.3, spočítá se číslo Al :

$$Al = \left(2\frac{X}{n} - 1\right) + 2u_{1-\beta} \sqrt{\frac{nX - X^2}{n^3}},$$

kde X je počet správných, resp. shodných, odpovědí a $u_{1-\beta}$ je $(1 - \beta)$ -kvantil standardizovaného normálního rozdělení.

Vyhodnocení zkoušky souhrnně pro oba případy (i) a (ii):

- *Počet správných, resp. shodných, odpovědí je nejvýše ℓ , resp. $Al < p_{d_0}$:*

Výsledek zkoušky je ten, že jsme prokázali, že intenzity zkoumané vlastnosti výrobků jsou dostatečně podobné na hladině významnosti β .

- *Počet správných, resp. shodných, odpovědí je větší než ℓ , resp. $Al \geq p_{d_0}$:*

Výsledek zkoušky je ten, že jsme neprokázali, že intenzity zkoumané vlastnosti výrobků jsou dostatečně podobné na hladině významnosti β .

4.4.3 Statistické vysvětlení párové zkoušky

Předpokládá se, že posuzovatel odpoví správně s pravděpodobností p , kterou neznáme. V normách je vztah mezi p a p_d dán jako $p = p_d + (1 - p_d)/2$. Zřejmě se vychází z následující úvahy: Označme „správná odpověď“ jako SO, „posuzovatel pozná rozdíl“ jako PPR, „posuzovatel nepozná rozdíl“ jako PNR. Věta o podmíněné pravděpodobnosti říká: $p = P(\text{SO}) = P(\text{SO}|\text{PPR})P(\text{PPR}) + P(\text{SO}|\text{PNR})P(\text{PNR})$.

Pravděpodobnost, že posuzovatel rozezná výrobky, je $P(\text{PPR}) = p_d$. Předpokládáme, že tento posuzovatel vždy odpoví správně, tedy $P(\text{SO}|\text{PPR}) = 1$. Pravděpodobnost, že posuzovatel nerozezná výrobky, je $P(\text{PNR}) = 1 - p_d$. Předpokládáme, že tento posuzovatel hádá a správně uhádne s pravděpodobností $P(\text{SO}|\text{PNR}) = 1/2$. Odvodili jsme vztah $p = p_d + (1 - p_d)/2$.

- *Případ (i), tj. je známo, který výrobek má intenzivnější vlastnost:*
Označme $\mathbf{1}(\text{SO})$ indikátor, že posuzovatel odpověděl správně. Předpokládáme, že veličina $\mathbf{1}(\text{SO})$ má alternativní rozdělení s parametrem $p = p_d + (1 - p_d)/2$. Pro $i \in \{1, \dots, n\}$, označme $\mathbf{1}(\text{SO})_i$ indikátor, že i -tý posuzovatel odpověděl správně. Předpokládáme, že $\mathbf{1}(\text{SO})_1, \dots, \mathbf{1}(\text{SO})_n$ tvoří náhodný výběr rozsahu n z alternativního rozdělení s parametrem p . Počet správných odpovědí X z odpovědí n posuzovatelů se řídí binomickým rozdělením s parametry n a $p = p_d + (1 - p_d)/2$.

Test pro rozdíl

Uvažujeme hladinu testu α . Hypotézy a test jsou stejné jako ve zkoušce duo-trio (sekce 4.2.3, test pro rozdíl).

Jediná odlišnost proti zkoušce duo-trio je v tom, že v asymptotickém testu je v normě ČSN EN ISO 5495 (2009) použita korekce na spojitost. Místo pravidla

$$\frac{X - n/2}{\sqrt{n(1/2)(1 - 1/2)}} > u_{1-\alpha}, \quad \text{tj. když } X > \frac{n}{2} + u_{1-\alpha} \sqrt{\frac{n}{4}},$$

které je ve zkoušce duo-trio použito, je v párové zkoušce pravidlo

$$\frac{X - n/2 - 1/2}{\sqrt{n(1/2)(1 - 1/2)}} > u_{1-\alpha}, \quad \text{tj. když } X > \frac{n+1}{2} + u_{1-\alpha} \sqrt{\frac{n}{4}}.$$

Korekce na spojitost je opatrnější verzí asymptotického testu pro diskrétní data.

Minimální počty správných odpovědí k pro vybraná n a α , při kterých se zamítá nulová hypotéza, jsou uvedeny v normě ČSN EN ISO 5495 (2009) v tabulce A.1.

Test pro podobnost

Uvažujeme hladinu testu β (značení β opět použijeme z důvodu shody s normou). Hypotézy a test jsou stejné jako ve zkoušce duo-trio (sekce 4.2.3, test pro podobnost).

Maximální počty správných odpovědí ℓ pro vybraná n , β a p_{d_0} , při kterých se zamítá nulová hypotéza, jsou uvedeny v normě ČSN EN ISO 5495 (2009) v tabulce A.3.

Odvození oboustranného intervalu spolehlivosti pro p_d , který je uvedený v příloze B.5 v normě ČSN EN ISO 5495 (2009), je naprosto stejné jako ve zkoušce duo-trio (sekce 4.2.3).

Volba počtu posuzovatelů, tabulky A.4

Při výběru počtu posuzovatelů n v případě znalosti toho, který výrobek má intenzivnější vlastnost, se norma ČSN EN ISO 5495 (2009) odkazuje na tabulku A.4. V tabulce A.4 jsou poskytnuty počty posuzovatelů n pro zkoušku podle stejného pravidla a stejných hodnot jako v tabulce A.3 u testu duo-trio (sekce 4.2.3).

- *Případ (ii), tj. není známo, který výrobek má intenzivnější zkoumanou vlastnost:*

Předpokládáme, že posuzovatel označí výrobek A jako ten s intenzivnější zkoumanou vlastností s pravděpodobností p . Parametr p zde nevystupuje jako pravděpodobnost správné odpovědi. Pravděpodobnost správné odpovědi označme jako p_c a (stále jako v případě (i)) předpokládejme vztah $p_c = p_d + (1 - p_d)/2$.

Označme $\mathbf{1}(A)$ indikátor, že posuzovatel vybral výrobek A jako ten s intenzivnější vlastností. Předpokládáme, že veličina $\mathbf{1}(A)$ má alternativní rozdělení s parametrem p a že $\mathbf{1}(A)_1, \dots, \mathbf{1}(A)_n$ tvoří náhodný výběr. Celkový počet výsledků X „posuzovatel vybral výrobek A jako ten s intenzivnější vlastností“, tj. $X = \sum_{i=1}^n \mathbf{1}(A)_i$, se řídí binomickým rozdělením s neznámým parametrem p . Pouze víme, že $p \in [0,1]$ a vztah mezi p a p_d je:

$$p = \begin{cases} p_d + (1 - p_d)/2, & \text{když } A \text{ je výrobek s intenzivnější vlastností,} \\ 1 - [p_d + (1 - p_d)/2], & \text{když } B \text{ je výrobek s intenzivnější vlastností.} \end{cases}$$

Test pro rozdíl

Zvolíme hladinu testu α . Mezi výrobky A a B není vnímatelný rozdíl, pokud je panel posuzovatelů nerozliší, tj. $p_d = 0$. Naopak, mezi výrobky je vnímatelný rozdíl, pokud je (alespoň) část populace rozliší, tj. $p_d > 0$. Norma ČSN EN ISO 5495 (2009) zřejmě vychází z hypotéz o p_d ve tvaru:

$$H_0 : p_d = 0 \quad \text{proti} \quad H_1 : p_d > 0,$$

což je ekvivalentní

$$H_0 : p_c = 1/2 \quad \text{proti} \quad H_1 : p_c > 1/2,$$

což je ekvivalentní

$$H_0 : p = 1/2 \quad \text{proti} \quad H_1 : p \neq 1/2. \quad (4.11)$$

Za test hypotéz (4.11) při malých hodnotách n (přibližně $n < 100$) je v normě zvolen přesný oboustranný test pro binomické rozdělení. H_0 zamítneme, pokud $X \leq k_1$ nebo $X \geq k_2$, kde k_1, k_2 jsou určeny:

$$k_1 = \max\{j; j \in \mathbb{N}, \mathbf{P}(X \leq j | n, p = 1/2) \leq \alpha/2\},$$

$$k_2 = \min\{j; j \in \mathbb{N}, \mathbf{P}(X \geq j | n, p = 1/2) \leq \alpha/2\}.$$

Ze „symetrie“ pravděpodobností v případě binomického rozdělení s parametry n a $1/2$, tj. $\mathbf{P}(X = j | n, p = 1/2) = \mathbf{P}(X = n - j | n, p = 1/2)$, $j \in \{0, \dots, \lfloor n/2 \rfloor\}$, vyplývá pravidlo pro zamítnutí H_0 v (4.11) uvedené v normách: H_0 zamítneme, když počet shodných odpovědí $Y = \max(X, n - X) \geq k$, kde

$$k = \min\{j; j \in \mathbb{N}, \mathbf{P}(X \geq j | n, p = 1/2) \leq \alpha/2\}$$

a

$$P(X \geq j | n, p = 1/2) = \sum_{i=j}^n \binom{n}{i} (1/2)^i (1 - 1/2)^{n-i} = (1/2)^n \sum_{i=j}^n \binom{n}{i}.$$

Hodnoty k pro vybraná n a α jsou uvedeny v normě ČSN EN ISO 5495 (2009) v tabulce A.2.

Za test hypotéz (4.11) při velkých hodnotách n je v normě zvolen asymptotický oboustranný test založený na aproximaci normálním rozdělením s korekcí na spojitost. Zamítáme H_0 , když

$$\frac{|X - n/2 - \operatorname{sgn}(X - n/2)/2|}{\sqrt{n(1/2)(1 - 1/2)}} > u_{1-\alpha/2},$$

kde $-\operatorname{sgn}(X - n/2)/2$ je korekce na spojitost. Vzhledem k symetričnosti rozdělení X kolem bodu $n/2$ za platnosti H_0 (v (4.11)) lze pravidlo pro zamítnutí H_0 vyjádřit „v řeči“ $Y = \max(X, n - X)$. Zamítneme H_0 , když

$$\frac{Y - n/2 - 1/2}{\sqrt{n(1/2)(1 - 1/2)}} > u_{1-\alpha/2}, \quad \text{tj. když } Y > \frac{n+1}{2} + u_{1-\alpha/2} \sqrt{\frac{n}{4}},$$

což je pravidlo uvedené v normě ČSN EN ISO 5495 (2009).

