UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Testování statistických hypotéz na základě fuzzy dat

Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **RNDr. Ondřej Pavlačka, Ph.D.**
Vypracoval: **Bc. Lukáš Konopka**
Studijní program: B1103 Aplikovaná matematika
Studijní obor Aplikace matematiky v ekonomii
Forma studia: prezenční
Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Lukáš Konopka

Název práce: Testování statistických hypotéz na základě fuzzy dat

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: RNDr. Ondřej Pavlačka, Ph.D.

Rok obhajoby práce: 2017

Abstrakt: Tato diplomová práce se zabývá testováním statistických hypotéz na základě fuzzy dat. V první kapitole jsou zavedeny základní pojmy z teorie fuzzy množin, a to především definice fuzzy množiny, fuzzy čísla, standardní fuzzy aritmetiky, podmíněné fuzzy aritmetiky a různých způsobů defuzzifikace. V druhé kapitole je nejdříve popsána teorie testování statistických hypotéz s obecným postupem při testování. Následně je zde představena hlavní problematika testování statistických hypotéz na základě fuzzy dat. Poté následuje uvedení konkrétních vybraných parametrických testů, jmenovitě jednovýběrový t-test, párový t-test, test hypotéz o rozptylu normálního rozdělení, F-test shody rozptylů dvou nezávislých normálně rozdělených znaků a přibližný t-test. Každý vybraný test je nejdříve popsán pro reálná data a poté pro fuzzy data. Na konci každého testu je uveden příklad s postupem jeho řešení.

Klíčová slova: statistika, testování hypotéz, fuzzy množina, fuzzy číslo, vstupní fuzzy data

Počet stran: 65

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Lukáš Konopka

Title: Statistical tests for fuzzy data

Type of thesis: Diploma's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: RNDr. Ondřej Pavlačka, Ph.D.

The year of presentation: 2017

Abstract: This Master's thesis deals with the statistical tests for fuzzy data. The first chapter of the thesis consists of the basic terms from fuzzy sets theory, mainly of the definitions of fuzzy sets, fuzzy number, standard fuzzy arithmetics, constrained fuzzy arithmetic and defuzzification. In the second chapter there is firstly explained the theory of testing hypothesis with basic procedure of testing. Subsequently, the main aim of this thesis which is statistical testing based on the fuzzy data is introduced. After that, this thesis focuses on particular parametric tests such as t-test, t-test for paired samples, Z-test, F-test and independent two-sample test. Each test is firstly described for real data then for the fuzzy data. In the end of each test is introduced an example with their solving.

Key words: statistics, testing hypothesis, fuzzy set, fuzzy number, fuzzy data

Number of pages: 65

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracoval samostatně pod vedením pana RNDr. Ondřeje Pavlačky, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne ....

podpis

Obsah

Poděkování

Chtěl bych poděkovat svému vedoucímu RNDr. Ondřeji Pavlačkovi, Ph.D. za všechen věnovaný čas a pomoc, kterou mi věnoval. Dále bych rád poděkoval své rodině a přátelům za jejich podporu během studia a psaní této práce.

Úvod

Teorie testovaní statistických hypotéz a teorie fuzzy množin, to jsou na první pohled úplně odlišné disciplíny. Při testování statistických hypotéz máme vstupní data reálná, zato v případě fuzzy dat počítáme s vágností nebo-li neurčitostí. Můžeme vlastně získat vstupní data, která budou fuzzy? Je možné na základě fuzzy dat ověřovat různé hypotézy? Jak posléze správně rozhodnout o závěru testování? Nejen těmito otázkami se budeme v této práci zabývat. Pro lepší představu si uvedeme motivační příklad. Představme si, že nás někdo na ulici zastaví s anketou týkající se průměrné měsíční útraty v korunách. Pravděpodobně nevíme úplně přesnou částku, tak odpovíme nějakou přibližnou částku. Náhle už budeme mít vstupní data fuzzy, ale i tato data bychom chtěli nějakým způsobem otestovat, a proto se tedy v této diplomové práci budeme právě zabývat testováním statistických hypotéz na základě fuzzy dat.

Na začátku této práce si nejdříve představíme základní pojmy z teorie fuzzy množin. Poté se v následující kapitole nejprve budeme zabývat základními pojmy a postupy při klasickém testování statistických dat a posléze už obecným postupem testování statistických hypotéz v případě fuzzy dat. Následně se detailně zaměříme na vybrané statistické testy, a to konkrétně na parametrické. Ke každému vybranému testu si uvedeme řešený příklad k pochopení problematiky.

Kapitola 1

Teorie fuzzy množin

V této kapitole se seznámíme se základními pojmy z teorie fuzzy množin. Zavedení těchto pojmů je důležité z hlediska následného pochopení další kapitoly této práce. Tato kapitola byla zpracována podle [6],[8],[11],[12],[14],[15],[16],[19]. Obrázky jsou vytvořeny v programu MATLAB.

1.1. Fuzzy množiny

Definice 1.1 *Nechť je daná neprázdná množina U, tzv. univerzum. Pak fuzzy množina A na univerzu U je definována zobrazením $\mu_A : U \rightarrow \langle 0, 1\rangle$. Funkci μ_A nazýváme funkcí příslušnosti fuzzy množiny A. Pro každé $x \in U$ nazveme hodnotu $\mu_A(x)$ stupněm příslušnosti prvku x k fuzzy množině A.*

Funkce příslušnosti fuzzy množiny může nabývat i hodnot mezi krajními hodnotami 0 a 1. To znamená, že pokud máme například fuzzy množinu "vysoký člověk" a pro konkrétního člověka je stupeň příslušnosti k této fuzzy množině 0,5, tak do dané fuzzy množiny "vysoký člověk" patří napůl.

Nyní je potřeba si zavést nezbytné pojmy, které popisují fuzzy množinu. S těmito pojmy budeme posléze pracovat.

Poznámka 1.1 *Pro zjednodušení zápisu budeme dříve značenou funkci příslušnosti μ_A zapisovat jako $A(.)$ a následně stupeň příslušnosti $\mu_A(x)$ jako $A(x)$, kde $x \in U$.*

Definice 1.2 *Nechť je dána fuzzy množina A definovaná na univerzu U a reálné číslo $\alpha \in \langle 0, 1\rangle$. Pak α-řezem fuzzy množiny A nazýváme (ostrou) množinu*

$$A_\alpha = \{x \in U \mid A(x) \geq \alpha\}.$$

Definice 1.3 *Jádrem fuzzy množiny A na univerzu U rozumíme (ostrou) množinu*

$$Ker A = \{x \in U \mid A(x) = 1\}.$$

Definice 1.4 *Nosičem fuzzy množiny A na univerzu U nazýváme (ostrou) množinu*

$$Supp\ A = \{x \in U \mid A(x) > 0\}.$$

Definice 1.5 *Výška $hgt(A)$ fuzzy množiny A na univerzu U je definována následovně:*

$$hgt(A) = sup_{x \in U} A(x).$$

Definice 1.6 *Fuzzy množina A na univerzu U se nazývá normální, jestliže*

$$Ker A \neq \emptyset.$$

V opačném případě se nazývá subnormální.

Tomuto způsobu, kdy se fuzzy množina zadefinuje pomocí funkce příslušnosti, se také někdy říká vertikální reprezentace. Druhý způsob zadefinování fuzzy množiny je pomocí α-řezů. Princip spočívá v tom, že máme systém řezů fuzzy množiny A, kde každému $\alpha \in \langle 0, 1\rangle$ přiřazujeme tzv. α-řez. Zápisu pomocí α-řezů se taky někdy říká horizontální reprezentace. Jednotlivé pojmy jsou pro lepší pochopení zobrazeny na obrázku 1.1.

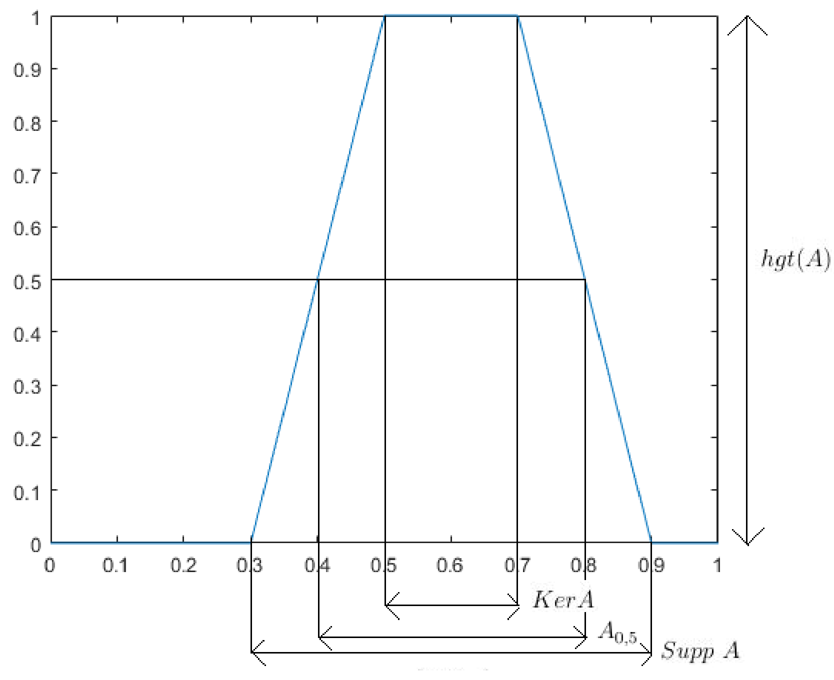
1.2. Fuzzy čísla

Kvantitativní údaje, které nám vyjadřují přibližná, neurčitá množství, jako například „přibližně 5," „něco mezi 100-150" atd. můžeme reprezentovat pomocí fuzzy čísel.

Definice 1.7 *Fuzzy množina C definovaná na množině reálných čísel $\mathbb{R}$, která má následující vlastnosti:*

1. *C je normální fuzzy množina,*
2. *α-řezy C_α představují pro všechna $\alpha \in (0,1\rangle$ uzavřené intervaly,*
3. *nosič $Supp\, C$ je ohraničený,*

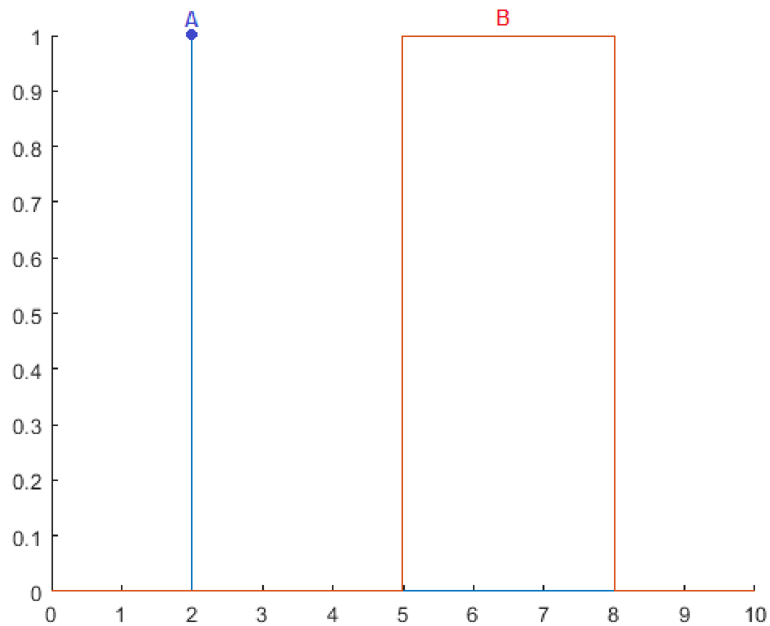
se nazývá fuzzy číslem.



Obrázek 1.1: Grafická reprezentace základních charakteristik fuzzy množin

Poznámka 1.2 *Symbolem $F_N(\mathbb{R})$ budeme označovat množinu všech fuzzy čísel.*

Poznámka 1.3 *Za speciální případ fuzzy čísla lze považovat reálné číslo a interval. Odpovídající funkce příslušností jsou znázorněné na obrázku 1.2. Můžeme tedy chápat fuzzy modely jako rozšíření reálných modelů a také pracovat dohromady s reálnými i fuzzy čísly.*



Obrázek 1.2: Funkce příslušnosti fuzzy čísel reprezentujících reálné číslo 2 a interval $\langle 5, 8\rangle$.

Následující věta nám umožní si udělat lepší představu o charakteru funkcí příslušnosti fuzzy čísel.

Věta 1.1 *Nechť C je fuzzy množina na $\mathbb{R}$. Pak C je fuzzy číslo právě tehdy, když existují $x_1, x_2, x_3, x_4 \in \mathbb{R}$, taková že $x_1 \leq x_2 \leq x_3 \leq x_4$, a pro funkci příslušnosti fuzzy čísla C platí*

$$C(x) = \begin{cases} L(x), & \text{pro } x \in (-\infty, x_2) \\ 1, & \text{pro } x \in \langle x_2, x_3\rangle \\ P(x), & \text{pro } x \in (x_3, \infty), \end{cases}$$

kde:

a) $L(x)$ je neklesající a zprava spojitá funkce na $(-\infty, x_2)$ a $L(x) = 0$ pro $(-\infty, x_1)$,

b) $P(x)$ je nerostoucí a zleva spojitá funkce na (x_3, ∞) a $P(x) = 0$ pro (x_4, ∞).

Bodům x_1, x_2, x_3, x_4 z předchozí věty říkáme význačné hodnoty fuzzy čísla C a musí splňovat následující podmínky:

a) $\langle x_1, x_4\rangle = \overline{Supp}\, C$, kde $\overline{Supp}\, C$ značí uzávěr nosiče,

b) $\langle x_2, x_3\rangle = Ker\ C$.

Každé fuzzy číslo je jednoznačně určeno pomocí dvojice reálných funkcí $\underline{c}(\alpha)$, $\overline{c}(\alpha)$, které popisují nejmenší a největší hodnoty jednotlivých α-řezů, jak je ukázáno v následujících dvou větách.

Věta 1.2 *Nechť C je fuzzy číslo a funkce $\underline{c}(\alpha)$, $\overline{c}(\alpha)$ na $\langle 0,1\rangle$ jsou definovány následovně: $\langle \underline{c}(\alpha), \overline{c}(\alpha)\rangle = C_\alpha$, pro $\alpha \in (0,1\rangle$ a $\langle \underline{c}(0), \overline{c}(0)\rangle = \overline{Supp}\, C$. Pak pro funkce $\underline{c}$, $\overline{c}$ platí:*

a) $\forall \alpha \leq \beta, kde\ \alpha, \beta \in \langle 0,1\rangle : \underline{c}(\alpha) \leq \underline{c}(\beta) \leq \overline{c}(\beta) \leq \overline{c}(\alpha)$,

b) funkce $\underline{c}$, $\overline{c}$ jsou spojité zleva na $(0,1\rangle$ a zprava v 0.

Věta 1.3 *Nechť $\underline{c}(\alpha)$, $\overline{c}(\alpha)$ na $\langle 0,1\rangle$ jsou spojité zleva na $(0,1\rangle$, zprava spojité v 0 a splňují nerovnost $\forall \alpha \leq \beta : \underline{c}(\alpha) \leq \underline{c}(\beta) \leq \overline{c}(\beta) \leq \overline{c}(\alpha)$. Pak $\langle \underline{c}(\alpha), \overline{c}(\alpha)\rangle$ pro $\forall \alpha \in (0,1\rangle$ jsou α-řezy nějakého fuzzy čísla a $\langle \underline{c}(0), \overline{c}(0)\rangle$ je jeho uzávěr nosiče.*

Poznámka 1.4 *První věta je dokázána v [7]. Druhá a třetí v [13].*

Fuzzy číslo C budeme zapisovat pomocí této dvojice funkcí $\underline{c}, \overline{c}$ následovně:

$$C = \{\langle \underline{c}(\alpha), \overline{c}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}.$$

Když dostaneme neurčité vstupní hodnoty, je našim cílem použít co nejjednodušší matematický model a zároveň získat co nejpřesnější výsledky. Nejjednodušší fuzzy číslo se nazývá lineární.

