

**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

---

**Fakulta elektrotechnická**

**Katedra kybernetiky**

**Zpracování videookulografických dat se zaměřením na detekci dyslexie**

**VOG Data Processing with Aim to Dyslexia Detection**

Diplomová práce

**Studijní program:** Biomedicínské inženýrství a informatika

**Studijní obor:** Biomedicínské inženýrství

**Vedoucí práce:** Ing. Vratislav Fabián, Ph. D.

**Bc. Alena Juklová**

---

**Praha, květen 2015**

České vysoké učení technické v Praze  
Fakulta elektrotechnická

Katedra kybernetiky

## ZADÁNÍ DIPLOMOVÉ PRÁCE

**Student:** Bc. Alena J u k l o v á

**Studijní program:** Biomedicínské inženýrství a informatika (magisterský)

**Obor:** Biomedicínské inženýrství

**Název tématu:** Zpracování videookulografických dat se zaměřením na detekci dyslexie

### Pokyny pro vypracování:

1. Proveďte podrobnou rešerši patentové a odborné literatury v oblasti diagnostiky dyslexie.
2. Proveďte předzpracování naměřených VOG dat a vyberte reprezentativní skupiny osob.
3. Navrhněte a implementujte vhodné metody pro statistické vyhodnocení naměřených dat.

### Seznam odborné literatury:

- [1] Jošt, J. (2009). Oční pohyby, čtení a dyslexie. Praha : Fortuna, 176 s.  
ISBN 978-80-7373-055-0.
- [2] Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51, 1457-1483.
- [3] Andreassi, J. L. (2007). *Psychophysiology, Human behavior & Physiological responses*. Psychological Press, New York. ISBN 978-0-8058-4951-6.

**Vedoucí diplomové práce:** Ing. Vratislav Fabián, Ph.D.

**Platnost zadání:** do konce letního semestru 2015/2016

L.S.

doc. Dr. Ing. Jan Kybic  
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.  
děkan

V Praze dne 9. 2. 2015

## **Anotace**

Diplomová práce seznamuje čtenáře s problematikou včasné diagnostiky dyslexie u předškolních dětí. Podává stručný přehled metod sledování očních pohybů, které jsou důležitou součástí vyšetření zaměřujících se na poruchy čtení. Dále hledá vhodné statistické a klasifikační metody pro analýzu dat naměřených na skupině dyslektických a zdravých dětí v rámci výzkumného projektu č. *TA 01011138: Sledování očních pohybů pro diagnostiku v neurovědách*. Praktická část je zaměřena na hodnocení kvality záznamů očních pohybů, po níž následuje korekce a výběr záznamů relevantních pro statistické vyhodnocení. Ze zpracovaných dat jsou dále statickými a klasifikačními metodami analyzovány hodnoty vybraných příznaků pro dyslektickou a kontrolní skupinu. Statistickou a klasifikační analýzou dat z úlohy *fixační stabilita* jsou nalezeny signifikantní rozdíly mezi skupinami zejména pro příznaky hodnotící počet patologických „odskoků“ způsobených nekontrolovanými pohyby oka vyšetřovaného během úlohy.

## **Klíčová slova**

Dyslexie, oční pohyby, statistická analýza dat, Matlab.

## **Annotation**

Theoretical part of the diploma thesis focuses on dyslexia diagnosis of preschool children. General overview of eye tracking methods, which are important part of examination focused on reading disabilities, follows. In the last theoretical part appropriate statistical and classification methods for data analysis have been found. These data come from research project n. *TA 01011138: Eye tracking for diagnosis in neuroscience*, where dyslectic as well as healthy children were recorded. Practical part of the thesis focuses on evaluation of records quality and selection of relevant records for further statistical analysis. From the processed records values of selected data features are analysed by selected statistical and classification methods. Significant differences between dyslectic and healthy (control) group in *fixation stability* task have been found for features evaluating pathological „leaps“ caused by uncontrolled eye movements during examination.

## **Key words**

Dyslexia, eye movements, statistical data analysis, Matlab.

## **Poděkování**

Touto cestou bych ráda poděkovala vedoucímu mé práce, panu Ing. Vratislavu Fabiánovi, Ph. D. a také panu Ing. Jaromíru Doležalovi, Ph. D. za konzultace a čas, který mi v souvislosti s vytvářením mé diplomové práce věnovali. Dále bych chtěla poděkovat všem vyučujícím, kteří mi během mého studia oboru Biomedicínské inženýrství na Fakultě elektrotechnické ČVUT v Praze předali cenné vědomosti a zkušenosti.

Velký vděk patří také mé rodině a mým blízkým, u kterých jsem vždy našla potřebnou podporu během svého studia.

## **Prohlášení autora práce**

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 11. 5. 2015

.....

Podpis autora práce

# Obsah

1	Úvod.....	1
1.1	Cíle práce.....	1
2	Rešerše .....	3
2.1	Dyslexie.....	3
2.1.1	Okno a oční pohyby .....	3
	Anatomie lidského oka .....	3
	Typy očních pohybů.....	5
2.1.2	Původ dyslexie .....	7
2.1.3	Dyslexie jako porucha mozku .....	8
2.2	Sledování očních pohybů .....	9
2.2.1	Techniky sledování očních pohybů .....	9
2.2.2	Aplikace využití sledování očních pohybů .....	11
2.3	Metody statistické analýzy a klasifikace .....	11
2.3.1	Statistická analýza dat.....	11
2.3.2	Testování dat .....	17
2.3.3	Základní klasifikační techniky dat.....	20
3	Metodika .....	24
3.1	Úvod.....	24
3.2	Získání dat.....	24
3.2.1	Použitá technologie k získání dat .....	24
3.2.2	Popis okulometrických testů .....	25
3.3	Hodnocení kvality záznamů a výběr záznamů .....	26
3.4	Předzpracování záznamů.....	27
3.5	Zpracování záznamů .....	28
3.5.1	Fixační stabilita.....	28
3.6	Statistické vyhodnocení a klasifikace dat .....	30
3.6.1	Průzkumová a popisná analýza dat .....	30
3.6.2	Statistické vyhodnocení - ANOVA.....	34
3.6.3	Použití klasifikační analýzy <i>mean-mean</i> a <i>median-median</i> .....	41
3.6.4	Další metody klasifikace .....	42
4	Výsledky a diskuze .....	45
5	Závěr .....	46
	BIBLIOGRAFIE.....	48
	PŘILOŽENÉ CD .....	52

## Seznam obrázků

Obrázek 2.1: Anatomický popis oční bulvy [17] .....	4
Obrázek 2.2: Svaly oka [16] .....	5
Obrázek 2.3: Princip elektrookulografie [17] .....	9
Obrázek 2.4: I4Tracking® .....	10
Obrázek 2.5: Tepelná mapa jako výsledek sledování webové stránky metodou eye tracking [19] ...	11
Obrázek 2.6: Krabicový graf a jeho význam [20] .....	13
Obrázek 2.7: Histogram [21] .....	14
Obrázek 2.8: Histogram a normální pravděpodobnostní graf (N-P plot) normálního rozložení (vpravo) a rozložení s kladnou šikmostí (vlevo) [26]. .....	16
Obrázek 2.9: Wrapper - Algoritmus má přívlastek indukční díky své schopnosti indukovat obecné závěry z konkrétních příkladů [34] .....	23
Obrázek 3.1: Proces získávání a zpracování dat .....	24
Obrázek 3.2: Rozložení dat podle věku a pohlaví u kontrolní skupiny zdravých dětí .....	24
Obrázek 3.3: Rozložení dat podle věku a pohlaví u skupiny dyslektiků .....	25
Obrázek 3.4: Červeně vyznačené artefakty po korekci. Fyziologické „odskoky“ v záznamu se vyznačují delším trváním. ....	27
Obrázek 3.5: Chvilé, kdy dochází vlivem mrkání k nesprávné detekci pupily. V signálu je vyznačena červená oprava tohoto artefaktu.....	28
Obrázek 3.6: Fixační úloha. Vlevo varianta uniformní, vpravo varianta optokinetická .....	29
Obrázek 3.7: Průběh očních pohybů v úloze „fixace bodu“ [33] .....	29
Obrázek 3.8: Naměřené záznamy uniformní a optokinetické úlohy [33] .....	30
Obrázek 3.9: Krabicový graf znázorňující hodnoty příznaku „Počet odskoků“ pro obě skupiny .....	31
Obrázek 3.10: Histogram znázorňující rozložení hodnot příznaku „Počet odskoků“ pro obě skupiny .....	31
Obrázek 3.11: Krabicový graf a histogram znázorňující hodnoty příznaku „Počet odskoků“ pro obě skupiny .....	32
Obrázek 3.12: Krabicový graf a histogram znázorňující hodnoty příznaku „Průměrná délka sakád“ pro obě skupiny .....	33
Obrázek 3.13: Ukázka výstupu funkce Levenetest.m .....	34
Obrázek 3.14: Normální pravděpodobnostní graf pro příznak „Odchylka od ideálu v první části úlohy“ před (nalevo) a po (napravo) logaritmické transformaci .....	36
Obrázek 3.15: Příznak „Odchylka od ideálu v první části úlohy“ před (nalevo) a po (napravo) logaritmické transformaci.....	37
Obrázek 3.16: Tabulka ANOVA jako výstup funkce anova1 .....	38
Obrázek 3.17: Struktura stats a krabicový diagram jako výstup funkce anova1.....	38
Obrázek 3.18: Tabulka ANOVA [38] .....	39

