Univerzita Karlova

Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

2021 **Karolína Švarcová**

MATEMATICKO-FYZIKÁLNÍ
FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Karolína Švarcová

Sensometrické diskriminační testování – porovnání párové porovnávací zkoušky a pořadové zkoušky

Katedra pravděpodobnosti a matematické statistiky

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

Studijní program: Matematika

Studijní obor: Finanční matematika

Praha 2021

Chtěla bych poděkovat panu profesoru Jaromíru Antochovi za zapůjčení literatury, za cenné připomínky a za ochotu a čas, který mi věnoval při vedení této práce. Ráda bych dále poděkovala svému příteli za motivaci a velikou podporu a děkuji také své rodině a přátelům za ochotu pomoci s mým praktickým příkladem i v nelehké době.

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

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

V Praze dne 6. ledna 2021 *podpis*

Název práce: Sensometrické diskriminační testování – porovnání párové porovnávací zkoušky a pořadové zkoušky

Autor: Karolína Švarcová

Katedra / Ústav: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: prof. RNDr. Jaromír Antoch, CSc., katedra pravděpodobnosti a matematické statistiky

Abstrakt: Sensometrické zkoušky jsou vhodné k rozhodnutí, zda mezi dvěma či více výrobky existuje vnímatelný senzorický rozdíl. Zkoušky se dají rozdělit do dvou hlavních skupin – první pro určení existence rozdílu na základě konkrétní známé senzorické vlastnosti, a druhé pro toto určení, když konkrétní rozdíl v senzorické vlastnosti není známý. V této práci se zabýváme sensometrickými zkouškami z první skupiny a popisujeme rozdíl ve statistickém přístupu zkoušky hodnotící dva vzorky a zkoušky hodnotící více vzorků najednou. Zejména v případě párové porovnávací zkoušky jsou výpočty založeny na binomickém rozdělení, zatímco pro pořadovou zkoušku se využijí statistické testy založené na pořadí náhodných výběrů. K ilustraci průběhu sensometrické zkoušky jsme uskutečnili jeden párový test a dále naznačili, jak se vybírá vhodná zkouška pro daný problém.

Klíčová slova: sensometrická analýza, párová porovnávací zkouška, pořadová zkouška

Title: Sensometric discriminant testing – comparison of paired comparison test and ranking test

Author: Karolína Švarcová

Department: Department of Probability and Mathematical Statistics

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

Abstract: Sensometric tests are useful for deciding if there exists a perceivable sensory difference between two or more products. The tests can be divided into two main groups – first for determining the existence of a difference based on a known and specified sensory attribute, and second for determination when the variance in the sensory attribute is not known. In this thesis we deal with sensometric tests from the first group and describe the contrast in the statistical approach of a test method that evaluates two samples and a test method evaluating more samples at once. Particularly, within the paired comparison test the calculations are based on the binomial distribution, whereas for the ranking test, statistical methods working with the ranks of random samples are used. To illustrate the process of a sensometric test, we execute a paired comparison test, and we show how a good test method for a given problem is chosen.

Keywords: sensometric analysis, paired comparison test, ranking test

Obsah

Úvod

Sensometrie je ta část vědeckého poznání, která se zabývá hodnocením takových vlastností výrobků, které jsou vnímatelné lidskými smysly. Souvisejícím oborem je psychofyzika, která studuje konkrétní vztah mezi měřitelnými podněty a odpovídající smyslovou odezvou.

Jedním z cílů sensometrických zkoušek je rozhodnutí, zda mezi dvěma nebo více výrobky existuje na základě nějakého smyslu vnímatelný rozdíl – jde o tak zvané rozdílové zkoušky. Nelze pomocí nich však určit, jak velký tento rozdíl je.

Mezi známé typy sensometrických zkoušek patří:

- *párová porovnávací zkouška* (Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009c)) – na základě definovaného kritéria se posuzuje vnímatelnost rozdílu mezi párem vzorků,
- *pořadová zkouška* (Český normalizační institut. (2008)) – posouzení rozdílu probíhá pomocí seřazení dvou nebo více vzorků podle intenzity definovaného kritéria,
- *trojúhelníková zkouška* (Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009a)) – vnímatelnost rozdílu se testuje předložením tří vzorků, z nichž má být odlišen jeden, rozdílný od ostatních dvou, které jsou identické,
- *zkouška dva z pěti* – předkládá se celkem pět vzorků, z nich dva jsou stejné vzorky jednoho výrobku a tři vzorky odlišného výrobku, které mají být seskupeny podle vnímané podobnosti do dvou skupin o dvou a třech vzorcích,
- *tetrádová zkouška* – jde o modifikaci předchozí zkoušky, předkládány jsou 4 vzorky, které mají být dle podobnosti seskupeny do dvou skupin po dvou vzorcích,
- *duo-trio zkouška* (Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2018)) – jsou předloženy tři vzorky, z nichž jeden je určený k přirovnání (jako kontrolní vzorek, viz Slovník 1.1.8) a z ostatních dvou se má vybrat ten vzorek, který je vnímaný jako jemu podobný, resp. od něj odlišný,

- *A – ne A zkouška* (Český normalizační institut. (2001)) – je znám výrobek „A“ a ze série vzorků se určuje, zda jde o daný výrobek A, či nikoli,
- a další.

Tato bakalářská práce se bude zabývat podrobnějším představením a porovnáním zkoušky párové porovnávací a zkoušky pořadové.

1 Základní informace

V této kapitole nejdříve formou slovníku zavedeme základní pojmy, s nimiž budeme dále pracovat. Vycházíme přitom zejména ze znění českých norem, připravených Úřadem pro technickou normalizaci, metrologii a státní zkušebnictví a Českým normalizačním institutem, a také z dalších evropských a mezinárodních norem. Ve slovníku uvedeme zároveň definice některých statistických pojmů, které budeme potřebovat, a to především pro upřesnění významu, ve kterém je chápeme v sensometrických zkouškách.

Následně přiblížíme rámec provedení sensometrické zkoušky. Ten je obecný pro všechny zkoušky a není přímo předmětem této práce, uvádíme ho však pro úplnost představy o průběhu těchto zkoušek.

1.1 Slovník

1.1.1 výrobek

Produkt, který lze posuzovat pomocí sensometrické analýzy.

1.1.2 senzorická vlastnost, deskriptor

Specifická vlastnost produktu vnímatelná lidskými smysly. Tj. například chuť, pach, barva, tvar, textura, zvuk, …

1.1.3 intenzita

Velikost senzorické vlastnosti. Například jak moc sladká je chuť výrobku nebo jak moc hrubý je na dotek povrch výrobku.

1.1.4 (senzorický) posuzovatel

Osoba, která hodnotí vlastnost výrobků v rámci sensometrické zkoušky.

1.1.5 (senzorický) panel

Skupina posuzovatelů, kteří se účastní sensometrické zkoušky.

1.1.6 vzorek (výrobku)

Výrobek nebo část výrobku, který se předkládá posuzovatelům a který je v rámci sensometrické zkoušky posuzován.

1.1.7 referenční vzorek

Výrobek, který definuje a ilustruje vlastnost nebo úroveň vlastnosti, která se pomocí sensometrické zkoušky posuzuje. Jde o stávající standard senzorické vlastnosti, nikoliv o samotný vzorek posuzovaný při zkoušce.

Jinými slovy, testovaná vlastnost výrobku se posuzovatelům demonstruje na produktu, který je standardem trhu.

1.1.8 kontrolní vzorek

Vzorek, se kterým jsou ostatní vzorky porovnávány. Jde o výrobek, který je přímo posuzován při sensometrické zkoušce.

Jinými slovy, posuzovatelům je představen konkrétní testovaný výrobek.

1.1.9 rozdíl

Stav, kdy lze jednotlivé vzorky rozlišit na základě sledované senzorické vlastnosti nebo úrovně této vlastnosti.

1.1.10 podobnost

Dva vzorky jsou si podobné, pokud mezi nimi není vnímatelný rozdíl. „Podobný“ ovšem nemusí znamenat „identický“, identičnost výrobků je nemožné sensometrickou zkouškou prokázat. Lze pouze ukázat, že rozdíl mezi nimi je natolik nepodstatný, že mohou být použity zaměnitelně.

1.1.11 preference

Výběr vzorku na základě vyšší „příjemnosti“. Jinými slovy, posuzovatel vybírá ten vzorek, který (v mezích testované vlastnosti) subjektivně hodnotí jako lepší než ostatní vzorky.

1.1.12 citlivost (posuzovatele)

Úroveň schopnosti vnímat a rozlišovat senzorické vlastnosti výrobků a jejich intenzitu pomocí smyslů.

1.1.13 chyba (hodnocení)

Rozdíl v hodnocení posuzovatele vůči skutečné hodnotě.

1.1.14 *nulová hypotéza / alternativní hypotéza, alternativa*

Při sensometrických zkouškách se ne vždy využívá klasická statistika tak, jak ji známe z klasických učebnic. Často se zde využívají testy ekvivalence[1] a inferiority. Nulová a alternativní hypotéza zde jednoduše označují dvě tvrzení, mezi kterými se rozhoduje.

Nulová hypotéza zpravidla vyjadřuje, že mezi vzorky není vnímatelný rozdíl, a alternativní hypotéza naopak vyjadřuje, že mezi vzorky je vnímatelný rozdíl.

Nulovou hypotézu značíme H_0, alternativu H_1.

Příklady jsou uvedeny v tabulce 1.

Tabulka 1: Příklady hypotéz. Symbol p značí podíl správných/shodných odpovědí z celkového počtu odpovědí n. Symbol s_i značí teoretický součet pořadí (viz Slovník 1.1.22) vzorku i a symbol v značí celkový počet vzorků.

Zkouška	**Nulová hypotéza H_0**	**Alternativa H_1**
Párová – jednostranný test	$p = \frac{1}{2}$	$p > \frac{1}{2}$
Pořadová – skupinové hodnocení	$s_1 = s_2 = \cdots = s_v$	$s_1 \leq s_2 \leq \cdots \leq s_v$

1.1.15 *α riziko*

Statistické riziko (pravděpodobnost), že bude rozhodnuto o existenci vnímatelného rozdílu mezi vzorky, ačkoliv doopravdy neexistuje. Jde o tzv. falešně pozitivní výsledek. Ve statistice se uvádí také jako chyba prvního druhu.

Hodnotu α chápeme jako hladinu významnosti rozdílového testu, resp. ve tvaru $1 - \alpha$ jako sílu testu podobnosti.

1.1.16 *β riziko*

Statistické riziko (pravděpodobnost), že bude rozhodnuto o neexistenci vnímatelného rozdílu mezi vzorky, ačkoliv doopravdy existuje. Jde o tzv. falešně negativní výsledek. Ve statistice se uvádí jako chyba druhého druhu.

Hodnotu β chápeme jako hladinu významnosti testu podobnosti, resp. ve tvaru $1 - \beta$ jako sílu rozdílového testu.