Definice 1.8 *Nechť C je fuzzy množina na $\mathbb{R}$ a $x_1 \leq x_2 \leq x_3 \leq x_4$ jsou význačné hodnoty fuzzy čísla. Fuzzy číslo C se nazývá lineární, jestliže jeho funkce příslušnosti má pro $\forall x \in \mathbb{R}$ tvar*

$$C(x; x_1, x_2, x_3, x_4) = \begin{cases} \frac{x-x_1}{x_2-x_1}, & pro \quad x_1 \leq x \leq x_2 \\ 1, & pro \quad x_2 \leq x \leq x_3 \\ \frac{x_4-x}{x_4-x_3}, & pro \quad x_3 \leq x \leq x_4 \\ 0 & \text{jinak.} \end{cases}$$

V této práci budeme označovat lineární fuzzy číslo pomocí čtveřice významných bodů $C = \langle x_1, x_2, x_3, x_4 \rangle$. Obě funkce $\underline{c}, \overline{c}$ budou tedy lineární a pro všechny $\alpha \in \langle 0, 1 \rangle$ je můžeme spočítat následovně:

$$\underline{c}(\alpha) = x_1 + \alpha(x_2 - x_1),$$

$$\overline{c}(\alpha) = x_4 + \alpha(x_4 - x_3).$$

Mezi nejznámější lineární fuzzy čísla patří lichoběžníkové fuzzy číslo, kde $x_1 \neq x_2 \neq x_3 \neq x_4$ a trojúhelníkové fuzzy číslo s význačnými body $x_1 \neq x_2 = x_3 \neq x_4$. My se ovšem během práce setkáme i se situací, kdy známe pouze krajní hodnoty určitých α-řezů. Výsledné fuzzy číslo je pak aproximováno pomocí takzvaného *po částech lineárního fuzzy čísla.*

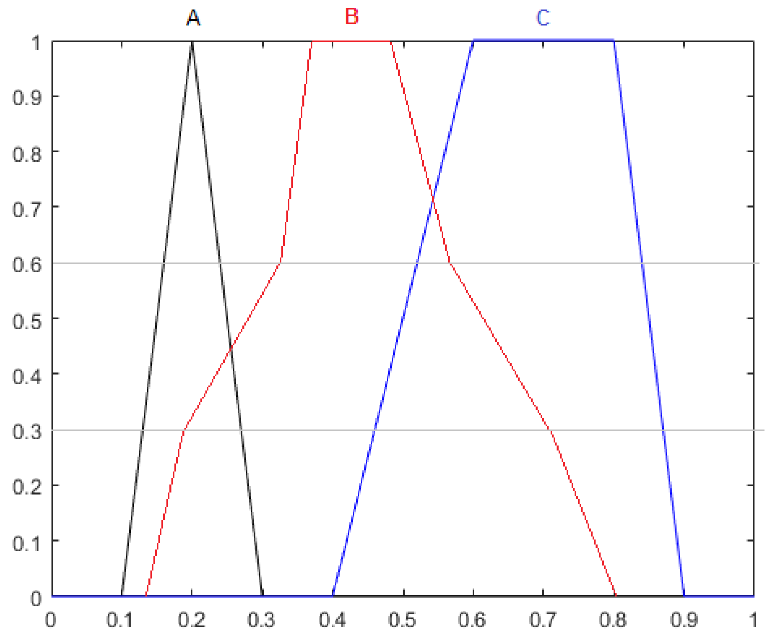
Definice 1.9 *Po částech lineárním fuzzy číslem určeným pomocí bodů* $(\underline{c}(\alpha_i), \alpha_i)$ *a* $(\overline{c}(\alpha_i), \alpha_i), i = 1, \ldots, n,$ *kde* $\alpha_i \leq \alpha_{i+1}$ *pro* $i = 1, \ldots, n-1,$ *a platí* $\alpha_1 = 0$ *a* $\alpha_n = 1,$ *nazveme fuzzy číslo C, jehož funkci příslušnosti můžeme definovat následujícím způsobem*

$$C(x) = \begin{cases} 0 & x < \underline{c}(\alpha_1) \\ (\alpha_2 - \alpha_1) \cdot \frac{x - \underline{c}(\alpha_1)}{\underline{c}(\alpha_2) - \underline{c}(\alpha_1)} + \alpha_1 & \underline{c}(\alpha_1) \leq x < \underline{c}(\alpha_2) \\ \ldots & \\ (\alpha_n - \alpha_{n-1}) \cdot \frac{x - \underline{c}(\alpha_{n-1})}{\underline{c}(\alpha_n) - \underline{c}(\alpha_{n-1})} + \alpha_{n-1} & \underline{c}(\alpha_{n-1}) \leq x < \underline{c}(\alpha_n) \\ 1 & \underline{c}(\alpha_n) \leq x < \overline{c}(\alpha_n) \\ (\alpha_n - \alpha_{n-1}) \cdot \frac{\overline{c}(\alpha_{n-1}) - x}{\overline{c}(\alpha_{n-1}) - \overline{c}(\alpha_n)} + \alpha_{n-1} & \overline{c}(\alpha_n) < x \leq \overline{c}(\alpha_{n-1}) \\ \ldots & \\ (\alpha_2 - \alpha_1) \cdot \frac{\overline{c}(\alpha_1) - x}{\overline{c}(\alpha_1) - \overline{c}(\alpha_2)} + \alpha_1 & \overline{c}(\alpha_2) < x \leq \overline{c}(\alpha_1) \\ 0 & \overline{c}(\alpha_1) < x. \end{cases}$$

Na obrázku 1.3 můžeme pak v praxi vidět jak vypadá takové trojúhelníkové, po částech lineární a lichoběžníkové fuzzy číslo.

1.3. Standardní intervalová a fuzzy aritmetika

S fuzzy čísly můžeme provádět jednotlivé základní aritmetické operace jako sčítání, odečítání, násobení a dělení. Abychom ovšem mohli vůbec provádět tyto



Obrázek 1.3: Trojúhelníkové fuzzy číslo A, po částech lineární fuzzy číslo B a lichoběžníkové fuzzy číslo C

jednotlivé aritmetické operace, je nutné si zadefinovat tzv. *princip rozšíření*, jehož autorem je profesor Lotfi Askar Zadeh [19]. Symbolem $F(U)$ označujeme systém všech fuzzy množin na U. $A \in F(U)$ znamená, že fuzzy množina A je definována na U.

Definice 1.10 *Fuzzifikací zobrazení $f : U \to V$ rozumíme zobrazení:*

$$f_F : F(U) \to F(V),$$

které každé fuzzy množině $A \in F(U)$ přiřazuje fuzzy množinu $f_F(A) \in F(V)$ s funkcí příslušnosti definovanou pro každé $y \in V$ vztahem

$$f_F(A)(y) = \begin{cases} \sup\{A(x) \,|\, f(x) = y, x \in U\}, & \textit{když } \{x \in U \,|\, f(x) = y\} \neq \emptyset, \\ 0 & \textit{jinak.} \end{cases}$$

Princip rozšíření je základem pro veškeré operace s fuzzy množinami a říká nám, jak je reálnou funkcí f a fuzzy množinou A na $\mathbb{R}$ indukována fuzzy množina $B = f_F(A)$ na $\mathbb{R}$.

Fuzzy aritmetika nám vychází z intervalové aritmetiky. Připomeňme si, že fuzzy číslo představuje fuzzy množinu s omezeným nosičem a každé fuzzy číslo je jednoznačně určeno pomocí dvojice reálných funkcí, které jsme si zadefinovali v minulé podkapitole. Tedy veškeré aritmetické operace, které provádíme s fuzzy čísly, se při vyjádření fuzzy čísla v podobě α-řezů dají definovat pomocí intervalové aritmetiky, jelikož vlastně v praxi provádíme operace s intervaly.

Základní operace intervalové aritmetiky pro dva intervaly $\langle a, b\rangle$ a $\langle c, d\rangle$ jsou následující:

- sčítání

$$\langle a, b\rangle + \langle c, d\rangle = \langle a + c, b + d\rangle$$

- odečítání

$$\langle a, b\rangle - \langle c, d\rangle = \langle a - d, b - c\rangle$$

- násobení

$$\langle a, b\rangle \cdot \langle c, d\rangle = \langle min\{ac, ad, bc, bd\}, max\{ac, ad, bc, bd\}\rangle$$

- dělení

$$\frac{\langle a, b\rangle}{\langle c, d\rangle} = \left\langle min\left\{\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right\}, max\left\{\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right\}\right\rangle, \text{ kde } 0 \notin \langle \text{c}, \text{d}\rangle.$$

Uveďme si nyní konkrétní příklady jednotlivých operací.

Příklad 1.1 *Mějme zadefinované dva intervaly $a = \langle 3, 5\rangle$ a $b = \langle -2, 3\rangle$. Nyní si napíšeme výsledky jednotlivých operací.*

$$\langle 3, 5\rangle + \langle -2, 3\rangle = \langle 1, 8\rangle$$

$$\langle 3, 5\rangle - \langle -2, 3\rangle = \langle 0, 7\rangle$$

$$\langle 3, 5\rangle \cdot \langle -2, 3\rangle = \langle -10, 15\rangle$$

$$\frac{\langle 3, 5\rangle}{\langle -2, 3\rangle} = \text{nedefinováno.}$$

Ve fuzzy aritmetice místo intervalů obsahující reálné čísla počítáme s fuzzy čísly, která jsou vyjádřená pomocí α-řezů, kde $\alpha \in (0, 1\rangle$. Mějme nyní dvě fuzzy čísla A a B a označme si jejich α-řezy následovně: $A_\alpha = \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle$ a $B_\alpha = \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle$. S těmito fuzzy čísly můžeme provádět uvedené čtyři základní aritmetické operace, kde α-řezy těchto operací, pro všechna $\alpha \in (0, 1\rangle$, jsou dány následovně:

- sčítání

$$(A + B)_\alpha = \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle + \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle = \langle \underline{a}(\alpha) + \underline{b}(\alpha), \overline{a}(\alpha) + \overline{b}(\alpha)\rangle$$

- odečítání

$$(A - B)_\alpha = \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle - \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle = \langle \underline{a}(\alpha) - \overline{b}(\alpha), \overline{a}(\alpha) - \underline{b}(\alpha)\rangle$$

- násobení

$$(A \cdot B)_\alpha = \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle \cdot \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle = \langle l(\alpha), m(\alpha)\rangle$$

  kde

$$l(\alpha) = \min\{\underline{a}(\alpha)\underline{b}(\alpha), \underline{a}(\alpha)\overline{b}(\alpha), \overline{a}(\alpha)\underline{b}(\alpha), \overline{a}(\alpha)\overline{b}(\alpha)\}$$
$$m(\alpha) = \max\{\underline{a}(\alpha)\underline{b}(\alpha), \underline{a}(\alpha)\overline{b}(\alpha), \overline{a}(\alpha)\underline{b}(\alpha), \overline{a}(\alpha)\overline{b}(\alpha)\}$$

- dělení (uvažujeme podmínku, že $0 \notin \langle \underline{b}(0), \overline{b}(0)\rangle$)

$$(A/B)_\alpha = \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle / \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle = \langle n(\alpha), o(\alpha)\rangle$$

  kde

$$n(\alpha) = \min\{\underline{a}(\alpha)/\underline{b}(\alpha), \underline{a}(\alpha)/\overline{b}(\alpha), \overline{a}(\alpha)/\underline{b}(\alpha), \overline{a}(\alpha)/\overline{b}(\alpha)\}$$
$$o(\alpha) = \max\{\underline{a}(\alpha)/\underline{b}(\alpha), \underline{a}(\alpha)/\overline{b}(\alpha), \overline{a}(\alpha)/\underline{b}(\alpha), \overline{a}(\alpha)/\overline{b}(\alpha)\}.$$

Nyní se podíváme na jednotlivé operace v konkrétních příkladech.

Příklad 1.2 *Mějme zadané lichoběžníkové fuzzy číslo* $A = \langle 1, 2, 3, 4\rangle$ *a trojúhelníkové fuzzy číslo* $B = \langle 3, 5, 6\rangle$. *Obě fuzzy čísla si vyjádříme pomocí funkcí* $\underline{c}(\alpha), \overline{c}(\alpha)$, *tedy* $A = \{\langle 1 + \alpha, 4 - \alpha\rangle, \alpha \in \langle 0, 1\rangle\}$, $B = \{\langle 3 + 2\alpha, 6 - \alpha\rangle, \alpha \in \langle 0, 1\rangle\}$. *Nyní budeme počítat součet, rozdíl, součin a podíl těchto dvou fuzzy čísel podle zadaných definic.*

- *sčítání*

$$(A+B)_\alpha = \langle 1+\alpha, 4-\alpha\rangle + \langle 3+2\alpha, 6-\alpha\rangle = \langle 4+3\alpha, 10-2\alpha\rangle.$$

Pokud například budeme chtít zjistit $0,5$*-řez, tak za* α *pouze dosadíme číslo* $0,5$ *a dostaneme výsledek* $\langle 5,5;9\rangle$.

- *odečítání*

$$(A-B)_\alpha = \langle 1+\alpha, 4-\alpha\rangle - \langle 3+2\alpha, 6-\alpha\rangle = \langle -5+2\alpha, 1-3\alpha\rangle.$$

Výsledek pro $0,5$*-řez teď bude vypadat* $\langle -4;-0,5\rangle$.

- *násobení*

$$(A\cdot B)_\alpha = \langle 1+\alpha, 4-\alpha\rangle \cdot \langle 3+2\alpha, 6-\alpha\rangle = \langle 3+5\alpha+2\alpha^2, 24-10\alpha+\alpha^2\rangle,$$

kde:

$$3+5\alpha+2\alpha^2 = \min\{(1+\alpha)\cdot(3+2\alpha), (1+\alpha)\cdot(6-\alpha), (4-\alpha)\cdot(3+2\alpha), (4-\alpha)\cdot(6-\alpha)\},$$

$$24-10\alpha+\alpha^2 = \max\{(1+\alpha)\cdot(3+2\alpha), (1+\alpha)\cdot(6-\alpha), (4-\alpha)\cdot(3+2\alpha), (4-\alpha)\cdot(6-\alpha)\}.$$

Nyní si na ukázku rozepíšeme opět výsledek pro $0,5$*-řez:*

$$(A\cdot B)_{0,5} = \langle 1,5;3,5\rangle \cdot \langle 4;5,5\rangle = \langle 6;19,25\rangle.$$

- *dělení*

$$(A/B)_\alpha = \langle 1+\alpha, 4-\alpha\rangle / \langle 3+2\alpha, 6-\alpha\rangle = \langle (1+\alpha)/(6-\alpha), (4-\alpha)/(3+2\alpha)\rangle,$$

kde:

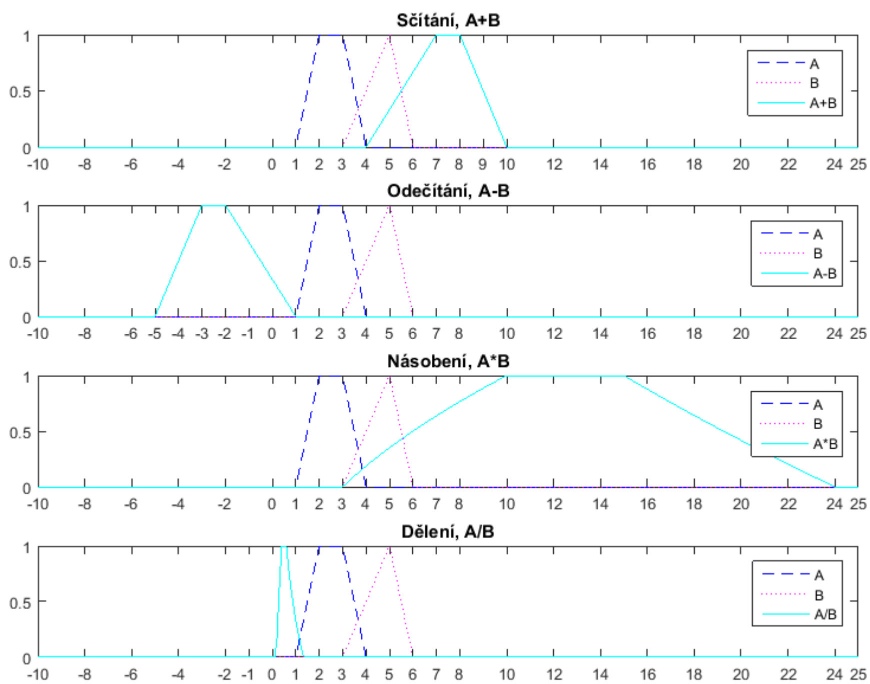
$$(1+\alpha)/(6-\alpha) = \min\{(1+\alpha)/(3+2\alpha), (1+\alpha)/(6-\alpha), (4-\alpha)/(3+2\alpha), (4-\alpha)/(6-\alpha)\},$$

$$(4-\alpha)/(3+2\alpha) = \max\{(1+\alpha)/(3+2\alpha), (1+\alpha)/(6-\alpha), (4-\alpha)/(3+2\alpha), (4-\alpha)/(6-\alpha)\}.$$

Naposledy si rozepíšeme výsledek pro $0,5$*-řez:*

$$(A/B)_{0,5} = \langle 1,5; 3,5\rangle / \langle 4; 5,5\rangle = \langle 0,27; 0,875\rangle.$$

Na závěr si všechny výsledky vykreslíme do jednoho grafu. Grafickou reprezentaci výsledků najdeme na obrázku 1.4.



Obrázek 1.4: Grafická reprezentace výsledků.

Poznámka 1.5 *Podle výsledku jednotlivých operací a obrázku 1.4 si můžeme všimnout, že operace sčítání a odečítání zachovávají linearitu funkcí popisujících krajní hodnoty jednotlivých* α*-řezů, zatímco operace násobení a dělení nikoliv. Výsledkem ovšem bude vždy fuzzy číslo.*

1.4. Podmíněná fuzzy aritmetika

Standardní fuzzy aritmetika se používá, když není mezi fuzzy čísly žádná závislost. My se ovšem někdy setkáváme se situacemi, kdy bude závislost mezi fuzzy čísly přítomna. K řešení tohoto problému se využije podmíněná fuzzy aritmetika [6], kdy si určíme dodatečnou podmínku, která nám popíše závislost mezi fuzzy čísly.

Uvažujme, že máme n proměnných. Mějme zadanou ostrou relaci D definovanou na prostoru R^n, kde $D \subseteq R^n$. Relace D nám vyjadřuje všechny přípustné kombinace hodnot. Symbol * nám bude značit jednu ze čtyř základních aritmetických operací, a to konkrétně sčítání, odečítání, násobení a dělení. Zápis pro jednu z těchto čtyř operací dvou fuzzy čísel A a B zadaných podle α-řezů, bez zavedené podmínky by vypadal následovně:

$$(A * B)_\alpha = \{a * b | a \in A_\alpha, b \in B_\alpha\}.$$

Pokud budeme uvažovat naši zadanou relaci $D \subseteq R^n$, tak zde už se jedná o *podmíněnou fuzzy aritmetiku* a obecný předpis vypadá následovně:

$$(A *_D B)_\alpha = \{a * b | a \in A_\alpha, b \in B_\alpha, (a, b) \in D\}.$$

Podmínka je vždy dána externě a vychází z povahy problému.