## Seznam tabulek

Tabulka 2.1: Typické hodnoty nejběžnějších očních pohybů [4] .....	6
Tabulka 2.2: Klasifikace .....	21
Tabulka 3.1: Základní statistiky pro příznak Počet odskoků .....	33
Tabulka 3.2: Základní statistiky pro příznak Průměrná délka sakád .....	34
Tabulka 3.3: Příznaky vykazující normální rozdělení před a po transformaci .....	35
Tabulka 3.4: ANOVA pro příznaky vykazující normální rozdělení .....	40
Tabulka 3.5: ANOVA pro příznaky vykazující exponenciální rozdělení .....	40
Tabulka 3.6: Matice záměn confmat .....	43
Tabulka 3.7: Tabulka: Výsledky klasifikace jednotlivých klasifikátorů .....	43

## Seznam příloh

Příloha 1: Seznam příznaků s hodnotami statistik průzkumové analýzy

Příloha 2: Krabicové diagramy a histogramy příznaků

Příloha 3: Histogramy příznaků před a po logaritmické transformaci

Příloha 4: Výsledky klasifikátorů *median-median* a *mean-mean*



# 1 Úvod

Dyslexie je vrozenou či získanou poruchou mozku, která pro mnoho lidí na celém světě představuje vážnou komplikaci v jejich běžném životě. Proto je velmi důležité tuto poruchu včas odhalit a její možné důsledky kompenzovat. Stěžejní metodou při vyšetřování dyslektického postižení se v posledních letech stává hodnocení očních pohybů. Vzhledem k tomu, že hodnocení očních pohybů nemusí být zaměřeno pouze na oblast čtení, je možné vyšetřovat i děti, které ještě neumí číst, tedy děti předškolního věku. Pokud by bylo možné dyslexii v tomto věku diagnostikovat, může se dítěti dostat správné odborné péče, která napomůže předejít negativním psychickým důsledkům, které tato porucha přináší. Vyšetření očních pohybů by mohlo významně doplnit standardní vyšetření prováděné v pedagogicko-psychologických poradnách a speciálních pedagogických centrech, a mohlo by tak přinést přesnější diagnostiku dyslexie.

Naměřená data, která jsou v tuto chvíli v rámci výzkumného projektu č. *TA 01011138: Sledování očních pohybů pro diagnostiku v neurověděch* k dispozici, je nutné zpracovat a analyzovat. Z naměřených dat budou dle známé metodiky vypočítány hodnoty příznaků. Ve své práci budu hledat signifikantní rozdíly v hodnotách těchto příznaků mezi dyslektickými dětmi a dětmi v kontrolní skupině. Výsledkem bude doporučení, které příznaky jsou pro vyšetření dyslexie u dětí signifikantní.

## 1.1 Cíle práce

Úvodním cílem diplomové práce je provést rešerši v oblasti diagnózy dyslexie. V tomto ohledu je důležité získat zejména bližší povědomí o původu dyslexie. Rešerše se bude dále věnovat souvislosti dyslexie s očními pohyby. Na otázku očních pohybů naváže rešerše v oblasti eye-trackingu, přičemž budou zmíněny zejména nejčastější techniky měření očních pohybů. Poslední částí rešerše je vyhledání a popsání vhodných statistických a klasifikačních metod, které budou v praktické části práce aplikovány na zpracovaná data.

Druhým cílem práce je provedení výběru reprezentativní skupiny dat pro další zpracování a korekce naměřených dat. Výběr dat bude zahrnovat hodnocení kvality všech záznamů a výběr relevantních záznamů pro skupinu dyslektických dětí a pro kontrolní skupinu. Předzpracování bude zahrnovat ruční korekci signálů od artefaktů za pomoci dostupných metod.

Posledním cílem práce je návrh a implementace vhodných metod pro statistické vyhodnocení naměřených dat. Zpracovaná data budou podrobena průzkumové analýze za použití základních grafických metod, dále popisné analýze, kde budou použity základní statistické charakteristiky. Před další analýzou budou data podrobena testování na normální rozdělení, a pokud bude nutné, budou data transformována. Následně budou data analyzována vhodnou vybranou metodou a výsledky této analýzy budou diskutovány. Ke statistickému zpracování dat bude náležet také část klasifikační, kde budou navrženy základní metody pro klasifikaci dat.

## 2 Rešerše

### 2.1 Dyslexie

Pojem dyslexie byl v minulosti v odborné literatuře definován různými způsoby. Český psychiatr a neurolog A. Heveroch již v roce 1904 definoval dyslexii ve svém článku „O jednostranné neschopnosti naučiti se čísti při znamenité řečové paměti“, který vyšel v časopise Česká škola, přičemž vyslovil domněnku, že neschopnost naučit se číst a psát má příčinu ve změnách v mozku, konkrétně v mozkové kůře. Dále v roce 1960 světově uznávaní psychologové J. Langmeier a Z. Matějček definovali vývojovou dyslexii jako „specifický defekt čtení, podmíněný nedostatkem některých primárních schopností, jež skládají komplexní schopnost pro učení za dané výukové metody“ [2], dále zmiňují, že se obvykle objevuje u dětí od samých počátků výuky. O bližší definici dyslexie se v minulosti dále zaslouhovali i další M. B. Rawsonová ve své publikaci Development Language Disability nebo M. Rutter a W. Yule, kteří pojem dyslexie nahradili pojmem „specifická retardace ve čtení“.

[2]

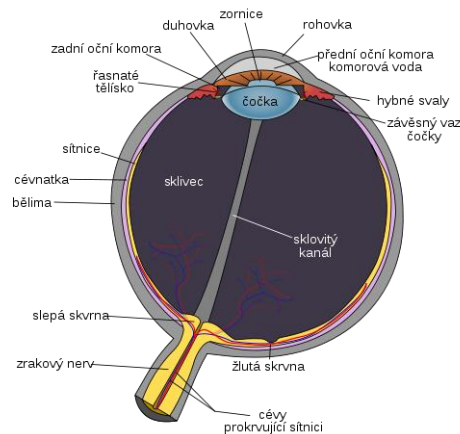
V posledních dvaceti letech se pojem dyslexie dostává stále na vyšší úroveň. V roce 2003 ho G. Reid Lyon ve své publikaci [11] popsal jako specifickou poruchu učení, která je neurobiologického původu. Je charakterizována potížemi s přesným a plynulým rozpoznáváním slov a slabou schopností hláskovat. Tato definice je považována za platnou i v dnešní době.

#### 2.1.1 Oko a oční pohyby

##### **Anatomie lidského oka**

Lidské oko propouští světlo dovnitř skrz zornici, pomocí čočky převrátí obraz a promítá ho na zadní část oka - sítnici. Sítnice je vyplněna světlo-citlivými buňkami zvanými čípkami a tyčinkami, které převádí přicházející světlo na elektrické signály, které dále postupují po optickém nervu do korového zrakového centra, kde dochází k dalšímu zpracování. Čípkami jsou citlivé na vysoké prostorové frekvence (zrakové detaily) a umožňují nám barevné vidění. Tyčinky jsou velmi citlivé na světlo, a tak podporují vidění za šera.

V zadní části oka se nachází místo zvané *fovea centralis*. Fovea se nachází uprostřed žluté skvrny (Obrázek 2.1). V této malé oblasti s rozpětím menším než 2° zorného pole se nachází extrémní množství čípků, na rozdíl od periferie sítnice, kde se čípkami vyskytují jen řídce. Důsledkem výše zmíněného je, že plnou ostrost vidění zajišťuje pouze tato malá oblast, na kterou musí světlo ze sledovaného ostrého objektu přímo dopadat, chceme-li zajistit ostré vidění.



Obrázek 2.1: Anatomický popis oční bulvy [17]

Pro měření založených na pořizování video záznamů očních pohybů je velmi podstatná již zmíněná zornice neboli *pupila*. Další důležitou částí je rohovka. Rohovka kryje oko z vnějšku a odráží světlo. Při snímání oka je obvykle požadován pouze jeden odraz, používá se nejčastěji infračervené záření pořízené z jednoho nebo více zdrojů. Výsledný odraz od rohovky je známý jako první Purkyňův odraz a je nejjasnější, světlo se dále odráží i od čočky, tentokrát však s menším jasnem. [1]

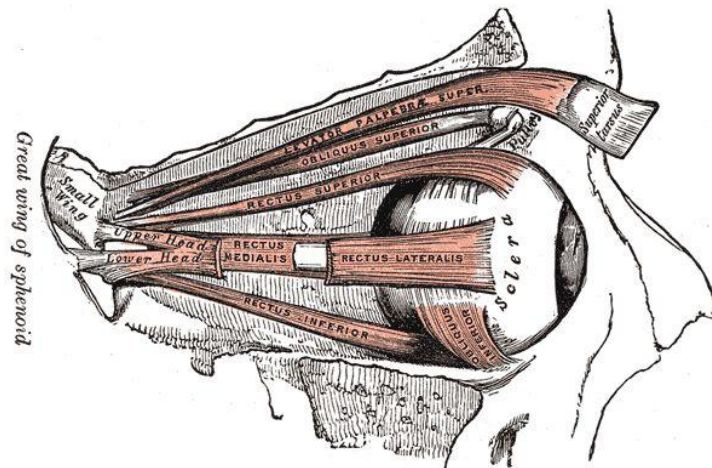
### Řízení očních pohybů

Oční pohyby umožňují zrakovému ústrojí získávat informace pozorováním důležitých aspektů v okolním prostředí. Rozpoznávání objektů, rozlišení a další informace přijaté zrakovým ústrojím jsou uskutečněny podvědomými očními pohyby. K nejdůležitější části zpracování nové informace dochází ve chvíli, kdy oči vykonávají krátké pauzy. [1]