---

[1] Testy ekvivalence jsou založeny na dvou hlavních myšlenkách:

- Ekvivalencí se nemyslí přesná shoda, ale „dostatečná“ podobnost.
- V testech ekvivalence je nulová hypotéza stanovena jako „neekvivalence“, alternativa je formulovaná jako ekvivalence. Formulace hypotéz u testů ekvivalence je „obráceně“ proti klasické formulaci, kde za nulové hypotézy platí požadovaná vlastnost.

1.1.17 podíl hodnocení vnímaného rozdílu p_d

Hodnota p_d může být interpretována minimálně dvěma následujícími způsoby:

1. Hodnota p_d je podíl členů panelu posuzovatelů, kteří jsou vždy schopni určit rozdíl mezi vzorky, zatímco zbytek (tj. $1 - p_d$) je podíl členů panelu posuzovatelů, kteří naopak nikdy nejsou schopni určit rozdíl, a tedy pouze hádají.
2. Hodnota p_d je pravděpodobnost, s jakou je posuzovatel schopen poznat rozdíl mezi vzorky. To znamená, že s pravděpodobností $1 - p_d$ rozdíl nepozná.

Ani jedna z interpretací není zcela reálná. Jedná se o dva extrémní pohledy na problém. V sensometrických zkouškách se hodnota p_d používá zpravidla v prvním smyslu.

1.1.18 citlivost (zkoušky)

Charakterizace prováděného testu, v normách zpravidla síla testu proti konkrétní alternativě. Shrnuje charakteristiku zkoušky definovanou hodnotami α, β a p_d.

1.1.19 správné odpovědi, očekávané odpovědi (v párové zkoušce)

Počet správných odpovědí, tj. počet posuzovatelů, kteří vybrali jako intenzivnější vzorek, který je doopravdy intenzivnější v testované senzorické vlastnosti.

Značíme x_0.

1.1.20 shodné odpovědi (v párové zkoušce)

Máme vzorky A a B. Označíme x_A počet těch hodnocení posuzovatelů, kdy byl zvolen vzorek A, a x_B počet těch hodnocení, kdy byl zvolen vzorek B. Shodné odpovědi značíme stejně jako správné odpovědi také x_0, které je v tomto případě rovno maximu odpovědí pro jednotlivé vzorky, tj. $x_0 = \max\{x_A, x_B\}$.

Jinými slovy, jde o počet odpovědí pro ten vzorek, který byl vybrán více posuzovateli.

1.1.21 shodná pořadí (v pořadové zkoušce)

Je-li více vzorků hodnoceno shodně, přiděluje se každému vzorku shodné pořadí, tzv. mid-rank. Např. jsou-li druhým pořadím hodnoceny dva vzorky, bere se pořadí 2 a 3 a oběma vzorkům se píše zprůměrované pořadí 2,5.

1.1.22 součet pořadí (v pořadové zkoušce)

Součet jednotlivých pořadí udělených danému vzorku jednotlivými posuzovateli.

Značíme r_i pro i -tý vzorek. Je-li pořadí pro vzorek i určené j -tým posuzovatelem R_{ji}, pak $r_i = \sum_{j=1}^{n} R_{ji}$.

1.1.23 předurčené pořadí (v pořadové zkoušce)

Pořadí předem určené organizátorem zkoušky, jež buď vychází ze znalosti předložené sady výrobků, nebo je určené např. na základě předchozích testů.

1.1.24 celá sada (v pořadové zkoušce)

Každý z členů panelu posuzovatelů hodnotí všechny vzorky.

1.1.25 vyvážená dílčí sada (v pořadové zkoušce)

Každý z posuzovatelů hodnotí pouze danou část všech vzorků. Každý vzorek musí být hodnocen stejným počtem posuzovatelů a každý pár vzorků musí být hodnocen stejným počtem posuzovatelů. Aby se dosáhlo požadované úrovně citlivosti, je někdy nutné zkoušku pro vyváženou dílčí sadu několikrát opakovat.

1.2 Obecné požadavky zkoušek

1.2.1 Posuzovatelé

Posuzovateli (viz Slovník 1.1.4) mohou být experti nebo vybraní posuzovatelé, a to s ohledem na povahu testovaného produktu a potřebu přesnosti. Například, je-li předmětem sensometrické zkoušky chuť drahého vína, budou k provedení zkoušky vybráni experti s vysokým stupněm citlivosti. Naopak, jde-li o test pohodlnosti křesla určeného do domácností nebo o test chuti produktu určeného běžným spotřebitelům, bude vedoucího zkoušky zajímat jejich názor.

Vybrané posuzovatele můžeme dále rozdělit na nezkušené, kteří nesplňují žádné zvláštní požadavky, a na zaškolené, kteří se již nějakých sensometrických zkoušek účastnili. Pro každou zkoušku by všichni posuzovatelé měli mít stejnou kvalifikaci, a pro zvýšení pravděpodobnosti nalezení významných rozdílů je vhodné, aby byli s produktem obeznámeni.

Otázkami výběru a výcviku posuzovatelů a jejich sledování se zabývají samostatné normy: ČSN ISO 8589-1 pro vybrané posuzovatele a ČSN ISO 8586-2 pro experty (Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009b)). Obecně se k výcviku posuzovatelů dají využít jednotlivé zkoušky. Příslušné normy obsahují informaci, jakým způsobem testy s tímto záměrem provádět a jak získaná data zpracovat.

1.2.2 Průběh zkoušky

Požadavky na průběh testu jsou vždy uvedeny v jednotlivých normách, s podrobnostmi vzhledem k typu zkoušky. Především jde o volbu vhodných prostor k provedení testu, výběr posuzovatelů, přípravu a podávání vzorků a jejich kódování, stejně jako přípravu odpovědního formuláře.

Jednotlivé vzorky (viz Slovník 1.1.6) by měly být připravovány všechny stejným způsobem, a to za stejných podmínek, není-li ovšem předmětem zkoušky právě vliv různého postupu výroby. Příprava by měla vždy proběhnout mimo dohled posuzovatelů. Zároveň způsob, jakým jsou vzorky podávány, nesmí napovídat, který je který. Proto jsou také vzorky kódovány zpravidla náhodnými trojmístnými čísly, protože označení po sobě jdoucími čísly by mohlo posuzovatele navádět k hodnocení například na základě předpokladu, že vzorek s nižším číslem je „lepší“, anebo podobně.

Po celý průběh zkoušky by se mělo zabránit možnosti posuzovatelů spolu komunikovat nebo se vzájemně vidět. Vzorky jsou poté všem předkládány ve stejných podmínkách, kdy je třeba uvážit zejména teplotu prostředí vhodnou pro testovaný produkt a osvětlení. Pokud vzhled není předmětem zkoušky, je třeba zamaskovat možné vizuální rozdíly mezi vzorky. Množství (popř. objem) všech vzorků musí být identické, stejně jako jejich teplota a způsob servírování. Posuzovatel by měl vzorky hodnotit v pořadí, v jakém mu byly předloženy, ale zpravidla je umožněno opakovat hodnocení vzorků, dovoluje-li to povaha výrobků.

Společně se vzorky posuzovatelé obdrží odpovědní formulář. Jeho vzor je opět součástí každé z norem, poukazující na odlišnosti jednotlivých zkoušek. V odpovědním formuláři musí být přesně specifikován cíl a postup zkoušky (např. jestli je vyžadován zvláštní přístup k hodnocení vzorků), uvedeno jméno posuzovatele, popřípadě jeho kód, a datum konání zkoušky. V příloze *1a* a *1b* jsou uvedeny vzory odpovědního formuláře pro párovou porovnávací zkoušku a pořadovou zkoušku.

Vhodnými prostorami k provedení zkoušky jsou takové prostory, které splňují výše uvedené požadavky na teplotu a zabránění kontaktu posuzovatelů. Nicméně je vhodné též přihlédnout k povaze testovaného produktu pro možnou volbu mezi interiérem a exteriérem. Například, má-li produkt testovaný na chuť silné aroma, které by mohlo zkreslit hodnocení chuti, může být vhodné provést zkoušku v přírodě a nikoli v uzavřeném prostoru.

1.2.3 Protokol o zkoušce

Protokol o zkoušce musí vždy obsahovat přesný cíl zkoušky, výsledky a jejich interpretaci. Dle typu zkoušky je nutné také specifikovat povahu testovaného produktu, popis vzorků a přijaté parametry zkoušky, což jsou zejména přijaté hladiny rizik a dále zkušenosti posuzovatelů a jejich počet.

Dalšími doporučenými (ne však nutnými) informacemi uvedenými v protokolu jsou například účel zkoušky, prostředí zkoušky a použité příslušenství, specifické informace a požadavky na posuzovatele a v neposlední řadě místo a datum provedení zkoušky.

Vzorový zápis protokolu o zkoušce je uveden v rámci kapitoly 5.

2 Párová porovnávací zkouška

2.1 Norma

Popisem metodologie párové porovnávací zkoušky se zabývá norma ČSN EN ISO 5495, Senzorická analýza – Metodologie – Párová porovnávací zkouška. Jde o překlad evropské normy EN ISO 5495:2007, jenž byl připraven Úřadem pro technickou normalizaci, metrologii a státní zkušebnictví.

Poznamenejme, že v případě jakýchkoliv nedorozumění je závazný text a dikce evropské normy.

2.2 Metodologie

Párová porovnávací zkouška je ze statistického hlediska rozhodování mezi dvěma možnostmi. Hladina citlivosti odpovídajícího testu se stanoví na základě zvolených rizik α, β a hodnoty p_d (viz Slovník 1.1.15, 1.1.16 a 1.1.17).

Pomocí této zkoušky můžeme určit, zda existuje vnímatelný senzorický rozdíl mezi vzorky dvou výrobků. Neslouží však k určení velikosti tohoto rozdílu.

Mimo možnosti použít zkoušku pro testování nebo výběr posuzovatelů, lze dle záměru zkoušky využít dvěma způsoby, a to jako:

1. párovou zkoušku podobnosti nebo
2. párovou rozdílovou zkoušku.

Párová zkouška podobnosti slouží k rozhodnutí, že mezi vzorky vnímatelný senzorický rozdíl neexistuje. Použije se v případě, že výrobce změní složení výrobku, výrobní postup či způsob skladování (např. obal výrobku), a chce se ujistit, že tato změna nebude konzumenty u výrobku rozpoznatelná. Na rozdíl od toho párová rozdílová zkouška bude použita v případě, kdy výrobce chce, aby provedená změna byla vnímatelná, například snaží-li se produkt vylepšit.

Zkoušku lze popřípadě také provést tak, že jeden ze vzorků je určen jako kontrolní (viz Slovník 1.1.8). V tom případě posuzovatelé znají identitu tohoto vzorku a pouze rozhodují, zda mu druhý vzorek je anebo není podobný.

Dále je nutné stanovit, zda se jedná o test jednostranný nebo oboustranný. Jednostranným testem je myšlena taková sensometrická zkouška, při které organizátor zkoušky zná odpověď, tj. ví, který z výrobků má danou senzorickou vlastnost. V případě oboustranného testu jde o porovnání dvou výrobků ve významu preference, tj. organizátor zkoušky nemá k dispozici žádnou správnou odpověď.