Poznámka 1.6 *Jelikož se zabýváme čtyřmi základními aritmetickými operacemi, uvažujeme tedy binární relaci D na $R \times R$.*

Teď mějme zadanou relaci $D \subseteq R^2$ a dvě fuzzy čísla, která si označíme takto: $A = \{\langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle, \alpha \in \langle 0, 1\rangle\}$ a $B = \{\langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle, \alpha \in \langle 0, 1\rangle\}$, potom výsledky aritmetických operací s těmito fuzzy čísly vypadají následovně:

$$A *_D B = \{\langle \underline{a *_D b}(\alpha), \overline{a *_D b}(\alpha), \alpha \in \langle 0, 1\rangle\},$$

kde

$$\underline{a *_D b}(\alpha) = \min\{a * b | a \in \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle, b \in \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle, (a, b) \in D\},$$

$$\overline{a *_D b}(\alpha) = \max\{a * b | a \in \langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle, b \in \langle \underline{b}(\alpha), \overline{b}(\alpha)\rangle, (a, b) \in D\}.$$

U podmíněné fuzzy aritmetiky ovšem nemusí nutně platit, že výsledkem musí být fuzzy číslo. Aby výsledek bylo právě fuzzy číslo, tak jsou postačující následující dvě podmínky:

1. Pro alespoň jednu dvojici $(x, y) \in D$ musí platit $A(x) = B(y) = 1$.
2. D je uzavřená konvexní podmnožina R^2.

Uveďme si nyní konkrétní příklad na podmíněnou fuzzy aritmetiku.

Příklad 1.3 *Mějme dvě trojúhelníková fuzzy čísla A, B zadefinovaná následovně:*

$$A = \{\langle 1 + 2\alpha, 5 - 2\alpha\rangle, \alpha \in \langle 0, 1\rangle\},$$

$$B = \{\langle 2 + 2\alpha, 6 - 2\alpha\rangle, \alpha \in \langle 0, 1\rangle\}.$$

Podívejme se podrobněji na operací odečítání. V rámci standardní fuzzy aritmetiky by se výsledek spočítal následovně:

$$(A - B)_\alpha = \langle 1 + 2\alpha, 5 - 2\alpha\rangle - \langle 2 + 2\alpha, 6 - 2\alpha\rangle = \langle -5 + 4\alpha, 3 - 4\alpha\rangle.$$

Zadejme si nyní konkrétní relaci $D = \{(a - b) \in R^2 | a \leq b\}$. Nyní výsledek s ohledem na zadanou podmínku bude vypadat následovně:

$$(A -_D B)_\alpha = \left\langle \underline{a -_D b}(\alpha), \overline{a -_D b}(\alpha) \right\rangle,$$

kde

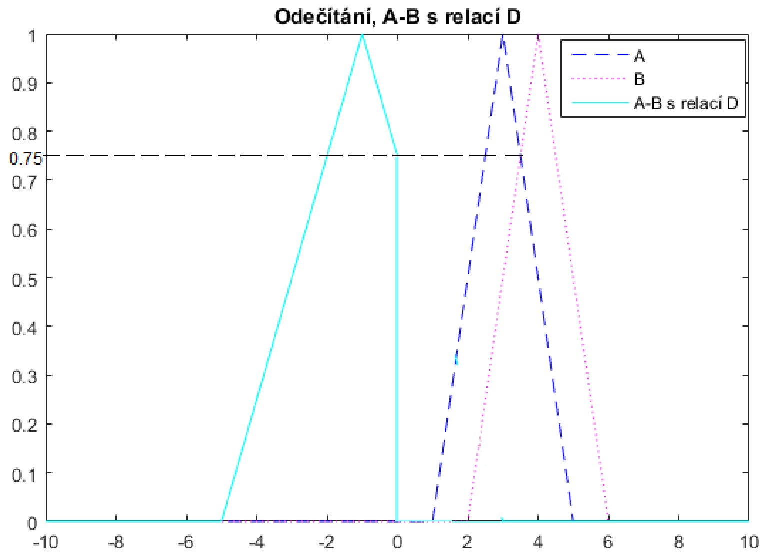
$$\underline{a -_D b}(\alpha) = -5 + 4\alpha \text{ } pro \text{ } \alpha \in \langle 0, 1\rangle$$

$$\overline{a -_D b}(\alpha) = \begin{cases} 3 - 4\alpha & pro \ \alpha \in \langle 0,75; 1\rangle \\ 0 & pro \ \alpha \in \langle 0; 0,75). \end{cases}$$

Výsledek je znázorněn na obrázku 1.5.

1.5. Defuzzifikace fuzzy čísel

Někdy je zapotřebí reprezentovat fuzzy číslo jedním reálným číslem. Tato reprezentace by měla, pokud to jde, co nejlépe charakterizovat dané fuzzy číslo. Tomuto postupu se říká defuzzifikace. Existuje však více metod, jak lze fuzzy číslo defuzzifikovat. To jakou metodu zvolit je individuální a záleží pouze na rozhodovateli, aby podle situace zvolil tu nejvhodnější. Defuzzikaci budeme velmi potřebovat v následující kapitole. Pojďme se však teď podívat na konkrétní vybrané metody defuzifikace.



Obrázek 1.5: Grafická reprezentace odečítání dvou fuzzy čísel s relací D.

Těžiště

Tato metoda je jedna z nejznámějších a také nejpoužívanějších. Definice těžiště je následující:

Definice 1.11 *Nechť $A \in F_N(\langle a, b \rangle)$, které není reprezentací čísla reálného. Potom těžiště fuzzy čísla A je definováno vztahem*

$$t_A = \frac{\int_a^b A(x) x dx}{\int_a^b A(x) dx}.$$

Poznámka 1.7 *Pokud se jedná o reálné číslo, tak poté je hodnota těžiště rovna přímo danému reálnému číslu.*

Poznámka 1.8 *Těžiště fuzzy čísla je analogické jako střední hodnota spojité náhodné veličiny, jejíž rozdělení pravděpodobností je popsáno funkcí hustoty.*

Střed maxima (MAX - Mean of Maxima)

V této metodě se v případě fuzzy čísel bere hodnota středu z prvků, ve kterých má fuzzy množina maximální stupeň příslušnosti. Obecně platí následující definice.

Definice 1.12 *Nechť* $A \in F_N(\langle a, b\rangle) \subset \mathbb{R}$. *Potom středem maxima fuzzy čísla* A *nazveme reálné číslo* MAX_A*, které je definováno následovně*

$$MAX_A = \frac{\int_M x dx}{\int_M dx},$$

kde $M = \{x | A(x) = hgt(A)\}$.

Střed průměrného intervalu (MOM - Middle Point of the Mean Interval)

Tato metoda je založena na následující definici.

Definice 1.13 *Nechť je dáno fuzzy číslo* $A = \{\langle \underline{a}(\alpha), \overline{a}(\alpha)\rangle, \alpha \in \langle 0, 1\rangle\}$. *Potom střed průměrného intervalu fuzzy čísla* A *nazveme reálné číslo* MOM_A*, které je definováno následovně*

$$MOM_A = \int_0^1 \frac{\underline{a}(\alpha) + \overline{a}(\alpha)}{2} d\alpha.$$

Nyní už máme zadefinované podstatné pojmy z teorie fuzzy množin a můžeme přejít na hlavní kapitolu této práce.

Kapitola 2

Testování statistických hypotéz na základě fuzzy dat

V této kapitole si nejdříve uvedeme základní pojmy z teorie testování statistických hypotéz a posléze se podíváme na obecný postup, který použijeme při testování statistických hypotéz. Dále se zaměříme na to, jak by se postupovalo při testování v případě, že vstupní data budou fuzzy čísla. Nakonec se podíváme na vybrané testy a uvedeme si konkrétní příklady pro lepší pochopení teorie. Tato kapitola byla zpracována podle [1], [3], [5], [12], [16], [18], [20].

2.1. Testování statistických hypotéz

Testování statistických hypotéz nám umožňuje posoudit, zda data, která jsme získali nějakým experimentem, vyhovují určitému předpokladu, který jsme učinili ještě před provedením testování.

Statistickou hypotézou tedy rozumíme určité tvrzení o rozdělení náhodných veličin. V případě, že se tato tvrzení týkají hodnot parametrů rozdělení náhodné veličiny, tak mluvíme o takzvaných *parametrických hypotézách*. Jinak se jedná o *neparametrické hypotézy*.

Při testování statistických hypotéz vždy porovnáváme dvě hypotézy. Nulovou hypotézu, kterou značíme H_0 a alternativní hypotézu, kterou značíme H_A. *Test hypotézy* je tedy postup, kterým na základě výsledků experimentu dospějeme k rozhodnutí o nulové hypotéze.

Rozhodnutí o nulové hypotéze H_0 může nastat dvojí, a to:

a) H_0 se zamítá ve prospěch alternativy,

b) H_0 nelze zamítnout.

Je očividné, že se při testování můžeme dopustit chyb, které dělíme na chyby prvního nebo druhého druhu. V následující tabulce je zobrazeno, kdy se které chyby dopustíme.

	Nezamítáme H_0	**Zamítáme** H_0
Platí H_0	Správné rozhodnutí	Chyba 1. druhu
Platí H_A	Chyba 2. druhu	Správné rozhodnutí

Tabulka 2.1: Tabulka rozhodnutí o hypotézách

K otestování nulové hypotézy H_0 proti alternativní hypotéze H_A použijeme statistiku T, kterou označujeme jako *testovací kritérium. Testovací kritérium* je funkce náhodného výběru, která má vztah k nulové hypotéze, a jejíž rozdělení pravděpodobnosti za předpokladu platnosti nulové hypotézy známe.

Kritickým oborem $W \subset R^1$ označíme množinu všech hodnot testového kritéria, při kterých budeme H_0 zamítat. Výběrem vhodného kritického oboru se snažíme omezit pravděpodobnost chyby 1. druhu, a to tak, že pevně zvolíme pravděpodobnost α, $0<\alpha<1$, které říkáme *hladina testu.* Platnou hypotézu tedy zamítáme nejvýše s pravděpodobností α, kde α se nejčastěji volí $\alpha = 0,05$ nebo $\alpha = 0,01$.

Jestliže hodnota testovací statistiky T náleží do kritického oboru W, tak naši hypotézu H_0 zamítáme ve prospěch alternativy. Pokud hodnota testovací statistiky nenáleží do příslušného kritického oboru, tak nulovou hypotézu nelze zamítnout. Hypotézu nemůžeme prokázat, ale pouze vyvrátit.

Poznámka 2.1 *Je potřeba si dát pozor na to, že symbolem α se všeobecně označuje hladina testu i úroveň řezu fuzzy čísla. V této práci bude vždy patrné, o kterém pojmu se zrovna bavíme.*

2.2. Obecný postup testování statistických hypotéz

Nyní když už jsme se seznámili se základními pojmy, tak si popíšeme, jak se postupuje při testování.

1. Stanovíme si nulovou a alternativní hypotézu.
2. Zvolíme hladinu testu α neboli pravděpodobnost chyby 1. druhu.
3. Zvolíme vhodnou testovací statistiku T a na základě získaných dat vypočítáme hodnotu realizace testovací statistiky.
4. Stanovíme si kritický obor W.
5. Provedeme závěr testování. Jestliže hodnota testovací statistiky T náleží do kritického oboru W, tak naši nulovou hypotézu H_0 zamítáme na hladině testu α. V opačném případě prohlásíme, že nulovou hypotézu H_0 nelze zamítnout na hladině testu α.

Při testování můžeme dojít ke třem následujícím situacím:

1. Učinili jsme správné rozhodnutí, což znamená, že jsme nezamítli platnou hypotézu nebo jsme zamítli neplatnou hypotézu.
2. Zamítli jsme platnou hypotézu a tím jsme se dopustili chyby 1. druhu.
3. Nezamítli jsme neplatnou hypotézu a tím jsme se dopustili chyby 2. druhu.

Jestliže jsme zamítli hypotézu, můžeme následně tvrdit, že data svědčí o tom, že hypotéza neplatí. Pokud jsme hypotézu nezamítli, pak je to buď tím, že hypotéza platí, anebo tím, že hypotéza sice neplatí, ale získaná data nám nedávají dostatečné důvody k jejímu zamítnutí [3].

2.3. Testování statistických hypotéz na základě fuzzy dat

Jak už jsme si v úvodu této práce nastínili, ne vždy máme k dispozici vstupní data, která by byla reálná. Většina odpovědí respondentů na číselné údaje v různých průzkumech bývá neurčitých. Také se může stát, že měřící stroj nám poskytuje pouze přibližné, zaokrouhlené hodnoty nebo začíná dosluhovat a měřit nepřesně. V těchto situacích je tedy vhodné tyto vstupní data modelovat pomocí fuzzy čísel. Tato problematika je naznačena v publikaci [16], ze které jsem se inspirovali.

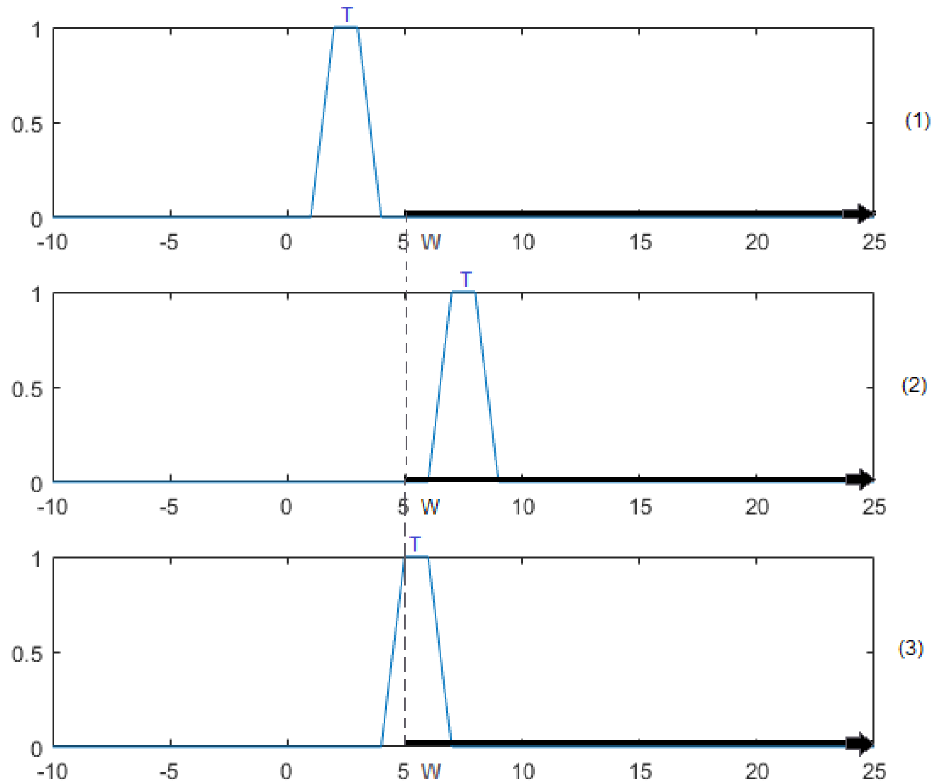
Pojďme se nyní podívat na to, jak se nám změní obecný postup při testování statistických hypotéz. První dva kroky (stanovení nulové, alternativní hypotézy a zvolení hladiny testu α) zůstávají stejné. V třetím kroku si vybíráme testovací statistiku T a spočítáme hodnotu realizace T. Zde už nastane změna, jelikož v reálném případě je hodnota testovací statistiky reálné číslo. Nyní máme ovšem vstupní data fuzzy čísla, tudíž při počítání s fuzzy čísly dostaneme výsledek také ve formě fuzzy čísla. Zároveň nám vznikne první problém, a to využití správné fuzzifikace T. Testovací statistiky mohou mít různé podoby a může se nám stát, že si nevystačíme pouze se standardní fuzzy aritmetikou, jelikož bude potřeba uvažovat závislost mezi jednotlivými fuzzy čísly. Tento problém se dá vyřešit za použití podmíněné fuzzy aritmetiky, s kterou jsme se již setkali v kapitole 1.4. Její využití nám ovšem povede k poměrně složitým optimalizačním úlohám, které se pokusíme vyřešit v programu MATLAB.

Stanovení kritického oboru W zůstává stejné, ale druhý problém nastane při finálním kroku, a to rozhodnutí o zamítnutí, či nezamítání naší hypotézy. Můžou totiž nastat tři následující situace:

1. $Supp\ T \cap W = \emptyset$,
2. $Supp\ T \subset W$,
3. $Supp\ T \not\subset W$ *a zároveň* $Supp\ T \cap W \neq \emptyset$.

Tyto tři situace situace jsou graficky zobrazeny na obrázku 2.1.

Pokud nastane první případ, tak nelze naši hypotézu zamítnout, jelikož celý nosič fuzzy čísla testovací statistiky leží mimo kritický obor. V druhém případě naši hypotézu zamítáme, protože celý nosič fuzzy čísla leží v kritickému oboru. Jak ovšem rozhodneme v situaci, kdy část nosiče leží v kritickému oboru a zby-



Obrázek 2.1: Grafická reprezentace tří situací při rozhodnutí o hypotéze.

tek nikoliv? V této práci si zkusíme vyřešit tento problém tak, že pak výsledné fuzzy čísla defuzzifikujeme pomocí různých metod, s kterými jsme se již seznámili v kapitole 1.5 a podle jejich výsledků provedeme rozhodnutí.

Poznámka 2.2 *V [16] je naznačeno rozhodnutí o hypotéze řešit pomocí takzvané fuzzy p-value. P-value je číslo, které udává nejmenší hladinu, při které bychom ještě hypotézu zamítli. Fuzzy p-value je fuzzy číslo založené na myšlence klasické p-value. Pro zájemce odkazuji na [16].*

2.4. Vybrané statistické testy

Budeme se zabývat parametrickými testy. Jak už bylo zmíněno, pokud se hypotézy týkají hodnot parametrů rozdělení náhodné veličiny, tak mluvíme o parametrických testech. Nyní se podíváme na vybrané testy hypotéz o parametrech jednorozměrného normálního rozdělení. Každý z testů si nejdříve popíšeme pro případ s reálnými daty a následně budeme zkoumat, co se změní, jestliže začneme uvažovat fuzzy data. U každého testu si uvedeme konkrétní příklad s komplexním řešením.

Před tím, než si představíme jednotlivé testy, je nezbytné si zadefinovat pojem náhodný výběr, který prakticky znamená náhodné vybrání vzorku o rozsahu n z celkové populace.

Definice 2.1 *Náhodný vektor $\boldsymbol{X} = (X_1, \ldots, X_n)^{'}$, jehož složky jsou nezávislé náhodné veličiny, které mají stejné rozdělení pravděpodobností jako zkoumaná náhodná veličina X, se nazývá náhodný výběr rozsahu n příslušný statistickému znaku X.*

Nakonec je ještě si potřeba zadefinovat několik základních výběrových funkcí, které se budou v předpisech pro testovací statistiky objevovat.