Oční pohyby jsou řízeny kortikální a subkortikální částí mozku ve spolupráci s hlavovými nervy a očními svaly napojenými na vnější stranu obou očních bulv. Oční fixace je řízena čelní (frontální) a týlní (okcipitální) částí kůry mozku. Řízení očních pohybů je přisuzováno čelní části mozkové kůry, konkrétní zásluhu má Brodmannova oblast 8. [1]

Důležitost používání očních pohybů jako objektivní prostředek pro studium procesu čtení je podložena faktem, že oční pohyby jsou vykonávány podvědomě. Pohyby očí jsou nejrychlejšími a nejčastějšími pohyby, které vykonáváme, a to i během spánku a dokonce v kómatu [3]. Pohyblivost oka v různých směrech je dána 6 okohybnými svaly (Obrázek 2.2). *Musculus rectus superior* a *inferior* pohybují bulbem ve vertikální rovině nahoru,

respektive dolů. Musculus lateralis a medialis pohybují okem laterálně, respektive mediálně v horizontální rovině. Musculus obliquus inferior a superior pohybují očním bulbem šikmo nahoru, respektive dolů [16]. Pohyby očí jsou navzájem propojeny (konjugovány) a jsou spárovány. Při pohybu týmž směrem jde o konjugované pohyby, při pohybu v protisměru (divergence, konvergence) jde o disjugované pohyby. [2]



Obrázek 2.2: Svaly oka [16]

### Typy očních pohybů

Oči se projevují různými očními pohyby; při sledování očních pohybů však převládají okamžiky, kdy zůstávají oči v klidu, lépe řečeno jsou fixovány. Pojem „fixace“ je však poněkud zavádějící, jelikož oko není nikdy úplně nehybné, ale vykazuje tři typy mikropohybů: tremor (třes, chvění, pohyb o frekvenci přibližně 90 Hz), drift (pomalý oční pohyb unášející oko ze středu fixace) a mikrosakády (rychlé pohyby navracející oko do jeho původní polohy) [4]. Dále jsou popsány základní oční pohyby, které se nejčastěji sledují.

#### Sakády

Rychlé kontrolované pohyby přemísťující oko z jedné fixace na druhou. Jejich funkcí je zachytit rychle se pohybující objekt a jeho obraz promítnout na foveu, místo nejostřejšího vidění na sítnici oka. Sakády mají jedinečný rys: invariantní poměr mezi maximální rychlostí a velikostí (amplitudou). Tento poměr se nazývá „hlavní sekvence“ a bývá užíván jakožto identifikační znak sakád u dosud neznámého očního pohybu. Sakadický systém kontroluje výsledek své činnosti, tj. dosažení objektu a jeho *foveaci* a pokud nalezne chybu, tj. objekt není *foveován*, vypočte novou tzv. korektivní sakádu. U velkých sakád (amplituda  $>15^\circ$ ) je běžné tzv. podstřelení či hypometrie. U malých sakád je naopak pozorováno příležitostně tzv. přestřelení či hypermetrie. Sakadický systém je schopen se učit či se přizpůsobovat změnám ve vizuálním poli. Tento adaptivní proces přitom probíhá nevědomě.

### *Plynulé sledovací oční pohyby*

Ve srovnání se sakádami jsou plynulé sledovací oční pohyby určeny pro hladké, kontinuální pozorování pohybujících se objektů. Aby tuto úlohu mohly plnit, pracují nikoli v režimu vzorkovacím jako sakády, nýbrž v režimu kontinuálním: nepřetržitě porovnávají rychlost pohybu oka s rychlostí objektu a zajišťují jejich rovnováhu, či minimalizují tzv. rychlostní chybu (neúnosný rozdíl v obou rychlostech). Tomu je podřízen druh podnětů, na který plynulé sledovací oční pohyby reagují, i reakční čas. Plynulé sledovací oční pohyby ve srovnání se sakádami reagují na podnět rychleji, v [41] je udána normální doba latence 130 ms oproti 200 ms u sakád. Podobně jako u sakád i u plynulých sledovacích očních pohybů je pozorována adaptační schopnost. Jsou-li plynulé sledovací oční pohyby vystaveny tréninku, např. denně v průběhu jednoho týdne sledovat kyvadlo, jejich schopnost hladce a plynule sledovat objekt narůstá [40]. Pravděpodobně každá funkční rodina navozuje – nezáměrně a zcela přirozeně – takovéto tréninkové situace u svého dítěte a zlepšuje tak kvalitu jeho okulomotorického systému.

### *Vestibulární a optokinetické oční pohyby*

Oba druhy očních pohybů patří k reflexům a jsou řízeny z nižších úrovní mozku, na rozdíl od plynulých sledovacích očních pohybů či dokonce sakád, které můžeme ovlivňovat svou vůlí. Protože jde o reflexní reakce, je latence (reakční doba) velmi nízká – kolem 10 – 15 ms [41] oproti komplikovaně programovaným sakádám, jejichž latence činí cca 200 ms. Je zajímavé, že u kongenitálně slepých bývá vestibulo-okulární reflex oslaben či dokonce absentuje [40]. Svědčí to o tom, že i relativně jednoduché oční pohyby mohou být formovány učením a dokonce pro svůj normální vývoj učení potřebují.[4]

Hodnoty parametrů nejběžnějších očních pohybů viz Tabulka 2.1.

Druh pohybu	Trvání [ms]	Amplituda	Rychlost
Fixace	200 – 300	-	-
Sakády	30 – 80	4 – 20°	30 – 500°/s
Plynulé oční pohyby	-	-	10 – 30°/s
Mikrosakády	10 – 30	10 – 40'	15 – 50°/s
Tremor	-	< 1'	20'/s
Drift	200 – 1000	1 – 60'	6 – 25'/s

Tabulka 2.1: Typické hodnoty nejběžnějších očních pohybů [4]

Mnoho dalších aktuálních informací s tematikou očních pohybů lze nalézt na [35].

### 2.1.2 Původ dyslexie

Pokud chceme dyslexii u dětí umět diagnostikovat a kompenzovat již v raném stadiu, je nutné se podrobněji zabývat otázkou původu této nemoci. Velké poznatky v oblasti očních pohybů a dyslexie přinesl G. Pavlidis, který ve své publikaci [12] vyvrací hypotézu, že nekontrolované pohyby očí u špatných čtenářů jsou důsledkem potíží při učení vizuálně-slovních asociací; naopak se kloní k názoru, že tyto pohyby jsou primární příčinou špatného čtení. Porovnává dyslexii s dalšími poruchami čtení, přičemž poukazuje na fakt, že většina těchto poruch může být předpověděna na základě neurologických, inteligenčních, socio-ekonomických, vzdělávacích a psychologických (motivačních, emočních) faktorů, o nichž je z dřívějších výzkumů [42] známo, že ovlivňují proces čtení. U dítěte, které vykazuje problémy v jedné či více z těchto oblastí, můžeme tedy předpokládat problémy se čtením. Naopak dítě, které tyto problémy nevykazuje, budeme považovat za normálního čtenáře. Dle [42] může být dítě považováno za dyslektika pouze tehdy, pokud jeho porucha čtení nemůže být predikována nedostatkem v některém ze zmíněných faktorů. Tyto faktory, ačkoli se nejedná o příčiny dyslexie, mohou mít vliv na její závažnost i zlepšení. Příčiny dyslexie jsou vrozené [12].

Dovednosti čtení se rozvíjejí postupně, přičemž se během let zlepšuje jejich přesnost a rychlost. Jejich vývoj probíhá paralelně s vývojem očních pohybů. Velká část tohoto vývoje probíhá během prvních tří až čtyř let, do deseti let věku dítěte dojde vývoj očních pohybů přibližně do dvou třetin. Je dokázáno, že existuje inverzní vztah mezi věkem a délkou fixace oka a mezi dopřednými a zpětnými očními pohyby. Kratší fixace je známkou rychlejšího zpracování informace.[3]

Existují studie, které sledují bludné (nekontrolované) oční pohyby dyslektických dětí při čtení většina jejich autorů (L. C. Gilbert a další) se shoduje na faktu, že mnoho dyslektických dětí vykazuje při čtení bludné oční pohyby, nesouhlasí však s podstatou a rozsahem souvislostí mezi bludnými očními pohyby a dyslexií. Pavlidis [9] na základě dříve provedených studií definuje 3 skupiny hypotéz objasňujících souvislosti mezi očními pohyby a dyslexií: 1) bludné oční pohyby jsou pouze dalším projevem problémů, které mají dyslektici s materiálem, který čtou; 2) bludné oční pohyby mohou někdy způsobovat dyslexii; 3) bludné oční pohyby a dyslexie jsou příznaky poruch centrálního nervového systému. Ve své studii [9] Pavlidis dále dokazuje, že dyslektické děti na rozdíl od normálních a pokročilých čtenářů a dětí s jinými poruchami čtení vykazují při čtení textu nadměrný počet zpětných očních pohybů a fixací, a to i v případě, kdy má text pro dyslektické děti nižší úroveň složitosti.