Síla přesného testu hypotéz (4.11) proti alternativě p , kde $p \in [0, 1], p \neq 1/2$, je

$$\sum_{i=0}^{n-k} \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=k}^n \binom{n}{i} p^i (1-p)^{n-i}. \quad (4.12)$$

Zajímá nás síla pro p dané hodnotou p_d . Víme, že hodnotě p_d přísluší právě jedna z hodnot $p = 1/2 + p_d/2$ nebo $p = 1/2 - p_d/2$, ale nevíme která. Protože je ale hodnota (4.12) stejná pro oba případy $p = 1/2 \mp p_d/2$, můžeme bez újmy na obecnosti psát, že síla testu proti p příslušícímu p_d , kde $p_d > 0$, je (4.12), kde $p = 1/2 + p_d/2$.

Test pro podobnost

Zvolíme hladinu testu β (značení opět dle shody s normou).

Řekneme, že výrobky si jsou dostatečně podobné, pokud je rozliší méně než $p_{d_0} \times 100$ % posuzovatelů. Norma ČSN EN ISO 5495 (2009) zřejmě vychází z hypotéz o p_d ve tvaru:

$$H_0 : p_d \geq p_{d_0} \quad \text{proti} \quad H_1 : p_d \leq p_{d_0},$$

což je ekvivalentní

$$H_0 : p_c \geq p_{c_0} \quad \text{proti} \quad H_1 : p_c \leq p_{c_0},$$

kde $p_{c_0} = p_{d_0} + (1 - p_{d_0})/2$, což je ekvivalentní

$$H_0 : p \notin (1/2 - \epsilon_0, 1/2 + \epsilon_0) \quad \text{proti} \quad H_1 : p \in (1/2 - \epsilon_0, 1/2 + \epsilon_0), \quad (4.13)$$

kde $\epsilon_0 = p_{c_0} - 1/2 = p_{d_0}/2$.

Pro malé hodnoty n (přibližně $n < 130$) je za test hypotéz (4.13) v normě uveden přesný test. Připomeňme, že $Y = \max(X, n - X)$ je počet shodných odpovědí. Pravidlo pro zamítnutí H_0 uvedené v normě je:

$$\begin{aligned} &\text{Zamítáme } H_0, \text{ když } Y \leq \ell, \text{ kde} \\ \ell &= \max\{j; j \in \mathbb{N}, \mathbf{P}(X \leq j | n, p = 1/2 + \epsilon_0) \leq \beta\}. \end{aligned} \quad (4.14)$$

Vysvětleme smysl tohoto pravidla. Ukážeme, že test (4.14) hypotéz (4.13) zamítá H_0 právě tehdy, když přesný jednostranný test s hladinou β hypotéz

$$H_{0_1} : p \leq 1/2 - \epsilon_0 \quad \text{proti} \quad H_{1_1} : p > 1/2 - \epsilon_0 \quad (4.15)$$

zamítá H_{0_1} a zároveň přesný jednostranný test s hladinou β hypotéz

$$H_{0_2} : p \geq 1/2 + \epsilon_0 \quad \text{proti} \quad H_{1_2} : p < 1/2 + \epsilon_0, \quad (4.16)$$

zamítá H_{0_2} . Takto sestrojenému testu hypotéz (4.13) se říká TOST procedura (blíže informace k TOST proceduře nalezneme v sekci 1.2). Výsledný test má hladinu nejvýše β (plyne z věty 1.11 v sekci 1.2).

Přesný test hypotéz (4.15) je daný pravidlem, že hypotézu H_{0_1} zamítáme, když $X \geq \ell_1$, kde $\ell_1 = \min\{j; j \in \mathbb{N}, \mathbf{P}(X \geq j | n, p = 1/2 - \epsilon_0) \leq \beta\}$.

Přesný test hypotéz (4.16) je daný pravidlem, že hypotézu H_{0_2} zamítáme, když $X \leq \ell_2$, kde $\ell_2 = \max\{j; j \in \mathbb{N}, \mathbf{P}(X \leq j | n, p = 1/2 + \epsilon_0) \leq \beta\}$.

Hypotézy H_{0_1} a H_{0_2} tedy současně zamítáme, když $\ell_1 \leq X \leq \ell_2$. V normě jsou uvažovány pouze případy, kdy $\ell_2 \geq n/2$. Ze „symetrie“

$$\mathbf{P}(X = j | n, p = 1/2 - \epsilon_0) = \mathbf{P}(X = n - j | n, p = 1/2 + \epsilon_0),$$

$j \in \{0, \dots, n\}$, vyplývá, že interval $[\ell_1, \ell_2]$ je symetrický kolem $n/2$. Tedy $[\ell_1, \ell_2]$ lze zapsat ve tvaru $[n - \ell, \ell]$, kde

$$\ell = \max\{j; j \in \mathbb{N}, \mathbf{P}(X \leq j | n, p = 1/2 + \epsilon_0) \leq \beta\}.$$

Abychom ukázali, že test (4.14) odpovídá současnému zamítnutí H_{0_1} a H_{0_2} , stačí ukázat, že $n - \ell \leq X \leq \ell$ je ekvivalentní $Y \leq \ell$. Platí

$$\begin{aligned} Y \leq \ell &\Leftrightarrow \max(X, n - X) \leq \ell \\ &\Leftrightarrow X \leq \ell \ \& \ n - X \leq \ell \Leftrightarrow n - \ell \leq X \leq \ell. \end{aligned} \quad (4.17)$$

Z (4.17) jsme získali požadovanou ekvivalenci $Y \leq \ell \Leftrightarrow n - \ell \leq X \leq \ell$. Hodnoty ℓ pro vybraná n , β a p_{d_0} , jsou uvedeny v normě ČSN EN ISO 5495 (2009) v tabulce A.3.

Poznámka. (Návrh na zlepšení testu pro podobnost hypotéz (4.13).)

Test pro podobnost (4.14) hypotéz (4.13) uvedený v normě je velmi jednoduchý na konstrukci, ale může mít hladinu daleko menší než stanovené β . Test, který lépe zachovává požadovanou hladinu významnosti testu a má větší sílu, vyplývá z věty 1.2. Ačkoliv věta 1.2 mluví o znáhodněných testech, myslíme „neznáhodněnou verzi“ takovou, že γ_1, γ_2 z věty 1.2 při testování položíme rovné

nule. Konstrukce testu vycházejícího z věty 1.2 je v tomto případě jednoduchá. Opět ze „symetrie“

$$P(X = j | n, p = 1/2 - \epsilon_0) = P(X = n - j | n, p = 1/2 + \epsilon_0),$$

plyne, že výsledný test založený na větě 1.2 má tvar:

$$\begin{aligned} &\text{Zamítáme } H_0 \text{ v (4.13), když } n - \tilde{\ell} \leq X \leq \tilde{\ell}, \text{ kde} \\ &\tilde{\ell} = \max\{j; j \in \mathbb{N}, j \geq n/2, P(X \in [n - j, j] | n, p = 1/2 + \epsilon_0) \leq \beta\}. \end{aligned}$$

Místo X ale pozorujeme $Y = \max(X, n - X)$. Vzhledem k tomu, že $n - \tilde{\ell} \leq X \leq \tilde{\ell} \Leftrightarrow Y \leq \tilde{\ell}$, je daný test ekvivalentní:

$$\begin{aligned} &\text{Zamítáme } H_0 \text{ v (4.13), když } Y \leq \tilde{\ell}, \text{ kde} \\ &\tilde{\ell} = \max\{j; j \in \mathbb{N}, j \geq n/2, P(X \in [n - j, j] | n, p = 1/2 + \epsilon_0) \leq \beta\}. \end{aligned} \quad (4.18)$$

Funkci pro výpočet hodnot $\tilde{\ell}$ jsme naprogramovali ve výpočetním prostředí R (R Core Team (2018)) a lze ji nalézt na příloženém CD nebo v příloze A.2. Spočítali jsme hodnoty $\tilde{\ell}$ pro všechny trojice (n, β, p_{d_0}) , které jsou uvedeny v tabulce A.3 v normě ČSN EN ISO 5495 (2009). Zjistili jsme, že z celkového počtu 500 trojic (n, β, p_{d_0}) došlo ke zlepšení testu v 77 případech oproti testu z normy.

Při velkých hodnotách n se v normách pro testování hypotéz (4.13) používá asymptotický test založený na CLV. Nulovou hypotézu zamítáme, když

$$\frac{Y - n(1/2 + \epsilon_0)}{\sqrt{n(Y/n)(1 - Y/n)}} < -u_{1-\beta}. \quad (4.19)$$

Dosazením $p_0 = p_{d_0} + (1 - p_{d_0})/2$ do (4.19) a úpravou získáme tvar pro zamítnutí H_0 , který je uvedený v normě ČSN EN ISO 5495 (2009):

$$\left(\frac{2Y}{n} - 1\right) + 2u_{1-\beta}\sqrt{\frac{nY - Y^2}{n^3}} < p_{d_0}.$$

Opět platí, že pravidlo (4.19) pro zamítnutí H_0 v (4.13) je ekvivalentní provedení dvou jednostranných asymptotických testů hypotéz (4.15) a (4.16) s hladinou β .

Asymptotické testy vypadají tak, že hypotézu H_{0_1} v (4.15), resp. hypotézu H_{0_2} v (4.16), zamítáme, když

$$\frac{X - n(1/2 - \epsilon_0)}{\sqrt{n(X/n)(1 - X/n)}} > u_{1-\beta}, \quad \text{resp.} \quad \frac{X - n(1/2 + \epsilon_0)}{\sqrt{n(X/n)(1 - X/n)}} < -u_{1-\beta}.$$

Všimněme si, že $n(Y/n)(1 - Y/n) = n(X/n)(1 - X/n)$. Dále můžeme psát

$$\begin{aligned} &\frac{Y - n(1/2 + \epsilon_0)}{\sqrt{n(Y/n)(1 - Y/n)}} < -u_{1-\beta} \quad \Leftrightarrow \quad \frac{\max(X, n - X) - n(1/2 + \epsilon_0)}{\sqrt{n(X/n)(1 - X/n)}} < -u_{1-\beta} \\ &\Leftrightarrow \max(X, n - X) < n(1/2 + \epsilon_0) - u_{1-\beta}\sqrt{n(X/n)(1 - X/n)}. \end{aligned}$$

Vzhledem ke vztahu (4.17) lze pokračovat úpravou

$$\begin{aligned} &\Leftrightarrow n - \left[n(1/2 + \epsilon_0) - u_{1-\beta} \sqrt{n(X/n)(1 - X/n)} \right] < X \\ &\quad \& X < \left[n(1/2 + \epsilon_0) - u_{1-\beta} \sqrt{n(X/n)(1 - X/n)} \right] \\ &\Leftrightarrow \frac{X - n(1/2 - \epsilon_0)}{\sqrt{n(X/n)(1 - X/n)}} > u_{1-\beta} \quad \& \quad \frac{X - n(1/2 + \epsilon_0)}{\sqrt{n(X/n)(1 - X/n)}} < -u_{1-\beta}. \end{aligned}$$

Odsud už plyne ekvivalence testu (4.19) a provedení dvou jednostranných testů.