Na základě výše zmíněných rizik α, β a podílu hodnocení p_d, určujících citlivost zkoušky, je zvolen počet posuzovatelů (resp. potřebný počet pozorování) n. Tímto výpočtem se budeme zabývat v kapitole 2.3. Využití většího množství posuzovatelů zvyšuje pravděpodobnost nalezení rozdílů, nicméně v praxi je počet posuzovatelů zpravidla vybrán na základě reálných a finančních možností. Pro rozdílovou zkoušku je to typicky 24–30 posuzovatelů, zatímco pro dosažení stejné citlivosti u zkoušky podobnosti je jich třeba přibližně dvakrát více.

Pro rozdílovou zkoušku je možné využít i menšího množství posuzovatelů a zkoušku několikrát opakovat, aby bylo dosaženo požadovaného množství odpovědí. V tom případě by každý posuzovatel měl hodnotit stejné množství párů vzorků. Pokud je to možné, je lepší se opakování vyhnout. V případě zkoušky podobnosti by hodnocení stejnými posuzovateli nemělo být opakováno.

Samotný princip zkoušky spočívá v tom, že každý z posuzovatelů hodnotí pár vzorků a vybírá z nich ten, který považuje za intenzivnější z hlediska testované vlastnosti.

Jelikož jde o zkoušku s nucenou volbou, není posuzovatelům dovoleno zvolit odpověď „bez rozdílu". Pokud posuzovatel není schopen určit intenzivnější vzorek, musí vybrat jeden z nich a popřípadě v odpovědním formuláři uvést jako poznámku, zda byla jeho volba pouze odhad. V případě, že je s posuzovateli zkouška opakována, nemohou již upravovat hodnocení z předchozích kol.

Pro interpretaci výsledků se porovnává počet správných odpovědí (v případě jednostranného testu; viz Slovník 1.1.19) nebo shodných odpovědí (v případě oboustranného testu; viz Slovník 1.1.20) s hodnotou uvedenou v normě s přihlédnutím k vybrané citlivosti zkoušky. Tímto výpočtem se budeme zabývat v kapitole 2.3.

V případě zkoušky podobnosti se chce, aby počet správných/shodných odpovědí byl menší nebo roven hodnotě tabelované v normě, protože záměrem je ukázat, že se počty odpovědí pro oba vzorky „zásadně" neliší. Společně s počtem pozorování n se pro vyhodnocení výsledků zkoušky podobnosti využijí hodnoty rizika β a p_d. Naopak v případě rozdílové zkoušky se využije hodnota rizika α, a chce se, aby počet správných/shodných odpovědí byl větší nebo roven hodnotě tabelované v normě, protože je záměrem ukázat, že se počty odpovědí pro jeden vzorek významně liší od počtu odpovědí pro druhý vzorek.

2.3 Metody výpočtů

V následujících podkapitolách popíšeme rozdílné přístupy k výpočtům pro párovou rozdílovou zkoušku a párovou zkoušku podobnosti, a nakonec uvedeme výpočet intervalu spolehlivosti pro parametr p_d.

Ilustrační příklady jsou uvedeny v kapitolách 4.1 a 5.2.

2.3.1 Párová rozdílová zkouška

Nechť $X_j, j = 1, \dots, n$, označuje indikátor toho, zda j-tý posuzovatel rozhodl správně. Tedy X_j má alternativní rozdělení $Alt(p)$, kde $p = p_d + \frac{1-p_d}{2} = \frac{p_d+1}{2}$ je pravděpodobnost správného rozhodnutí. Dle 1.1.17, bodu 1., se toto zpravidla interpretuje tak, že podíl p_d posuzovatelů vždy odpoví správně a zbytek (tedy podíl $1 - p_d$) odpoví náhodně (tj. správně odpoví s poloviční pravděpodobností).

Počet správných odpovědí $Y_n = \sum_{j=1}^{n} X_j$ se pak tedy řídí binomickým rozdělením $Bi(n, p)$.

Testovanou nulovou hypotézou je, že se dva vzorky od sebe neliší:

H_0: $p = p_0$, kde $p_0 = \frac{1}{2}$,

Alternativa říká, že se od sebe vzorky liší. V případě jednostranného testu:

H_1: $p > p_0$

a v případě oboustranného testu:

H_1': $p \neq p_0$.

V normě se pro $n < 120$ používá přesný jednostranný nebo oboustranný test pro binomické rozdělení, viz např. Hátle a Likeš (1974).

Nulová hypotéza zamítne, je-li $Y_n \geq x_{min}$, kde $x_{min} = \min\left\{k; P\left(Y_n \geq k \mid n, p = \frac{1}{2}\right) \leq \alpha\right\}$ je minimální počet správných/shodných odpovědí potřebných k zamítnutí H_0. Tedy, je-li x_0 počet správných/shodných odpovědí získaných při provedení zkoušky, nulová hypotéza se zamítne, pokud $x_0 \geq x_{min}$.

Riziko α se tedy spočítá následovně:

$$\alpha = P\left(Y_n \geq k \mid n, p = \frac{1}{2}\right) = \sum_{k=x_{min}}^{n} b_{n,p_0}(k)$$

kde $b_{n,p_0}(k) = \binom{n}{k} p_0^k (1 - p_0)^{n-k}$ je hustota binomického rozdělení $Bi(n, p_0)$.

Riziko β se určí ze silofunkce tohoto testu, jež má tvar:

$$\beta(p) = \sum_{k=0}^{x_{min}-1} b_{n,p}(k)$$

kde $b_{n,p}(k)$ je hustota binomického rozdělení $Bi(n,p)$.

K výpočtům byl použit SAS software a makro BINRISKS (Schlich (1993)). Na základě zvolených rizik α a β je vypočítán minimální počet posuzovatelů n potřebných pro test, který je definovaný hodnotou p_0, a příslušný minimální počet správných/shodných odpovědí x_{min}. Některé vybrané kritické hodnoty jsou uvedeny v tabulce T1 v apendixu.

Pro hodnoty n nezahrnuté v tabulkách A1 a A2 v normě se dají spočítat asymptotické aproximace kritických hodnot následovně:

$$x'_{min} = \frac{n+1}{2} + z_{\alpha} \cdot \sqrt{0{,}25n}$$

kde z_{α} je kritická hodnota standardního normálního rozdělení pro hladinu rizika α, příslušná jednostrannému či oboustrannému testu (viz tabulka T2 v apendixu). Volba $\frac{n+1}{2}$ je tzv. korekce na spojitost. Jedná se o opatrnější verzi asymptotického testu aplikovaného na diskrétní model a data.

Za x_{min} se volí nejbližší celé číslo větší než x'_{min}.

2.3.2 Párová zkouška podobnosti

Minimální počet posuzovatelů n pro párovou zkoušku podobnosti se vybere z tabulek vypočtených pro rozdílovou párovou zkoušku výše (tabulka A4 nebo A5 v normě).

Jak už bylo zmíněno, v případě párové zkoušky podobnosti je ale na rozdíl od rozdílové zkoušky požadováno určit maximální počet správných/shodných odpovědí, abychom mohli říct, že jsou si vzorky podobné – neboli chceme $x_0 \leq x_{max}$.

Nepoužije se proto stejný postup pro určení kritické hodnoty jako výše. Pro výpočet maximálního počtu správných/shodný odpovědí x_{max} pro příslušný počet posuzovatelů n a zvolené hodnoty rizika β a p_d byl využit program založený na výpočtu přesných intervalů spolehlivosti pro binomické rozdělení (MacRae (1995)).

Pro hodnoty n nezahrnuté v tabulce A3 v normě se počítá horní hranice intervalu spolehlivosti následovně:

$$x_{max} = \left(2\frac{x_0}{n} - 1\right) + 2 \cdot z_\beta \cdot \sqrt{\frac{nx_0 - x_0^2}{n^3}}$$

kde x_0 je počet správných/shodných odpovědí a z_β je kritická hodnota standardního normálního rozdělení pro hladinu rizika β pro jednostranný test (tabulka T2 v apendixu).

2.3.3 Interval spolehlivosti pro odhad parametru p_d

Je-li požadováno, může se také vypočítat interval spolehlivosti pro odhad parametru p_d. Ten nám zodpoví otázku, zda byl námi předpokládaný podíl hodnocení vnímaného rozdílu správný, a tedy i zda učiněný závěr byl správný.

K odhadu parametru použijeme podíl správných/shodných odpovědí $p_{x_0} = \frac{x_0}{n}$ následovně:

$$\widehat{p_d} = 2p_{x_0} - 1$$

Výpočet je založený na aproximaci standardním normálním rozdělením $N(0,1)$. V tabulce T2 v apendixu jsou uvedené příslušné kritické hodnoty z_α na hladinách rizika α (které se použije pro párovou rozdílovou zkoušku), resp. z_β na hladinách rizika β (pro párovou zkoušku podobnosti).

Určíme standardní chybu odhadu $\widehat{p_d}$:

$$s_d = 2\sqrt{\frac{p_{x_0}(1 - p_{x_0})}{n}} = 2\sqrt{\frac{nx_0 - x_0^2}{n^3}}$$

Oboustranný interval spolehlivosti má tvar $(\widehat{p_d} - z_\alpha s_d, \widehat{p_d} + z_\alpha s_d)$ pro párovou rozdílovou zkoušku. Pro párovou zkoušku podobnosti se kritická hodnota z_α nahradí kritickou hodnotou z_β.

Jednostranný interval spolehlivosti norma neuvádí, ale pro ilustraci a porovnání výsledků v kapitole 5 ho zde uvedeme. Pro párovou zkoušku podobnosti má jednostranný interval spolehlivosti tvar $(\widehat{p_d} - z_\beta s_d, \infty)$ a pro párovou rozdílovou zkoušku $(-\infty, \widehat{p_d} + z_\alpha s_d)$.

3 Pořadová zkouška

3.1 Norma

Popisem metodologie pořadové zkoušky se zabývá norma ČSN ISO 8587, Senzorická analýza – Metodologie – Pořadová zkouška. Jde o překlad mezinárodní normy ISO 8587:2006, jenž byl připraven Českým normalizačním institutem.

Poznamenejme, že v případě jakýchkoliv nedorozumění je závazný text a dikce mezinárodní normy.

3.2 Metodologie

Pořadová zkouška je metodou sensometrického hodnocení, která se použije pro hodnocení rozdílů mezi několika vzorky na základě intenzity jednoho nebo několika deskriptorů, popřípadě obecně na základě celkového dojmu. Pomocí této zkoušky se dá zjistit, zda takové rozdíly existují, opět ovšem nelze určit, jak jsou velké.

Mimo trénování posuzovatelů a určení prahu jejich citlivosti lze zkoušku využít k hodnocení výrobků následujícími způsoby:

1. pro předběžný výběr několika vzorků z většího množství,
2. pro určení pořadí na základě preference a celkové příjemnosti,
3. pro určení vlivů na intenzitu deskriptoru na základě popsaného kritéria (např. postup výroby, vliv různých surovin, balení a skladování).