Definice 2.2 *Nechť máme náhodný výběr rozsahu n příslušný statistickému znaku X.*

Výběrové funkce

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j,$$

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2,$$

$$S_n = \sqrt{S_n^2},$$

postupně nazýváme výběrový průměr, výběrový rozptyl a výběrová směrodatná odchylka.

2.4.1. Jednovýběrový t-test

Jednovýběrový t-test se používá v experimentálních situacích, kdy zkoumaná náhodná veličina X má normální rozdělení s parametry μ a σ^2. Uvažujeme, že parametr μ by mohl být roven (menší nebo roven, větší nebo roven) hypotetické hodnotě μ_0. Parametr σ^2 neznáme. Můžeme tedy testovat:

1. nulovou hypotézu $H_0 : \mu = \mu_0$ proti tzv. oboustranné alternativě $H_A : \mu \neq \mu_0$,
2. nulovou hypotézu $H_0 : \mu \leq \mu_0$ proti tzv. jednostranné alternativě $H_0 : \mu > \mu_0$,
3. nulovou hypotézu $H_0 : \mu \geq \mu_0$ proti tzv. jednostranné alternativě $H_0 : \mu < \mu_0$.

Testovacím kritériem pro jednovýběrový t-test je výběrová funkce

$$T = \frac{\overline{X}_n - \mu_0}{S_n}\sqrt{n}.$$

Za platnosti nulové hypotézy má tato funkce t-rozdělení o $n-1$ stupních volnosti. Dále platí, že $\overline{X}_n$ je bodovým odhadem $\mu = E(X)$.

Nyní se podívejme, jak bude vypadat kritický obor W, do kterého budou patřit takové hodnoty testového kritéria, které svědčí ve prospěch alternativy H_A.

1. V případě oboustranné alternativy budou kritický obor tvořit malé i velké hodnoty výběrové funkce T. Proto

$$W = (-\infty, -t_{n-1,1-\frac{\alpha}{2}}\rangle \cup \langle t_{n-1,1-\frac{\alpha}{2}}, \infty).$$

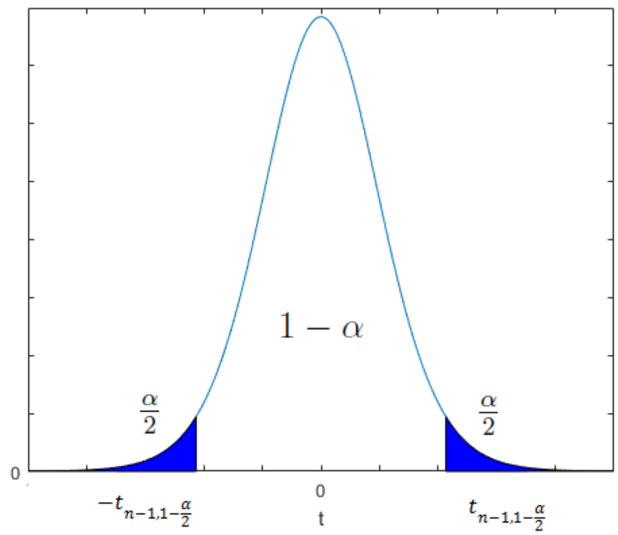
2. Ve prospěch alternativy svědčí velké hodnoty testového kritéria, proto

$$W = \langle t_{n-1,1-\alpha}, \infty).$$

3. Ve prospěch alternativy svědčí malé hodnoty testového kritéria, proto

$$W = (-\infty, -t_{n-1,1-\alpha}\rangle.$$

Poznámka 2.3 *Při vytváření kritického oboru vycházíme z charakteru grafu hustoty t-rozdělení.*



Obrázek 2.2: Hustota t-rozdělení o n-1 stupních volnosti s kritickým oborem v případě oboustranné alternativy.

Uvažujme nyní, že vstupní data budou fuzzy čísla. Pro potřeby výpočtu hodnoty realizace testovací statistiky T_F nyní musíme z těchto fuzzy čísel spočítat takzvaný výběrový fuzzy průměr a výběrovou fuzzy směrodatnou odchylku.

Definice 2.3 *Nechť $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $i = 1,\dots,n$, jsou fuzzy čísla. Pak výběrovým fuzzy průměrem fuzzy čísel $X_1,\dots,X_n$ nazveme fuzzy číslo $\overline{X}_{n_F} = \{\langle \underline{\overline{x}_{n_F}}(\alpha), \overline{\overline{x}_{n_F}}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, kde pro každé $\alpha \in \langle 0,1\rangle$ platí:*

$$\underline{\overline{x}_{n_F}}(\alpha) = \frac{\sum_{i=1}^{n} \underline{x}_i(\alpha)}{n},$$

$$\overline{\overline{x}_{n_F}}(\alpha) = \frac{\sum_{i=1}^{n} \overline{x}_i(\alpha)}{n}.$$

Definice 2.4 *Nechť $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $i = 1,\dots,n$, jsou fuzzy čísla. Pak výběrovou fuzzy směrodatnou odchylkou fuzzy čísel $X_1,\dots,X_n$ nazveme fuzzy číslo $S_{n_F} = \{\langle \underline{s_{n_F}}(\alpha), \overline{s_{n_F}}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, kde pro každé $\alpha \in \langle 0,1\rangle$ platí:*

$$\underline{s_{n_F}}(\alpha) = \sqrt{\min\{s_n^2(x_1,\dots,x_n) | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, i \in \{1,\dots,n\}\}},$$

$$\overline{s_{n_F}}(\alpha) = \sqrt{\max\{s_n^2(x_1,\dots,x_n) | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, i \in \{1,\dots,n\}\}}.$$

Testovací statistika pro fuzzy vstupy bude vypadat tedy následovně

$$T_F = \frac{\overline{X}_{n_F} - \mu_0}{S_{n_F}}\sqrt{n}.$$

Dejme si ovšem velký pozor na to, jakou fuzzy aritmetiku použijeme k výpočtu hodnoty realizace T_F. Nemůžeme použít standardní fuzzy aritmetiku z kapitoly 1.3, jelikož výběrová fuzzy směrodatná odchylka ve jmenovateli je závislá na hodnotách výběrového fuzzy průměru v čitateli. Napišme si nyní obecný předpis, jak se T_F spočítá.

Definice 2.5 *Nechť $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $i = 1,\ldots,n$, jsou fuzzy čísla. Pak fuzzy testovací statistikou pro jednovýběrový t-test fuzzy čísel $X_1,\ldots,X_n$ nazveme fuzzy číslo $T_F = \{\langle \underline{t_F}(\alpha), \overline{t_F}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, kde pro každé $\alpha \in \langle 0,1\rangle$ platí:*

$$\underline{t_F}(\alpha) = \min\left\{ \frac{\frac{1}{n}\sum_1^n x_i - \mu_0}{\sqrt{\frac{1}{n-1}\sum_1^n (x_i - \frac{1}{n}\sum_1^n x_i)^2}}\sqrt{n} | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, i \in \{1,\ldots,n\} \right\},$$

$$\overline{t_F}(\alpha) = \max\left\{ \frac{\frac{1}{n}\sum_1^n x_i - \mu_0}{\sqrt{\frac{1}{n-1}\sum_1^n (x_i - \frac{1}{n}\sum_1^n x_i)^2}}\sqrt{n} | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, i \in \{1,\ldots,n\} \right\}.$$

Hypotetickou hodnotu μ_0 budeme uvažovat pouze jako reálné číslo. Hodnota realizace testovací statistiky bude fuzzy číslo, a tedy při finálním rozhodování o zamítnutí/nezamítnutí hypotézy mohou nastat tři situace, které jsme si už popsali na začátku kapitoly. Kdybychom právě pro výpočet realizace použili standardní přístup k fuzzy aritmetice, výsledný nosič fuzzy čísla by byl více širší. Vznikla by pak větší šance na provedení chyby prvního nebo druhého druhu.

Poznámka 2.4 *Všechny příklady, které si uvedeme, mají pouze ilustrativní charakter a slouží k pochopení problematiky. V praxi by nemusel například rozsah výběru o velikosti deset být dostačující. Na druhou stranu, někdy prostě víc respondentů nemusí být. Také je důležité si uvědomit, že budeme automaticky předpokládat, že náhodné veličiny mají normální rozdělení. Opět v praxi by bylo vhodné* *nejdříve normalitu otestovat. V případě fuzzy dat by to byla další problematika ke zkoumání. Porušení normality ovšem obvykle příliš nemění výsledné rozdělení [3].*

Příklad 2.1 *V rámci průzkumu v roce 2016 jsme se zeptali náhodně deseti studentů navazujícího studia na následující otázku: Kolik průměrně jste utratili měsíčně v roce 2015 za potraviny a nealkoholické nápoje? Na tuto otázku nám přesně odpověděli pouze dva studenti, s tím, že jeden si to sám nedávno spočítal, a tedy průměrně měsíčně utratil 1953 Kč. Druhý student nám zase odpověděl, že utratí přesně každý měsíc 2500Kč, což je tedy i jeho roční průměr. Ostatních osmi studentů jsme se tudíž zeptali, zda-li nám mohou říci:*

1. *minimální částku, kterou si myslí, že průměrně utratili měsíčně za potraviny a nealkoholické nápoje za rok 2015,*
2. *částku, která podle nich představuje nejmožnější odhad toho, co utratili měsíčně za potraviny a nealkoholické nápoje za rok 2015,*
3. *maximální částku, kterou si myslí, že průměrně utratili měsíčně za potraviny a nealkoholické nápoje za rok 2015.*

Odpovědi všech deseti studentů jsme přehledně zaznamenali do tabulky 2.2.

Respondent	Minimální hod.	Nejočekávanější hod.	Maximální hod.	Přesná hod.
1	1900	2000	2100	n/a
2	1500	2500-3000	3500	n/a
3	n/a	n/a	n/a	1953
4	2000	2200	2240	n/a
5	3900	4000-5000	5100	n/a
6	1000	1200	1500	n/a
7	2000	2100	2200	n/a
8	n/a	n/a	n/a	2500
9	1800	2000	2200	n/a
10	1400	1500	1600	n/a

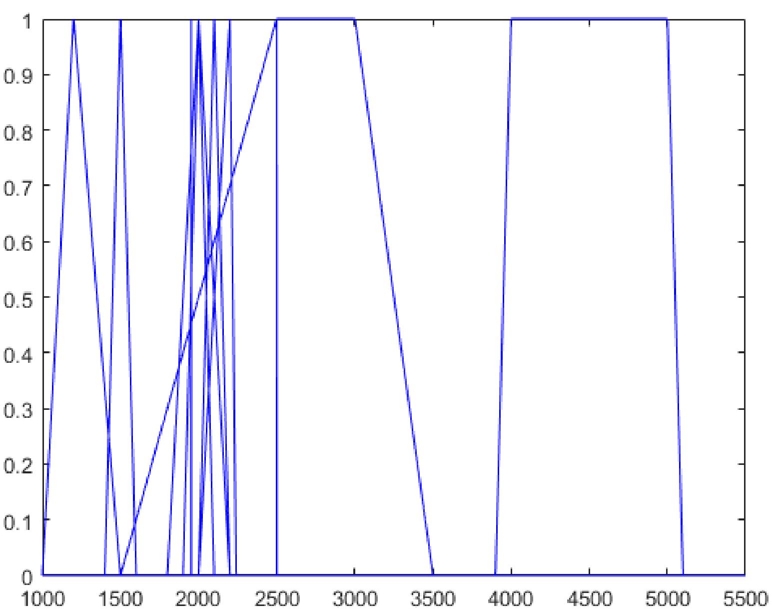
Tabulka 2.2: Odpovědi respondentů navazujícího studia

Podle českého statistického úřadu za rok 2015 v průměru utratila osoba za měsíc za potraviny a nealkoholické nápoje 2044 Kč [4]. Liší se výdaje vybraných studentů od průměrné hodnoty uváděné ČSÚ?

Jak je vidět, většina získaných vstupních dat jsou vágní (neurčitá), proto je budeme považovat za fuzzy čísla. Jak už bylo zmíněno, i reálné číslo (odpovědi 3 a 8) se dá reprezentovat pomocí fuzzy čísla. Odpovědi budeme tedy modelovat pomocí trojúhelníkových a lichoběžníkových fuzzy čísel. Pojďme nyní zkusit odpovědět na otázku položenou v zadání tohoto příkladu. Budeme předpokládat, že náhodná veličina X, která nabývá zodpovězených hodnot, bude mít normální rozdělení. Využijeme jednovýběrový t-test, kde budeme testovat

$$H_0 : \mu = 2044 \qquad \textit{proti} \qquad H_A : \mu \neq 2044.$$

Hladinu testu α si příhodně zvolíme $0,1$, abychom si ukázali rozdíl mezi využitím správného a špatného přístupu při počítání realizaci testovací statistiky. Rozsah výběru máme $n = 10$. Vykresleme si nyní vstupní data pomocí fuzzy čísel (obrázek 2.3).

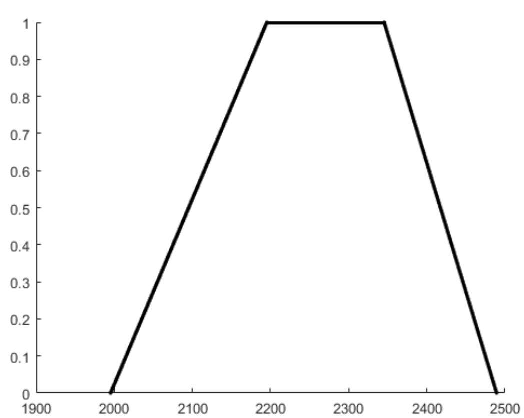


Obrázek 2.3: Vstupní fuzzy data.

Pro výpočet výběrového fuzzy průměru $\overline{X}_{n_F}$ nám poslouží následující skript, který jsem vytvořil v programu MATLAB. Tento i všechny následné skripty můžeme také naleznout na přiloženém CD.

```
%načtení vstupních dat maticově
y=[1900 2000 2000 2100; 1500 2500 3000 3500; 1953 1953 1953 1953; ...
    2000 2200 2200 2240; 3900 4000 5000 5100; 1000 1200 1200 1500;...
    2000 2100 2100 2200; 2500 2500 2500 2500; 1800 2000 2000 2200;...
    1400 1500 1500 1600];

x = -1000:0.1:5500;
n=size(y);
pocet_alfa_rezu1 =1000;

mean1=zeros(pocet_alfa_rezu1+1,1);
mean2=zeros(pocet_alfa_rezu1+1,1);

for j=0:pocet_alfa_rezu1
    alfa=j/pocet_alfa_rezu1;
    for i=1:n(1,1)
         yd1(i,1)=y(i,1)+alfa*(y(i,2)-y(i,1));
         yh1(i,1)=y(i,4)-alfa*(y(i,4)-y(i,3));
    end
mean1(j+1,1)=mean(yd1);
mean2(j+1,1)=mean(yh1);
    hold on
    plot(mean1(j+1,1),alfa,'.k',mean2(j+1,1),alfa,'.k')
    hold off
end
```

Výsledné fuzzy číslo výběrového fuzzy průměru je na obrázku 2.4.

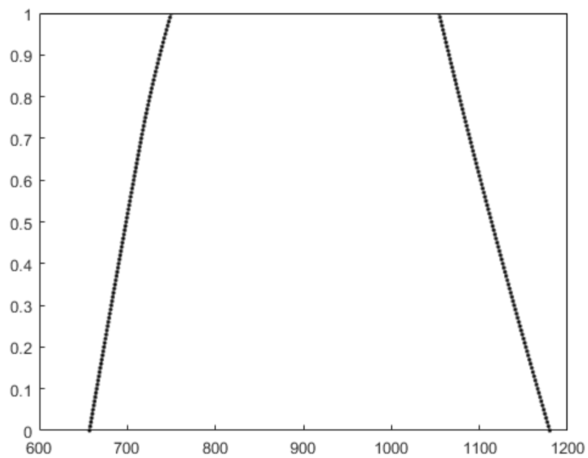


Obrázek 2.4: Výběrový fuzzy průměr.

Skript pro výpočet výběrové fuzzy směrodatné odchylky už je lehce složitější. Využívá se zde optimalizačních metod.

```
%načtení vstupních dat maticově
y=[1900 2000 2000 2100; 1500 2500 3000 3500; 1953 1953 1953 1953; ...
    2000 2200 2200 2240; 3900 4000 5000 5100; 1000 1200 1200 1500;...
    2000 2100 2100 2200; 2500 2500 2500 2500; 1800 2000 2000 2200;...
    1400 1500 1500 1600];
%zadáni optimalizovaných funkcí
FunkceRozptylu = @(x)var(x);
FunkceRozptyluKratMinus1 = @(x)-var(x);
%počáteční bod optimalizace, je dobré vyzkoušet různé možnosti

opts = optimset('Algorithm','active-set','Display','notify');
gs = GlobalSearch;

n=size(y);
pocet_alfa_rezu=100;
smodch=zeros(pocet_alfa_rezu+1,2);
yd=zeros(n(1,1),1);
yh=zeros(n(1,1),1);


for j=0:pocet_alfa_rezu
    alfa=j/pocet_alfa_rezu;
    for i=1:n(1,1)
         yd(i,1)=y(i,1)+alfa*(y(i,2)-y(i,1));
         yh(i,1)=y(i,4)-alfa*(y(i,4)-y(i,3));
    end
    y0 = (yd(:,1)+yh(:,1))/2; %%% počáteční bod
    problemMin = createOptimProblem('fmincon','objective',...
    FunkceRozptylu,'x0',y0,'lb',yd(:,1),'ub',yh(:,1),'options',opts);
    [~,varmin] = run(gs,problemMin);
    smodch(j+1,1)=sqrt(varmin);

    problemMax = createOptimProblem('fmincon','objective',...
    FunkceRozptyluKratMinus1,'x0',y0,'lb',yd(:,1),'ub',yh(:,1),'options',opts);
    [~,varmax_minus] = run(gs,problemMax);
    %varmax=-varmax_minus;
    smodch(j+1,2)=sqrt(-varmax_minus);

    hold on
    plot(smodch(j+1,1),alfa,'.k',smodch(j+1,2),alfa,'.k')
    hold off
end
```

Výsledné fuzzy číslo výběrové směrodatné fuzzy odchylky je na obrázku 2.5. Pro ukázku jsme si toto fuzzy číslo vykreslily pomocí sto α-řezů. Kvůli početní a časové náročnosti se dále v práci většinou omezíme na výpočet dvaceti α-řezů a následně budeme tyto fuzzy čísla modelovat jako po částech lineární fuzzy čísla.