Síla přesného testu (4.13) proti alternativě p , kde $p \in (-\epsilon_0 + 1/2, \epsilon_0 + 1/2)$, je

$$\sum_{i=n-\ell}^{\ell} \binom{n}{i} p^i (1-p)^{n-i}. \quad (4.20)$$

Zajímá nás síla pro p příslušící hodnotě p_d . Víme, že hodnotě p_d přísluší právě jedna z hodnot $p = 1/2 + p_d/2$ nebo $p = 1/2 - p_d/2$, ale nevíme která. Všimněme si, že hodnota (4.20) je stejná pro oba případy $p = 1/2 \mp p_d/2$. Tedy bez újmy na obecnosti můžeme psát, že síla testu proti p příslušícímu p_d , kde $p_d > 0$, je (4.20), kde $p = 1/2 + p_d/2$.

Volba počtu posuzovatelů pro zkoušku, tabulka A.5

Při výběru počtu posuzovatelů n v případě, kdy nevíme, který výrobek má intenzivnější vlastnost, se norma ČSN EN ISO 5495 (2009) odkazuje na tabulku A.5. Tabulka A.5 je sestavena z buněk, které přísluší trojicím (α, β, p_d) , kde $\alpha, \beta \in \{0.001, 0.01, 0.05, 0.1, 0.2, 0.5\}$ a $p_d \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. V buňkách jsou uvedeny počty posuzovatelů n , které odpovídají následujícím pravidlům:

- **test pro rozdíl**

Buňka v tabulce A.5 příslušná hodnotám α , β a p_d obsahuje počet posuzovatelů n pro zkoušku takový, že přesný test hypotéz (4.11) se stanovenou hladinou α má proti alternativě $p = 1/2 + p_d/2$ (resp. $p = 1/2 - p_d/2$) sílu alespoň $1 - \beta$.

Poznamenejme, že n není vždy nejmenší (i když většinou nejmenší je) počet posuzovatelů takový, že přesný test hypotéz (4.11) s hladinou α má proti alternativě $p = 1/2 + p_d/2$ sílu alespoň $1 - \beta$.

- **test pro podobnost**

Buňka v tabulce A.5 příslušná hodnotám α , β a p_d obsahuje nejmenší počet posuzovatelů n pro zkoušku takový, že přesný test hypotéz (4.9) se zvolenou hodnotou $p_0 = 1/2 + p_d/2$ a stanovenou hladinou β má proti alternativě $p = 1/2$ sílu alespoň $1 - \alpha$.

Upozorněme, že neustále máme na mysli pouze vybrané hodnoty α, β a p_d uvedené v normě. Obecně platný vztah mezi počty posuzovatelů stanovených podle pravidel výše pro hodnoty $\alpha, \beta, p_d \in (0,1)$ se nám nepodařilo nalézt. Dále upozorněme, že α a β v tabulce A.5 nabývají hodnoty 0.5. Tuto hodnotu lze použít pouze ve smyslu síly testu ($1 - \alpha$ nebo $1 - \beta$), nedoporučujeme ji použít jako hladinu testu. Test s hladinou významnosti 0.5 by byl velmi neopatrný.

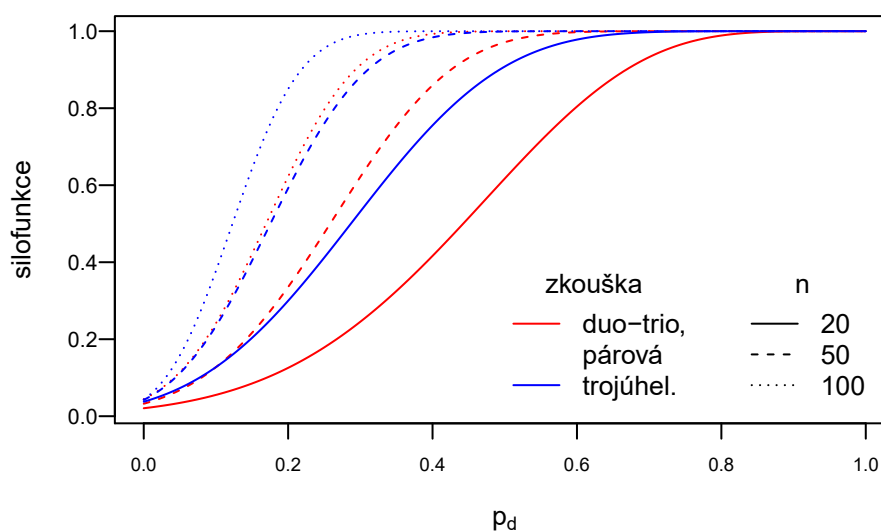
4.5 Porovnání zkoušek

Důležitou vlastností testu, kromě jeho hladiny, je jeho síla. Rozhodujeme-li se mezi několika testy, síla je jedním z faktorů, který rozhodne o naší volbě. Silnější testy potřebují méně pozorování k dosažení dané síly než testy s nižší silou. Pro podniky je finančně výhodnější použít silnější testy.

Podívejme se na silofunkce přesných testů ve zkouškách duo-trio, trojúhelníkové a párové porovnávací ve vybraných případech. Upozorníme, že zkoušky duo-trio a trojúhelníková se používají pro řešení stejných úloh, párová porovnávací zkouška se používá pro řešení trochu odlišných úloh. Ačkoliv si úlohy mohou být velmi podobné, je třeba správně rozhodnout, jak zní naše úloha, a tedy vybrat, zda zvolíme párovou porovnávací zkoušku nebo se budeme rozhodovat mezi zkouškou duo-trio a trojúhelníkovou zkouškou. V párové porovnávací zkoušce budeme uvažovat pouze případ, kdy víme, který výrobek má intenzivnější vlastnost (problematika v sekci 4.4). V tomto případě jsou testy ve zkouškách duo-trio a párové porovnávací stejné, tedy i silofunkce jsou stejné.

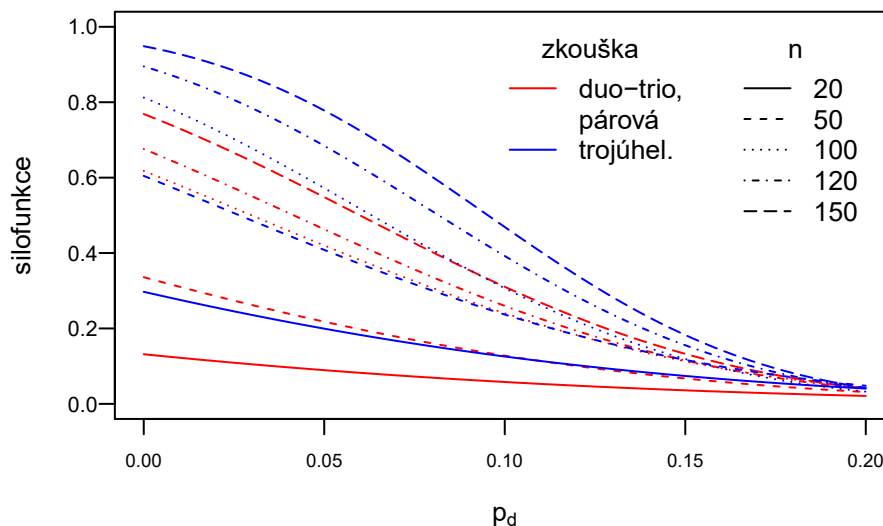
Všechny tři zkoušky jsou založené na pozorování počtu správných odpovědí posuzovatelů a v normách se testují hypotézy o parametru p , což je pravděpodobnost správné odpovědi. Primárně nás ale zajímá parametr p_d , což je podíl lidí v populaci posuzovatelů, kteří rozeznají výrobky. Vztah mezi p a p_d je v každé zkoušce jednoznačně daný a testy hypotéz o parametru p jsou ekvivalentní testům hypotéz o parametru p_d , odkážme se na sekci 4.2.3 pro zkoušku duo-trio, sekci 4.3.3 pro trojúhelníkovou zkoušku a sekci 4.4.3 pro párovou zkoušku. Proto budeme silofunkce vykreslovat s argumentem p_d , nikoliv p , a hovořit o síle testu proti p_d , nikoliv proti p . Všechny obrázky v této práci jsou vytvořeny ve výpočetním prostředí R (R Core Team (2018)).

Nejprve se podívejme na testy pro rozdíl. Stanovme hladinu významnosti 0.05 a počty posuzovatelů $n = 20, 50, 100$. Test v trojúhelníkové zkoušce má proti většině hodnot p_d , kde $p_d \in (0,1]$, větší sílu než testy ve zkouškách duo-trio a párové porovnávací. Silofunkce jsou vykresleny na obrázku 4.1.



Obrázek 4.1: Silofunkce testu pro rozdíl, když je stanovená hladina testu 0.05 a počet posuzovatelů n .

Dále se zaměříme na testy pro podobnost. Stanovme $p_{d_0} = 0.2$, hladinu významnosti 0.05 a počty posuzovatelů $n = 20, 50, 100, 120, 150$. Test v trojúhelníkové zkoušce má opět proti většině hodnot p_d , kde $p_d \in [0, p_{d_0})$, větší sílu než testy ve zkouškách duo-trio a párové porovnávací. Silofunkce jsou vykresleny na obrázku 4.2.



Obrázek 4.2: Silofunkce testu pro podobnost, když je $p_{d_0} = 0.2$, stanovená hladina testu 0.05 a počet posuzovatelů n .

Zavedme pojem psychometrická funkce (Ennis (1993)). Psychometrická funkce zachycuje vztah mezi odpovědí posuzovatele a daným podnětem. Definujme psychometrickou funkci v modelech uvedených v normách.