Pro určení vhodných posuzovatelů a jejich množství se přihlíží k účelu zkoušky. Zejména, pro hodnocení výrobků na základě popsaného kritéria je třeba užít zkušené posuzovatele, a s ohledem na hladinu přijatelného statistického rizika se doporučuje 12–15 posuzovatelů. Pro hodnocení výrobků na základě preference se pak doporučuje minimálně 60 běžných konzumentů. Pro výcvik a hodnocení posuzovatelů není dáno žádné omezení, ačkoliv větší šance na nalezení rozdílů je ve větším panelu posuzovatelů.

Principem zkoušky je, že každý z posuzovatelů na základě specifikované vlastnosti nebo celkové preference seřadí předložené vzorky dle jejich intenzity. Během jedné zkoušky budou vzorky hodnoceny s ohledem pouze na jeden deskriptor. Je-li nutné seřadit vzorky na základě několika deskriptorů, každý bude hodnocen v oddělené zkoušce.

Počet vzorků hodnocených v rámci pořadové zkoušky je zpravidla tři a více, neboť i když je možné v pořadové zkoušce řadit pouze dva vzorky, je v takovém případě vhodnější využít párovou porovnávací zkoušku. Horní hranice počtu vzorků by měla být zodpovědně určena na základě povahy výrobků. Např. pro vysoce tučné nebo velice kořeněné výrobky je stanoven maximální počet vzorků tři, zatímco pro výrobky s méně výraznou chutí to může být až 15 vzorků.

Při této zkoušce je podstatné se ujistit, že posuzovatelé správně rozumí zadání. Zejména je nutné upřesnit, zda mají seřadit vzorky od nejlepšího k nejhoršímu nebo naopak, a poučit je, že se mají vyhnout hodnocení vzorků shodným pořadím (viz Slovník 1.1.21). To mohou použít jenom tehdy, když doopravdy nejsou schopni vzorky seřadit. I v tomto případě je ale preferovaným postupem vzorky seřadit do pořadí, a uvést je jako nerozlišitelné pouze v poznámce.

Kódy vzorků by neměly být v odpovědním formuláři předem uvedeny, aby dané pořadí neovlivnilo hodnocení posuzovatele. Posuzovatel kódy zapíše sám dle toho, jak mu byly vzorky předloženy, do jednoho řádku, a poté zapisuje své hodnocení do druhého řádku. Vzorky jsou předkládány všechny najednou a posuzovatel je může hodnotit v libovolném pořadí. Vhodné může být, aby si posuzovatel vzorky předběžně seřadil a pořadí ověřil opakovaným hodnocením, dovoluje-li to povaha výrobků.

V případě, že není v závislosti na množství vzorků nebo povaze výrobků praktické řadit celou sadu vzorků, je možné využít hodnocení vyvážené dílčí sady (viz Slovník 1.1.25).

Pro interpretaci výsledků určíme součet pořadí (viz Slovník 1.1.22) pro jednotlivé vzorky. To nám ukáže, zda byla hodnocení posuzovatelů shodná (v tom případě budou součty pořadí rozdílné), anebo zda se hodnocení lišila (v tom případě se součty pořadí nebudou významně lišit).

V rámci pořadové zkoušky lze hodnotit:

1. individuální shodu mezi hodnoceními dvou posuzovatelů, nebo
2. skupinově shodu mezi všemi hodnoceními, a to buď
    a. v případě předurčeného pořadí (viz Slovník 1.1.23), nebo
    b. bez předurčeného pořadí, kde dále také určíme, které vzorky jsou podstatně odlišné od kterých.

Alternativně se vybírá rozdílný test či metoda pro zkoušku, kde:

1. se vyskytuje hodnocení shodným pořadím, nebo
2. byly řazeny pouze dva vzorky.

Konkrétními výpočty se budeme zabývat v kapitole 3.3.

3.3 Metody výpočtů

Poznamenejme, že jednotlivé zkoušky a postupy jejich vyhodnocení se v sensometrické literatuře většinou jmenují podle statistického testu, který se pro jejich hodnocení používá.

V závislosti na účelu zkoušky se vybere statistický test dle tabulky 2.

Tabulka 2: Statistické metody dle cíle pořadové zkoušky.

Cíl zkoušky	Statistická metoda
Individuální hodnocení dvou posuzovatelů	Spearmanova zkouška
Skupinové hodnocení s předurčeným pořadím	Zkouška dle Page
Skupinové hodnocení bez předurčeného pořadí	Friedmanova zkouška
Zkouška se stejnými pořadími	Friedmanova zkouška (s upraveným koeficientem)
Řazení dvou výrobků	Znaménkový test

Tyto metody budou stručně představeny v následujících podkapitolách a konkrétní ilustrační příklady jsou uvedeny v kapitole 4.2.

3.3.1 Individuální hodnocení dvou posuzovatelů

Pomocí Spearmanovy zkoušky (Anděl (1985)) se dá posoudit shoda (souvislost) individuálního hodnocení jednoho posuzovatele buď s hodnocením jiného posuzovatele nebo s předurčeným pořadím.

H_0: mezi pořadími neexistuje souvislost.

H_1: mezi pořadími je silná souvislost.

Tato zkouška využívá korelační koeficient, který nepracuje přímo s hodnotami pozorování, ale s jejich pořadími, tzv. Spearmanův korelační koeficient r_s.

Ten je vypočten následovně:

$$r_s = 1 - \frac{6\sum_{i=1}^{v} d_i^2}{v(v^2-1)} \tag{1}$$

kde v je počet řazených vzorků a d_i je rozdíl obou pořadí i-tého vzorku. Tj. nechť $R_1, \ldots, R_v$ značí pořadí určené prvním posuzovatelem (resp. předurčené pořadí) a $Q_1, \ldots, Q_v$ pořadí určené druhým posuzovatelem. Pak definujeme $d_i = R_i - Q_i$.

Mezi danými pořadími je velká shoda (souvislost), pokud se Spearmanův korelační koeficient blíží hodnotě 1. Blíží-li se naopak hodnotě -1, existuje mezi pořadími silná neshoda (což by zde mohlo znamenat, že si posuzovatel spletl zadání a hodnotil vzorky v opačném pořadí). Hodnota blízká 0 se zpravidla interpretuje tak, že mezi pořadími žádný vztah není.

V testu hypotézy H_0 s hladinou testu α to znamená, že hypotéza se zamítá (tedy je učiněn závěr, že mezi pořadími existuje souvislost) v případě, že vypočtený Spearmanův korelační koeficient $|r_s| \geq r_s(\alpha)$, kde $r_s(\alpha)$ je kritická hodnota Spearmanova korelačního koeficientu (viz tabulka T3 v apendixu).

Pro hodnoty $v > 30$, neuváděné ve statistických tabulkách, se nulová hypotéza zamítne, pokud $|r_s| > \frac{z_{\alpha/2}}{\sqrt{v-1}}$, kde $z_{\alpha/2}$ je kritická hodnota standardního normálního rozdělení.

3.3.2 Skupinové hodnocení s předurčeným pořadím

Zkouška dle Page (Page (1963)) se použije pro posouzení shody hodnocení všech n posuzovatelů, pokud známe předurčené pořadí sady v vzorků.

H_0: $s_1 = s_2 = \cdots = s_v$, kde $s_1, \ldots, s_v$ jsou teoretické součty pořadí vzorků v jejich předurčeném pořadí.

H_1: $s_1 \leq s_2 \leq \cdots \leq s_v$, přičemž alespoň jedna nerovnost je ostrá.

Pro výrobky se spočítá součet pořadí $r_1, \ldots, r_v$, kde r_i je pořadí i-tého vzorku v jejich předurčeném pořadí (viz Slovník 1.1.23). K testu nulové hypotézy se počítá Pageův koeficient L:

$$L = \sum_{i=1}^{v} i \cdot r_i \tag{2}$$

Statistika L bude největší, bude-li předurčené pořadí vzorků „shodné“ s hodnocením posuzovatelů.

Nulová hypotéza se zamítá, pokud $L \geq L_{n,v}(\alpha)$, kde $L_{n,v}(\alpha)$ je kritická hodnota pro zkoušku dle Page na základě hodnocení n posuzovatelů, počtu výrobků v a hladiny α (viz tabulka T4 v apendixu).

V případě, že pro hodnotu n nebo v není uvedena kritická hodnota v tabulce, spočítá se aproximace:

$$L' = \frac{12L - 3n \cdot v(v+1)^2}{v(v+1)\sqrt{n(v-1)}} \tag{3}$$

Statistika L' je asymptoticky normálně rozdělena. Kritické hodnoty normovaného normálního rozdělení z_α jsou uvedeny v tabulce T5 v apendixu.

Nulová hypotéza se zamítá, pokud $L' \geq z_\alpha$.

Statistika (2) nebo (3) se využije pro zkoušku, kde byla všemi posuzovateli hodnocena celá sada. V případě vyvážené dílčí sady se počítá statistika:

$$L'' = \frac{12L - 3n \cdot k(k+1)(v+1)}{\sqrt{n \cdot k(k-1)(k+1) \cdot v(v+1)}} \tag{4}$$

kde k je počet vzorků hodnocených každým posuzovatelem.

Statistika L'' je také asymptoticky normálně rozdělena, tedy nulová hypotéza se zamítá, pokud $L'' \geq z_\alpha$.

Je-li nulová hypotéza zamítnuta, výsledek je interpretován tak, že data z uskutečněné zkoušky odpovídají predikci. Jinými slovy, vzorky byly jednotlivými posuzovateli hodnoceny v předurčeném pořadí.

3.3.3 Skupinové hodnocení bez předurčeného pořadí

Pomocí Friedmanovy zkoušky (Anděl (1985)) se nejdříve určí, zda existují rozdíly mezi alespoň dvěma vzorky.

Mějme pozorování X_{ji}, vzorek i hodnocený j-tým posuzovatelem, $j = 1, \dots, n$, $i = 1, \dots, v$, a jejich pořadí R_{ji}. Friedmanův test zkoumá, zda rozdělení veličin X_{ji} je totožné pro každé i. Jinými slovy, ověřuje, zda každý vzorek i byl řazen stejně.

Spočítá se součet pořadí $r_i = \sum_{j=1}^{n} R_{ji}$ pro $i = 1, \dots, v$, kde r_i je pořadí i-tého vzorku. Označíme s_i teoretický součet pořadí i-tého nejlepšího vzorku neboli teoreticky jsou vzorky seřazené od nejlepšího (nejnižší součet pořadí) po nejhorší (nejvyšší součet pořadí).

$H_0\colon s_1 = s_2 = \dots = s_v$, neboli součty pořadí všech vzorků jsou stejné, a tedy mezi vzorky neexistují rozdíly.

$H_1\colon s_1 \leq s_2 \leq \dots \leq s_v$, přičemž alespoň jedna nerovnost je ostrá. Jinými slovy, existuje rozdíl v součtu pořadí mezi alespoň dvěma vzorky.

K testu nulové hypotézy se počítá Friedmanova statistika F:

$$F = \frac{12}{n \cdot v(v+1)} \sum_{i=1}^{v} r_i^2 - 3n(v+1) \tag{5}$$

kde n je počet posuzovatelů a v je celkový počet výrobků.