Obrázek 2.5: Výběrová fuzzy směrodatná odchylka.

Teď už známe vše potřebné k dosazení do testovací statistiky T_F. $\overline{Supp}\ \overline{X}_{n_F} = \langle 1995,3; 2489,3 \rangle$. $\overline{Supp}\ \overline{S}_{n_F} = \langle 650,051; 1179,8 \rangle$. *Pokud bychom použili špatný přístup a výsledný uzávěr nosiče fuzzy testovací statistiky* T_F *počítali standardní fuzzy aritmetikou, tak bychom dostali výsledek:*

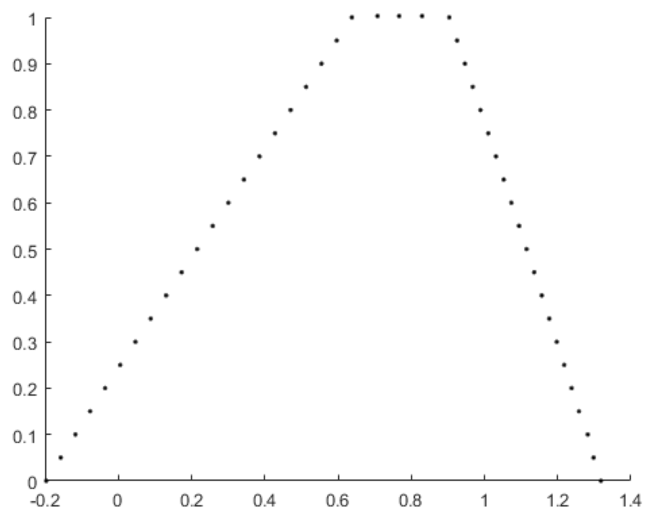
$$\overline{Supp}\ T_F = \frac{\overline{Supp}\ \overline{X}_{n_F} - 2044}{\overline{Supp}\ \overline{S}_{n_F}} \sqrt{10} = \langle -0,2369; 2,166 \rangle.$$

Musíme využít výpočet s uzávěry nosičů, jelikož standardní fuzzy aritmetika je definována pro uzavřené intervaly. Správný přístup je ovšem výpočet pomocí podmíněné fuzzy aritmetiky. K tomu použijeme následující skript, který je založen na vzorcích výpočtu T_F*, které jsme si zadefinovali výše. Posléze porovnáme výsledky obou přístupů.*

```
%načtení vstupních dat maticově
y=[1900 2000 2000 2100; 1500 2500 3000 3500; 1953 1953 1953 1953; ...
    2000 2200 2200 2240; 3900 4000 5000 5100; 1000 1200 1200 1500;...
    2000 2100 2100 2200; 2500 2500 2500 2500; 1800 2000 2000 2200;...
    1400 1500 1500 1600];
mu = 2044;
n = size(y);

%zadání optimalizovaných funkcí
FunkceT = @(x)(mean(x)-mu)/std(x)*sqrt(n(1,1));
```

```
FunkceTKratMinus1 = @(x)-(mean(x)-mu)/std(x)*sqrt(n(1,1));
%počáteční bod optimalizace, je dobré vyzkoušet různé možnosti

opts = optimset('Algorithm','active-set','Display','notify');
gs = GlobalSearch;

pocet_alfa_rezu=20;
T=zeros(pocet_alfa_rezu+1,2);
yd=zeros(n(1,1),1);
yh=zeros(n(1,1),1);


for j=0:pocet_alfa_rezu
    alfa=j/pocet_alfa_rezu;
    for i=1:n(1,1)
         yd(i,1)=y(i,1)+alfa*(y(i,2)-y(i,1));
         yh(i,1)=y(i,4)-alfa*(y(i,4)-y(i,3));
    end
    y0 = (yd(:,1)+yh(:,1))/2; %%% počáteční bod
    problemMin = createOptimProblem('fmincon','objective',...
    FunkceT,'x0',y0,'lb',yd(:,1),'ub',yh(:,1),'options',opts);
    [~,T(j+1,1)] = run(gs,problemMin);

    problemMax = createOptimProblem('fmincon','objective',...
    FunkceTKratMinus1,'x0',y0,'lb',yd(:,1),'ub',yh(:,1),'options',opts);
    [~,T_minus] = run(gs,problemMax);
    T(j+1,2)=-T_minus;

    hold on
    plot(T(j+1,1),alfa,'.k',T(j+1,2),alfa,'.k')
    hold off
end
```

Výsledné fuzzy číslo fuzzy testovací statistiky T_F můžeme vidět na obrázku 2.6. Nosič tohoto fuzzy čísla se rovná $(-0,1966; 1,3179)$. Kritický obor bude následující $W = (-\infty, -t_{10-1,1-\frac{0,1}{2}}\rangle \cup \langle t_{10-1,1-\frac{0,1}{2}}, \infty)$. V tabulkách si najdeme příslušnou hodnotu kvantilu a dostáváme $W = (-\infty; -1,833\rangle \cup \langle 1,833; \infty)$. Pojďme nyní zformulovat závěr celého našeho testování. Při správném přístupu podmíněné fuzzy aritmetiky, můžeme říci, že hypotézu nelze zamítnout, jelikož nastal případ Supp $T \cap W = \emptyset$. Nelze tedy zamítnout, že by se průměrné měsíční výdaje vybraných studentů navazujícího studia lišily od celorepublikového průměru. V případě klasické fuzzy aritmetiky tento závěr implicitně vyvodit nemůžeme, jelikož by část nosiče náležela do kritického oboru.



Obrázek 2.6: Výsledné fuzzy číslo testovací statistiky T_F vykresleno pomocí dvaceti α-řezů.

2.4.2. Párový t-test

Párový t-test se používá v situacích, kdy zkoumáme dvě náhodné veličiny, a navíc existuje závislost mezi těmito dvěma náhodnými veličinami. Jinými slovy provádíme dvě měření u jednoho výběrového souboru. Nejdříve se provede první měření, pak se aplikuje nějaký pokusný zásah a posléze se provede druhé měření na stejném výběrovém souboru. Takto získané hodnoty tvoří páry. V testu vycházíme z rozdílů naměřených párových hodnot a testujeme hypotézu, že střední hodnoty měření před pokusem a po pokusu se rovnají. Jinak řečeno, zda rozdíl středních hodnot párových měření je nulový.

Uvažujme tedy rozdíl našich dvou náhodných veličin Y a Z, který si označíme $X = Y - Z$. Za předpokladu normálního rozdělení náhodné veličiny X poté testujeme hypotézu:

$$H_0 : \mu = E(Y) - E(Z) = 0 \text{ proti } H_A : \mu \neq 0.$$

Testujeme tedy, zda pokusný zásah měl vliv na střední hodnotu sledovaného znaku. Uvedenou hypotézu poté ověříme klasickým jednovýběrovým t-testem, s kterým jsme se seznámili v minulé podkapitole. Tento jednovýběrový t-test použijeme na rozdíly párových hodnot. Stejný postup by se aplikoval i v případě fuzzy čísel, kde bychom použili už známou výběrovou funkci T_F. K výpočtu párových rozdílu bychom využili standardní fuzzy aritmetiku. Ukážeme si to na příkladu.

Příklad 2.2 *Zeptali jsme se sedmnácti studentů jednoho oboru na jejich vlastní představu o nástupní hrubé mzdě po vystudování tohoto oboru. Chceme popsat očekávanou hrubou mzdu studentů pomocí trojúhelníkových fuzzy čísel. Studentů jsme se tudíž zeptali, zda-li nám mohou říci:*

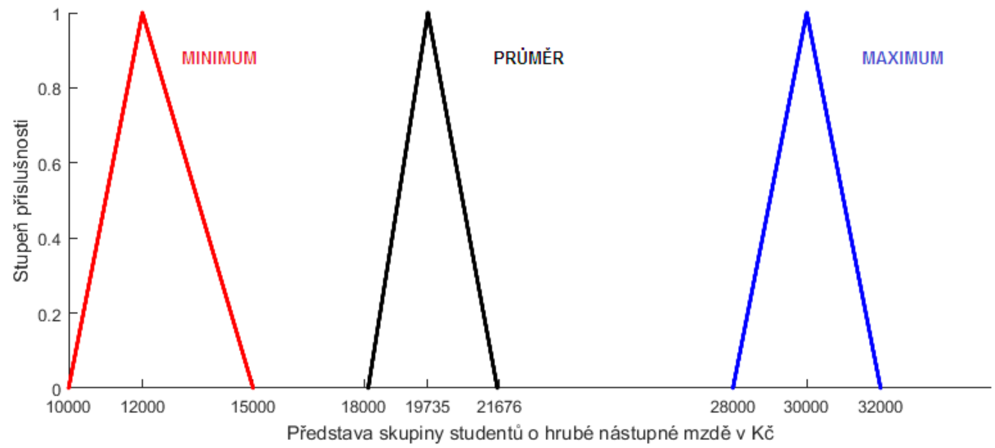
1. *minimální hrubou mzdu, pokud by nabídka byla menší než tato částka, tak by práci nepřijali,*
2. *nejočekávanější hrubou mzdu ve smyslu, jakou nabídku si myslí, že dostanou,*
3. *maximální hrubou mzdu, vyšší nabídku než tuto částku neočekávají.*

Odpovědi studentů jsou zaznamenány v tabulce 2.3.

Respondent	Minimální mzda	Nejočekávanější mzda	Maximální mzda
1	28 000	30 000	32 000
2	23 500	25 000	27 000
3	19 000	21 500	23 500
4	17 000	18 000	20 000
5	21 000	23 000	25 000
6	15 000	17 000	18 500
7	21 000	22 000	23 000
8	17 000	18 000	19 500
9	18 000	20 000	23 000
10	18 000	20 000	22 000
11	20 000	23 000	25 000
12	21 000	23 000	24 000
13	15 000	16 000	17 500
14	10 000	12 000	15 000
15	12 000	13 000	15 000
16	13 000	15 000	17 500
17	17 500	19 000	21 000

Tabulka 2.3: Odpovědi respondentů před pokusným zásahem

Vezměme si nyní tato data a spočítejme z nich fuzzy minimum, fuzzy průměr a fuzzy maximum. Tyto fuzzy charakteristiky jsou znázorněny na obrázku 2.7.



Obrázek 2.7: Fuzzy minimum, fuzzy průměr fuzzy maximum ze vstupních fuzzy dat.

Posléze jsme tento obrázek 2.7. ukázali našim sedmnácti respondentům a požádali je, aby poté, co viděli výsledky celé skupiny, znovu odpověděli na tři požadované údaje. Odpovědi studentů po provedení pokusného zásahu najdeme v tabulce 2.4.

Respondent	**Minimální mzda**	**Nejočekávanější mzda**	**Maximální mzda**
1	22 000	25 000	27 000
2	21 000	23 000	25 000
3	19 000	21 000	23 500
4	18 000	19 000	22 000
5	19 000	21 000	23 000
6	17 000	19 500	21 000
7	20 000	21 000	22 000
8	17 000	18 000	19 500
9	18 000	20 000	22 500
10	18 000	20 000	22 000
11	20 000	21 500	24 000
12	19 000	21 000	23 000
13	17 500	19 000	20 000
14	15 000	18 000	20 000
15	16 000	18 000	22 000
16	16 500	20 000	22 500
17	17 500	19 000	21 000

Tabulka 2.4: Odpovědi respondentů po pokusném zásahu

Nyní, když máme získaná data, které budeme modelovat pomocí trojúhelníkových fuzzy čísel, musíme vypočítat párový rozdíl mezi odpověďmi před a po pokusném zásahu pro každého respondenta. Při počítání párových rozdílů využíváme tedy odečítání fuzzy čísel pomocí standardní fuzzy aritmetiky. Párové rozdíly pro všechny respondenty jsou uvedeny v tabulce 2.5.

Respondent	**Párové rozdíly**		
1	1 000	5 000	10 000
2	-1 500	2 000	6 000
3	-4 500	500	4 500
4	-5 000	-1 000	2 000
5	-2 000	2 000	6 000
6	-6 000	-2 500	1 500
7	-1 000	1 000	3 000
8	-2 500	0	2 500
9	-4 500	0	5 000
10	-4 000	0	4 000
11	-4 000	1 500	5 000
12	-2 000	2 000	5 000
13	-5 000	-3 000	0
14	-10 000	-6 000	0
15	-10 000	-5 000	-1 000
16	-9 000	-5 000	1 000
17	-3 500	0	3 500

Tabulka 2.5: Fuzzy čísla párových rozdílů

Nyní otestujeme, zda ukázání jednotlivým studentům obrázek 2.7 s daty, jak si vedou jako skupina, nemělo vliv na střední hodnotu sledovaného znaku, což je představa studentů o nástupní hrubé mzdě, tedy

$$H_0 : \mu = 0 \textit{ proti } H_A : \mu \neq 0.$$

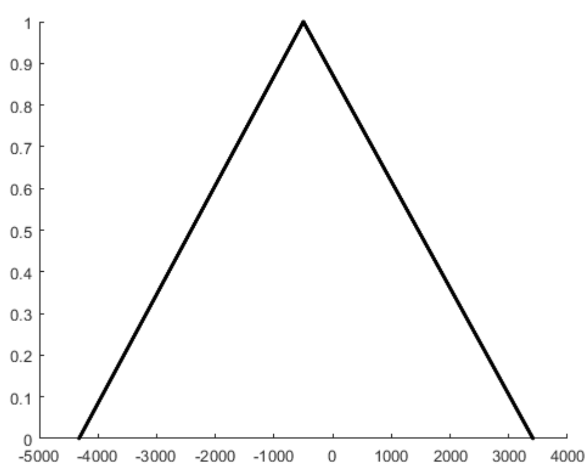
Na vypočítané trojúhelníková fuzzy čísla párových rozdílů, použijeme nám už známý jednovýběrový t-test s fuzzy testovací statistikou

$$T_F = \frac{\overline{X}_F - 0}{S_{n_F}} \sqrt{n},$$

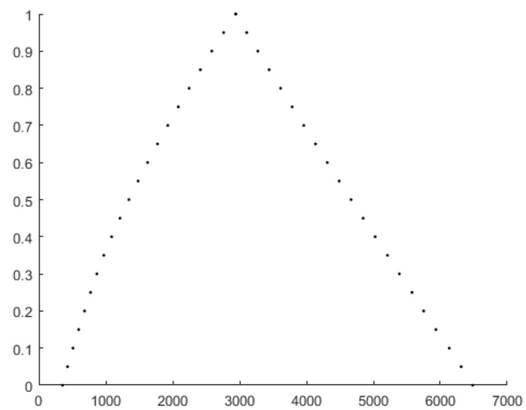
kde μ_0 rovno 0. Hladinu testu α si zvolíme $0,05$ a $n = 17$. Nyní postup výpočtu je identický jako v předchozím příkladě s tím rozdílem, že dosadíme do skriptů jiná vstupní data a pro výpočet výsledného fuzzy čísla testovací statistiky bude μ_0 rovno 0.

```
%načtení vstupních dat maticově
y=[1000 5000 5000 10000; -1500 2000 2000 6000; -4500 500 500 4500; ...
  -5000 -1000 -1000 2000; -2000 2000 2000 6000; -6000 -2500 -2500 1500;...
  -1000 1000 1000 3000; -2500 0 0 2500; -4500 0 0 5000; -4000 0 0 4000;...
  -4000 1500 1500 5000; -2000 2000 2000 5000; -5000 -3000 -3000 0;...
  -10000 -6000 -6000 0;-10000 -5000 -5000 -1000; -9000 -5000 -5000 1000;...
  -3500 0 0 3500];
mu = 0;
n = size(y);
% zadani optimalizovanych funkci
FunkceT = @(x)(mean(x)-mu)/std(x)*sqrt(n(1,1));
FunkceTKratMinus1 = @(x)-(mean(x)-mu)/std(x)*sqrt(n(1,1));
```

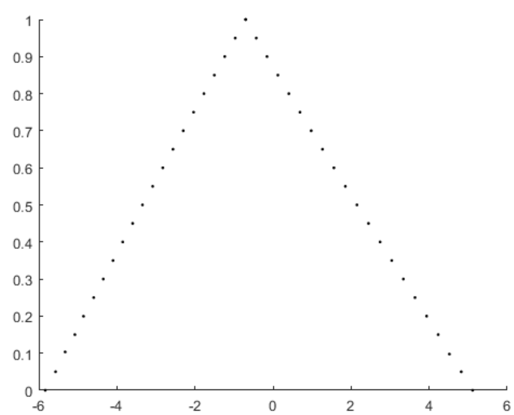
Pojďme se teď podívat na obrázek výběrového fuzzy průměru, výběrové fuzzy směrodatné odchylky a hlavně na výsledné fuzzy číslo realizace fuzzy testovací statistiky, které je důležité pro rozhodnutí o naší hypotéze. Kvůli časové náročnosti celého optimalizačního problému nám bude opět pro ukázku stačit si vykreslit dvacet α-řezů pro výběrovou fuzzy směrodatnou odchylku a výsledné fuzzy číslo fuzzy testovací statistiky.



Obrázek 2.8: Výběrový fuzzy průměr pro párové rozdíly.



Obrázek 2.9: Výběrová fuzzy směrodatná odchylka pro párové rozdíly.



Obrázek 2.10: Výsledná fuzzy testovací statistika T_F pro párové rozdíly vykreslena pomocí dvaceti α-řezů.