Definice 4.1. *Psychometrickou funkci pro modely v normách definujeme následovně:*

Pro zkoušku duo-trio

$$f_{DT}^N(p_d) = p_d + (1 - p_d)/2, \quad p_d \in [0,1].$$

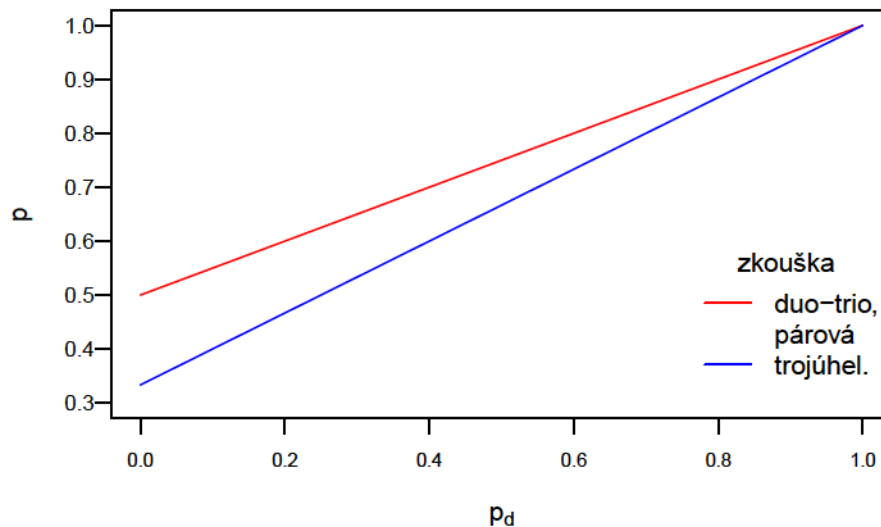
Pro trojúhelníkovou zkoušku

$$f_{\Delta}^N(p_d) = p_d + (1 - p_d)/3, \quad p_d \in [0,1].$$

Pro párovou porovnávací zkoušku

$$f_{PP}^N(p_d) = p_d + (1 - p_d)/2 \quad p_d \in [0,1].$$

Hodnoty $f_{DT}^N(p_d)$, $f_{\Delta}^N(p_d)$, $f_{PP}^N(p_d)$ jsou pravděpodobnosti správných odpovědí v příslušných zkouškách při dané hodnotě p_d . Parametr p_d je míra toho, jak moc se výrobky vnímatelně liší. Psychometrická funkce z definice 4.1 tedy popisuje vztah mezi p pravděpodobností správné odpovědi posuzovatele a p_d podílem posuzovatelů v populaci, kteří rozeznají výrobky. Psychometrické funkce jsou vykresleny na obrázku 4.3. Obě psychometrické funkce jsou afinními funkcemi p_d . Psychometrická funkce příslušící trojúhelníkové zkoušce roste rychleji než psychometrická funkce ve zkouškách duo-trio a párové porovnávací.



Obrázek 4.3: Psychometrické funkce v modelu založeném na p_d .

4.6 Thurstonovský přístup

Ve všech třech uvedených normách je v poznámce pod čarou zmíněno, že pro danou zkoušku může být použit Thurstone–Ura model. Vysvětleme, jak tento model vypadá (Bradley (1963)).

Modely použité v normách ve zkoušce duo-trio, trojúhelníkové a párové porovnávací se snaží zachytit chování posuzovatelů. Například práce Ennis (1993); Frijters (1979); Hopkins a Gridgeman (1955) se zmiňují o nesouladu modelu s chováním posuzovatelů. Thurstone–Ura model poskytuje alternativní model. Zjišťování, který z modelů je v praxi vhodnější, není součástí této práce.

4.6.1 Thurstone–Ura model

Thurstone–Ura model je založený na Thurstonově myšlence porovnávání podnětů (Thurstone (1927 a,b)).

Data získáváme stejným způsobem jako ve zkoušce duo-trio, trojúhelníkové a párové porovnávací zkoušce v normách. Posuzovatelům předkládáme vzorky výrobků A a B .

Oproti normám je rozdíl v pravděpodobnostním modelu. Předpokládá se, že hodnocení výrobku A se řídí normálním rozdělením $\mathcal{N}(0, \sigma^2)$, hodnocení výrobku B se řídí rozdělením $\mathcal{N}(\mu, \sigma^2)$, kde $\mu \geq 0$, $\sigma^2 > 0$, a hodnocení jsou nezávislá. Předpokládá se, že když posuzovatel hodnotí výrobek A , realizuje se veličina s rozdělením $\mathcal{N}(0, \sigma^2)$, a když hodnotí výrobek B , realizuje se veličina s rozdělením $\mathcal{N}(\mu, \sigma^2)$.

Vnímatelný rozdíl mezi výrobky vyjadřují parametry μ a σ^2 . Ukážeme, že vnímatelný rozdíl výrobků lze popsat jediným parametrem, a to $\delta = \mu/\sigma$.

Trojúhelníková zkouška

Bez újmy na obecnosti předpokládejme, že $X_1, X_2 \sim \mathcal{N}(0, \sigma^2)$ jsou nezávislá hodnocení stejného výrobku a $Y \sim \mathcal{N}(\mu, \sigma^2)$ je hodnocení odlišného výrobku.

Předpokládáme, že posuzovatel odpoví správně v trojúhelníkové zkoušce, když $|X_1 - X_2| < |X_1 - Y|$ a $|X_1 - X_2| < |X_2 - Y|$, tj. když se vzorky X_1 a X_2 jeví „více podobně“ než X_1 a Y i než X_2 a Y . Situace, kdy posuzovatel odpoví správně, jsou:

$$A_1 = \{X_1 < X_2 < Y \ \& \ X_2 - X_1 < Y - X_2\},$$

$$A_2 = \{Y < X_2 < X_1 \ \& \ X_1 - X_2 < X_2 - Y\},$$

$$A_3 = \{X_2 < X_1 < Y \ \& \ X_1 - X_2 < Y - X_1\},$$

$$A_4 = \{Y < X_1 < X_2 \ \& \ X_2 - X_1 < X_1 - Y\}.$$

Tyto čtyři disjunktní jevy lze souhrnně zapsat jako

$$\left\{ \left| Y - \frac{X_1 + X_2}{2} \right| > \frac{3}{2} |X_1 - X_2| \right\}.$$

Označme

$$U = \frac{1}{\sigma} \sqrt{\frac{2}{3}} \left(Y - \frac{X_1 + X_2}{2} \right) \sim \mathcal{N} \left(\delta \sqrt{\frac{2}{3}}, 1 \right), \quad V = \frac{1}{\sigma} \frac{3}{2} \sqrt{\frac{2}{3}} (X_1 - X_2) \sim \mathcal{N}(0, 3).$$

Dále označme pravděpodobnost správné odpovědi v trojúhelníkové zkoušce jako p_Δ . Potom

$$p_\Delta = \mathbf{P}(|U| > |V|) = \mathbf{P} \left(\frac{|U|}{|V|/\sqrt{3}} > \sqrt{3} \right) = \mathbf{P} \left(\frac{U^2}{V^2/3} > 3 \right)$$

Veličiny U a V jsou nezávislé, což lze snadno ukázat. Náhodná veličina $U^2/(V^2/3)$ má necentrální F-rozdělení s 1 a 1 stupněm volnosti a parametrem necentrality $2\delta^2/3$.

Uveďme ještě jiný výpočet pravděpodobnosti správné odpovědi. Rozdělme jev $\{|U| > |V|\}$ na čtyři disjunktní jevy:

$$B_1 = \{U > 0 \ \& \ V > 0 \ \& \ |U| > |V|\}, \quad B_2 = \{U > 0 \ \& \ V < 0 \ \& \ |U| > |V|\},$$

$$B_3 = \{U < 0 \ \& \ V > 0 \ \& \ |U| > |V|\}, \quad B_4 = \{U < 0 \ \& \ V < 0 \ \& \ |U| > |V|\}.$$

(4.21)

Protože V a $-V$ jsou stejně rozdělené a U a V jsou nezávislé, je $P(B_1) = P(B_2)$ a $P(B_3) = P(B_4)$. Označme ϕ , resp. ϕ_U , resp. ϕ_V hustotu standardizovaného normálního rozdělení, resp. hustotu U , resp. hustotu V a označme Φ , resp. Φ_U , resp. Φ_V příslušné distribuční funkce.

$$\begin{aligned} \mathbf{P}(B_1) &= \int_0^\infty \left[\phi_V(v) \int_v^\infty \phi_U(u) du \right] dv = \int_0^\infty \phi_V[1 - \Phi_U(v)] dv \\ &= \int_0^\infty \phi_V \Phi_{-U}(-v) dv \\ \mathbf{P}(B_3) &= \int_0^\infty \phi_V(v) \left[\int_{-\infty}^{-v} \phi_U(u) du \right] dv = \int_0^\infty \phi_V(v) \Phi_U(-v) dv. \end{aligned}$$

Pravděpodobnost správné odpovědi je

$$\begin{aligned} \mathbb{P}(|U| > |V|) &= 2[\mathbb{P}(B_1) + \mathbb{P}(B_3)] = 2 \int_0^\infty \phi_V(v) [\Phi_{-U}(-v) + \Phi_U(-v)] dv \\ &= 2 \int_0^\infty (1/\sqrt{3}) \phi(v/\sqrt{3}) \left[\Phi\left(-v + \delta\sqrt{2/3}\right) + \Phi\left(-v - \delta\sqrt{2/3}\right) \right] dv \\ &= 2 \int_0^\infty \phi(t) \left[\Phi\left(-t\sqrt{3} + \delta\sqrt{2/3}\right) + \Phi\left(-t\sqrt{3} - \delta\sqrt{2/3}\right) \right] dt. \end{aligned}$$

Zkouška duo-trio

Bez újmy na obecnosti předpokládejme, že $X_1, X_2 \sim \mathcal{N}(0, \sigma^2)$ jsou hodnocení dvou vzorků stejného výrobku, přičemž X_1 je referenční a $Y \sim \mathcal{N}(\mu, \sigma^2)$ je hodnocení odlišného vzorku. Předpokládáme, že posuzovatel odpoví správně, když $|X_1 - X_2| < |X_1 - Y_1|$, tj. když se vzorky X_1 a X_2 jeví „podobněji“ než vzorky X_1 a Y . Situace, kdy posuzovatel odpoví správně, jsou:

$$A_1 = \{X_1 < X_2 < Y\},$$

$$A_2 = \{Y < X_2 < X_1\},$$

$$A_3 = \{X_2 < X_1 < Y \text{ \& } X_1 - X_2 < Y - X_1\},$$

$$A_4 = \{Y < X_1 < X_2 \text{ \& } X_2 - X_1 < X_1 - Y\}.$$

Všimněme si, že

$$\mathbb{P}(X_1 < X_2 < Y \text{ nebo } Y < X_2 < X_1) = \mathbb{P}\left(\left|Y - \frac{X_1 + X_2}{2}\right| > \frac{|X_1 - X_2|}{2}\right)/2.$$