Nebylo dosud prokázáno, že by dyslexie mohla být přisuzována nízkému I. Q., emocionálním, vzdělávacím nebo sociálním faktorům. Naopak se ukázalo, že dyslexii způsobují vrozené vývojové vady [43]. Na této skutečnosti položil Pavlidis [9] svou další hypotézu, a sice že pokud je dyslexie způsobována špatnou funkcí mozku, pak by se měla tato nedostatečnost projevovat nejen při čtení, ale i při provádění dalších úloh, které jsou s procesem čtení spojeny a které jsou zpracovány stejnými částmi mozku. Pavlidis proto provedl další studii se stejným složením skupin dětí – normální a pokročilí čtenáři, dyslektici a děti s jinou poruchou čtení. Aplikováním úlohy založené na sledování světelných podnětů dokázal, že počet zpětných pohybů u dyslektiků je opět signifikantně větší než u dětí s jinými poruchami čtení, jejichž hodnoty se lišily pouze nevýznamně od hodnot normálních a pokročilých čtenářů. Výsledky této studie se tedy kloní k závěru, že dyslektici mají na rozdíl od dětí s jinými poruchami primární problém nezávislý na čtení a můžeme je od výše zmíněných skupin dětí odlišit i prostřednictvím úloh, které nejsou zaměřeny na čtení a u kterých lze analyzovat charakteristické oční pohyby.

Očními pohyby u dyslektiků se zabývá také [13], kde autoři zmiňují, že při výzkumu očních pohybů u dyslektiků se pravidelně nachází skupina dětí s patologickými očními pohyby – nystagmus<sup>1</sup>, opsoklonus<sup>2</sup>. U těchto subjektů je pravděpodobné, že porucha očních pohybů (způsobená u nich nejčastěji strukturální mozkovou lézí) může být přímo zodpovědná za poruchy čtení.

### 2.1.3 Dyslexie jako porucha mozku

O příčinách dyslexie více hovoří studie [13], v níž autoři uvádějí tři typy příčin dyslexie – fonologickou, magnocelulární a cerebelární. Fonologická teorie vysvětluje dyslexii v rovině neuropsychologické a za základní příčinu považuje abnormální mozkové zpracování řečové informace na různých úrovních. Mozečková (cerebelární) teorie vysvětluje poruchu čtení postižením mozečku, od kterého odvozuje jak fonologický sensorický deficit, tak poruchu motorické koordinace projevující se i při hlasitém čtení. [13] Protože dyslektik obtížně rozpoznává fonémy ve zvukových celcích slov, není schopen se orientovat ani v grafém-fonémové korespondenci a zdeformovaným fonologickým klíčem pak nemůže odemknout vstup k významu, který je zakódován v grafické podobě slova [14]. Magnocelulární teorie spatřuje příčinu dyslexie v poruše magnocelulárního systému CNS. [13] Magnocelulární systém je přednostně (či takřka výhradně) angažován v těch činnostech, které jsou slabinou dyslektiků. Jelikož jsou oba subsystemy odděleny už na sítnici a jejich oddělené vedení pokračuje asociačními oblastmi kůry mozkové, může porucha vznikat

---

<sup>1</sup> Nystagmus - je rytmický konjugovaný kmitavý pohyb očních bulbů [44]

<sup>2</sup> Opsoklonus – mimovolní svalový záškub očních svalů [45]



kdekoli na této dráze, takže je těžko definovatelná psychologickými testy. K jejich zjišťování se užívají metody fyziologické. Dle studie M. Livingstonové (1991) je magnetocelulární systém dyslektiků o poznání více „desorganizovaný“ než u ne-dyslektiků, na rozdíl od parvocelulárního systému, u kterého mezi dyslektiky a ne-dyslektiky není rozdíl. Další možnou příčinou výskytu dyslexie, příčinou genetickou, se zabýval profesor Galaburda. Ten se domnívá, že anomálie, které našel v buněčných vrstvách mozkové kůry, nemohly vzniknout poškozením mozkové tkáně, ale vytvořily se velmi pravděpodobně už v prvních týdnech a měsících vývoje plodu.[2] To potvrzuje teorii o vrozené příčině vzniku dyslexie.

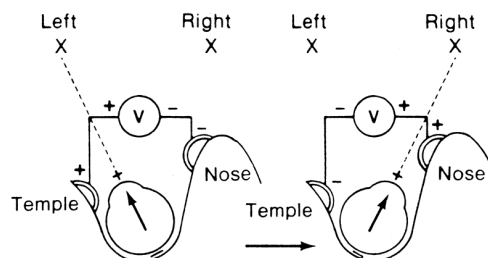
## 2.2 Sledování očních pohybů

### 2.2.1 Techniky sledování očních pohybů

Měřicí zařízení nejčastěji používané k měření očních pohybů je známé jako *eye-tracker* (z angl. *eye* – oko, *to track* – sledovat). Obecně existují dva způsoby monitorování očních pohybů: ty, které měří polohu oka vůči hlavě a ty, které měří orientaci oka v prostoru, či polohu bodu, na které měřená osoba upírá svou pozornost. Posledně jmenované měření je typicky využíváno pro identifikaci prvků v zorném poli, např. v (grafických) interaktivních aplikacích. Asi nejrozšířenějším zařízením pro měření bodu pozornosti je *eye-tracker* založený na detekci odrazu od rohovky na videozáznamu.

Existují čtyři obecné kategorie metod měření očních pohybů: elektrookulografie (EOG), sklerální<sup>3</sup> kontaktní čočka s měřicí cívkou, fotookulografie (POG) nebo videookulografie (VOG), a kombinovaný odraz rohovky a detekce středu zornice na videozáznamu.

*Elektrookulografie* spočívá v zaznamenávání rozdílů elektrických potenciálů na kůži obklopující oční jamku (Obrázek 2.3). Snímá potenciály 15 – 200  $\mu\text{V}$  s citlivostí 20  $\mu\text{V}$ /stupeň pohybu oka. Tato technika měří oční pohyby relativně k pozici hlavy, a proto není vhodná pro měření bodu pozornosti v případě, že poloha hlavy není měřena také.



Obrázek 2.3: Princip elektrookulografie [17]

<sup>3</sup> Skléra (bělma) – neprůhledná, vnější vrstva oka [31]

*Sklerální kontaktní čočka s měřicí cívkou* je velmi přesnou metodou, která zahrnuje využití optického nebo mechanického referenčního objektu připevněného na kontaktní čočku, která je poté vpravena přímo do oka. Proto je považována za nejvíce invazivní a ne příliš komfortní metodu. Tato technika taktéž měří oční pohyby relativně k pozici hlavy, a proto není vhodná pro měření bodu pozornosti.

*Fotookulografie (POG) nebo videookulografie (VOG)* dohromady reprezentuje širokou škálu technik nahrávání očních pohybů zahrnující měření různých příznaků během rotace/translace jako jsou sledování tvaru zornice či pozice limbu<sup>4</sup> nebo sledování odrazů přímého blízko umístěného (nejčastěji infračerveného) zdroje světla od rohovky.

Aby mohlo být sledováno místo, kam měřená osoba upírá svou pozornost, aniž by byla hlava fixována, je nutné měřit více prvků. Těmito prvky jsou odraz rohovky a střed zornice. Odraz od rohovky (typicky infračerveného světla) je měřen relativně k umístění středu zornice. [5]

### **Technologie ke sledování očních pohybů – I4Tracking®**

Pro sledování očních pohybů bylo v posledních letech vyvinuto mnoho zařízení. Jedním z nich je systém pro sledování očních pohybů založený na videookulografické metodě vyvíjené společností Medicton Group (Obrázek 2.4). Základem této metody je kamera, která snímá pohyby oka. V obraze z kamery je detekována zornice infračervenými odlesky odraženými od oka. Na základě těchto informací a provedené kalibrace je poté vyhodnocen směr pohledu měřené osoby.



Obrázek 2.4: I4Tracking®

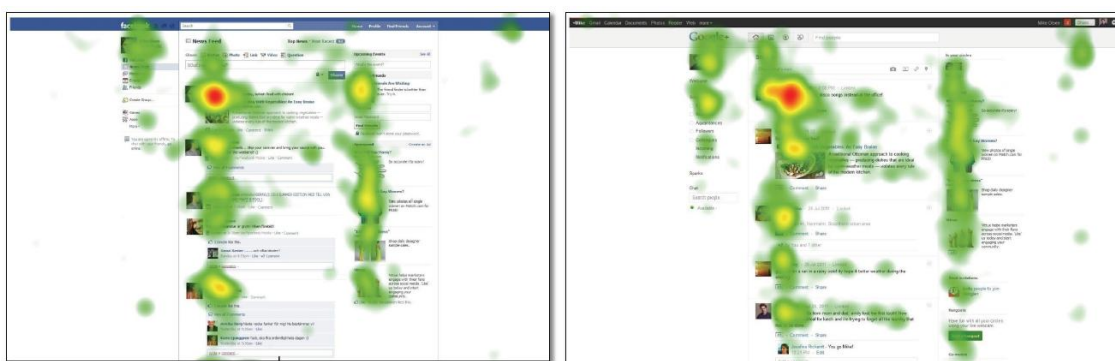
---

<sup>4</sup> Limbus – rozhraní rohovky a bělimy.

## 2.2.2 Aplikace využití sledování očních pohybů

Eye tracking má ve světě mnoho aplikací, ty je možné obecně rozdělit do dvou skupin, a sice diagnostické a interaktivní. V rámci diagnostických aplikací poskytuje *eye-tracker* objektivní a kvantitativní informace o vizuálních procesech.