Nosič výsledného fuzzy čísla testovací statistiky se rovná $(-5,8350; 5,1128)$. *Kritický obor bude následující* $W = (-\infty, -t_{17-1,1-\frac{0,05}{2}}\rangle \cup \langle t_{17-1,1-\frac{0,05}{2}}, \infty)$. *V tabulkách si najdeme příslušnou hodnotu kvantilu a po dosazení dostáváme* $W = (-\infty; -2,120\rangle \cup \langle 2,120; \infty)$.

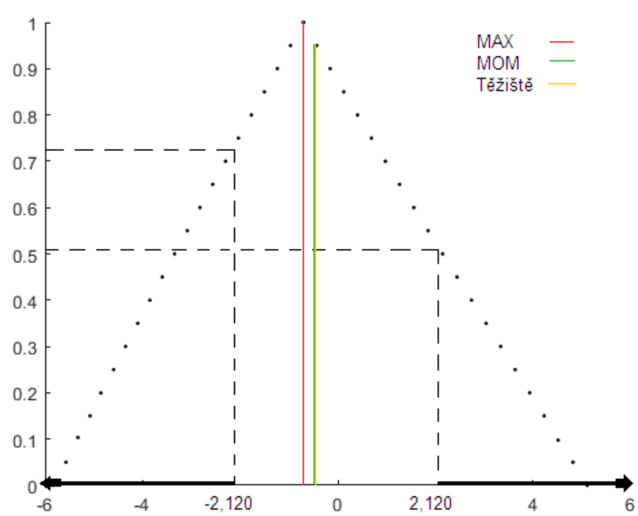
Nyní ovšem vidíme, že pouze část nosiče výsledné fuzzy testovací statistiky náleží do kritického oboru. Zkusíme si výsledné fuzzy číslo tedy defuzzifikovat pomocí tří přístupů, s kterými jsme se seznámili v kapitole 1.5. Konkrétně pomocí metody těžiště, metody středu maxim a metody MOM. V případě, když máme lineární fuzzy číslo, tak pro první dvě metody existuje přímo v programu MATLAB *příkaz na jejich vypočítání. MOM se dá posléze spočítat jako součet význačných hodnot fuzzy čísla x_1, x_2, x_3, x_4 děleno čtyřmi [11]. Dále si všimněme, že MATLAB zná metodu středu maxim pod pojmem mom, neplést tedy s naší definicí metody MOM.*

```
x = -10:0.1:10;
x1=-5.8350;
x2=-0.7007;
x3=-0.7007;
x4=5.1128;
D = trapmf(x,[x1 x2 x3 x4]);
teziste = defuzz(x,D,'centroid');
stredymaxim = defuzz(x,D,'mom');
MOM=(x1+x2+x3+x4)/4;
```

Po zadání do programu MATLAB nám vyjdou následující výsledky:

- *metoda těžiště = -0,47*
- *metoda středu maxim = -0,7*
- *metoda MOM = -0,53*

Pojďme si nyní celou situaci vykreslit do obrázku 2.11.



Obrázek 2.11: Fuzzy testovací statistika T_F s kritickým oborem a jednotlivými defuzzifikacemi.

Jak je tedy vidět ve všech třech případech bychom rozhodli, že hypotézu nelze zamítnout, jelikož hodnoty defuzzifikací nenáleží do kritického oboru. Nelze tedy zamítnout, že by ukázaní obrázku studentům, jak si vedou jako skupina mělo vliv na střední hodnotu představy studentů o jejich hrubé mzdě.

Poznámka 2.5 *Alternativně bychom mohli rozhodnutí o ne/zamítnutí hypotézy provést takto. Hodnoty funkce příslušnosti v hranicích kritického oboru jsou* $0,72$ *a* $0,51$*. Z těchto hodnot vezměme maximum, a tedy stupeň příslušnosti zamítnutí hypotézy je* $0,72$*. Rozhodovatel si může tedy stanovit určitou mez stupně příslušnosti, podle které se bude rozhodovat o ne/zamítnutí hypotézy. Tímto způsobem se dá taky vyřešit problém, kdy některé hodnoty defuzzifikací budou svědčit pro zamítnutí hypotézy a podle ostatních bychom hypotézu nezamítali.*

2.4.3. Test hypotéz o rozptylu normálního rozdělení

V tomto testu opět předpokládáme, že náhodná veličina X má normální rozdělení s parametry μ a σ^2. Symbol σ_0^2 značí hypotetickou hodnotu rozptylu normálního rozdělení. Testujeme tedy:

1. nulovou hypotézu $H_0 : \sigma^2 = \sigma_0^2$ proti tzv. oboustranné alternativě $H_A : \sigma^2 \neq \sigma_0^2$,
2. nulovou hypotézu $H_0 : \sigma^2 \leq \sigma_0^2$ proti tzv. jednostranné alternativě $H_0 : \sigma^2 > \sigma_0^2$,
3. nulovou hypotézu $H_0 : \sigma^2 \geq \sigma_0^2$ proti tzv. jednostranné alternativě $H_0 : \sigma^2 < \sigma_0^2$.

Testovacím kritériem je následující výběrová funkce

$$Z = \frac{(n-1)S_n^2}{\sigma_0^2}.$$

Tato funkce má za platnosti nulové hypotézy χ^2 rozdělení o $n-1$ stupních volnosti, S_n^2 značí výběrový rozptyl, který je nestranným odhadem $\sigma^2 = var(X)$. Nyní se opět podíváme na jednotlivé kritické obory.

1. V případě oboustranné alternativy budou kritický obor tvořit malé i velké hodnoty výběrové funkce Z. Proto

$$W = (0, \chi^2_{n-1,\frac{\alpha}{2}}\rangle \cup \langle\chi^2_{n-1,1-\frac{\alpha}{2}}, \infty).$$

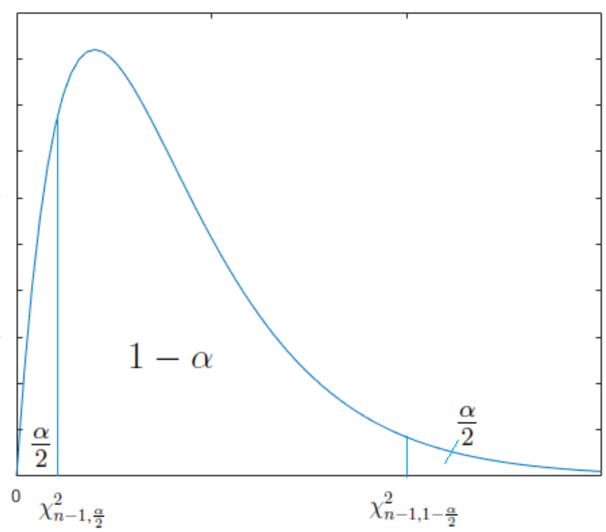
2. Ve prospěch alternativy svědčí velké hodnoty testového kritéria, proto

$$W = \langle\chi^2_{n-1,1-\alpha}, \infty).$$

3. Ve prospěch alternativy svědčí malé hodnoty testového kritéria, proto

$$W = (0, \chi^2_{n-1,\alpha}\rangle.$$

Poznámka 2.6 *Při vytváření kritického oboru vycházíme z charakteru grafu hustoty χ^2-rozdělení.*



Obrázek 2.12: Hustota χ^2-rozdělení o n-1 stupních volnosti.

V případě, že vstupní data budou fuzzy čísla, tak pro potřeby výpočtu hodnoty realizace testovací statistiky Z_F si musíme z těchto fuzzy čísel spočítat výběrový fuzzy rozptyl. Zadefinujme si tento pojem.

Definice 2.6 *Nechť* $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $i = 1,\dots,n$, *jsou fuzzy čísla. Pak výběrovým fuzzy rozptylem fuzzy čísel* $X_1,\dots,X_n$ *nazveme fuzzy číslo* $S^2_{n_F} = \{\langle \underline{s^2_{n_F}}(\alpha), \overline{s^2_{n_F}}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, *kde pro každé* $\alpha \in \langle 0,1\rangle$ *platí:*

$$\underline{s^2_{n_F}}(\alpha) = \min\{s^2_n(x_1,\dots,x_n) | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, i \in \{1,\dots,n\}\},$$

$$\overline{s^2_{n_F}}(\alpha) = \max\{s^2_n(x_1,\dots,x_n) | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, i \in \{1,\dots,n\}\}.$$

Testovací statistika pro fuzzy data bude tedy vypadat následovně

$$Z_F = \frac{(n-1)S^2_{n_F}}{\sigma^2_0}.$$

Výpočet realizace testovací statistiky Z_F provádíme podle následující definice.

Definice 2.7 *Nechť* $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $i = 1,\dots,n$, *jsou fuzzy čísla. Pak fuzzy testovací statistikou pro test hypotéz o rozptylu normálního rozdělení fuzzy čísel* $X_1,\dots,X_n$ *nazveme fuzzy číslo* $Z_F = \{\langle \underline{z_F}(\alpha), \overline{z_F}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, *kde pro každé* $\alpha \in \langle 0,1\rangle$ *platí:*

$$\underline{z_F}(\alpha) = \frac{(n-1)\underline{s^2_{n_F}}(\alpha)}{\sigma^2_0},$$

$$\overline{z_F}(\alpha) = \frac{(n-1)\overline{s^2_{n_F}}(\alpha)}{\sigma^2_0}.$$

Příklad 2.3 *Ve firmě se vyrábí jeden druh nealkoholického nápoje. Firma vyrábí tento nápoj pouze v dvoulitrovém vydání. Ve složení se vyskytuje jeden druh sladidla, který není v menších dávkách nebezpečný. Firma stanovila takovou směrnici, že číselná hodnota směrodatné odchylka* σ *obsahu tohoto sladidla ve vyráběných nápojích nesmí překročit hodnotu* $0,2$ *gramu. Po jednom roce firma náhodně vybrala* 20 *láhví svého vyráběného nápoje a změřila množství daného sladidla. Měřící přístroj je ovšem už značně zastaralý, a tudíž je vhodné počítat s odchylkou* $0,1$ *gramu v kladném i záporném směru u každé naměřené hodnoty. Naměřené hodnoty v gramech i s danými odchylkami jsou přehledně zobrazeny v tabulce 2.6.*

Láhev	**-0,1 gramu**	**Naměřená hodnota**	**+0,1 gramu**
1	0,51	0,61	0,71
2	1	1,1	1,2
3	0,75	0,85	0,95
4	0,63	0,73	0,83
5	0,57	0,67	0,77
6	0,7	0,8	0,9
7	0,66	0,76	0,86
8	0,92	1,02	1,12
9	0,58	0,68	0,78
10	0,8	0,9	1
11	0,77	0,87	0,97
12	0,81	0,91	1,01
13	0,66	0,76	0,86
14	1,02	1,12	1,22
15	0,88	0,98	1,08
16	0,74	0,84	0,94
17	0,69	0,79	0,89
18	0,89	0,99	1,09
19	0,63	0,73	0,83
20	0,9	1,0	1,1

Tabulka 2.6: Naměřené hodnoty sladidla v gramech

Naměřené hodnoty tedy budeme modelovat pomocí trojúhelníkových fuzzy čísel. Předpokládejme, že náhodná veličina značící obsah sladidla má normální rozdělení s parametry μ, σ^2. Bude nás zajímat, zda je důvod k obavám, že firma nedodržuje svoji stanovenou směrnici.

Jak už víme, tak druhá mocnina směrodatné odchylka je rozptyl. Budeme tedy testovat následující hypotézu:

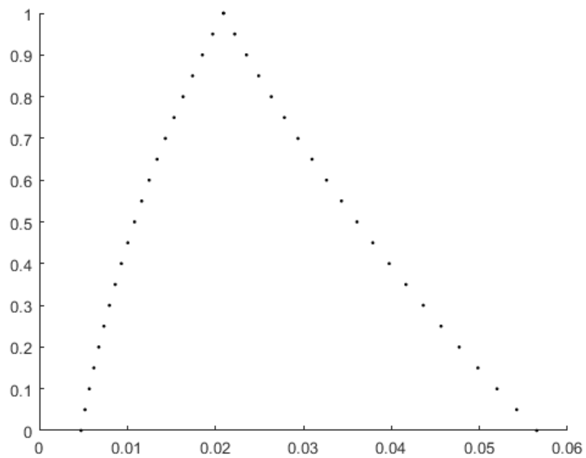
$$H_0 : \sigma^2 \leq 0,04 \ (= 0,2^2) \ \textit{proti} \ H_A : \sigma > 0,04.$$

Rozsah výběru máme $n = 20$ a hladinu testu si zvolíme $\alpha = 0,05$. Pro výpočet výběrového fuzzy rozptylu využijeme nám už známý skript pro výběrovou fuzzy směrodatnou odchylku, kde po lehké úpravě dostáváme skript pro výběrový fuzzy rozptyl.

```
%načtení vstupních dat maticově
y=[0.51 0.61 0.61 0.71; 1.0 1.1 1.1 1.2; 0.75 0.85 0.85 0.95; ...
    0.63 0.73 0.73 0.83; 0.57 0.67 0.67 0.77; 0.7 0.8 0.8 0.9;...
    0.66 0.77 0.77 0.87; 0.92 1.02 1.02 1.12; 0.58 0.68 0.68 0.78;...
    0.8 0.9 0.9 1; 0.77 0.87 0.87 0.97; 0.81 0.91 0.91 1.01;...
    0.66 0.76 0.76 0.86; 1.02 1.12 1.12 1.22; 0.88 0.98 0.98 1.08;...
    0.74 0.84 0.84 0.94; 0.69 0.79 0.79 0.89; 0.89 0.99 0.99 1.09;...
    0.63 0.73 0.73 0.83; 0.9 1.0 1.0 1.1;];

%zadání optimalizovaných funkcí
FunkceRozptylu = @(x)var(x);
FunkceRozptyluKratMinus1 = @(x)-var(x);
%počáteční bod optimalizace, nutno vyzkoušet různé možnosti

opts = optimset('Algorithm','active-set','Display','notify');
gs = GlobalSearch;

n=size(y);
pocet_alfa_rezu=20;
rozptyl=zeros(pocet_alfa_rezu+1,2);
yd=zeros(n(1,1),1);
yh=zeros(n(1,1),1);


for j=0:pocet_alfa_rezu
    alfa=j/pocet_alfa_rezu;
    for i=1:n(1,1)
        yd(i,1)=y(i,1)+alfa*(y(i,2)-y(i,1));
        yh(i,1)=y(i,4)-alfa*(y(i,4)-y(i,3));
    end
    y0 = (yd(:,1)+yh(:,1))/2; %%% počáteční bod
    problemMin = createOptimProblem('fmincon','objective',...
    FunkceRozptylu,'x0',y0,'lb',yd(:,1),'ub',yh(:,1),'options',opts);
    [~,varmin] = run(gs,problemMin);
    rozptyl(j+1,1)=varmin;

    problemMax = createOptimProblem('fmincon','objective',...
    FunkceRozptyluKratMinus1,'x0',y0,'lb',yd(:,1),'ub',yh(:,1),'options',opts);
    [~,varmax_minus] = run(gs,problemMax);
    %varmax=-varmax_minus;
    rozptyl(j+1,2)=-varmax_minus;

    hold on
    plot(rozptyl(j+1,1),alfa,'.k',rozptyl(j+1,2),alfa,'.k')
    hold off
end
```

Výběrový fuzzy rozptyl pro naše vstupní data můžeme vidět na obrázku 2.13.

Uzávěr nosiče tohoto fuzzy čísla je $\langle 0,0047; 0,0565\rangle$. *Teď už máme všechny potřebné výpočty a můžeme dosadit do testovací statistiky* Z_F *a spočítat její reali-*

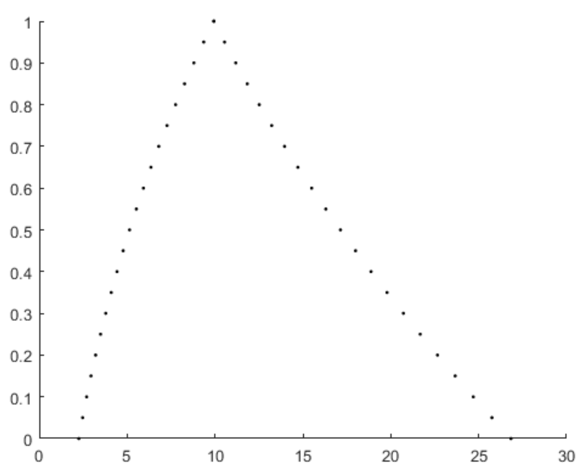


Obrázek 2.13: Výběrový fuzzy rozptyl pro vstupní data.

zaci pomocí standardní fuzzy aritmetiky, jelikož v testovací statistice máme pouze jedno fuzzy číslo. Dostáváme tudíž výsledek pro $\overline{Supp}\ Z_F$*:*

$$\overline{Supp}\ Z_F = \frac{(20-1)\langle 0,0047; 0,0565\rangle}{0,04} = \langle 2,2326; 26,8375\rangle.$$

Fuzzy číslo realizace testovací statistiky Z_F*, které bylo opět vykresleno pomocí dvaceti* $\alpha-$*řezů, můžeme vidět na obrázku 2.14.*



Obrázek 2.14: Fuzzy číslo realizace testovací statistiky Z_F vykresleno pomocí dvaceti $\alpha-$ řezů.

Podívejme se nyní na kritický obor, který bude následující $W = \langle \chi^2_{19;0,95}, \infty)$ a po dosazení tabulkové hodnoty dostáváme $W = \langle 30,1; \infty)$. Je tedy zřejmé, že nastal případ Supp $T_F \cap W = \emptyset$, a tudíž nelze naši hypotézu zamítnout. Nelze zamítnout, že by firma nedodržovala nastavenou směrnici.