Označme

$$U = \frac{1}{\sigma} \sqrt{\frac{2}{3}} \left(Y - \frac{X_1 + X_2}{2}\right) \sim N\left(\delta \sqrt{\frac{2}{3}}, 1\right), \quad V = \frac{1}{\sigma} \sqrt{\frac{2}{3}} \frac{(X_1 - X_2)}{2} \sim \mathcal{N}\left(0, \frac{1}{3}\right).$$

Potom

$$\mathbb{P}(X_1 < X_2 < Y \text{ nebo } Y < X_2 < X_1) = \mathbb{P}(|U| > |V|)/2 = \mathbb{P}\left(\frac{U^2}{3V^2} > \frac{1}{3}\right)/2.$$

Veličiny U a V jsou nezávislé, což lze snadno ověřit. Náhodná veličina $(U^2)/(3V^2)$ má necentrální F-rozdělení s 1 a 1 stupněm volnosti a parametrem necentrality $(2\delta^2/3)$. Všimněme si, že pravděpodobnost, že nastaneme jev A_3 nebo A_4 , je polovina pravděpodobnosti správné odpovědi v trojúhelníkové zkoušce. Označme p_{DT} pravděpodobnost správné odpovědi ve zkoušce duo-trio. Potom

$$p_{DT} = \mathbb{P}(Z > 1/3)/2 + \mathbb{P}(Z > 3)/2,$$

kde Z má necentrální F-rozdělení s 1 a 1 stupněm volnosti a parametrem necentrality $2\delta^2/3$.

Opět uvedme ještě další výpočet správné odpovědi. Rozdělme $\{|U| > |V|\}$ na čtyři disjunktní jevy B_1 – B_4 stejně jako v trojúhelníkové zkoušce, viz (4.21). Na základě stejné argumentace, tj. V a $-V$ mají stejné rozdělení a U a V jsou

nezávislé, je $P(B_1) = P(B_2)$ a $P(B_3) = P(B_4)$. Vyjádření $P(B_1)$ a $P(B_3)$ je obdobné jako v trojúhelníkové zkoušce.

$$\begin{aligned} P(|U| > |V|) &= 2[P(B_1) + P(B_3)] = 2 \int_0^\infty \phi_V(v) [\Phi_{-U}(-v) + \Phi_U(-v)] dv \\ &= 2 \int_0^\infty \sqrt{3} \phi(v\sqrt{3}) \left[\Phi\left(-v + \delta\sqrt{2/3}\right) + \Phi\left(-v - \delta\sqrt{2/3}\right) \right] dv \\ &= 2 \int_0^\infty \phi(t) \left[\Phi\left(-t/\sqrt{3} + \delta\sqrt{2/3}\right) + \Phi\left(-t/\sqrt{3} - \delta\sqrt{2/3}\right) \right] dt. \end{aligned}$$

Pravděpodobnost, že nastaneme jev A_1 nebo A_2 je $P(|U| > |V|)/2$ a pravděpodobnost, že nastaneme jev A_3 nebo A_4 , je $1/2$ pravděpodobnosti správné odpovědi v trojúhelníkové zkoušce. Proto

$$\begin{aligned} p_{\text{DT}} &= \int_0^\infty \phi(t) \left[\Phi\left(-t/\sqrt{3} + \delta\sqrt{2/3}\right) + \Phi\left(-t/\sqrt{3} - \delta\sqrt{2/3}\right) \right. \\ &\quad \left. + \Phi\left(-t\sqrt{3} + \delta\sqrt{2/3}\right) + \Phi\left(-t\sqrt{3} - \delta\sqrt{2/3}\right) \right] dt. \end{aligned}$$

Párová porovnávací zkouška

Označme $X \sim \mathcal{N}(0, \sigma^2)$ a $Y \sim \mathcal{N}(\mu, \sigma^2)$ hodnocení dvou různých výrobků. Posuzovatel odpoví správně v případě $X < Y$, tj. když se výrobek Y jeví, že má intenzivnější zkoumanou vlastnost. Označme p_{PP} pravděpodobnost správné odpovědi v párové porovnávací zkoušce. Potom

$$p_{\text{PP}} = P(X < Y) = P(X - Y < 0) = \Phi(\delta/\sqrt{2}).$$

4.6.2 Hypotézy a testování v Thurstone–Ura modelu

V thurstonovském modelu místo p_d vystupuje δ . Pravděpodobnost správné odpovědi v thurstonovském modelu závisí na jediném parametru $\delta = \mu/\sigma$. Pokud je $\delta = 0$, dva výrobky působí stejně. Čím vyšší je δ , tím větší je vnímatelný rozdíl mezi výrobky. Hypotézy v thurstonovském modelu sestavujeme o parametru δ , a to analogicky jako hypotézy o parametru p_d v modelu používaném v normách.

Při zjišťování, zda je mezi výrobky vnímatelný rozdíl, ve zkoušce duo-trio, trojúhelníkové a párové porovnávací zkoušce sestavujeme hypotézy jako

$$H_0 : \delta = 0 \quad \text{proti} \quad H_1 : \delta > 0,$$

což je ekvivalentní

$$\begin{aligned} H_0 : p_{\text{DT}} = 1/2, \quad \text{resp.} \quad p_{\Delta} = 1/3, \quad \text{resp.} \quad p_{\text{PP}} = 1/2 \\ \text{proti} \\ H_1 : p_{\text{DT}} > 1/2, \quad \text{resp.} \quad p_{\Delta} > 1/3, \quad \text{resp.} \quad p_{\text{PP}} > 1/2. \end{aligned} \tag{4.22}$$

Testy hypotéz (4.22) lze sestavit naprosto analogicky jako testy pro rozdíl (u příslušných zkoušek), které jsou uvedeny v normách. Pro „malé“ počty posuzovatelů použijeme přesný test založený na binomickém rozdělení a pro „velké“ počty posuzovatelů použijeme asymptotický test.

Při zjišťování, zda si jsou výrobky dostatečně podobné, sestavujeme hypotézy jako

$$H_0 : \delta \geq \delta_0 \quad \text{proti} \quad H_1 : 0 \leq \delta < \delta_0$$

pro pevně zvolené $\delta_0 > 0$, což je ekvivalentní

$$\begin{aligned} H_0 : p_{DT} \geq p_{DT_0}, \quad \text{resp.} \quad p_{\Delta} \geq p_{\Delta_0}, \quad \text{resp.} \quad p_{PP} \geq p_{PP_0} \\ \text{proti} \\ H_1 : p_{DT} < p_{DT_0}, \quad \text{resp.} \quad p_{\Delta} < p_{\Delta_0}, \quad \text{resp.} \quad p_{PP} < p_{PP_0}. \end{aligned} \quad (4.23)$$

Testy hypotéz (4.23) lze opět naprosto analogicky sestavit jako testy pro podobnost (u příslušných zkoušek), které jsou uvedené v normách. Pro „malé“ počty posuzovatelů použijeme přesný test založený na binomickém rozdělení a pro „velké“ počty posuzovatelů použijeme asymptotický test.

Vztahy mezi správnými odpověďmi $p_{DT}, p_{\Delta}, p_{PP}$ a δ jsou v thurstonovském modelu jiné než vztahy mezi $p_{DT}, p_{\Delta}, p_{PP}$ a p_d v modelu uvedeném v normách. Definujme psychometrické funkce v thurstonovském modelu.

Definice 4.2. *Psychometrickou funkci v Thurstone–Ura modelu definujeme následovně:*

Pro zkoušku duo-trio

$$f_{DT}^T(\delta) = P(Z > 1/3)/2 + P(Z > 3)/2, \quad \delta \in [0, \infty),$$

kde Z má necentrální F -rozdělení s 1 a 1 stupněm volnosti a parametrem necentrality $2\delta^2/3$.

Pro trojúhelníkovou zkoušku

$$f_{\Delta}^T(\delta) = P(Z > 3), \quad \delta \in [0, \infty),$$

kde Z má necentrální F -rozdělení s 1 a 1 stupněm volnosti a parametrem necentrality $2\delta^2/3$.

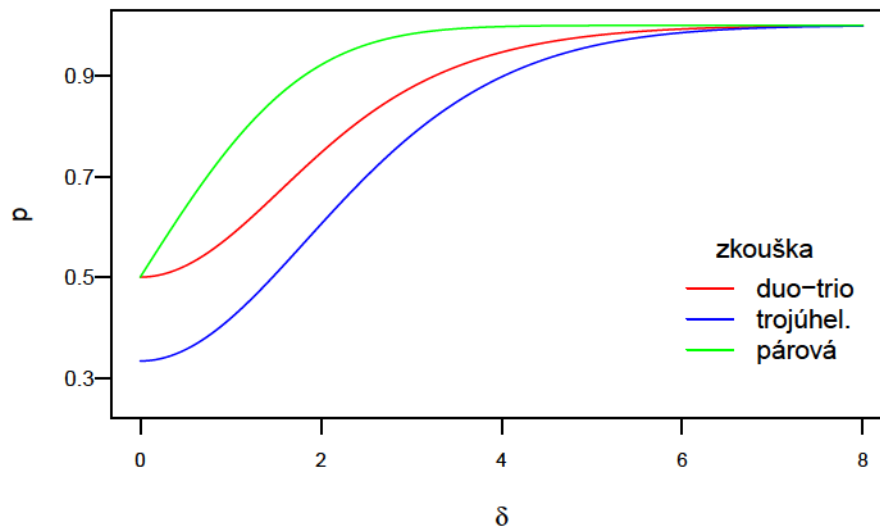
Pro párovou porovnávací zkoušku

$$f_{PP}^T(\delta) = \Phi(\delta/\sqrt{2}), \quad \delta \in [0, \infty),$$

kde Φ je distribuční funkce standardizovaného normálního rozdělení.