Využití sledování očních pohybů v psychologii a neurovědě bylo již nastíněno dříve v kapitole 2.1, kde se sledování očních pohybů diskutuje v souvislosti s dyslecií. Tato diplomová práce se právě takovou aplikací v oblasti psychologie zabývá. *Eye-tracking* se však využívá i v dalších oblastech, které dále stručně zmiňuji. V oblasti inženýrství lze najít aplikace například v letectví či v automobilovém průmyslu. Velké uplatnění nachází *eye-tracking* oblasti marketingu a reklamy, kde je hlavní motivací v oblasti marketingových výzkumů pochopení chování a jednání zákazníků. Pomocí *eye-trackingu* se hodnotí přehlednost webových stránek, ale i vzhled produktů (Obrázek 2.5). V počítačových vědách se *eye-tracking* využívá k úlohám, při kterých dochází k interakci člověka s počítačem prostřednictvím směru jeho pohledu (tedy prostřednictvím očních pohybů). V této oblasti existuje mnoho moderních aplikací, které stále procházejí vývojem. Více o aplikacích *eye-trackingu* napříč všemi zmíněnými oblastmi možno najít v [5].



Obrázek 2.5: Tepelná mapa jako výsledek sledování webové stránky metodou eye tracking [19]

## 2.3 Metody statistické analýzy a klasifikace

### 2.3.1 Statistická analýza dat

Statistická analýza dat se typicky skládá ze tří hlavních kroků: průzkum dat, popis dat a analýza dat.

#### 2.3.1.1 Popisná a průzkumová analýza dat

Ačkoli je průzkumová analýza dat často ve studiích opomíjena, jedná se o velmi důležitou část analýzy. Jejím účelem je bližší seznámení se s daty, což umožňuje lepší rozhodování o krocích v dalších částech analýzy. Dalším důležitým bodem průzkumové analýzy je kontrola výskytu chybných hodnot v datech. Chyby mohou být zapříčiněny chybným

kódováním dat či nesprávným měřením během experimentu. Postoupení dat k další analýze bez kontroly chyb může vést k devastujícím následkům, které mohou přinést neexistující signifikantní výsledky, nebo naopak mohou správné signifikantní výsledky skrýt.

Prvním cílem průzkumu dat je kontrola, zda je *kvalita dat dostatečná* pro další vyhodnocení. Dalším hlavním cílem je *určení rozdělení pravděpodobnosti jednotlivých proměnných*. Obvyklým požadavkem pro statistické testy je normalita rozdělení. Data z mnoha *eye-tracking* měření však normální rozdělení nevykazují. Měření, zahrnující měření doby fixace či většinu sakadických měření, mívají často zešikmené rozdělení, kdy je jeden konec histogramu užší než druhý. Takto zešikmená rozdělení lze na normální převést pomocí transformace. Možnou transformací je převedení hodnot pomocí logaritmu. Pozitivně zešikmené rozdělení tak získá více podobu normálního rozdělení. Dalším a robustnějším řešením pro zešikmené rozdělení je použití metod vyvinutých pro rozdělení gamma.[4]

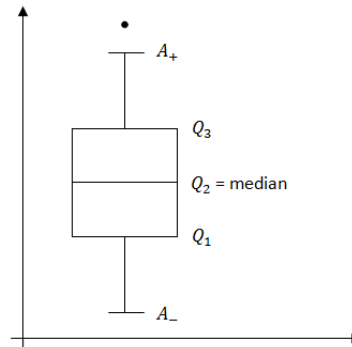
Dalším bodem průzkumové analýzy dat je identifikace odlehlých hodnot, tzv. *outlierů*. Outliery mohou být důsledkem chyb při nahrávání dat nebo při nesprávné detekci události, ale mohou být i skutečnými naměřenými hodnotami, proto je nutné s outliery pracovat velmi opatrně. Pokud se jedná o chybné hodnoty, je tyto nutné opravit či odstranit. Pokud se jedná o naměřené hodnoty, je potřeba zvážit, zda tyto hodnoty zachovat nebo také odstranit.

Nezbytnou součástí dalších kroků průzkumové analýzy je vizuální znázornění dat. Existuje mnoho nástrojů, pomocí kterých vykreslování dat obvykle uskutečňujeme, dále si uvedeme ty nejběžnější z nich. [4]

#### Boxplot (Krabicový graf)

Boxplot neboli krabicový graf je užitečným nástrojem, který poskytuje souběžně informaci o rozložení dat, ale také informaci o potenciálních *outlierech*. V deskriptivní statistice je krabicový graf jedním ze způsobů grafické vizualizace numerických dat pomocí jejich kvartilů. Střední "krabicová" část diagramu je shora ohraničena 3. kvartilem, zesponu 1. kvartilem a mezi nimi se nachází linie vymežující medián (Obrázek 2.6). Boxploty mohou obsahovat také linie vycházející ze střední části diagramu kolmo nahoru a dolů, tzv. *vousy*, vyjadřující rozptyl dat pod prvním a nad třetím kvartilem. Odlehlé hodnoty, tzv. outliery, pak mohou být vykresleny jako jednotlivé body.

Boxploty zobrazují rozdíly mezi datovými soubory bez jakýchkoli předpokladů normálního rozdělení dat, jsou tedy neparametrické. Rozteče mezi jednotlivými prvky střední části diagramu indikují stupeň disperze (rozptylu) a šikmosti dat.[20]



Obr. 7

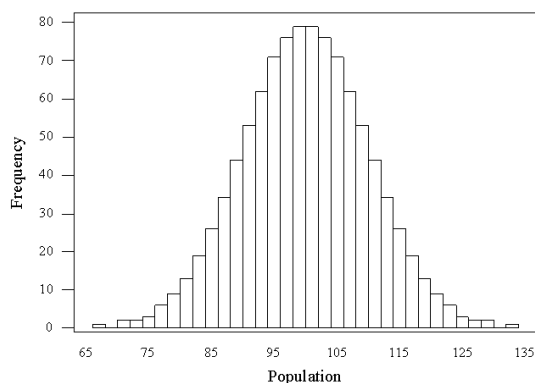
Obrázek 2.6: Krabicový graf a jeho význam [20]

### Histogram

Dalším významným bodem průzkumové analýzy je vykreslení dat pomocí histogramu (Obrázek 2.7), kde na vodorovné ose vidíme hodnoty příznaku v populaci, na svislé ose frekvence výskytu dané hodnoty. Histogram slouží jako nástroj pro grafický popis souboru dat. Vizualně vyjadřuje rozložení hodnot, odhaluje významné skoky, poskytuje informaci o relativních výskytech (frekvencích) pozorování [6]. Histogramy zobrazují absolutní nebo relativní četnost výskytu pozorování. Pokud chceme, aby byl histogram odhadem hustoty rozdělení pravděpodobnosti, je nutné použít histogram s relativními četnostmi. Poté hledáme funkci:

$$\int \hat{f}(x) dx = 1 \quad (2.1)$$

Musíme dále určit počáteční hodnotu pro sloupce  $t_0$  a šířku sloupců  $h$ . Pomocí těchto parametrů je histogram konstruován. Šířka sloupce ovlivňuje hladkost histogramu. [6] Histogram je důležitou součástí průzkumové analýzy dat, jelikož nám podává základní informaci o rozdělení pravděpodobnosti dané veličiny.



Obrázek 2.7: Histogram [21]

Dalším užitečným nástrojem pro vykreslení dat je korelační diagram nebo též bodový graf (*scatterplot*). Použitím výše zmíněných vykreslení lze při průzkumové analýze určit přibližné rozdělení dané populace, či odhalit potenciálně odchýlená měření, která lze následně vyloučit z další analýzy.

### 2.3.1.2 Popisná analýza dat

Popis dat zahrnuje vyhodnocení souhrnných statistik (střední hodnota, medián, maximální a minimální hodnota, směrodatná odchylka, atd.) pro stručné představení výsledků studie. Aby tyto statistiky mohly být vyhodnoceny, musí být data formátována do podoby vhodné pro zpracování zvoleným softwarovým prostředkem. Volba vhodných statistik záleží na rozsahu proměnných. Dále uvádím některé významné statistiky a způsob jejich výpočtu. [4]

#### Střední hodnota

Má-li náhodná veličina  $X$  diskrétní rozdělení, kde  $P[X = s_i] = p_i$  pro  $i \in I$  nejvýše spočetnou množinu různých výsledků, pak [24]:

$$EX = \sum_I s_i p_i \quad (2.2)$$

Odhad střední hodnoty z naměřených dat nejčastěji počítáme z aritmetického průměru:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.3)$$

## Rozptyl

Jedná se o charakteristiku variability rozdělení pravděpodobnosti náhodné veličiny, která vyjadřuje variabilitu rozdělení souboru náhodných hodnot kolem její střední hodnoty [23].

$$\sigma^2 = E(X^2) - [E(X)]^2 \quad (2.4)$$

Odhadem rozptylu je výběrový rozptyl dat a může být vypočítán jako:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (2.5)$$

## Směrodatná odchylka

Směrodatnou odchylku lze vypočítat pomocí střední hodnoty následovně:

$$\sigma = \sqrt{E(X^2) - (E(X))^2} \quad (2.6)$$

Odhadem směrodatné odchylky je výběrová směrodatná odchylka těchto dat a může být vypočítána jako [22]:

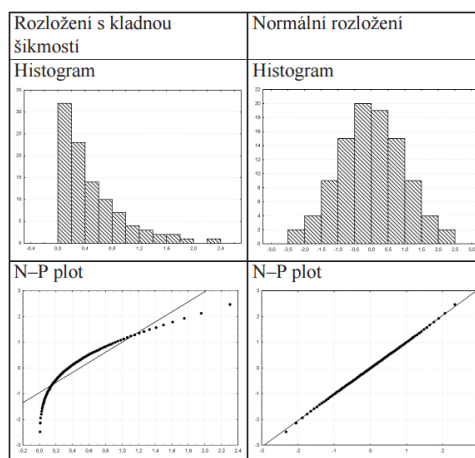
$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2.7)$$

### **2.3.1.3 Testy normality a transformace dat**

#### **2.3.1.3.1 Ověření normality dat**

Postupy statistického hodnocení se liší především podle toho, jaké znalosti máme o typu rozdělení sledované náhodné veličiny v základním souboru. Proto je nutné provést jako jeden z prvních kroků při statistickém testování tzv. test normality, tj. zjištění, zda soubor dat sledované náhodné veličiny odpovídá Gaussovu normálnímu rozdělení pravděpodobnosti, či nikoli (v tomto případě pak pracujeme s neznámým rozdělením). [25] Vizualní ověření normality se provádí například pomocí již zmíněného histogramu nebo normálního pravděpodobnostního grafu, jenž se konstruuje následovně: na vodorovnou osu jsou vynášeny uspořádané hodnoty  $x(1) \leq \dots \leq x(n)$  a na svislou osu kvantily  $\mu_{\alpha_j}$ ,

kde  $\alpha_j = 3j-1/3n+1$ , přičemž  $j$  je pořadí  $j$ -té uspořádané hodnoty. Pocházejí-li data z normálního rozdělení, pak všechny dvojice  $x(j)$  a  $\mu_{\alpha_j}$  budou ležet na přímce.[26] (Obrázek 2.8).