Statistika (5) se použije v případě celé sady. Pro vyváženou dílčí sadu se počítá modifikovaná Friedmanova statistika:

$$F'' = \frac{12}{t \cdot g \cdot v(k+1)} \sum_{i=1}^{v} r_i^2 - \frac{3t \cdot m^2(k+1)}{g} \tag{6}$$

kde t je počet opakování zkoušky,

k je počet vzorků hodnocených každým posuzovatelem,

m značí, kolikrát byl každý vzorek hodnocen,

g značí, kolikrát byl každý pár vzorků hodnocen dohromady.

Nulová hypotéza se zamítá, pokud statistika F, případně F'', překročí kritickou hodnotu Friedmanovy zkoušky na stanovené hladině α, $F_{n,v}(\alpha)$, viz tabulka T6 v apendixu.

V případě, že pro počet vzorků v není $F_{n,v}(\alpha)$ ve statistických tabulkách uvedena, dá se statistika F, resp. F'' aproximovat rozdělením χ^2_{v-1}. Kritické hodnoty χ^2 rozdělení jsou uvedeny v tabulce T7 v apendixu.

Jestliže byla nulová hypotéza zamítnuta a tedy konstatováno, že mezi součty pořadí vzorků existují rozdíly, zjistí se dále, které vzorky se významně liší od kterých.

Již jsou známy součty pořadí r_i, $i = 1, \dots, v$. Nejdříve se spočítá absolutní rozdíl mezi součty pořadí pro dvojice vzorků $|r_i - r_k|$, $1 \leq i < k \leq v$. Následně se určí nejmenší významný rozdíl LSD (Least Significant Difference).

Pro celou sadu se počítá kritická hodnota:

$$LSD_{\alpha'} = z_{\alpha'} \cdot \sqrt{\frac{n \cdot v(v+1)}{6}} \tag{7}$$

a pro vyváženou dílčí sadu:

$$LSD''_{\alpha'} = z_{\alpha'} \cdot \sqrt{\frac{t(k+1)(m \cdot k - m + g)}{6}} \tag{8}$$

kde $z_{\alpha'}$ je kritická hodnota normálního rozdělení pro oboustranný test na hladině α'.

Kdyby nám stačilo použít zvolenou hladinu rizika α na každý pár zvlášť, pak $\alpha' = \alpha$. Zpravidla je však požadováno hladinu rizika α aplikovat na celý test, počítá se tedy nejdříve riziko pro každou dvojici výrobků zvlášť následovně: $\alpha' = \frac{2\alpha}{v(v-1)}$.

Na závěr se porovnají jednotlivé vypočtené rozdíly $|r_i - r_k|$ s příslušnou kritickou hodnotou $LSD_{\alpha'}$ nebo $LSD''_{\alpha'}$. V případě, že $|r_i - r_k| \geq LSD$, resp. LSD'', pro nějaká konkrétní i,k, pak lze říct, že tyto dva výrobky jsou významně odlišné. Pokud naopak $|r_i - r_k| < LSD$, resp. LSD'', pak lze říct, že tyto dva vzorky se od sebe významně neliší.

3.3.4 Stejná pořadí

V případě, že některé vzorky byly hodnoceny stejným pořadím a místo pořadí určených posuzovateli se použije mid-rank, upraví se statistika Friedmanovy zkoušky z předešlé kapitoly následovně:

$$F' = \frac{F}{1 - \{\sum_{i=1}^{l}(t_i^3 - t_i) / [n \cdot v(v^2 - 1)]\}} \tag{9}$$

kde l značí, kolikrát bylo použito hodnocení stejným pořadím,
t_i je počet vzorků hodnocených stejným pořadím v hodnocení $i = 1, \dots, l$.

V případě vyvážené dílčí sady se statistika F v čitateli rovnice (9) zamění za statistiku F''.

Nulová hypotéza se zamítá pro $F' > F_{n,v}(\alpha)$, resp. $F' > \chi^2_{v-1}(\alpha)$, uvedených v tabulkách T6 a T7 v apendixu.

3.3.5 Dva výrobky

Pro zkoušku, při které jsou posuzovány pouze dva vzorky, je obecně vhodnější využít párovou porovnávací zkoušku, ale lze využít také analogii znaménkového testu.

Pokud máme pouze dva vzorky (A a B) a pro každý z nich známe počet odpovědí, kdy byl řazen jako první, x_A a x_A (tj. vzorek A byl jako první řazen $x_A$krát a vzorek B byl jako první řazen $x_B$krát), pak určíme $x_0 = \min\{x_A, x_B\}$.

Nulová hypotéza o tom, že vzorky A a B byly řazeny stejně celým panelem posuzovatelů, se zamítá tehdy, pokud $x_0 < k(\alpha)$, kde $k(\alpha)$ je kritická hodnota znaménkového testu, viz tabulka T8 v apendixu.

4 Příklady na simulovaných datech

4.1 Příklady párové porovnávací zkoušky

Příklad 1. *Párová zkouška podobnosti, jednostranný test*

Výrobce si je vědom, že jeho výrobek obsahuje složku, která může výrobku dodávat nepříjemnou bylinnou chuť. Přeje si tedy určit maximální koncentraci této složky ve výrobku tak, aby rozdíl oproti vzorku neobsahující tuto složku nebyl znatelný.

V této zkoušce je nutné co nejvíce snížit riziko, že nebudou detekovány rozdíly, které mezi vzorky existují. Hodnota rizika β je proto zvolena $0{,}05$. Na druhé straně riziko, že budou detekovány rozdíly, které neexistují, není tak důležité, neboť by to vedlo pouze k opatrnější specifikaci těchto rozdílů. Je tedy zvolena hodnota rizika $\alpha = 0{,}50$. Podíl posuzovatelů správně rozlišujících rozdíly p_d je zvolen 20 %.

Z tabulky A4 v normě se zjistí, že za těchto parametrů je k jednostrannému párovému testu podobnosti nutné přizvat minimálně 67 posuzovatelů. Avšak pro požadované parametry $\beta = 0{,}05$ a $p_d = 20\ \%$ nejsou v tabulce, která se používá pro vyhodnocení výsledků, uvedeny maximální počty správných odpovědí (z důvodu, že hodnoty nepřesahují $n/2$) až do $n = 78$. Proto je výrobcem rozhodnuto najmout 78 posuzovatelů.

Pro test se připraví dva roztoky: vzorek A neosahující složku způsobující nepříjemnou chuť a vzorek C obsahující koncentraci této složky, definovanou na základě předchozí znalosti. Vzorky jsou posuzovatelům podávány současně ve stejných skleničkách. Prvním 39 posuzovatelům jsou předkládány v pořadí AC a zbylým 39 posuzovatelům v pořadí CA.

Jako více bylinný označilo vzorek A 37 posuzovatelů, tedy máme $x_A = 37$. Zbývajících 41 posuzovatelů označilo za více bylinný vzorek C, tedy máme $x_C = 41$. Podle 1.1.19 určíme počet správných odpovědí $x_0 = x_C = 41$. Z tabulky A3 v normě se se vyčte, že při stanovené hodnotě rizika $\beta = 0{,}05$ a $n = 78$ nesmí být x_0 větší než 39, aby se o výrobcích dalo říct, že mezi nimi nejsou vnímatelné rozdíly.

Protože $x_0 > 39$, je rozhodnuto, že nelze zkoušku uzavřít se závěrem, že jsou si vzorky dostatečně podobné. Koncentrace složky ve vzorku C je moc silná na to, aby mohla být použita ve výrobě. Je proto doporučeno opakovat zkoušku pro vzorek s nižší koncentrací.

Příklad 2. *Párová rozdílová zkouška, oboustranný test*

Výrobce chce začít vyrábět nové solené bramborové lupínky. Z důvodu vysokých nákladů je požadováno pomocí sensometrické zkoušky porovnat dvě různé složky poskytující slanou chuť. Není předem známo, která složka poskytne při stejné koncentraci intenzivnější slanou chuť, a je třeba určit, zda existuje vnímatelný rozdíl. Pokud nebudou při zkoušce pozorovány rozdíly, výrobce bude chtít otestovat další složku.

Je požadováno co nejvíce snížit riziko, že budou vnímány rozdíly, které neexistují, tedy je zvolena hodnota rizika $\alpha = 0{,}05$ a $p_d = 50$ %. Zároveň je potřeba co nejvíce snížit i riziko, že nebudou vnímány existující rozdíly, protože by to vedlo k vyšším nákladům kvůli nutnosti provádět další test. Je proto zvolena hodnota rizika $\beta = 0{,}10$.

Z tabulky A5 v normě se vyčte, že za těchto podmínek je potřeba zkoušku provést s nejméně 42 posuzovateli. Je rozhodnuto najmout jich $n = 44$.

Oba vzorky A a B jsou po přípravě na 3 dny uskladněny ve stejných vakuovaných obalech, ve kterých chce výrobce finální výrobek prodávat. Poté jsou při zkoušce podávány posuzovatelům současně v identických keramických miskách v množství 3 gramů. Barva ani tvar vzorků není pro test relevantní a není třeba vzhled nijak maskovat. Vzorky jsou předloženy 22 posuzovatelům v pořadí AB a zbylým 22 posuzovatelům v pořadí BA.

Jako slanější označilo vzorek A 31 posuzovatelů, tedy máme $x_A = 31$. Zbylých 13 posuzovatelů označilo jako slanější vzorek B, máme tedy $x_B = 13$. Dle 1.1.20 se určí počet shodných odpovědí $x_0 = \max(x_A, x_B) = 31$. Z tabulky A2 v normě se vyčte, že při hodnotě rizika $\alpha = 0{,}05$ a množství přizvaných posuzovatelů $n = 44$, musí být x_0 větší nebo rovna 29, aby bylo možné rozhodnout, že mezi vzorky existují jasné rozdíly. Z tabulky zároveň vyplývá, že je-li $x_0 \geq 31$ pro $n = 44$, pak lze zkouška uzavřít se závěrem, že existují rozdíly, dokonce na hladině $\alpha = 0{,}01$.

To znamená, že výrobce může zkoušku uzavřít s tím, že slaná složka použitá ve vzorku A je vnímaná jako slanější na 99% hladině spolehlivosti a může tedy být použita pro budoucí výrobu.

4.2 Příklady pořadové zkoušky

Příklad 3. *Individuální shoda dvou hodnocení*

Chtějme posoudit individuální shodu hodnocení dvou posuzovatelů na hladině testu $\alpha = 0{,}05$. Mějme tato hodnocení pro pořadovou zkoušku pěti vzorků ($v = 5$) v tabulce 3.

Tabulka 3: Hodnocení posuzovatelů (individuální)

Posuzovatel j	Vzorek i				
	1	2	3	4	5
1	1	3	4	2	5
2	2	3	5	1	4
Rozdíl pořadí d_i	–1	0	–1	1	1

H_0: mezi pořadími neexistuje souvislost.

H_1: mezi pořadími je silná souvislost.

Spočítáme Spearmanův korelační koeficient (1):

$$r_s = 1 - \frac{6\sum_{i=1}^{v} d_i^2}{v(v^2-1)} = 1 - \frac{6}{5(5^2-1)}((-1)^2 + 0^2 + (-1)^2 + 1^2 + 1^2) = 0{,}8$$

V tabulce T3 v apendixu najdeme, že pro $v = 5$ a $\alpha = 0{,}05$ je kritická hodnota $r_s(\alpha) = 0{,}9$.