2.4.4. F-test shody rozptylů dvou nezávislých normálně rozdělených znaků

Teď budeme uvažovat pro změnu opět dvě náhodné veličiny X, Y a navíc požadujeme jejich vzájemnou nezávislost. To je zásadní rozdíl oproti již námi známého párové t-testu, kde dvě náhodné veličiny byly na sobě závislé. Dále předpokládejme, že náhodná veličina X má normální rozdělení s parametry μ_1 a σ_1^2 a náhodná veličina Y má taktéž normální rozdělení, ovšem s parametry μ_2 a σ_2^2. Navíc $\sigma_1^2 \neq 0$ a $\sigma_2^2 \neq 0$.

Testujeme následující hypotézy:

1. nulovou hypotézu $H_0 : \sigma_1^2 = \sigma_2^2$ proti alternativě $H_A : \sigma_1^2 \neq \sigma_2^2$,
2. nulovou hypotézu $H_0 : \sigma_1^2 \leq \sigma_2^2$ proti alternativě $H_0 : \sigma_1^2 > \sigma_2^2$,
3. nulovou hypotézu $H_0 : \sigma_1^2 \geq \sigma_2^2$ proti alternativě $H_0 : \sigma_1^2 < \sigma_2^2$.

Testujeme tedy shodnost rozptylů. Testovací kritérium je výběrová funkce

$$V = \frac{S_n^2}{S_m^2} \frac{\sigma_2^2}{\sigma_1^2}.$$

Tato funkce má za platnosti nulové hypotézy F rozdělení o $n-1, m-1$ stupních volnosti. Navíc při určování hodnoty testového kritéria dosazujeme do vzorce $\frac{\sigma_2^2}{\sigma_1^2} = 1$. S_n^2 značí výběrový rozptyl náhodné veličiny X o rozsahu výběru n a S_m^2 značí výběrový rozptyl náhodné veličiny Y o rozsahu výběru m.

Příslušné kritické obory jsou následující:

1. V případě oboustranné alternativy budou kritický obor tvořit malé i velké hodnoty výběrové funkce F. Proto

$$W = (0, F_{n-1,m-1,\frac{\alpha}{2}}\rangle \cup \langle F_{n-1,m-1,1-\frac{\alpha}{2}}, \infty).$$

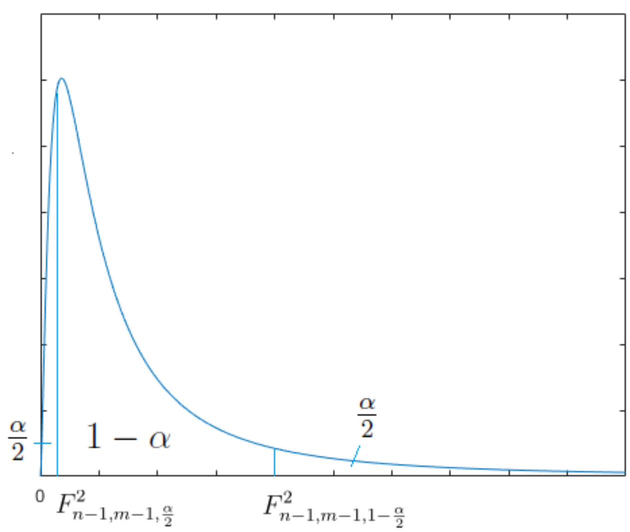
2. Ve prospěch alternativy svědčí velké hodnoty testového kritéria, proto

$$W = \langle F_{n-1,m-1,1-\alpha}, \infty).$$

3. Ve prospěch alternativy svědčí malé hodnoty testového kritéria, proto

$$W = (0, F_{n-1,m-1,\alpha}\rangle.$$

Poznámka 2.7 *Při vytváření kritického oboru vycházíme z charakteru grafu hustoty F-rozdělení, který můžeme vidět na obrázku 2.15.*



Obrázek 2.15: Hustota F-rozdělení o n-1, m-1 stupních volnosti.

Pojďme se nyní opět opět podívat, jak se nám změní testovací statistika pokud vstupní data budou fuzzy čísla. Testovací statistika bude následující výběrová funkce

$$V_F = \frac{S^2_{n_F}}{S^2_{m_F}} \frac{\sigma_2^2}{\sigma_1^2}.$$

Pro potřeby výpočtu hodnoty realizace testovací statistiky si musíme z těchto vstupních fuzzy čísel spočítat výběrový fuzzy rozptyl $S^2_{n_F} = \{\langle \underline{s^2_{n_F}}(\alpha), \overline{s^2_{n_F}}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, resp. $S^2_{m_F} = \{\langle \underline{s^2_{m_F}}(\alpha), \overline{s^2_{m_F}}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$. S tímto pojmem jsme se už seznámili v minulé podkapitole. Navíc je si také potřeba uvědomit, že zde nemusíme při výpočtu dělení dvou výběrových fuzzy rozptylů používat podmíněnou fuzzy aritmetiku, jelikož dané dvě fuzzy čísla jsou na sobě nezávislá. Také dosazujeme do vzorce $\frac{\sigma_2^2}{\sigma_1^2} = 1$.

Výpočet realizace testovací statistiky V_F tedy provádíme podle následující definice.

Definice 2.8 *Nechť* $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $i = 1,\dots,n$, *a* $Y_j = \{\langle \underline{y}_j(\alpha), \overline{y}_j(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, $j = 1,\dots,m$ *jsou fuzzy čísla. Pak fuzzy testovací statistikou pro F-test shody rozptylů dvou nezávislých normálně rozdělených znaků fuzzy čísel* $X_1,\dots,X_n$ *a* $Y_1,\dots,Y_m$ *nazveme fuzzy číslo* $V_F = \{\langle \underline{v_F}(\alpha), \overline{v_F}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$, *kde pro každé* $\alpha \in \langle 0,1\rangle$ *platí:*

$$\underline{v_F}(\alpha) = \min\left\{\frac{\underline{s^2_{n_F}}(\alpha)}{\underline{s^2_{m_F}}(\alpha)}, \frac{\underline{s^2_{n_F}}(\alpha)}{\overline{s^2_{m_F}}(\alpha)}, \frac{\overline{s^2_{n_F}}(\alpha)}{\underline{s^2_{m_F}}(\alpha)}, \frac{\overline{s^2_{n_F}}(\alpha)}{\overline{s^2_{m_F}}(\alpha)}\right\},$$

$$\overline{v_F}(\alpha) = \max\left\{\frac{\underline{s^2_{n_F}}(\alpha)}{\underline{s^2_{m_F}}(\alpha)}, \frac{\underline{s^2_{n_F}}(\alpha)}{\overline{s^2_{m_F}}(\alpha)}, \frac{\overline{s^2_{n_F}}(\alpha)}{\underline{s^2_{m_F}}(\alpha)}, \frac{\overline{s^2_{n_F}}(\alpha)}{\overline{s^2_{m_F}}(\alpha)}\right\}.$$

Příklad 2.4 *Vezměme si opět data z příkladu 2.1. Máme tedy odpovědi deseti studentů navazujícího studia, kteří odpovídali na otázku, kolik průměrně měsíčně utratili za jídlo a nealkoholické nápoje za rok 2015. Odpovědi respondentů navazujícího studia tudíž nalezneme v tabulce 2.2. Posléze jsme se dotázali šestnácti studentů bakalářského studia na tu stejnou otázku, a znovu, pokud studenti nevěděli přesnou odpověď, tak jsme se zeptali na minimální, nejmožnější a maximální odhad utracené částky. Přesně nám odpověděl pouze jeden studen, jehož odpověď byla 2000 Kč. Odpovědi všech studentů bakalářského studia najdeme v tabulce 2.7. Zajímalo by nás, jestli je podstatný rozdíl mezi výdaji za potraviny a nealkoholické nápoje mezi studenty navazujícího a bakalářského studia.*

Náhodná veličina X nám bude určovat průměrné měsíční výdaje studentů navazujícího studia, kde rozsah výběru n=10. Náhodná veličina Y nám zase určuje průměrné měsíční výdaje studentů bakalářského studia, kde rozsah výběru m=16.

Respondent	Minimální hod.	Nejočekávanější hod.	Maximální hod.	Přesná hod.
1	1900	2100	2300	n/a
2	3000	3200	3500	n/a
3	1500	1700	1850	n/a
4	1900	2000	2100	n/a
5	2200	2500	3000	n/a
6	2000	2200	2300	n/a
7	1700	1850	2000	n/a
8	1700	2000-2200	2500	n/a
9	n/a	n/a	n/a	2000
10	2300	2400	2500	n/a
11	2300	2500-2800	3000	n/a
12	2500	2650	2800	n/a
13	2350	2500	2750	n/a
14	2500	3000	3500	n/a
15	1500	1600	1800	n/a
16	1850	2100	2300	n/a

Tabulka 2.7: Odpovědi respondentů bakalářského studia

Je zřejmé, že X a Y jsou nezávislé náhodné veličiny. Předpokládáme, že průměrné výdaje studentů mají normální rozdělení. Hladinu testu si zvolíme $\alpha = 0,05$. Budeme testovat nulovou hypotézu, zda rozptyl obou náhodných veličin je stejný:

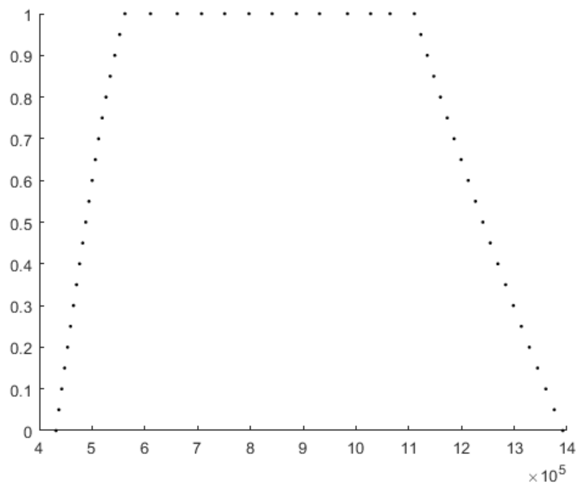
$$H_0 : \sigma_1^2 = \sigma_2^2 \textit{ proti } H_A : \sigma_1^2 \neq \sigma_2^2.$$

Potřebujeme nyní vypočítat výběrový fuzzy rozptyl pro oba výběry. K tomu využijeme nám už známý skript, který spustíme nejdříve pro vstupní data od studentů navazujícího studia a posléze pro studenty bakalářského studia.

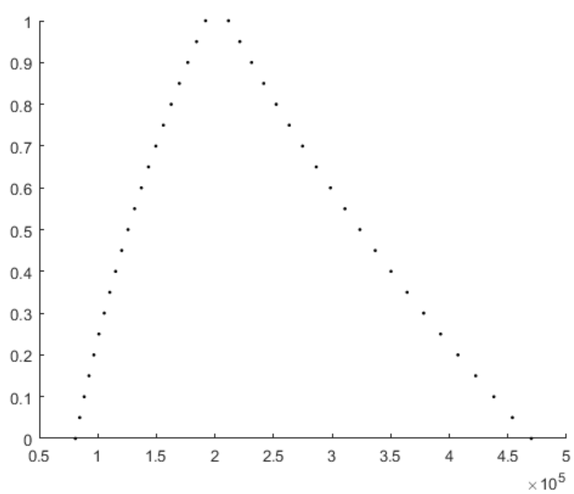
```
%načtení vstupních dat maticově (studenti navazujícího studia)
y=[1900 2000 2000 2100; 1500 2500 3000 3500; 1953 1953 1953 1953; ...
    2000 2200 2200 2240; 3900 4000 5000 5100; 1000 1200 1200 1500;...
    2000 2100 2100 2200; 2500 2500 2500 2500; 1800 2000 2000 2200;...
    1400 1500 1500 1600];

%načtení vstupních dat maticově (studenti bakalářského studia)
y=[1900 2100 2100 2300; 3000 3200 3200 3500; 1500 1700 1700 1850; ...
    1900 2000 2000 2100; 2200 2500 2500 3000; 2000 2200 2200 2300;...
    1700 1850 1850 2000; 1700 2000 2200 2500; 2000 2000 2000 2000;...
    2300 2400 2400 2500; 2300 2500 2800 3000; 2500 2650 2650 2800;...
    2350 2500 2500 2750; 2500 3000 3000 3500; 1500 1600 1600 1800;...
    1850 2100 2100 2300];
```

```
% zadani optimalizovanych funkci
FunkceRozptylu = @(x)var(x);
FunkceRozptyluKratMinus1 = @(x)-var(x);
```

Dostáváme výběrové fuzzy rozptyly pro oba výběry, které můžeme vidět na obrázcích 2.16 a 2.17 (opět vykresleno po dvaceti α-řezech).



Obrázek 2.16: Výběrový fuzzy rozptyl pro data studentů navazujícího studia.



Obrázek 2.17: Výběrový fuzzy rozptyl pro data studentů bakalářského studia.

$\overline{Supp}\ S^2_{n_F} = \langle 431\,000; 1\,392\,000\rangle$ *a* $\overline{Supp}\ S^2_{m_F} = \langle 81\,000; 469\,000\rangle$. *Dosadíme* *do testovací statistiky a dostáváme pro* $\overline{Supp}\ V_F$*:*

$$\overline{Supp}\ V_F = \frac{\langle 431\,000; 1392\,000\rangle}{\langle 81\,000; 469\,000\rangle} = \langle 0,919; 17,185\rangle,$$

kde

$$0,919 = \min\left\{\frac{431\,000}{81\,000}, \frac{431\,000}{469\,000}, \frac{1\,392\,000}{81\,000}, \frac{1\,392\,000}{469\,000}\right\},$$

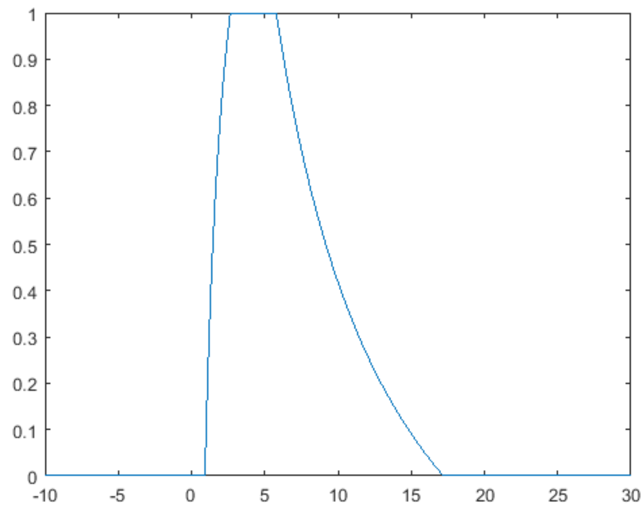
$$17,185 = \max\left\{\frac{431\,000}{81\,000}, \frac{431\,000}{469\,000}, \frac{1\,392\,000}{81\,000}, \frac{1\,392\,000}{469\,000}\right\}.$$

Pro připomenutí výpočet realizace testovací statistiky se zde provádí pomocí standardní fuzzy aritmetiky, jelikož není zde nutné si zavádět nějakou dodatečnou podmínku. Výběrové fuzzy rozptyly jsou na sobě nezávislé. Vykresleme si výsledné fuzzy číslo realizace testovací statistiky pomocí programu MATLAB. Jak je vidět i z obrázků 2.16 a 2.17 tyto fuzzy rozptyly nejsou lineárními fuzzy čísly, ale jelikož linearita není porušena nějakým zásadním způsobem, budeme pro jednoduchost výpočtu modelovat tyto fuzzy čísla jako lineární lichoběžníkové fuzzy čísla. Na to nám v programu MATLAB slouží, jak už víme příkaz trapmf a pro dělení fuzzy čísel příkaz fuzarith s parametrem div. Výpočet hodnoty realizace testovací statistiky v programu MATLAB vypadá tedy následovně:

```
x=-10:0.001:30;
A = trapmf(x,[431000 562000 1111000 1392000]);
B = trapmf(x,[81000 192000 212000 469000]);
Cdiv = fuzarith(x,A,B,'div');
plot(x,Cdiv)
```

Příkazem plot jsme si vykreslili toto výsledné fuzzy číslo, které můžeme vidět na obrázku 2.18.

Co se týče kritického oboru, tak ten je následující: $W = (0, F_{10-1,16-1,\frac{0,05}{2}}\rangle \cup \langle F_{10-1,16-1,1-\frac{0,05}{2}}, \infty) = (0; 0,265\rangle \cup \langle 3,123; \infty)$. *Jak jsme si vypočítali, tak nosič fuzzy čísla realizace testovací statistiky* V_F *je* $(0,919; 17,208)$. *Opět nám tedy nastává situace, kdy nosič fuzzy čísla náleží do kritického oboru jenom zčásti.*



Obrázek 2.18: Výsledné fuzzy číslo realizace testovací statistiky V_F.

Zde nastává ještě jeden problém, a to ten, že jak vidíme, tak výsledné fuzzy číslo není lineární a při linearizaci bychom ztratili už velkou část informace. Defuzzifikovat takovéto fuzzy číslo by ovšem velice náročné, musíme si tedy pomoci a použít modelaci tohoto fuzzy čísla jako takzvané po částech lineární fuzzy číslo. S tímto typem fuzzy čísla jsme se seznámili už v první kapitole. Vezmeme si dvacet α-řezů. V [12] se už defuzzifikací po částech lineárních fuzzy čísel zabývali, proto využijeme již vytvořené skripty (v příloze pod názvem teziste.m a MOM.m) a zjistíme hodnoty defuzzifikací pro metody těžiště, středů maxim a MOM.

Získáváme tyto výsledky:

- *metoda těžiště = 6,1917*
- *metoda středu maxim = 4,22*
- *metoda MOM = 5,7271*

Hodnota maximálního stupně příslušnosti, s jakým by šlo hypotézu zamítnout, je rovna 1. *Výsledky všech metod náleží kritickému oboru, proto naši hypotézu zamítáme a výběry nemají stejný rozptyl. Tím ovšem jsme ještě neodpověděli na naši otázku, zda je podstatný mezi výdaji za potraviny a nealkoholické nápoje mezi studenty navazujícího a bakalářského studia. Odpovědět budeme moci po* *seznámení se s dalším testem.*

2.4.5. Přibližný t-test

Přibližný t-test se používá v případě testování shodnosti středních hodnot dvou nezávisle normálně rozdělených znaků a platí, že $\sigma_1^2 \neq \sigma_2^2$. Přibližných t-testů je více, my v této práci využijeme takzvaný Aspinové-Welchův test [17].