Obrázek 2.8: Histogram a normální pravděpodobnostní graf (N-P plot) normálního rozložení (vpravo) a rozložení s kladnou šikmostí (vlevo) [26].

### Lillieforsův Test

Lillieforsův test je modifikací Kolmogorova-Smirnovova testu. Kolmogorovův-Smirnovův test je vhodný v situaci, kdy jsou parametry distribuce, na kterou hypotézu provádíme, zcela známy. Nicméně někdy je těžké dopředu kompletně specifikovat parametry, když je rozdělení neznámé. V tomto případě musejí být parametry odhadnuty na základě dat.

Máme-li výběr o  $n$  pozorování, Lilliefors statistika je definována [27] jako:

$$D = \max_x |F^*(X) - S_n(X)| \quad (2.8)$$

Kde  $S_n(X)$  je distribuční funkce výběru a  $F^*(X)$  je (kumulativní) distribuční funkce s  $\mu = \bar{X}$ , výběrovým průměrem a  $s^2$ , výběrovým rozptylem, definovaná s denominátorem  $n - 1$ . [27]

#### 2.3.1.3.2 Transformace dat

##### Logaritmická transformace

Některé proměnné nebývají normálně rozložené, a tak nenaplnují předpoklady pro parametrické statistické testy. Použití takových testů jako jsou t-test, ANOVA nebo



lineární regrese proto může přinést mylné výsledky. V některých případech je možné data transformovat, aby lépe odpovídala předpokladům. Mnoho medicínských dat vykazuje logaritmicke-normální rozdělení, taková data lze na normální rozdělení převést pomocí logaritmicke transformace. Logaritmicke transformace je široce používaná metoda pro úpravu vychýlených dat v biomedicínských a psychologických výzkumech. [28]

Logaritmicke transformace je logaritmus každého pozorování. Můžeme použít logaritmus o základu 10 nebo přirozený logaritmus. Ve statistických testech mezi nimi není rozdíl, jelikož se liší pouze konstantou. Při hodnocení dat v závěru je nutné uvést typ logaritmu, který byl při transformaci použit. Zpětná transformace se provádí umocněním dat číslem 10 nebo  $e$ . Problémem, se kterým se při použití logaritmicke transformace můžeme setkat, je výskyt nulových či záporných hodnot v datech. Vzhledem k tomu, že logaritmus pro tyto hodnoty není definován, je nutné tyto hodnoty upravit – konvencí je přičtení konstanty 0,5. [29]

### **2.3.2 Testování dat**

Princip statistického testování je dle [4] následující. Účastníci studie tvoří vzorek populace, jež nás zajímá, například normální čtenáři, dyslektické děti apod. Taková populace je příliš velká na to, aby byli změřeni všichni její členové. Vzorek je tedy nedokonalý obraz reality, a tudíž se ve výsledcích dopouštíme jistého stupně nejistoty. Tato nejistota se zmenšuje s rostoucí velikostí výběru a nazývá se „výběrová chyba“.

#### **2.3.2.1 Analýza rozptylu (ANOVA)**

Existují různé typy statistické analýzy v závislosti na proměnných zahrnutých ve studii. Jejich výběr spočívá v typech a množství proměnných. Analýza, která se ve studiích zabývajících se vyhodnocováním očních pohybů, používá [10] a bude použita i v této práci, se nazývá analýza rozptylu neboli ANOVA (anglicky *Analysis of Variance*). ANOVA je zobecněnou obdobou dvou-výběrového nepárového t-testu<sup>5</sup>, který porovnává střední hodnoty dvou výběrů. ANOVA umožňuje ověřit, zda na hodnotu náhodné veličiny pro určitého jedince má statisticky významný vliv hodnota některého znaku, který se u jedince dá pozorovat. Tento znak musí nabývat jen konečného počtu možných hodnot (nejméně dvou) a slouží k rozdělení jedinců do vzájemně porovnávaných skupin. Kvantitativní hodnota znaku přitom nemá povahu míry. Je-li třeba vzít v úvahu i konkrétní kvantitativní hodnotu jako míru určitého znaku, použije se místo analýzy rozptylu-

---

<sup>5</sup> t-test – matematická statistika porovnávající střední hodnoty dvou výběrů.

lu lineární model. Analýza rozptylu je pro víc než jeden znak značně výpočetně náročná metoda a je pro ni téměř vždy potřeba počítač se speciálním statistickým softwarem.

### Jednofaktorová ANOVA

Jedná se o postup pro testování hypotéz, že střední hodnota  $K$  populací je shodná, přičemž  $K \geq 2$ . Jednofaktorová ANOVA porovnává střední hodnoty výběrů nebo skupin a následně učiní závěr o středních hodnotách těchto populací. Jednofaktorová ANOVA svůj název nese podle toho, že pracuje vždy s jednou nezávislou proměnnou či faktorem, neboli s jedním příznakem.

Pro testování, zda je rozdíl dvojice středních hodnot větší, než náhodně způsobený rozdíl, lze nejprve provést sérii t-testů na  $K$  středních hodnot vzorků – nicméně tento postup se potýká s důležitým problémem (např. nárůst chyby I. druhu<sup>6</sup>). Postup vyvinutý panem R. A. Fisherem zvaný ANOVA umožňuje testovat hypotézu rovnosti středních hodnot  $K$  populací, zatímco udržuje chybu I. druhu na předem stanovené (apriorní)  $\alpha$  úrovni. ANOVA pracuje pouze s jednou nezávislou proměnnou. Nezávislou proměnnou můžeme u výzkumu řídit nebo ovládat; jedná se typicky o kategorickou proměnnou, což znamená, že rozděluje jednotlivce do dvou či více skupin.

### **Hypotéza pro jednofaktorovou ANOVA**

Nulová hypotéza ( $H_0$ ) při testování jednofaktorové ANOVA říká, že střední hodnoty populací jsou si rovné.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_K \quad (2.9)$$

Alternativní hypotéza ( $H_a$ ) říká, že střední hodnota alespoň jedné skupiny je signifikantně odlišná od středních hodnot zbývajících skupin.

$$H_a: \mu_i = \mu_k \quad (2.10)$$

Pro některá  $i, k$ , kde  $i$  a  $k$  zastupují unikátní skupinu.

### **Předpoklady**

**1. Předpoklad nezávislosti** říká, že pozorování jsou náhodnými a nezávislými vzorky populace. Nulová hypotéza říká, že vzorky pocházejí z populací se stejnými středními hodnotami. Vzorky musejí být náhodné a nezávislé, pokud mají být reprezentativními výběry populace. Hodnota jednoho pozorování nemá souvislost s jiným pozorováním.

---

<sup>6</sup> Chyba I. druhu - chybné rozhodnutí učiněno poté, co test odmítne pravdivou nulovou hypotézu.

**2. Předpoklad normality** říká, že rozdělení populací, ze kterých vzorky pocházejí, je normální. Pro testování předpokladu normality lze například použít běžně užívaný Shapiro-Wilkův test, nebo Lillieforsův test, který ověřuje hypotézu, že data byla vybrána z normálně rozložené populace. Z tohoto testu je porovnána  $p$  hodnota je porovnána s apriorní hladinou  $\alpha$  (statistická hladina významnosti) – výsledkem je zamítnutí ( $p < \alpha$ ) nebo přijetí ( $p > \alpha$ ) nulové hypotézy. Hladinu  $\alpha$  volíme podle velikosti vzorku – běžně používanými hodnotami jsou 0.05 a 0.01.

**3. Test homogeneity rozptylu** je dalším krokem ANOVA, kde nulová hypotéza předpokládá, že neexistuje rozdíl v rozptylech výběrů. Pro testování může být použit Levenův  $F$ -test rovnosti rozptylů. Levenův test používá pro testování předpokladu homogeneity rozptylu hladinu významnosti nastavenou *apriorně* pro ANOVA (např.  $\alpha = 0.05$ ).