Jelikož neplatí nerovnost $|r_s| \geq r_s(\alpha)$, nezamítáme nulovou hypotézu. Spearmanův test nám tedy dává závěr, že mezi pořadími neexistuje na hladině $\alpha = 0{,}05$ souvislost.

Příklad 4. *Skupinové hodnocení s předurčeným pořadím, celá sada*

Chtějme na hladině testu $\alpha = 0{,}05$ posoudit shodu hodnocení $n = 10$ posuzovatelů s předurčeným pořadím $v = 5$ vzorků. Mějme hodnocení posuzovatelů v tabulce 4, kde vzorky jsou seřazeny v jejich předurčeném pořadí.

Tabulka 4: Hodnocení posuzovatelů (celá sada)

Posuzovatel j	Vzorek i				
	1	2	3	4	5
1	1	3	2	4	5
2	2	1	3	4	5
3	1	2	4	3	5
4	3	2	1	5	4
5	1	2	3	5	4
6	1	2	5	3	4
7	2	4	1	3	5
8	1	2	4	5	3
9	1	3	2	4	5
10	2	1	3	5	4
Součet pořadí r_i	15	22	28	41	44

H_0: $s_1 = s_2 = \cdots = s_v$.

H_1: $s_1 \leq s_2 \leq \cdots \leq s_v$, přičemž alespoň jedna nerovnost je ostrá.

Spočítáme Pageův koeficient (2):

$$L = \sum_{i=1}^{v} i \cdot r_i = 1 \cdot 15 + 2 \cdot 22 + 3 \cdot 28 + 4 \cdot 41 + 5 \cdot 44 = 527$$

V tabulce T4 v apendixu najdeme, že pro $n = 10$, $v = 5$ a $\alpha = 0{,}05$ je kritická hodnota $L(\alpha) = 477$.

Protože platí $L \geq L(\alpha)$, zamítáme nulovou hypotézu. Pageův test nás vede k závěru, že vzorky byly řazeny v předurčeném pořadí. Jinými slovy, posuzovatelé při zvoleném riziku $\alpha = 0{,}05$ identifikovali rozdíly mezi vzorky, výsledek testu tedy lze interpretovat tak, že předurčené pořadí bylo ověřeno.

Příklad 5. *Skupinové hodnocení bez předurčeného pořadí, vyvážená dílčí sada*

Chtějme na hladině testu $\alpha = 0{,}05$ posoudit shodu hodnocení $n = 10$ posuzovatelů, kteří v pořadové zkoušce hodnotili $k = 3$ vzorky z celkových $v = 5$. Mějme hodnocení posuzovatelů v tabulce 5.

Tabulka 5: Hodnocení posuzovatelů (vyvážená dílčí sada)

Posuzovatel j	Vzorek i				
	1	2	3	4	5
1	1	2	3		
2	1	2		3	
3	2	3			1
4	1		2	3	
5	2		3		1
6	1			3	2
7		1	3	2	
8		2	3		1
9		3		2	1
10			1	3	2
Součet pořadí r_i	8	13	15	16	8

H_0: $s_1 = s_2 = \cdots = s_v$.

H_1: $s_1 \leq s_2 \leq \cdots \leq s_v$, přičemž alespoň jedna nerovnost je ostrá.

Každý vzorek byl hodnocen šestkrát ($m = 6$) a každý pár vzorků byl hodnocen třikrát ($g = 3$). Zkouška byla opakována jedenkrát ($t = 1$).

Spočítáme Friedmanovu statistiku pro vyváženou dílčí sadu (6):

$$F'' = \frac{12}{t \cdot g \cdot v(k+1)} \sum_{i=1}^{v} r_i^2 - \frac{3t \cdot m^2(k+1)}{g}$$

$$= \frac{12}{1 \cdot 3 \cdot 5(3+1)} (8^2 + 13^2 + 15^2 + 16^2 + 8^2) - \frac{3 \cdot 1 \cdot 6^2(3+1)}{3}$$

$$= \frac{12 \cdot 778}{60} - \frac{432}{3} = 11{,}6$$

V tabulce T6 v apendixu najdeme, že pro $n = 10$, $v = 5$ a $\alpha = 0{,}05$, je kritická hodnota $F(\alpha) = 9{,}25$.

Protože platí $F'' \geq F(\alpha)$, zamítáme nulovou hypotézu, že všechny vzorky byly řazeny stejně. Friedmanův test dává závěr, že mezi součty pořadí vzorků existují rozdíly. Jinými slovy, těchto pět vzorků může být vnímáno jako odlišné.

Určíme tedy dále, jaké vzorky se liší od jakých. V tabulce 6 spočítáme absolutní hodnotu rozdílů součtu pořadí jednotlivých vzorků $|r_i - r_k|$.

Tabulka 6: Rozdíly součtu pořadí

Vzorek i	Vzorek k			
	2	3	4	5
1	5	7	8	0
2		2	3	5
3			1	7
4				8

Při hladině $\alpha = 0{,}05$ pro každý pár vzorků zvlášť, máme $\alpha' = \alpha$ a příslušná kritická hodnota standardního normálního rozdělení je $z_{\alpha'} = 1{,}96$.

Nyní spočítáme nejmenší významný rozdíl LSD pro vyváženou dílčí sadu (8):

$$LSD'' = z_{\alpha'} \cdot \sqrt{\frac{t(k+1)(m \cdot k - m + g)}{6}} = 1{,}96 \cdot \sqrt{\frac{1 \cdot (3+1)(6 \cdot 3 - 6 + 3)}{6}} = 6{,}2$$

Významně odlišné jsou vzorky takové, pro které platí $|r_i - r_k| \geq LSD$. Z tabulky 6 vidíme, že rozdílné jsou tedy páry vzorků 1 a 3, 1 a 4, 3 a 5, 4 a 5. Konkrétně, jak lze nahlédnout v tabulce 5, součty pořadí vzorků 1 a 5 jsou významně nižší, než součty pořadí vzorků 3 a 4. Vzorek 2 se významně neliší od žádného z ostatních vzorků.

Příklad 6. *Skupinové hodnocení s povoleným hodnocením stejným pořadím*

Chtějme na hladině testu $\alpha = 0{,}05$ posoudit shodu hodnocení $n = 10$ posuzovatelů, kteří v pořadové zkoušce hodnotili $v = 5$ vzorků a bylo jim povoleno vzorky hodnotit stejným pořadím. Mějme hodnocení posuzovatelů v tabulce 7.

Tabulka 7: Hodnocení posuzovatelů (shodná pořadí)

Posuzovatel j	Vzorek i				
	1	2	3	4	5
1	3	2	5	1	4
2	2*	2*	2*	4	5
3	4	5	2	3	1
4	5	2	3	4	1
5	1	3	2	4,5*	4,5*
6	2,5*	4	2,5*	1	5
7	5	2	3	1	4
8	3	1	4,5*	2	4,5*
9	4	3	1	5	2
10	2	5	3	4	1
Součet pořadí r_i	31,5	29	28	29,5	32

H_0: $s_1 = s_2 = \cdots = s_v$.

H_1: $s_1 \leq s_2 \leq \cdots \leq s_v$, přičemž alespoň jedna nerovnost je ostrá.

Spočítáme nejdříve Friedmanovu statistiku pro celou sadu (5):

$$F = \frac{12}{n \cdot v(v+1)} \sum_{i=1}^{v} r_i^2 - 3n(v+1)$$

$$= \frac{12}{10 \cdot 5(5+1)} (31{,}5^2 + 29^2 + 28^2 + 29{,}5^2 + 32^2) - 3$$

$$\cdot 10(5+1) = \frac{12 \cdot 4511{,}5}{300} - 180 = 0{,}46$$

Shodná hodnocení jsou v tabulce 7 označena hvězdičkou, máme tedy $l = 4$, $t_1 = 3$, $t_2 = 2$, $t_3 = 2$ a $t_4 = 2$.

Upravíme Friedmanovu statistiku dle (9):

$$F' = \frac{F}{1 - \{\sum_{i=1}^{l}(t_i^3 - t_i)\,/[n \cdot v(v^2 - 1)]\}}$$

$$= \frac{0{,}46}{1 - \{[(3^3 - 3) + (2^3 - 2) + (2^3 - 2) + (2^3 - 2)]/[10 \cdot 5(5^2 - 1)]\}}$$

$$= \frac{0{,}46}{1 - \frac{42}{1200}} = 0{,}477$$

V tabulce T6 v apendixu najdeme, že pro $n = 10$, $v = 5$ a $\alpha = 0{,}05$ je kritická hodnota $F(\alpha) = 9{,}25$.

Protože neplatí $F' \geq F(\alpha)$, nemůžeme zamítnout nulovou hypotézu, že všechny vzorky byly řazeny stejně. Výsledek interpretujeme tak, že při zvoleném riziku $\alpha = 0{,}05$ mezi vzorky neexistují rozdíly.

4.3 Porovnání zkoušek v rozhodnutí o podobnosti dvou výrobků

Na následujícím příkladu porovnáme vlastnosti a rozdíl ve vyhodnocení výsledků párové zkoušky podobnosti (oboustranný test) a pořadové zkoušky, ve které byly řazeny pouze dva výrobky.

Příklad 7. Výrobce chce začít prodávat ovocné sušenky. V rámci ušetření nákladů na výrobu by chtěl místo pravého másla používat v receptu levnější pečící tuk (Hera). Cílem zkoušky je zjistit, zda spotřebitelé tuto záměnu nepoznají a zda sušenky budou vnímat chuťově stejně lahodně s ohledem na máslovou chuť, neboť v takovém případě si výrobce bude moct tuto změnu receptu dovolit.

Pro párovou porovnávací zkoušku se zvolí hodnoty rizik $\alpha = 0{,}05$ a $\beta = 0{,}10$. Pro výrobce je důležité minimalizovat jak riziko α, že bude chybně rozhodnuto o rozdílech, které neexistují, tak riziko β, že existující rozdíly nebudou detekovány. Procento hodnotitelů, kteří rozdíly správně detekují, p_d je stanoveno na 40 %. Dle tabulky A5 v normě je minimální počet posuzovatelů 65, takže se výrobce rozhodne jich najmout 66.

Jsou připraveny dva vzorky – vzorek M s pravým máslem a vzorek T s levnějším pečícím tukem. Na jeden den jsou oba vzorky uskladněny v obalech, ve kterých výrobce bude sušenky prodávat. Dalšího dne jsou vzorky současně podávány posuzovatelům přímo v plastových táccích, které jsou součástí balení. Prvním 33 posuzovatelům jsou předkládány v pořadí MT a zbylým 33 posuzovatelům v pořadí TM.

Jako máslovější označilo vzorek M 39 posuzovatelů, tedy máme $x_M = 39$, a zbylých 27 posuzovatelů označilo jako máslovější vzorek T, tedy máme $x_T = 27$. Dle 1.1.20 je počet shodných odpovědí $x_0 = \max(x_M, x_T) = 39$. Pro stanovené hodnoty $\beta = 0{,}10$, $p_d = 40$ % a pro velikost panelu $n = 66$, se v tabulce A3 v normě zjistí, že maximální počet shodných odpovědí pro rozhodnutí, že jsou si vzorky podobné, je 40.