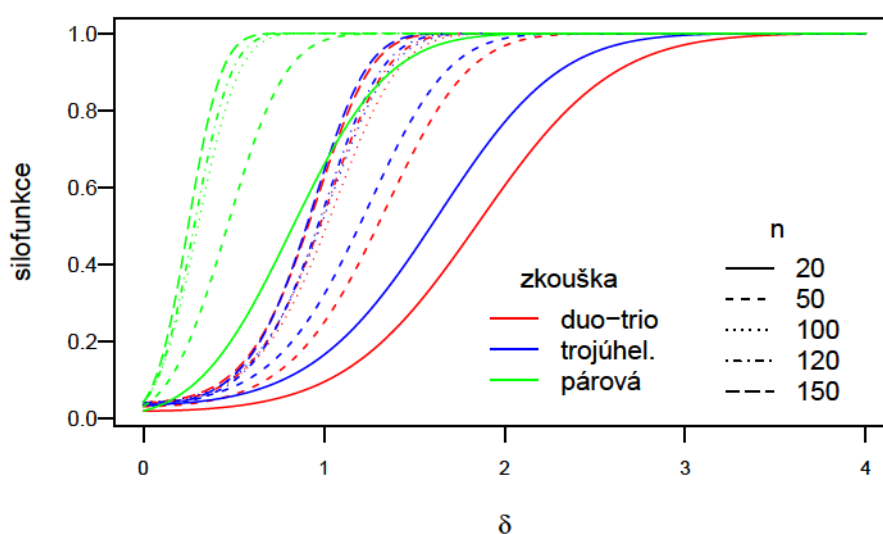
Hodnoty $f_{DT}^T(\delta), f_{\Delta}^T(\delta), f_{PP}^T(\delta)$ jsou pravděpodobnosti správných odpovědí pro danou hodnotu δ v příslušných zkouškách. Zde je δ míra toho, jak moc jsou výrobky vnímatelně odlišné. Pravděpodobnost správné odpovědi (v libovolné zkoušce) označme p . Psychometrické funkce z definice 4.2 popisují vztahy mezi pravděpodobnostmi správné odpovědi p a δ . Psychometrické funkce jsou vykresleny na obrázku 4.4. Pro porovnání vztahů p a p_d oproti p a δ se odkažme na obrázky 4.3 a 4.4.

Rozdílné vztahy mezi p a δ zapříčiňují rozdílnou sílu testů v Thurstone–Ura modelu oproti modelu v normách. Obdobně jako když jsme porovnávali sílu testů v modelech z norem, mluvíme o síle testu proti alternativě δ , nikoliv p . Podívejme se na vybrané případy.



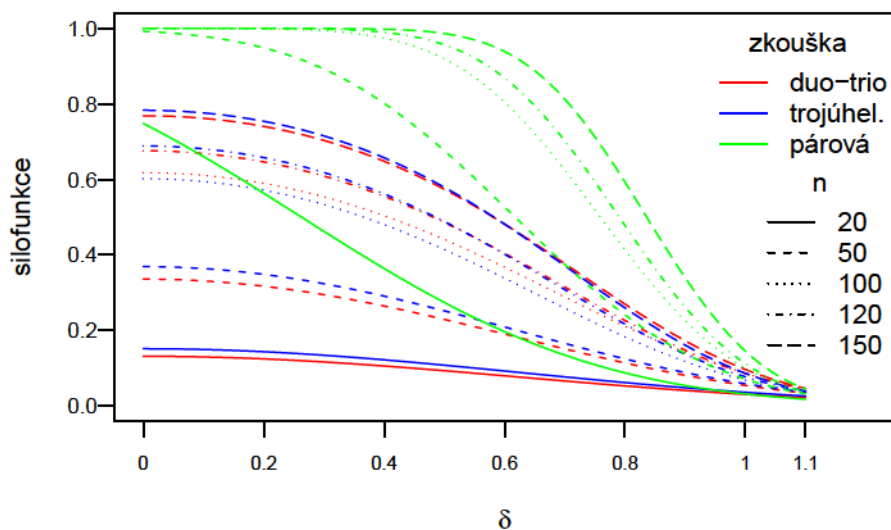
Obrázek 4.4: Psychometrické funkce v modelu založeném na δ .

Zaměříme se na testy pro rozdíl hypotéz (4.22). Stanovme hladinu významnosti 0.05 a počty posuzovatelů $n = 20, 50, 100, 120, 150$. Test v párové porovnávací zkoušce má proti většině hodnot δ , kde $\delta \in (0, 4)$, větší sílu než zbylé dva testy. Při hodnotách $n = 20, 50, 100$ má test v trojúhelníkové zkoušce proti většině hodnot δ , kde $\delta \in (0, 4)$, větší sílu než test ve zkoušce duo-trio. v případech, kdy $n = 120, 150$ a δ je přibližně z intervalu $(0, 0.75)$, je silnější test ve zkoušce duo-trio než test v trojúhelníkové zkoušce a pro hodnoty δ , které jsou přibližně z intervalu $(0.75, 4)$, je pořadí testů naopak. Silofunkce testů jsou vykresleny na obrázku 4.5. Sílu jsme zkoumali pro hodnoty $\delta \in (0, 4)$, protože pro daná n je síla všech tří testů pro $\delta \geq 4$ velmi podobná a blízká jedné.



Obrázek 4.5: Silofunkce testu pro rozdíl v thurstonovské modelu se stanovenou hladinou významnosti 0.05 a počtem posuzovatelů n .

Nyní se podíváme na testy pro podobnost hypotéz (4.23). Stanovme hodnotu $\delta_0 = 1.1$, hladinu významnosti 0.05 a počty posuzovatelů $n = 20, 50, 100, 120, 150$. Test v párové porovnávací zkoušce má opět proti většině hodnot δ , kde $\delta \in [0, \delta_0)$, největší sílu. Další pořadí testů není tak jasné. V případě $n = 20, 50$ je pro většinu hodnot δ silnější test trojúhelníkové zkoušky. V případě $n = 100$ je pro většinu hodnot δ silnější test zkoušky duo-trio. V případech, kdy $n = 120, 150$, a δ je přibližně z intervalu $(0.55, 1.1)$, je test zkoušky duo-trio silnější než test trojúhelníkové zkoušky. Naopak v případě, kdy je δ přibližně z intervalu $[0, 0.55)$, je silnější test zkoušky duo-trio než trojúhelníkové. Silofunkce testů jsou vykresleny na obrázku 4.6.



Obrázek 4.6: Silofunkce testu pro podobnost v thurstonovské modelu se stanovenou hladinou významnosti 0.05, zvoleným $\delta_0 = 1.1$ a počtem posuzovatelů n .

Závěr

V této práci jsme se věnovali několika tématům vztaženým k úloze, zda zákazníci poznají rozdíl mezi výrobky.

K tomu bylo třeba věnovat se pozměněné úloze testování hypotéz, a to testům ekvivalence a noninferiority. V první kapitole jsme čtenáře seznámili s myšlenkou těchto testů a shrnuli jsme teorii o stejnoměrně nejsilnějších testech ekvivalence a noninferiority v jednoparametrických rodinách exponenciálního typu a rodinách hustot striktně totálně pozitivních řádu tři. Ačkoliv je známo, jak tyto testy mají vypadat, v praxi není snadné jejich tvar nalézt. Proto jsme popsali dva algoritmy pro spojitá a diskrétní rozdělení inspirované knihou Wellek (2003), pomocí kterých je možné tvary testů (alespoň přibližně) nalézt. Algoritmy jsme doplnili větou ukazující jejich správnost. Zjistili jsme přitom, že algoritmus pro spojitá rozdělení má v praxi v některých případech špatné numerické vlastnosti a nemusí dobře fungovat. Popsali jsme také jiné testy ekvivalence založené na tzv. TOST proceduře, které se hojně užívají v praxi. Velkou výhodou těchto testů je jejich snadná konstrukce, naopak slabou stránkou je, že mohou být velmi konzervativní.

Úloha, zda zákazníci poznají rozdíl mezi výrobky, patří mezi úlohy tzv. senzorní analýzy. V senzorní analýze lidé hodnotí výrobky svými smysly. Abychom v těchto úlohách mohli použít statistické metody, je nutné popsat, jak lidé výrobky (podněty) svými smysly hodnotí. Jak navrhnout pravděpodobnostní model v oblasti senzorní analýzy, není vůbec zřejmé. Jeden z prvních, kdo se snažil popsat působení podnětů na člověka (či hodnocení podnětů člověkem) pomocí pravděpodobnostního modelu, byl americký psycholog, statistik a psychometr L.L. Thurstone. Jeho myšlenka porovnávání podnětů je věnována druhé kapitole. Parametry v thurstonovském modelu poskytují možnost, jak podněty porovnávat. Thurstone (1927a) vytvořil metodu pro odhad parametrů založenou na párových porovnáváních. Navrhl soustavu částečně odhadnutých rovnic, ze které se odhady parametrů získávají. Částečným odhadnutím rovnic jsou do soustavy rovnic vneseny chyby a přesné řešení soustavy nemusí existovat. V případě, kdy v Thurstonově modelu uvažujeme rovnost rozptylů a nezávislost hodnocení, jsme odvodili explicitní tvar odhadů parametrů, který je založen na metodě nejmenších čtverců. Odhady parametrů v případech, kdy jsou rozptyly různé či v modelu vystupují nenulové korelační koeficienty, mohou být námětem pro další bádání.

Ve třetí kapitole jsme se věnovali odhadování vektoru priorit. Vektor priorit je užitečný, když vybíráme z n objektů nebo když chceme objekty seřadit či porovnat, a je také základem pro komplikovanější úlohy, které v této práci neřešíme. Pole uplatnění je široké a využití má i v senzorní analýze, což jsme v této práci pouze naznačili. Podrobné studium této problematiky by mohlo být zajímavým návrhem další diplomové práce. V průběhu studijního pobytu jsme se s profesorem Genestem zabývali novým způsobem odhadování vektoru priorit, který je inspirován Saatyho návrhem (Saaty (1977)) a je založen na metodě nejmenších čtverců. Avšak v simulační studii, kterou jsme provedli, se nový způsob chová v průměru o trochu hůře než Saatyho původní návrh.