### **Porušení předpokladu jednofaktorové ANOVA**

Pokud je statistický postup ovlivněn porušením některého předpokladu, říká se, že postup je robustní s ohledem na daný předpoklad. Jednofaktorová ANOVA je robustní s ohledem na porušení předpokladů s výjimkou případu nerovnosti rozptylů s odlišnou velikostí vzorků. To znamená, že ANOVA může být použita pro rozptyly jen přibližně stejné pouze, když počet subjektů ve všech skupinách je rovný (kde rovný může být definován tak, aby větší skupina nebyla větší než 1½krát než menší skupina). ANOVA je také robustní v případě, že mají data pouze přibližné normální rozdělení. Můžeme tedy použít jednofaktorovou ANOVA i v případě, že předpoklad homogeneity rozptylu (pokud rozptyl větší skupiny není více než 4 nebo 5násobný oproti menší skupině) nebo předpoklad normality není zcela dodržen. Obecně porušení předpokladů mění hodnotu chyby I. druhu. Místo toho, abychom pracovali na stanovené hladině významnosti, může být současná hodnota chyby I. druhu menší nebo větší v závislosti na tom, který předpoklad byl porušen. Pokud populace vzorku není normální, je efekt chyby I. druhu minimální. Pokud se liší rozptyly populací, může nastat problém v případě, že velikosti výběrů nejsou stejné. Pokud větší rozptyl souvisí s větším výběrem, pak  $F$  test příliš konzervativní. Pokud je menší rozptyl spojen s větším výběrem,  $F$  test bude příliš liberální. (Pokud je hladina  $\alpha$  0.05, „konzervativní“ znamená, že současná hodnota je menší než 0.05.) Pokud jsou velikosti vzorků stejné, efekt heterogenity rozptylů (např. porušení předpokladu homogeneity rozptylu) na chybu I. druhu je minimální. Jinými slovy, efekt porušení předpokladů svým způsobem souvisí s tím, jaký předpoklad byl porušen. Pokud jsou porušení extrémní – s ohledem na normalitu a homogenitu rozptylu – je alternativním testem místo jednofaktorové analýzy rozptylu *Kruskal-Wallisův* test. Jedná se o neparametrický test, který je použit pro  $K$  nezávislých skupin a nepředpokládá normalitu populace. [30]

### 2.3.3 Základní klasifikační techniky dat

#### 2.3.3.1 *Statistické rozpoznávání vzorů*

Statistické rozpoznávání vzorů se uplatňuje v mnoha oborech, jako jsou medicína, počítačové vidění, robotika, vojenství, průmysl a v mnohých dalších. Prvním krokem v rozpoznávání vzorů je výběr příznaků, které budou použity pro rozlišení tříd. Správný výběr příznaků je nezbytný pro vytváření přesných klasifikátorů. [6]

Když jsou příznaky vybrány, obdržíme vzorek těchto příznaků (*features*) pro různé třídy. Neboli my nacházíme objekty náležící do jednotlivých tříd a měříme příznaky. Každý pozorovaný soubor měření příznaku (*case, pattern, vzor*) má označení třídy (*class label*). Když máme data, o kterých je známé jejich přiřazení k jednotlivým třídám, můžeme použít tuto informaci k vytvoření metodologie, která bude jako vstup přijímat náměry příznaků a jejímž výstupem bude třída, ke které tato měření náleží.

Obecně lze říci, že mezi původní podobou matice příznaků a konečným označením tříd, se může podoba této původní matice několikrát změnit: příznaková matice se upravuje a výběrem – selekcí příznaků, normalizací příznaků (*scaling*), lineárním nebo nelineárním mapováním (extrakce příznaků), klasifikací množinou klasifikátorů nebo kombinací klasifikátorů a finálním označením. V každém tomto kroku jsou data transformována mapováním. [39]

#### 2.3.3.2 *Výběr klasifikátoru*

Aspekty, které při výběru klasifikátorů sledujeme, jsou apriorní pravděpodobnost, rozdělení dat, množství trénovacích dat a funkce ceny. Nejprve je nutné zvolit metodu, pomocí které hodnotíme generalizační výkon klasifikátoru. Jednou z možností je odhad tohoto výkonu pomocí středních hodnot z nezávisle vybraných testovacích dat datasetu. Při porovnávání klasifikátorů může tato metoda vykazovat nepředvídatelné nedostatky (více v [8]).

#### 2.3.3.3 *Učení a přizpůsobení klasifikátoru*

V nejširším slova smyslu každá metoda, která obsahuje informaci z trénovacích vzorků při návrhu klasifikátoru, využívá učení. Vytváření klasifikátorů pak zahrnuje použití nějakého typu modelu nebo typu klasifikátoru a použití trénovacích vzorků k naučení či odhadu neznámých parametrů modelu. Učení je formou algoritmizace, která zmenšuje chybu na trénovacích datech. Učení může mít několik základních forem – učení s učitelem (*supervised learning*), učení bez učitele (*unsupervised learning*) a tzv. posílené učení (*reinforcement learning*).

Při učení s učitelem učitel poskytne označení tříd nebo cenu pro každý vzorek trénovací množiny. Učení bez učitele nebo *clustering* využívá jakési „přírodní seskupování“ vstupních vzorků. U posledního typu učení (*reinforcement learning*) dochází výpočtem ke stanovení předběžných tříd na trénovacích datech a následné použití známé cílové třídy pro vylepšení klasifikátoru. [8]

#### 2.3.3.4 Hodnocení a porovnávání klasifikátorů – Přesnost klasifikačních modelů

Přesnost klasifikačního modelu vyjadřuje míru schopnosti modelu klasifikovat neznámá data (tzn. data, na která model nebyl trénován). Použití trénovacích dat k výpočtu přesnosti modelu by vedlo k chybným výsledkům, neboť se jedná o data, na kterých byl systém trénován a která se při reálném nasazení klasifikačního systému nebudou téměř vyskytovat [46].

Klasifikací získáme předpokládané zařazení jednotlivých vzorků do tříd, které pak porovnáváme se skutečným zařazením, podle toho jsou všechny klasifikované prvky zařazeny do 4 skupin (Tabulka 2.2).

	Předpovídaný pozitivní	Předpovídaný negativní
Skutečně pozitivní	TP	FN
Skutečně negativní	FP	TN

Tabulka 2.2: Klasifikace

*TP (True Positive)*: správně zařazené vzorky do pozitivní třídy

*FN (True Negative)*: nesprávně zařazené vzorky do negativní třídy

*FP (False Positive)*: nesprávně zařazené vzorky do pozitivní třídy

*TN (True Negative)*: správně zařazené vzorky do negativní třídy

Z těchto údajů poté zjišťujeme následující parametry:

- *Přesnost*: Část správně klasifikovaných subjektů ( $TP + TN$ ) z celkového počtu subjektů v testovací množině ( $TP + TN + FP + FN$ ).
- *Senzitivita*: Poměr správně klasifikovaných dyslektiků ( $TP$ ) k celkovému počtu dyslektiků v testovací množině ( $TP + FN$ ).

- *Specificita*: Poměr správně klasifikovaných subjektů kontrolní skupiny (TN) k celkovému počtu subjektů kontrolní skupiny v testovací množině (TN + FP).

### Křížová validace

Křížová validace se typicky využívá pro určení míry klasifikační chyby pro aplikace využívající rozpoznávání vzorů nebo pro predikci chyby při vytváření klasifikačních modelů. [6]

V jednoduché validaci náhodně rozdělíme množinu označených trénovacích vzorků  $D$  na dvě části: z nichž jedna je použita jako tradiční trénovací množina pro úpravu parametrů modelu. Druhá množina dat – *validační* – je použita pro odhad generalizační chyby. Vzhledem k tomu, že naším cílem je co nejmenší generalizační chyba, trénujeme klasifikátor do chvíle, kdy je chyba minimální. Je nezbytné, aby validační (testovací) množina nezahrnovala body používané pro trénování parametrů v klasifikátoru – metodologická chyba známá jako „testování na trénovací množině“. [8]

Jednoduché zobecnění výše popsané metody je *m-fold cross-validace*. (Pro  $m$ -fold validaci se vstupní množina dat rozdělí na  $m$  podmnožin a proces se  $m$ -krát opakuje. Limitní případ, kdy  $m$  je rovno počtu vzorků-1 se nazývá *leave-one-out*.) V tomto případě je trénovací množina náhodně rozdělena do  $m$  disjunktních podmnožin stejné velikosti  $m/n$ , kde  $n$  je počet vzorů v  $D$ . Klasifikátor je trénován  $m$ -krát, přičemž je vždy jedna množina „držena“ jako validační set. Odhadovaný výkon klasifikátoru se vypočte ze středních hodnot  $m$  chyb. Tato technika může být využita prakticky pro každou klasifikační metodu.

Validace je heuristická a nemusí v každém případě klasifikátory vylepšit. Nicméně validace je extrémně jednoduchá a pro mnoho skutečných problémů vylepšuje generalizační přesnost. Existuje několik metod pro výběr části  $\gamma$  z datasetu  $D$ , která bude použita jako validační množina ( $0 < \gamma < 1$ ). Skoro vždy by měla validační množina představovat menší část datasetu ( $\gamma < 0.5$ ), tradičně se volí validační množina  $\gamma = 0.1$ , tato hodnota byla prokázána jako efektivní v mnoha případech.

Křížová validace je od základu empirický postup, který testuje klasifikátor experimentálně. Pokud jednou natrénujeme klasifikátor za použití křížové validace, validační chyba nám dá odhad přesnosti finálního klasifikátoru na neznámé testovací množině. Pokud skutečná, avšak neznámá míra chyby je  $p$  a pokud  $k$  z  $n'$  nezávislých, náhodně vybraných testovacích vzorků je chybně klasifikovaných, poté  $k$  má binomické rozdělení:

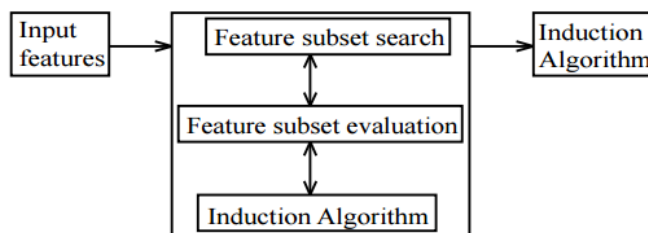
$$P(k) = \binom{n'}{k} p^k (1-p)^{n'-k} \quad (2.11)$$

### 2.3.3.5 Výběr příznaků

Výběr rozpoznávacích příznaků je rozhodujícím krokem a závisí na charakteristikách oblasti problému. Přístup k vzorovým datům je jistě hodnotný pro výběr množiny příznaků. Nicméně důležitou roli hraje také apriorní vědomost. [8] Nejčastějšími metodami výběru příznaků je extrakce, kdy dochází ke zvýšení počtu příznaků, dále selekce (výběr) příznaků. Nejběžnějšími přístupy k selekci příznaků jsou *forward search* – začíná s prázdným souborem příznaků a *backward search* – začíná s plným souborem příznaků.