Protože platí nerovnost $x_0 \leq 40$, může být zkouška uzavřena s rozhodnutím, že na 90% hladině spolehlivosti mezi výrobky existuje podobnost a výrobce tedy může začít sušenky vyrábět s použitím levnějšího pečícího tuku.

Předpokládejme nyní, že byla k tomuto rozhodnutí použita pořadová zkouška. Všechny podmínky jsou stejné a ze zkoušky si výrobce odnesl stejná data. Tedy, počet posuzovatelů, kteří řadili jako první (máslovější) vzorek M, je $x_1 = 39$, a počet posuzovatelů, kteří řadili jako první vzorek T, je $x_2 = 27$.

Určí se $x_0 = \min(x_1, x_2) = 27$. Pro zvolenou hladinu $\alpha = 0{,}05$ se z tabulky kritických hodnot znaménkového testu (T8 v apendixu) vyčte $k(\alpha) = 24$. Jelikož neplatí $x_0 < k(\alpha)$, zkouška nabízí závěr, že nelze zamítnout nulovou hypotézu, že vzorky byly řazeny stejně. I z této zkoušky si tedy výrobce může odnést závěr, že vzorky jsou si podobné, a to se sílou testu 95 %.

5 Příklad párové porovnávací zkoušky na reálných datech

5.1 Úvodní komentář

Zápis protokolu o zkoušce měl být dle prvního plánu této práce ukázán na reálném příkladu párové zkoušky podobnosti. Testovaným výrobkem měly být ovocné sušenky a cílem testu by bylo ověřit, že když se v jejich složení zamění máslo za levnější pečící tuk (Hera), rozdíl v chuti nebude rozpoznatelný.

K párové zkoušce podobnosti je ovšem potřeba nejméně 50–60 posuzovatelů, resp. více, chceme-li se řídit pro test zvolenými přísnějšími hodnotami rizik α, β a hodnotou p_d. Vzhledem k tomu, že je třeba zkoušku provádět v jeden čas na jednom místě, tento příklad by v rámci současné epidemiologické situace 2020 a platných nařízení vlády nemohl být uskutečněn poctivě.

Proto jsem se nakonec rozhodla příklad upravit. Jako výrobek si ponecháme sušenky, ale budeme chtít naopak ověřit, že sušenky obsahující pravé máslo jsou rozpoznatelně chutnější než sušenky pečené s levnějším tukem. Půjde tedy o párovou rozdílovou zkoušku, která bude v současné situaci lépe proveditelná, a to z následujících důvodů:

1. Párová rozdílová zkouška vyžaduje pro dosažení stejné hladiny citlivosti zkoušky přibližně dvakrát méně posuzovatelů než párová zkouška podobnosti.
2. Na rozdíl od párové zkoušky podobnosti, párová rozdílová zkouška navíc dovoluje opakování testu se stejnými posuzovateli.

V následující kapitole bude popsán protokol o zkoušce. Komentáře a poznámky navíc, které se běžně v protokolu neuvádějí a slouží pouze k ilustraci postupu, budou uvedeny v hranatých závorkách.

5.2 Zápis protokolu o zkoušce

Účel a cíl zkoušky

Pečeme sušenky pro běžné spotřebitele. Doposud byly sušenky vyráběny za použití levného pečícího tuku (Hera s máslovou příchutí) – vzorek T. Se záměrem zlepšit jejich máslovou chuť bylo rozhodnuto začít v novém receptu místo Hery používat máslo – vzorek M.

Zkouškou chceme ověřit, zda je tato změna v chuti rozpoznatelná a sušenky vyráběné dle nového receptu jsou z hlediska máslové chuti vnímány jako lahodnější. [Jedná se tedy o párovou rozdílovou zkoušku, a navíc o jednostranný test, protože víme, že máslovější jsou nové sušenky M.]

Popis výrobku

Křupavé sušenky s kousky brusinek.

Jsou připravovány stejným způsobem, v receptu pouze zaměňujeme tuk Hera za stejné množství másla. Sušenky jsou vyrobeny stejně velké (8 gramů) a pečené ve stejných podmínkách na jednom místě.

Tvar sušenek není pro test relevantní a může se lehce lišit, stejně jako množství brusinek v jednotlivých sušenkách. Barva sušenek je pro oba vzorky stejná a není třeba nijak maskovat jejich vzhled.

Vzorky jsou po vychladnutí skladovány v plastových dózách v pokojové teplotě (22 °C), a ke zkoušce jsou určeny následující den.

Hladina citlivosti

[Ukažme nejdříve pro ilustraci v tabulce 8, jaký je požadovaný počet posuzovatelů n při některých hodnotách p_d v závislosti na hladinách rizik α a β.

Tabulka 8: Minimální celkový počet posuzovatelů n pro jednostranný párový test.

p_d	α	β			
		0,50	0,20	0,10	0,5
10 %	0,50	–	75	167	271
	0,20	81	294	451	618
	0,10	170	461	658	861
	0,05	281	620	866	1092
30 %	0,50	–	–	23	33
	0,20	–	32	49	68
	0,10	21	53	72	96
	0,05	30	69	93	119
40 %	0,50	–	–	9	20
	0,20	–	19	30	39
	0,10	14	28	39	53
	0,05	18	37	53	67

Pozn.: Prázdné buňky v této tabulce značí v praxi neuvažované kombinace hladin rizik α, β a hodnoty p_d. Například, kdybychom uvažovali $p_d = 100$ % (rozdíl bude vždy rozeznán správně), rizika α a β by nutně vždy byla nulová, tudíž vyšší hladiny nepředstavují praktický význam. Obdobně i pro nižší hodnoty p_d některé kombinace hladin rizik α a β nedávají smysl.

Z tabulky můžeme vyčíst, že pro příliš nízkou hladinu p_d je požadovaný počet posuzovatelů nerealistický. Připomeňme, že v praxi firmy s ohledem na reálné a finanční možnosti vybírají pro jednostrannou párovou zkoušku přibližně 24–30 posuzovatelů.

Tohoto rozmezí dosáhneme až na úrovni $p_d = 30$ %. Vidíme ale, že realistické jsou pouze hodnoty při příliš vysoké hladině alespoň jednoho z rizik α a β. Zjevně bychom dosáhli mnohem nižších čísel i při nižších hladinách obou rizik pro hodnoty $p_d \geq 50$ %. Takový předpoklad ovšem není pro praxi reálný, norma obsahuje tabelované hodnoty pro nejvýše $p_d = 50$ %. Zejména pro panel posuzovatelů složeného z vybraných posuzovatelů bude hodnota p_d jistě nižší než pro experty.

Pokud tedy chceme zůstat v rozumné mezi počtu posuzovatelů přizvaných ke zkoušce při realistické hodnotě p_d, budeme se muset smířit s vyššími riziky α a β.]

Koeficient p_d je zvolen 40 %.

Riziko, že bude vnímán rozdíl, který neexistuje, chceme co nejvíce snížit. Zvolíme $\alpha = 0{,}05$.

Vyšší riziko, že nebude vnímán rozdíl, který existuje, jsme ochotni přijmout. Zvolíme $\beta = 0{,}20$.

Panel posuzovatelů

Minimální počet posuzovatelů (resp. potřebných pozorování) byl určen na základě určených koeficientů α, β a p_d – 37. Celkem bylo ke zkoušce nakonec přizváno 10 posuzovatelů, se kterými byla zkouška $4 \times$ opakována, což dalo dohromady 40 uskutečněných pozorování.

Všichni posuzovatelé neměli zkušenost se senzorickými zkouškami, ale byli obeznámeni s výrobkem.

Místo a datum zkoušky

Domácnost s pokojovou teplotou 22 °C.

Datum provedení zkoušky: 6. 12. 2020.

Předkládání vzorků

Vzorky byly předkládány současně na keramických talířcích, jedna sušenka od každého výrobku. Každému z posuzovatelů byly $2 \times$ předloženy v pořadí TM a $2 \times$ v pořadí MT. V každém opakování byly vzorky předloženy 5 posuzovatelům v pořadí TM a zbylým 5 v pořadí MT.

[Odpovědní formulář s konkrétními pokyny je přiložen v příloze *2*.]

Výsledky a závěr

[Testována je nulová hypotéza $H_0\colon p = \frac{1}{2}$. Alternativou je $H_1\colon p > \frac{1}{2}$.]

Počet správných odpovědí (tj. jako máslovější byl posuzovateli zvolen vzorek M) je $x_0 = 19$.

K závěru, že je mezi vzorky podstatný rozdíl, je pro $n = 40$ a $\alpha = 0{,}05$ potřeba alespoň 25 správných odpovědí.

Tedy, protože neplatí $x_0 \geq 25$, můžeme usoudit, že rozdíl mezi sušenkami není dostatečně vnímatelný. Z některých poznámek ve vyplněných odpovědních formulářích naopak vyplynulo, že vzorek T je preferovaný pro svoji větší jemnost a křehkost, na rozdíl od vzorku M, který posuzovatelé hodnotili jako křupavější. Velká část hodnocení byly navíc pouze odhady.

Vyrábět sušenky dle nového receptu se tedy nevyplatí.

Interval spolehlivosti

Pro odhad parametru p_d spočítáme na základě podílu správných odpovědí jednostranný interval spolehlivosti.

Máme $x_0 = 19$, $n = 40$, a z toho $p_{x_0} = \frac{x_0}{n} = \frac{19}{40} = 0{,}475$.

Parametr p_d odhadujeme jako $\widehat{p_d} = 2p_{x_0} - 1 = 2 \cdot 0{,}475 - 1 = -0{,}05$.

Standardní chyba je $s_d = 2\sqrt{\frac{nx_0 - x_0^2}{n^3}} = 2\sqrt{\frac{40 \cdot 19 - 19^2}{40^3}} = 0{,}15792$.

Interval spolehlivosti má tvar $(-\infty, \widehat{p_d} + z_\alpha s_d)$.

Spočítáme horní hranici intervalu:

$$\widehat{p_d} + z_\alpha s_d = -0{,}05 + 1{,}64 \cdot 0{,}15792 = 0{,}201.$$

Z toho vyplývá, že s 95% pravděpodobností, podíl konzumentů schopných rozlišit tyto dva vzorky nepřesáhne 20 %.

To podporuje výsledek, že mezi vzorky není dostatečný rozdíl.

Doslov

K testování senzorických rozdílů mezi dvěma nebo více vzorky lze využít mnoho přístupů. Některé z těchto přístupů byly popsány v existujících normách. V současnosti se stávají populárními i některé další postupy, pro které se normy teprve připravují, např. tetrádová zkouška.

Senzoričtí analytici při řešení problému často spoléhají na ustálené oblíbené metody. Aby se zvolená metoda dala využít, řešený problém zjednodušují nebo pozměňují. To ale znamená nepřesnost testu a může vést i ke snížení pravděpodobnosti odhalení rozdílů, protože některá jiná metoda by k tomu mohla být vhodnější.