K úloze senzorní analýzy typu, zda zákazníci poznají rozdíl mezi výrobky, existuje řada nepovinných norem. Zaměřili jsme se na tři běžně používané normy,

a to zkoušku duo-trio, trojúhelníkovou a párovou porovnávací zkoušku. Ukázalo se, že normy mají dobrý statistický podklad, který ale v normách není dobře vysvětlen. Podrobně jsme statistický základ norem popsali. Vysvětlili jsme, co znamenají číselné tabulky v normách, na jejichž základě se rozhoduje o závěru zkoušek a také slouží k doporučení počtu účastníků zkoušek. Ačkoliv se jednalo o přesné a asymptotické testy o binomickém rozdělení, zpětná konstrukce byla někdy velmi pracná. K práci jsme přiložili CD, na kterém jsou tři R skripty s naprogramovanými funkcemi, které slouží k výpočtu tabulek z norem a k počítání hladiny a síly příslušných testů. Zjistili jsme, že v normě, která přísluší párové porovnávací zkoušce, se používá test (test pro podobnost, případ (ii)), který není optimální. Pomocí teorie stejnoměrně nejsilnějších testů ekvivalence, kterou jsme zpracovali v první kapitole, jsme navrhli alternativní test, který je méně konzervativní a silnější než test použitý v normě. Alternativní test jsme také přidali do R skriptu příslušícímu párové porovnávací zkoušce, který je na přiloženém CD.

Testy ve zkouškách duo-trio, trojúhelníkové a párové porovnávací, které jsou uvedeny v normách, jsme na základě silofunkcí porovnali. Také jsme v daných zkouškách popsali psychometrickou funkci, která ukazuje vztah mezi odpovědí posuzovatele a podnětem. Dále jsme ukázali alternativní modely k modelům z norem, které jsou založeny na Thurstonově myšlence měření podnětů a které byly zmíněny v normách jako další možné modely. V těchto modelech jsme opět porovnali testy na základě síly a ukázali psychometrickou funkci.

Dodatek

Definice D.1. Necht rozdělení náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)^\top$ patří do množiny pravděpodobnostních rozdělení \mathcal{P} . Bud $\theta \in \mathbb{R}^k$ nějaká charakteristika rozdělení \mathbf{X} , nazývejme ji parametr. Předpokládejme, že $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, kde $\Theta \subseteq \mathbb{R}^k$ je borelovská množina a P_θ značí pravděpodobnost rozdělení \mathbf{X} , pokud je θ skutečná hodnota parametru. Předpokládejme, že pro každé $\theta \in \Theta$ je P_θ absolutně spojitá míra vzhledem k nějaké σ -konečné míře μ a příslušnou hustotu $dP_\theta/d\mu$ označme p_θ . Množinu Θ nazývejme parametrický prostor a množinu rozdělení $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ model.

Vždy předpokládejme podmínku identifikovatelnosti $P_{\theta_1} = P_{\theta_2} \Rightarrow \theta_1 = \theta_2$. V celé práci uvažujme parametrický prostor $\Theta \subseteq \mathbb{R}^k$ a dvě jeho neprázdné podmnožiny $\Theta_0, \Theta_1 \subset \Theta$, které splňují $\Theta_0 \cap \Theta_1 = \emptyset$ a $\Theta_0 \cup \Theta_1 = \Theta$. Ve shodě s literaturou budeme Θ_0 nazývat (nulovou) hypotézou a Θ_1 alternativou.

Definice D.2. Necht rozdělení náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)^\top$ patří do rodiny rozdělení $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. Mějme statistické hypotézy $H_0 : \theta \in \Theta_0$ a $H_1 : \theta \in \Theta_1$. Zvolme množinu $W \subset \mathbb{R}^n$. Postup takový, že H_0 zamítneme, když $\mathbf{X} \in W$, nazývejme statistický test a W kritický obor.

Definice D.3. Necht rozdělení náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)^\top$ patří do rodiny rozdělení $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. Uvažujme statistický problém testování hypotézy $H_0 : \theta \in \Theta_0$ proti $H_1 : \theta \in \Theta_1$ a statistický test s kritickým oborem $W \subset \mathbb{R}^n$.

- Necht $\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in W) = \alpha \in (0,1)$. Potom α nazývejme hladina významnosti testu.
- Funkci $\beta(\theta) = P_\theta(\mathbf{X} \in W), \theta \in \Theta$, nazývejme silofunkce. Pro $\theta \in \Theta_1$ říkejme, že $\beta(\theta)$ je síla testu proti alternativě θ .

Někdy se místo silofunkce $\beta(\theta)$ používá funkce $1 - \beta(\theta)$, které se říká operační charakteristika testu.

Definice D.4. Necht rozdělení náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)^\top$ patří do rodiny rozdělení $\mathcal{P} = \{P_\theta, \theta \in \Theta \subseteq \mathbb{R}^k\}$. Mějme statistický problém testování hypotézy $H_0 : \theta \in \Theta_0$ proti $H_1 : \theta \in \Theta_1$. Necht $\phi : \mathbb{R}^n \rightarrow [0,1]$ je měřitelná funkce a $\mathbf{x} \in \mathbb{R}^n$ realizace náhodného vektoru \mathbf{X} . Postup takový, že pro dané \mathbf{x} zamítáme nulovou hypotézu s pravděpodobností $\phi(\mathbf{x})$, nazývejme znáhodněný test a ϕ kritická (nebo testová) funkce.

Poznámka.

- Pokud je P_θ skutečné pravděpodobnost rozdělení náhodného vektoru \mathbf{X} , potom je pravděpodobnost zamítnutí nulové hypotézy rovna

$$E_\theta \phi(\mathbf{X}) = \int \phi(\mathbf{x}) dP_\theta(\mathbf{x}).$$

- Pokud kritická funkce ϕ nabývá pouze hodnot $\{0,1\}$, jedná se o statistický test z definice D.2. Statistické testy jsou podmnožinou znáhodněných statistických testů.

Definice D.5 je rozšířením definice D.3 na znáhodněné testy.

Definice D.5. *Nechť rozdělení náhodného vektoru $\mathbf{X} = (X_1, \dots, X_n)^\top$ patří do rodiny rozdělení $\mathcal{P} = \{P_\theta, \theta \in \Theta \subseteq \mathbb{R}^k\}$. Mějme statistický problém testování hypotézy $H_0 : \theta \in \Theta_0$ proti $H_1 : \theta \in \Theta_1$ a znáhodněný test s kritickou funkcí $\phi : \mathbb{R}^n \rightarrow [0,1]$.*

- *Nechť $\sup_{\theta \in \Theta_0} E_\theta \phi(\mathbf{X}) = \alpha \in (0,1)$. Potom α nazýváme hladina významnosti znáhodněného testu.*
- *Funkci $\beta(\theta) = E_\theta \phi(\mathbf{X})$ pro $\theta \in \Theta$ nazýváme silofunkce. Pro $\theta \in \Theta_1$ říkáme, že $\beta(\theta)$ je síla znáhodněného testu proti alternativě θ .*

Poznamenejme, že (znáhodněný) test má hladinu významnosti α právě tehdy, když $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$. Při konstrukci testu stanovíme $\alpha \in (0,1)$ a požadujeme, aby pro každé $\theta \in \Theta_0$ platilo $\beta(\theta) \leq \alpha$.

Definice D.6. *Znáhodněný test hypotézy $H_0 : \theta \in \Theta_0$ proti $H_1 : \theta \in \Theta_1$ s kritickou funkcí ϕ^* a hladinou α se nazývá stejnoměrně nejsilnější (zkráceně UMP) test, pokud je nejsilnější proti každé alternativě mezi všemi znáhodněnými testy s hladinou α .*

Poznámka. Definice D.6 jinými slovy: Pro nějakou kritickou funkci ϕ označme příslušnou silofunkci β_ϕ . Znáhodněný test hypotézy $H_0 : \theta \in \Theta_0$ proti $H_1 : \theta \in \Theta_1$ s kritickou funkcí ϕ^* a hladinou α je stejnoměrně nejsilnější, když platí

$$\beta_\phi(\theta) \leq \beta_{\phi^*}(\theta)$$

pro libovolný znáhodněný test ϕ s hladinou α a každé $\theta \in \Theta_1$.

Stejněměrně nejsilnější testy obecně nemusí existovat. Někdy se ale lze omezit na menší třídu testů, vzhledem k níž již stejnoměrně nejsilnější test existuje.

Definice D.7. *Znáhodněný test s kritickou funkcí ϕ^* a hladinou α se nazývá nestranný, pokud má proti každé alternativě sílu alespoň α .*

Definice D.8. *Znáhodněný statistický test s testovou hladinou $\alpha \in (0,1)$ se nazývá stejnoměrně nejsilnější nestranný (zkráceně UMPU) test, pokud je stejnoměrně nejsilnější mezi nestrannými testy.*

Seznam použité literatury

- BRADLEY, R. A. (1963). Some relationships among sensory difference tests. *Biometrics*, **19**(3), 385–397.
- ČSN EN ISO 10399 (2010). *Senzorická analýza - Metodologie - Zkouška duo-trio*. Úřad pro technickou normalizaci metrologii a státní zkušebnictví.
- ČSN EN ISO 4120 (2009). *Senzorická analýza - Metodologie - Trojúhelníková zkouška*. Úřad pro technickou normalizaci metrologii a státní zkušebnictví.
- ČSN EN ISO 5495 (2009). *Senzorická analýza - Metodologie - Párová porovnávací zkouška*. Úřad pro technickou normalizaci metrologii a státní zkušebnictví.
- ENNIS, D. M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, **8**(4), 353–370.
- FRIJTERS, J. E. R. (1979). The paradox of discriminatory nondiscriminators resolved. *Chemical Senses*, **4**(4), 355–358.
- HOPKINS, J. W. a GRIDGEMAN, N. T. (1955). Comparative sensitivity of pair and triad flavor intensity difference tests. *Biometrics*, **11**(1), 63–68.
- JOHNSON, C. R., BEINE, W. B. a WANG, T. J. (1979). Right-left asymmetry in an eigenvector ranking procedure. *Journal of Mathematical Psychology*, **19**(1), 61–64.
- KALLENBERG, W. C. M. a JANSSEN, P. (1984). *Testing Statistical Hypotheses: Worked Solutions*. Amsterdam : Centrum voor Wiskunde en Informatica.
- KARLIN, S. (1968). *Total Positivity*, volume I. Stanford University Press.
- LEHMANN, E. L. (1997). *Testing Statistical Hypotheses*. Springer, 2nd edition. ISBN 0-387-94919-4.
- PARDO, S. (2014). *Equivalence and Noninferiority Tests for Quality, Manufacturing and Test Engineers*. Chapman & Hall/CRC. ISBN 978-1-4665-8688-8.
- R CORE TEAM (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- SAATY, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, **15**(3), 234–281.
- SCHUIRMANN, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, **15**(6), 657–680.
- THURSTONE, L. L. (1927a). A law of comparative judgment. *Psychological Review*, **34**(4), 273–286.