Nechť tedy náhodná veličina X má normální rozdělení s parametry μ_1 a σ_1^2 a náhodná veličina Y má taktéž normální rozdělení, ovšem s parametry μ_2 a σ_2^2. Nechť jsou tyto dvě náhodné veličiny mezi sebou nezávislé.

Testujeme následující hypotézy:

1. nulovou hypotézu $H_0 : \mu_1^2 = \mu_2^2$ proti alternativě $H_A : \mu_1^2 \neq \mu_2^2$,
2. nulovou hypotézu $H_0 : \mu_1^2 \leq \mu_2^2$ proti alternativě $H_0 : \mu_1^2 > \mu_2^2$,
3. nulovou hypotézu $H_0 : \mu_1^2 \geq \mu_2^2$ proti alternativě $H_0 : \mu_1^2 < \mu_2^2$.

Testovacím kritériem je následující výběrová funkce:

$$T^* = \frac{\overline{X}_n - \overline{Y}_m}{S},$$

kde

$$S = \sqrt{\frac{S_n^2}{n} + \frac{S_m^2}{m}}.$$

Za platnosti nulové hypotézy má tato funkce přibližně t-rozdělení o f stupních volnosti, kde

$$f = \frac{S^4}{\frac{1}{n-1}(\frac{S_n^2}{n})^2 + \frac{1}{m-1}(\frac{S_m^2}{m})^2}.$$

Počet stupňů volnosti f většinou nevyjde jako celé číslo, proto provedeme zaokrouhlení a posléze najdeme v tabulkách příslušný kvantil t-rozdělení pro naši zaokrouhlenou hodnotu f.

Nyní se podívejme, jak bude vypadat kritický obor W.

1. V případě oboustranné alternativy budou kritický obor tvořit malé i velké hodnoty výběrové funkce T. Proto

$$W = (-\infty, -t_{f,1-\frac{\alpha}{2}}\rangle \cup \langle t_{f,1-\frac{\alpha}{2}}, \infty).$$

2. Ve prospěch alternativy svědčí velké hodnoty testového kritéria, proto

$$W = \langle t_{f,1-\alpha}, \infty).$$

3. Ve prospěch alternativy svědčí malé hodnoty testového kritéria, proto

$$W = (-\infty, -t_{f,1-\alpha}\rangle.$$

V případě, kdy vstupní data budou fuzzy čísla, bude výpočet testovací statistiky u tohoto testu značně komplikovaný. Testovací statistika bude následující výběrová funkce:

$$T_F^* = \frac{\overline{X}_{n_F} - \overline{Y}_{m_F}}{S_F},$$

kde

$$S_F = \sqrt{\frac{S_{n_F}^2}{n} + \frac{S_{m_F}^2}{m}}.$$

Problémem je, že S_F ve jmenovateli závisí na výběrových fuzzy průměrech, co jsou v čitateli, takže se nedá uplatnit standardní fuzzy aritmetika. Výpočet realizace testovací statistiky T_F^* tedy provádíme podle následující definice.

Definice 2.9 *Nechť* $X_i = \{\langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}, i = 1,\dots,n,$ *a* $Y_j = \{\langle \underline{y}_j(\alpha), \overline{y}_j(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}, j = 1,\dots,m,$ *jsou fuzzy čísla. Pak fuzzy testovací statistikou pro Aspinové-Welchův test fuzzy čísel* $X_1,\dots,X_n$ *a* $Y_1,\dots,Y_m$ *nazveme fuzzy číslo* $T_F^* = \{\langle \underline{t_F^*}(\alpha), \overline{t_F^*}(\alpha)\rangle, \alpha \in \langle 0,1\rangle\}$*, kde pro každé* $\alpha \in \langle 0,1\rangle$ *platí:*

$$\begin{gathered} t_F^*(\alpha) = \min\{\tfrac{\frac{1}{n}\sum_1^n x_i - \frac{1}{m}\sum_1^m y_j}{S_F} | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \\ i \in \{1,\dots,n\}, y_j \in \langle \underline{y}_j(\alpha), \overline{y}_j(\alpha)\rangle, j \in \{1,\dots,m\}\}, \end{gathered}$$

$$\begin{gathered} t_F^*(\alpha) = \max\{\tfrac{\frac{1}{n}\sum_1^n x_i - \frac{1}{m}\sum_1^m y_j}{S_F} | x_i \in \langle \underline{x}_i(\alpha), \overline{x}_i(\alpha)\rangle, \\ i \in \{1,\dots,n\}, y_j \in \langle \underline{y}_j(\alpha), \overline{y}_j(\alpha)\rangle, j \in \{1,\dots,m\}\}, \end{gathered}$$

kde

$$S_F = \sqrt{\frac{\frac{1}{n-1}\sum_1^n (x_i - \frac{1}{n}\sum_1^n x_i)^2}{n} + \frac{\frac{1}{m-1}\sum_1^m (y_j - \frac{1}{m}\sum_1^m y_j)^2}{m}}.$$

Optimalizovat takovou úlohu je poměrně náročné a i MATLAB při ní naráží na své početní limity. Při zadání optimalizačního problému nebyl tento program schopen spočítat správné hodnoty minim a maxim pro jednotlivé $\alpha-$řezy. Proto musíme použít nouzové řešení a spočítáme hodnotu testovací statistiky pomocí standardní fuzzy aritmetiky. Výsledek bude tedy lehce nepřesný. Tato nepřesnost znamená, že výsledek bude více neurčitý. Skutečné fuzzy číslo bude jeho podmnožinou. Nepřesnost výsledku můžeme zkusit kompenzovat větší přísností a zvolit si hladinu testu například $\alpha = 0,01$, a to i za cenu větší pravděpodobnosti provedení chyby druhého druhu.

Další problém u tohoto testu nastává s výpočtem stupně volnosti f. Totiž správně by f mělo být taky fuzzy číslo. Jenže jak určit ten správný zaokrouhlený počet stupňů volnosti, když nosič klidně může být široký přes například padesát celých čísel a navíc i optimalizace f je poměrně náročná. Nabízí se opět spočítat f pomocí standardní fuzzy aritmetiky a následně toto fuzzy číslo defuzzifikovat pomocí nám známých metod. To ovšem znamená, že s jiným typem defuzzifikace vyjde jiné zaokrouhlené f, a tudíž i možnost jiného závěru testování. Jako druhou možnost si můžeme vzít ze všech vstupních fuzzy čísel jejich nejočekávanější hodnotu (tedy budeme mít čísla reálná) a z nich spočítat hodnotu f jako jedno reálné číslo.

Příklad 2.5 *Pojďme pokračovat s příkladem 2.4. Zajímalo by nás, jestli je podstatný rozdíl mezi výdaji za potraviny a nealkoholické nápoje mezi studenty navazujícího a bakalářského studia. Nejdříve jsme otestovali shodu rozptylů, kde jsme hypotézu zamítli. Proto tedy teď při testování středních hodnot využijeme přibližný t-test, se kterým jsme se výše seznámili. Víme, že* $n = 10, m = 16, \overline{Supp}\ S^2_{n_F} = \langle 431\,000; 1\,392\,000\rangle, \overline{Supp}\ S^2_{m_F} = \langle 81\,000; 469\,000\rangle$ *a hladinu testu* α *si zvolíme* $0,01$.

Nyní ještě potřebujeme spočítat pro oba výběry výběrový fuzzy průměr. Na to už máme skript z minulých příkladů, takže jen dosadíme naše vstupní data a dostaneme $\overline{Supp}\ \overline{X}_n = \langle 1995,3; 2489,5\rangle$ *a* $\overline{Supp}\ \overline{Y}_m = \langle 2075; 2512,5\rangle$.

Jak už bylo naznačeno spočítat hodnotu testovací statistiky T_F^* *pomocí optimalizace je poměrně náročné. Proto si zde rozepíšeme postup pomocí standardní fuzzy aritmetiky, i když správný přístup by byl pomocí podmíněné. Postup si rozepíšeme pouze pro uzávěry nosičů* $\overline{Supp}\ S_F$ *a* $\overline{Supp}\ T_F^*$.

$$\overline{Supp}\ S_F = \sqrt{\frac{\langle 431\,000; 1\,392\,000\rangle}{10} + \frac{\langle 81\,000; 469\,000\rangle}{16}}$$

$$\overline{Supp}\ S_F = \sqrt{\langle 43\,100; 139\,200\rangle + \langle 5062,5; 29\,312,5\rangle}$$

$$\overline{Supp}\ S_F = \sqrt{\langle 48\,162,5; 168\,512,5\rangle} = \langle 219,46; 410,5\rangle$$

$$\overline{Supp}\ T_F^* = \frac{\langle 1995,3; 2489,5\rangle - \langle 2075; 2512,5\rangle}{\langle 219,46; 410,5\rangle} = \frac{\langle -517,2; 414,5\rangle}{\langle 219,46; 410,5\rangle} = \langle -2,36; 1,88\rangle$$

Nám k rozhodnutí stačí pouze nosič, a tedy $Supp\ T_F^* = (-2,36; 1,88)$.

Druhým problémem bylo určení počtu stupně volností f. Zkusíme si ten způsob, že f vypočítáme pomocí reálných čísel, které reprezentují nejočekávanější hodnoty daných vstupních fuzzy čísel. Po dosazení do vzorečku nám po zaokrouhlení vyjde $f = 177$. Kritický obor vypadá tedy následovně $W = (-\infty, t_{177;0,995}\rangle \cup \langle t_{177,0;995}, \infty)$. Platí, že při počtu stupních volnosti větších jak 100, se postupně tvar t-rozdělení blíží aproximativně k normálnímu rozdělení [1]. Najdeme tedy v tabulkách pro normální rozdělení příslušnou hodnotu, která je $2,576$. Máme tedy kritický obor $W = (-\infty, -2,576\rangle \cup \langle 2,576, \infty)$. Jak je tedy vidět celý nosič T_F^* nenáleží do kritického oboru, proto naši hypotézu nemůžeme zamítnout. Navíc při počítání podmíněnou fuzzy aritmetikou bývá výsledný nosič užší než u klasické fuzzy aritmetiky, tudíž bychom určitě také hypotézu nezamítali na dané hladině testu. Nelze říci, že by byl podstatný rozdíl mezi výdaji studentů navazujícího a bakalářského studia za potraviny a nealkoholické nápoje.

Poznámka 2.8 *Pokud bychom zvolili* $\alpha = 0,05$, *byla by hodnota kvantilu* $1,96$. *Zde by pouze jenom malá část nosiče přesahovala do kritického oboru a navíc* *při předpokládaném zúžení nosiče by se dalo očekávat, že i na této hladině nelze hypotézu zamítnout.*

Poznámka 2.9 *Pokud by nám v předchozím testu vyšlo, že nelze zamítnout shodnost rozptylů, tak bychom použili místo přibližného testu, takzvaný dvou-výběrový t-test. Zde se opět testuje shoda středních hodnot. Pro test se využívá výběrová funkce:*

$$T_2 = \frac{\overline{X}_n - \overline{Y}_n - (\mu_1 - \mu_2)}{\sqrt{(n-1)S_n^2 + (m-1)S_m^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Tato funkce má za platnosti nulové hypotézy t-rozdělení o $n+m-2$ stupni volnosti. Příslušné kritické obory jsou následující:

1. *V případě oboustranné alternativy budou kritický obor tvořit malé i velké hodnoty výběrové funkce T. Proto*

$$W = (-\infty, -t_{n+m-2,1-\frac{\alpha}{2}}\rangle \cup \langle t_{n+m-2,1-\frac{\alpha}{2}}, \infty).$$

2. *Ve prospěch alternativy svědčí velké hodnoty testového kritéria, proto*

$$W = \langle t_{n+m-2,1-\alpha}, \infty).$$

3. *Ve prospěch alternativy svědčí malé hodnoty testového kritéria, proto*

$$W = (-\infty, -t_{n+m-2,1-\alpha}\rangle.$$

V případě vstupních fuzzy dat vypadá testovací statistika následovně

$$T_{2_F} = \frac{\overline{X}_{n_F} - \overline{Y}_{m_F} - (\mu_1 - \mu_2)}{\sqrt{(n-1)S_{n_F}^2 + (m-1)S_{m_F}^2}} \sqrt{\frac{nm(n+m-2)}{n+m}}.$$

Poté se postupuje ve výpočtu obdobně jako v přibližném t-testu. Navíc odpadá problém s výpočtem stupně volnosti f.

Seznámili jsme se tedy se základními parametrickými testy a ukázali si, jak postupovat v případě vstupních dat v podobě fuzzy čísel. Existují i další parametrické testy nejen za předpokladu normálního rozdělení, ale i třeba alternativního nebo Poissonova, těmi se už v práci zabývat nebudeme.

Závěr

Cílem této diplomové práce bylo popsat obecný postup testování statistických hypotéz na základě fuzzy dat a také se posléze detailně zaměřit na vybrané parametrické statistické testy a popsat, jak probíhá testování u jednotlivých testů, když získaná data jsou fuzzy čísla. Cílem také bylo u každého testu uvést vlastní konkrétní příklad pro pochopení problematiky.

V první kapitole této práce jsme si zavedli základní pojmy z teorie fuzzy množin. Tyto pojmy byly důležité k pochopení následující kapitoly. V té jsme si nejdříve uvedli základní pojmy z testování statistických hypotéz, poté jsme si představili obecný postup při testování. Pak už následovala podkapitola o tom, jak vypadá obecný postup testování hypotéz, když máme vstupní data ve formě fuzzy čísel. Dále jsme si i rozebrali problémy, které mohou při takovémto testování nastat a navrhli jsme si jejich možné řešení. Dále jsme si představili konkrétní vybrané parametrické testy. U každého testu jsme si nejdřív ukázali, jak vypadá v případě reálných dat a co se změní, když uvažujeme data fuzzy. Poté jsme si uvedli konkrétní příklad a postup jeho vyřešení.

Téma této práce mě zaujalo, jelikož jsem se už aritmetikou fuzzy čísel zabýval v mé bakalářské práci a zde se mi naskytl způsob, jak tyto znalosti aplikovat více v praxi. Zpracovat toto téma byla pro mě obrovská výzva, jelikož, jak jsem později zjistil, k tomuto tématu není moc literatury, a tudíž většina problémů, s kterými jsme se v práci setkali, je vyřešena podle mých myšlenek a nápadů. Navíc jsem se musel zdokonalit v programu MATLAB, jelikož bylo potřeba vytvořit v tomto programu optimalizační úlohy sloužící k řešení našich úloh. Doufám, že tato práce poslouží i dalším lidem k seznámení se s touto problematikou.

Literatura

[1] Anděl, J.:. Statistické metody. MATFYZPRESS, Praha, 2003.

[2] Arefi, M., Taheri, S.: Testing Fuzzy Hypotheses Using Fuzzy Data Based on Fuzzy Test Statistic, Journal of Uncertain Systems, 2011, 45-61.

[3] Hron, K., Kunderová, P.: Základy počtu pravděpodobnosti a metod matematické statistiky. Univerzita Palackého v Olomouci, Olomouc, 2013, 1. vydání.

[4] Holanová, T., Ve výdajích českých domácností pořád vede bydlení a jídlo. Méně ale utrácí za vzdělání nebo kulturu [online]. Hospodářské noviny, 2016, [citováno 16.10. 2016]. Dostupné z: http://byznys.ihned.cz/c1-65195930-ve-vydajich-ceskych-domacnosti-porad-vede-bydleni-skoro-stejne-pak-cesi-utrati-za-jidlo-a-piti.

[5] Kacrová, P.: Statistické zpracování fuzzy dat. Diplomová práce, Univerzita Palackého v Olomouci, Olomouc, 2011.

[6] Klir, G.J., Pan, Y.: Constrained fuzzy arithmetic: Basic questions and some answers. Soft Computing 2, Springer-Verlag, 1998, 100-108.

[7] Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic Theory and Applications. Preatice Hall, New Jersey, 1996.

[8] Konopka, L.: Srovnání různých přístupů k intervalové a fuzzy aritmetice. Bakalářská práce, Univerzita Palackého v Olomouci, Olomouc, 2015.

[9] Mareš, M.: Algebraické vlastnosti fuzzy veličin. Robust's, 2002, 224-239.

[10] Navara, M., Olšák, P.: Základy fuzzy množin. ČVUT, Praha, 2002.

[11] Novosádová, L.: Metody porovnání fuzzy čísel. Bakalářská práce, Univerzita Palackého v Olomouci, Olomouc, 2010.

[12] Novosádová, L.: Metody defuzzifikace fuzzy čísel. Diplomová práce, Univerzita Palackého v Olomouci, Olomouc, 2012.

[13] Pavlačka, O., Talašová, J., Fuzzy vectors as a tool for modeling uncertain multidimensional quantities. Fuzzy Sets and Systems 161 (2010) 1585-1603.

[14] Talášek, T.: Fuzzy modely založené na bázích pravidel a jejich aplikace. Diplomové práce, Univerzita Palackého v Olomouci, Olomouc, 2012.

[15] Talašová, J.: Fuzzy metody vícekriteriálního hodnocení a rozhodování. Univerzita Palackého, Olomouc, 2003, 1. vydání.

[16] Viertl, R.: Statistical methods for fuzzy data. Wiley, Chinchester, 2011.

[17] Welch, B. L.: The generalization of "Student's"problem when several different population variances are involved, Biometrika, 1947.

[18] Přispěvatelé Wikipedie, Testování statistických hypotéz [online], Wikipedie: Otevřená encyklopedie, c2017, Datum poslední revize 12. 03. 2017, 19:02 UTC, [citováno 08. 01. 2017]

[19] Zadeh, L. A.: The concept of a linguistic variable and its application to approximate reasoning - II. Information s, sciences, 1975, 8.4: 301-357.

[20] Ženčák, P.: MATLAB pro začátečníky i mírně pokročilé. Univerzita Palackého v Olomouci, Olomouc, 2013.