Párová zkouška podobnosti, která je popsána v této práci, je jednou z nejpoužívanějších pro testování rozdílů intenzity jednoho konkrétního deskriptoru ve dvou vzorcích. Otázku, zda mezi dvěma vzorky existují nějaké rozdíly, je-li rozdíl v senzorické vlastnosti neznámý, ale lépe zodpoví například trojúhelníková zkouška.

Druhá ze zkoušek popsaných v této práci, pořadová zkouška, patří do stejné skupiny testů jako párová zkouška podobnosti. Testuje rozdíl v intenzitě jednoho deskriptoru mezi dvěma a více výrobky, typicky 3–6.

Postup určení ideálního testu k řešení problému popisuje například Meilgaard, Carr & Civille (2006).

Mimo povahy výrobku a počtu vzorků záleží zejména na účelu testu. Jde o vývoj nového produktu, pokus o vylepšení stávajícího produktu, snížení nákladů na výrobu stávajícího produktu, nebo změnu ve výrobním, či skladovacím procesu produktu? Je požadováno testovat kvalitu výrobku, názor a preferenci konzumentů nebo trénovat posuzovatele?

Při výběru vhodného testu je vhodné též přihlížet k povaze testované vlastnosti. Pravděpodobně nejčastější využití sensometrických zkoušek je v potravinářském průmyslu k testování chuti pokrmů, či nápojů. Lze ovšem samozřejmě testovat rozličné vlastnosti vnímané ostatními smysly, a k tomu mohou být vhodné jiné typy zkoušek. Například může jít o vůni výrobku, kdy není příliš praktické hodnotit mnoho vzorků najednou, nebo o texturu výrobku či pohodlnost jeho užívání, kdy naopak může být vhodnější drobné rozdíly pozorovat na více vzorcích.

Nedá se říct, že některý postup pro testování rozdílů mezi výrobky je vyloženě špatný. Výběr testu bude vždy, stejně jako hodnocení posuzovatelů, velmi subjektivní. V každém případě by ale analytik měl využít postup co možná nejvhodnější, ačkoliv jeho realizace může být složitější nebo ekonomicky nákladnější. Na druhou stranu nelze přehlížet, že kvalitní sensometrické testování může být opravdu velice nákladnou záležitostí, a je tedy nutné vyvážit kvalitu s finančními možnostmi.

Seznam použité literatury

Anděl, J. (1985). *Matematická statistika*. SNTL, Praha.

Anděl, J. (2007). *Statistické metody*. Matfyzpress, Praha.

Český normalizační institut. (2001). *ČSN ISO 8588 (560038) Senzorická analýza – Metodologie – Zkouška „A"–"ne A"*.

Český normalizační institut. (2008). *ČSN ISO 8587 (560033) Senzorická analýza – Metodologie – Pořadová zkouška*.

Hátle, J., Likeš, J. (1974). *Základy počtu pravděpodobnosti a matematické statistiky*. SNTL, Praha.

Lawless, H. T. (2012). *Laboratory Exercises for Sensory Evaluation*. Springer Science & Business Media, New York.

MacRae, A. W. (1995). Confidence intervals for the triangular test. *Food Quality and Preference*, *Volume 6, Issue 2*, 61–67.

Meilgaard, M. C., Carr, B. T., & Civille, G. V. (2006). *Sensory Evaluation Techniques, Fourth Edition*. CRC Press, Boca Raton.

Næs, T., Brockhoff, P. B., & Tomic, O. (2011). *Statistics for Sensory and Consumer Science*. John Wiley & Sons, New York.

Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association*, *Volume 58, Number 301*, 216–230.

Schlich, P. (1993). Risk Tables for Discrimination Tests. *Food Quality and Preference*, *Volume 4, Issue 3*, 141–151.

Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009a). *ČSN EN ISO 4120 (560032) Senzorická analýza – Metodologie – Trojúhelníková zkouška*.

Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009b). *ČSN EN ISO 5492 (560030) Senzorická analýza – Slovník*.

Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009c). *ČSN EN ISO 5495 (560032) Senzorická analýza – Metodologie – Párová porovnávací zkouška*.

Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2015). *ČSN EN ISO 8586 (560037) Senzorická analýza – Obecná směrnice pro výběr, výcvik a sledování činnosti vybraných posuzovatelů a odborných senzorických posuzovatelů*.

Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2018). *ČSN EN ISO 10399 (560032) Senzorická analýza – Metodologie – Zkouška duo-trio*.

Seznam tabulek

Apendix

Následující tabulky slouží pro ilustraci a zobrazují pouze vybrané hodnoty. Tabulky byly převzaté z příslušných norem (Úřad pro technickou normalizaci, metrologii a státní zkušebnictví. (2009c), Český normalizační institut. (2008)), nebo ze zdrojů příslušných jednotlivým metodám (Schlich (1993), Anděl (1985), Page (1963)), kde je lze nalézt v jejich rozsáhlejší verzi.

T1: Minimální celkový počet pozorování n *a příslušný minimální počet správných odpovědí* x_{min} *pro jednostranný rozdílový test, pro vybrané hodnoty* α *a* β *(pro párovou porovnávací zkoušku)*

p_d	α	β					
		0,20		0,05		0,01	
		n	x_{min}	n	x_{min}	n	x_{min}
50 %	0,20	12	8	26	16	39	23
	0,10	19	13	33	21	48	29
	0,05	23	16	42	27	58	36
	0,01	40	28	59	39	107	70
40 %	0,20	19	12	39	23	60	34
	0,10	28	18	53	32	79	46
	0,05	37	24	67	41	93	55
	0,01	64	42	96	60	130	79
30 %	0,20	32	19	68	38	110	60
	0,10	53	32	96	55	145	81
	0,05	69	42	119	69	173	98
	0,01	112	69	174	103	235	136
20 %	0,20	77	43	158	85	253	134
	0,10	115	65	214	117	322	173
	0,05	158	90	268	148	392	213
	0,01	252	145	391	219	535	295
10 %	0,20	294	115	618	320	1 006	517
	0,10	461	245	861	450	1 310	679
	0,05	620	331	1 092	574	1 583	825
	0,01	1 007	541	1 582	838	2 170	1 140

T2: Kritické hodnoty standardního normálního rozdělení (pro párovou porovnávací zkoušku).

Hladina $1 - \alpha$	**Kritická hodnota** z_a	
	Jednostranný test	Oboustranný test
0,80	0,84	1,28
0,90	1,28	1,64
0,95	1,64	1,96
0,99	2,33	2,58
0,999	3,10	3,29

T3: Kritické hodnoty Spearmanova korelačního koeficientu $r_s(\alpha)$*, pro vybrané hodnoty* v.

Počet výrobků v	**Hodnota** α	
	0,05	0,01
5	0,900	–
6	0,829	0,943
7	0,745	0,893
8	0,691	0,857
9	0,683	0,817
10	0,636	0,782
11	0,609	0,755
12	0,580	0,727
13	0,555	0,698
14	0,534	0,675
15	0,518	0,654

T4: Kritické hodnoty Pageova koeficientu $L_{n,v}(\alpha)$, pro vybrané hodnoty n a v.

Počet posuzovatelů n	**Počet výrobků** v					
	3	5	10	3	5	10
	Hodnota $\alpha = 0{,}05$			**Hodnota** $\alpha = 0{,}01$		
10	128	477	*3 169*	131	487	*3 228*
15	190	*707*	*4 714*	194	*721*	*4 786*
20	251	*937*	*6 253*	256	*953*	*6 337*
50	*617*	*2 309*	*15 446*	*624*	*2 333*	*15 579*

Pozn.: Hodnoty uvedené kurzívou jsou aproximované normálním rozdělením.

T5: Kritické hodnoty standardního normálního rozdělení (pro pořadovou zkoušku).

Hodnota α	**Kritická hodnota** z_α
0,05	1,64
0,01	2,33

T6: Kritické hodnoty pro Friedmanovu zkoušku $F_{n,v}(\alpha)$, pro vybrané hodnoty n a v.

Počet posuzovatelů n	**Počet výrobků** v							
	3	5	10	15	3	5	10	15
	Hodnota $\alpha = 0{,}05$				**Hodnota** $\alpha = 0{,}01$			
10	6,20	*9,25*	*16,44*	*23,10*	9,60	*12,38*	*20,53*	*27,90*
15	6,40	*9,33*	*16,60*	*23,20*	8,93	*12,68*	*20,90*	*28,20*
20	*5,99*	*9,37*	*16,70*	*23,30*	*8,87*	*12,83*	*21,10*	*28,40*
∞	*5,99*	*9,49*	*16,92*	*23,69*	*9,21*	*13,28*	*21,67*	*29,14*

Pozn.: Hodnoty uvedené kurzívou jsou aproximované χ^2 rozdělením.

T7: Kritické hodnoty χ^2 rozdělení $\chi^2_{v-1}(\alpha)$, pro vybrané hodnoty v.

Počet výrobků v	**Počet stupňů volnosti $v-1$**	**Hodnota α**	
		0,05	0,01
5	4	9,49	13,28
6	5	11,07	15,09
7	6	12,59	16,81
8	7	14,07	18,48
9	8	15,51	20,09
10	9	16,92	21,67
11	10	18,31	23,21
12	11	19,68	24,72
13	12	21,03	26,22
14	13	22,36	27,69
15	14	23,68	29,14

T8: Kritické hodnoty znaménkového testu $k(\alpha)$, pro vybrané hodnoty n.

Počet posuzovatelů n	**Hodnota α**	
	0,05	0,01
10	1	0
15	3	2
20	5	3
50	17	15
100	39	36

Přílohy

Příloha 1

a – Odpovědní formulář pro párovou porovnávací zkoušku

Párová porovnávací zkouška

Jméno: .... Kód posuzovatele: ....

Instrukce: Datum: ....

Ochutnejte dva vzorky, začněte vzorkem vlevo. Níže označte křížkem kód vzorku, který je sladší. Pokud si volbou nejste jisti, označte i přesto jednu možnost a do poznámek uveďte, že šlo jen o odhad.

☐ 524 ☐ 679

Poznámky: ....
....

b – Odpovědní formulář pro pořadovou zkoušku s 5 vzorky

Pořadová zkouška

Jméno: .... Kód posuzovatele: ....

Instrukce: Datum: ....

Prosím, zapište kódy obdržených vzorků zleva doprava:

Ochutnejte všechny vzorky, začněte vzorkem vlevo. Před každým hodnocením se napijte vody. Do rámečků níže uveďte kódy vzorků podle sladkosti vzestupně, tj. nejméně sladký úplně vlevo a nejvíce sladký úplně vpravo:

	Nejméně				Nejvíce
Kód:					

Poznámky: ....

Příloha 2: Odpovědní formulář příkladu na reálných datech

Párová porovnávací zkouška

Jméno: .... Datum: ....

Instrukce:

Ochutnejte dva vzorky, začněte vzorkem vlevo. Před každým hodnocením se napijte vody. Hodnocení můžete opakovat. Níže označte křížkem kód vzorku, který chutná máslověji. Pokud si volbou nejste jisti, označte i přesto jednu možnost a do poznámek uveďte, že šlo jen o odhad.

☐ 963 ☐ 704

Poznámky: ....
....