- THURSTONE, L. L. (1927b). Psychophysical analysis. *The American Journal of Psychology*, **38**(3), 368–389.
- TSUKIDA, K. a GUPTA, M. R. (2011). How to analyze paired comparison data [online]. Tutoriál. URL <http://mayagupta.org/publications/ PairedComparisonTutorialTsukidaGupta.pdf>.
- VÍTOVÁ, E. (2011). Senzorická analýza – důležitý nástroj pro zvyšování kvality potravin [online]. URL <http://www.chempoint.cz/vitova>.
- WELLEK, S. (2003). *Testing Statistical Hypotheses of Equivalence*. Chapman & Hall/CRC. ISBN 1-58488-160-7.

Přílohy

A.1 Ukázka zdrojového kódu

Všechny výpočty v této diplomové práci jsme prováděli ve výpočetním prostředí R (R Core Team (2018)). K diplomové práci je přiloženo CD, na kterém jsou tři R skripty se zdrojovými kódy ke kapitole 4. Každý R skript přísluší jedné ze zkoušek duo-trio, trojúhelníkové a párové porovnávací (sekce 4.2, 4.3, 4.4). V každém R skriptu jsou naprogramované funkce k výpočtu tabulek A.*, které jsou uvedeny v normách, hladiny a síly příslušných přesných testů a některé další související výpočty. Ukažme vybrané části R skriptu příslušnému zkoušce duo-trio. Zbylé dva R skripty jsou podobné.

```
##### ZKOUŠKA DUO-TRIO #####
##### norma ČSN EN ISO 10399 (2010) #####

##### test pro rozdíl #####

##### tabulka A.1
k_duo_trio_rozdil <- function(n, alpha){
  # n - počet posuzovatelů
  # alpha - požadovaná hladina testu
  # Funkce vrátí hodnotu z~tabulky A.1 příslušnou buňce (n,alpha).
  # Pokud by test dodržující předepsanou hladinu nikdy nezamítal
  # nulovou hypotézu, funkce vrátí hodnotu -1.
  k~<- qbinom(1-alpha, n , 1/2) + 1
  ifelse(((0 <= k) & (k~<= n)), return(k), return(-1))
}

## vytiskne tabulku A.1
alphaset <- c(0.2, 0.1, 0.05, 0.01, 0.001)
nset <- c(6:30, seq(32, 88, 4))
for (n in nset) {
  cat("n = ", n, "|", sapply(alphaset,
    function(a) k_duo_trio_rozdil(n, a)), "\n") }
}
```

```
##### skutečná hladina testu pro rozdíl
hladina_duo_trio_rozdil <- function(n,alpha){
  # n - počet posuzovatelů
  # alpha - požadovaná hladina testu
  # Funkce vrátí skutečnou hladinu testu.
  k~<- k_duo_trio_rozdil(n, alpha)
  if(k != -1){
    h <- 1 - pbinom(k-1, n, 1/2)
    return(h)
  }else{return(0)} }
}
```

```
##### silofunkce testu pro rozdíl
silfce_duo_trio_rozdil <- function(n, alpha, pd){
  # n - počet posuzovatelů
  # alpha - hladina testu
  # pd - alternativa
  # Funkce vrátí sílu testu pro rozdíl proti alternativě pd.
  p <- pd + 1/2*(1-pd)
  k~<- k_duo_trio_rozdil(n, alpha)
  if(k == -1){
    return(0)
  }else{
    s~<- 1 - pbinom(k-1, n, p)
  }
  return(s)
}
```

```
##### test pro podobnost #####

#### tabulka A.2
l_duo_trio_podobnost <- function(n, beta, pd0){
  # n - počet posuzovatelů
  # beta - požadovaná hladina testu
  # pd0 - hraniční mez u~hypotéz
  # Funkce vrátí hodnotu z~tabulky A.2 příslušnou
  # buňce (n, beta, pd), kde pd = pd0. Pokud
  # by test s~předepsanou hladinou beta nikdy
  # nezamítal nulovou hypotézu, funkce vrátí
  # hodnotu -1.
  p0 <- pd0 + (1-pd0)*1/2
  a <- qbinom(beta, n, p0)
  if(pbinom(a, n, p0) == beta){
    l <- a
  }else{
    l <- a - 1
  }
  ifelse(((0 <= l) & (l <= n)), return(l), return(-1))
}

## vytiskne tabulku A.2
nset <- seq(20, 112, by = 4)
betaset <- c(0.001, 0.01, 0.05, 0.1, 0.2)
pdset <- c(0.1, 0.2, 0.3, 0.4, 0.5)
for (n in nset) {
  cat("n =", n, "\n")
  for (b in betaset) {
    print(c(sapply(pdset,
      function(p) l_duo_trio_podobnost(n,b,p)))) }
}
```

```
##### skutečná hladina testu pro podobnost
hladina_duo_trio_podobnost <- function(n, beta, pd0){
  # n - počet posuzovatelů
  # beta - požadovaná hladina testu
  # pd0 - hraniční mez u-hypotéz
  # Funkce vrací skutečnou hladinu testu pro podobnost.
  l <- l_duo_trio_podobnost(n, beta, pd0)
  if(l != -1){
    p0 <- pd0 + (1-pd0)*1/2
    h <- pbinom(l, n, p0)
    return(h)
  }else{return(0)}
}
```

```
##### tabulka A.3
# doporučený počet posuzovatelů na základě síly testu
# pro rozdíl nebo podobnost

## tabulka A.3 založená na síle testu pro rozdíl
tabA3_duo_trio_rozdil <- function(alpha, beta, pd){
  # pd - alternativa
  # alpha - požadovaná hladina testu
  # 1 - beta - požadovaná minimální síla testu proti pd
  # Funkce vrátí hodnotu v~tabulce A.3 příslušnou buňce
  # (alpha, beta, pd).
  n <- 1
  repeat{
    n <- n+1
    if(silfce_duo_trio_rozdil(n, alpha, pd) >= (1-beta)){break}
  }
  return(n)
}

## vytiskne tabulku A.3 založenou na testu pro rozdíl
pdset <- c(0.5, 0.4, 0.3, 0.2, 0.1)
alphaset <- c(0.2, 0.1, 0.05, 0.01, 0.001)
betaset <- c(0.2, 0.1, 0.05, 0.01, 0.001)

for (p in pdset) {
  cat("pd =", p, "\n")
  for (a in alphaset) {
    print(sapply(betaset,
      function(b) tabA3_duo_trio_rozdil(a, b, p)))
  }
}
```

```

## tabulka A.3 založená na síle testu pro podobnost
tabA3_duo_trio_podobnost <- function(alpha, beta, pd){
  # pd - alternativa
  # beta - požadovaná hladina testu
  # 1 - alpha - požadovaná minimální síla testu proti pd
  # Funkce vrátí hodnotu v~tabulce A.3 příslušnou buňce
  # (alpha, beta, pd).
  n <- 1
  repeat{
    n <- n+1
    if(silfce_duo_trio_podobnost(n, beta, pd, 0) >= (1-alpha)){
      break}
  }
  return(n)
}

## vytiskne tabulku A.3 založenou na síle testu pro podobnost
pdset <- c(0.5, 0.4, 0.3, 0.2, 0.1)
alphaset <- c(0.2, 0.1, 0.05, 0.01, 0.001)
betaset <- c(0.2, 0.1, 0.05, 0.01, 0.001)
for (p in pdset) {
  cat("pd =", p, "\n")
  for (a in alphaset) {
    print(sapply(betaset,
      function(b) tabA3_duo_trio_podobnost(a, b, p)))  }
}

```

A.2 Zdrojový kód k návrhu testu pro podobnost

Ukažme zdrojový kód k testu pro podobnost, který jsme navrhli v sekci 4.4.3 v poznámce na straně 46 jako zástupný test k původnímu testu z norem. Vstupními parametry funkce `l_ump_par_podobnost_ii` je trojice (n, β, pd_0) , kde n je počet posuzovatelů pro zkoušku, β je požadovaná hladina testu a pd_0 hraniční mez u hypotéz (v sekci 4.4.3 značená jako p_{d_0}). Funkce `l_ump_par_podobnost_ii` spočítá hodnotu $\tilde{\ell}$ z (4.18). Funkce `hladina_ump_par_podobnost_ii` se vstupními parametry (n, β, pd_0) spočítá skutečnou hladinu významnosti testu. Funkce `silfce_ump_par_podobnost_ii` se vstupním parametry (n, β, pd_0, pd) spočítá hodnotu silofunkce v bodě pd .

```

##### Návrh zástupného testu pro podobnost v~případě (ii),
##### když vedoucí zkoušky neví, který výrobek má
##### intenzivnější zkoumanou vlastnost.

l_ump_par_podobnost_ii <- function(n, beta, pd0){
  # n - počet posuzovatelů
  # beta - požadovaná hladina testu

```

```

# pd0 - hraniční mez u-hypotéz
# Funkce vrátí hodnotu z~tabulky A.2 příslušnou buňce
# (n, beta, pd), kde pd = pd0. Pokud by test
# s~předeřsanou hladinou beta nikdy nezamítal
# nulovou hypotézu, funkce vrátí hodnotu -1.
p0 <- 1/2 + 1/2 * pd0
l <- floor((n+1)/2)-1
repeat{
  l <- l + 1
  if(l >= n){
    l <- -1
    break}
  if( (pbinom(l, n, p0) - pbinom(n-l-1, n, p0)) <= beta &
      (pbinom(l+1, n, p0) - pbinom(n-l-2, n, p0)) > beta){break}
}
return(l)
}

#Skutečná hladina testu
hladina_ump_par_podobnost_ii <- function(n, beta, pd0){
  # n - počet posuzovatelů
  # beta - požadovaná hladina testu
  # Funkce vrátí skutečnou hladinu testu.
  l <- l_ump_par_podobnost_ii(n, beta, pd0)
  if(l == -1){
    return(0)
  }else{
    p0 <- 1/2 + 1/2 * pd0
    h <- pbinom(l, n, p0) - pbinom(n-l-1, n, p0)
    return(h)
  }
}

# síla testu
silfce_ump_par_podobnost_ii <- function(n, beta, pd0, pd){
  # n - počet posuzovatelů
  # beta - hladina testu
  # pd0 - hraniční mez u-hypotéz
  # pd - alternativa
  # Funkce vrátí hodnotu silofunkce v~bodě pd.
  l <- l_ump_par_podobnost_ii(n, beta, pd0)
  if(l == -1){ return(0) }else{
    p <- pd + 1/2*(1-pd)
    s~<- pbinom(l, n, p) - pbinom(n-l-1, n, p)
    return(s)
  }
}

```