#### Výběr příznaků metodou Wrapper

V metodě *wrapper* je výběr množiny příznaků provedena algoritmem strojového učení, který představuje „černou skříňku“. Algoritmus pro výběr příznaků hledá vhodnou podmnožinu příznaků za použití samotného algoritmu strojového učení, který je součástí vyhodnocovací funkce. Vyhodnocovací funkce hodnotí přesnost klasifikátoru, podle níž zvolí nejvhodnější kombinaci příznaků. Touto vyhodnocovací funkcí může být křížová validace (Obrázek 2.9). [37]



Obrázek 2.9: Wrapper - Algoritmus má přívlastek indukční díky své schopnosti indukovat obecné závěry z konkrétních příkladů [34]

## 3 Metodika

### 3.1 Úvod

Na (Obrázek 3.1) jsou nastíněny jednotlivé kroky získávání a zpracování záznamů z vyšetření očních pohybů. Předmětem této diplomové práce jsou kroky znázorněné zelenou barvou, tedy hodnocení kvality záznamů a výběr záznamů, předzpracování záznamů a statistické vyhodnocení a klasifikace dat. Dále se budu jednotlivými kroky zabývat podrobněji.



Obrázek 3.1: Proces získávání a zpracování dat

### 3.2 Získání dat

V rámci výzkumného projektu č. TA 01011138: *Sledování očních pohybů pro diagnostiku v neurověděch* bylo změřeno 378 žáků SZŠ (217 chlapců/161 dívek, níže vidíme věkové rozložení skupiny zdravých jedinců (Obrázek 3.2) a skupiny dyslektiků (Obrázek 3.3). Žáci podstoupili baterii sedmi diagnostických okulometrických testů pořízených systémem snímání očních pohybů.

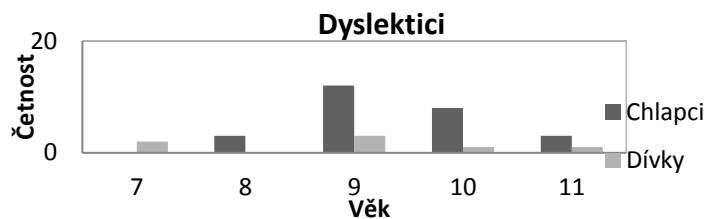
#### 3.2.1 Použitá technologie k získání dat

Pro účely snímání očních pohybů byl použit systém I4Tracking®, který byl vyvinut společností Medicton Group s.r.o., za přispění ČVUT v Praze. Tento systém pracuje na základě videookulografické metody, která používá ke sledování polohy oka a jeho stavu oční kameru.



Obrázek 3.2: Rozložení dat podle věku a pohlaví u kontrolní skupiny zdravých dětí





Obrázek 3.3: Rozložení dat podle věku a pohlaví u skupiny dyslektiků

### 3.2.2 Popis okulometrických testů

Diagnostika specifické vývojové dyslexie je založena na vizuální stimulaci probanda prostřednictvím speciálních stimulů. Jako vizuální podněty se používají neverbální úlohy (např. dítě sleduje sekvenčně body ve statickém či dynamickém provedení) z důvodu diagnostiky i předškolních dětí. V rámci projektu byla v konzultaci s odborníky vytvořena sada úloh, která tvoří baterii okulometrických testů. Baterie okulomotorických úloh byla navržena a ověřována tak, aby zachytila nejčastější vývojové poruchy učení (dyslexie, dysortografie, dysgrafie, dyskalkulie aj.), ADHD, vývojové poruchy jazyka (dysfázie), poruchy citové vazby, poruchy autistického spektra, intelektovou disabilitu. Zahrnuje úlohy očních pohybů, které jsou kontrolovány jak z nižších etáží CNS (reflexy), tak z vyšších etáží CNS (voluntární oční pohyby). Jedná se především o tyto dílčí vyšetření:

*Vyšetření fixační stability* představuje úlohu s cílem fixovat pohled na terčík uprostřed obrazovky. Obsahuje variantu optokinetická, stimulující optokinetický nystagmus a znesnadňující fixaci. Doba úlohy opět 10 s. Přejít z jedné varianty do druhé bude plynulý: s nástupem jedenácté vteřiny naskočí na obrazovce během jedné - dvou vteřin svíslé světle a tmavě šedé pruhy, které se začnou ihned při svém zjevení pohybovat ve směru zprava doleva. Rychlost pohybu  $5^\circ/s$ . [32]

*Vyšetření plynulých sledovacích očních pohybů (PSOP)* představuje úlohu, u které je cílem plynule sledovat stimul pohybující se horizontálně na obrazovce. Více o úloze *fixační stability* dále v kapitole 3.5.1.

*Vyšetření sakád I* představuje úlohu sekvenčního sledování, při které se na obrazovce postupně objevují terčíky a cílem je na ně přesunovat pohled. V úvodní, lehčí fázi jsou doby, kdy jsou jednotlivé terčíky zobrazeny, konstantní a trvají cca 350 ms. Rovněž vzdálenosti mezi jednotlivými stimuly jsou v úvodní fázi konstantní (ne méně než  $3^\circ$ ). V navazující druhé, těžší fázi se proměňují doby podnětů, a současně se proměňují i vzdálenosti mezi terčíky.

*Vyšetření sakád II* představuje u úlohy sekvenčního sledování, při které si dítě samo určuje tempo. Na obrazovce naskočí 6 řádek terčků po šesti terčících v každé řádce a cílem je se podívat na všechny terčíky ve směru jako by se jednalo o čtení textu.

*Vyšetření pro-sakád* představuje klasickou sakadickou úlohu. Vyšetřovaná osoba fixuje nejprve bod ve středu obrazovky, a jakmile se objeví sakadický podnět na jedné či druhé straně obrazovky, přesouvá svůj pohled k němu. Strana, na které se sakadický podnět objeví, se vybírá náhodně.

*Vyšetření anti-sakád* představuje anti-úlohu ke klasické sakadické úloze. Vyšetřovaná osoba je instruována, aby se podívala na opačnou stranu, než na které se objevil sakadický podnět, a zhruba stejně daleko od fixačního bodu.

*Vyšetření gravitačního středu* představuje úlohu obdobnou klasické sakadické úloze, ale v některých časových okamžicích se namísto jednoho terčiku zobrazí dva terčíky blízko u sebe a zkoumá se, kam míří prvotní sakadická reakce. [32]

### **3.3 Hodnocení kvality záznamů a výběr záznamů**

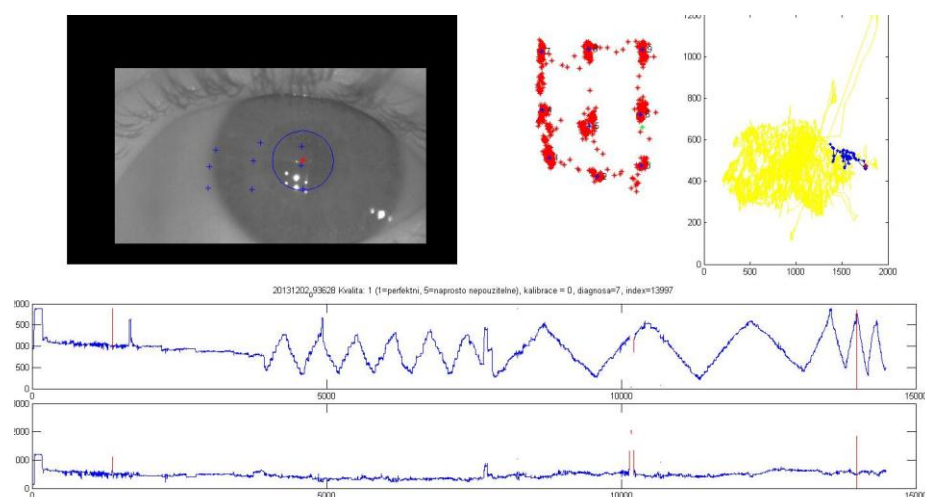
Nejprve byla vyhodnocena kvalita jednotlivých záznamů. Toto hodnocení bylo provedeno pomocí vyhodnocovacího nástroje (Obrázek 3.4), který je součástí programového balíčku *Eye Movements Signal Analysis* (EMSA, dále jen toolbox) v prostředí Matlab. Tento nástroj postupně prochází záznamy podle uživatelem zvolených kritérií – věk, diagnóza, úloha, a umožňuje ke každému záznamu přiřadit hodnocení. U každého jednotlivého záznamu byla hodnocena jeho celková kvalita (plynulost signálu) a množství artefaktů, které se v signálu vyskytovaly. V záznamech se vyskytovaly artefakty technické způsobené nesprávnou kalibrací, špatnou detekcí zornice nebo chybou při samotném měření, ale také artefakty biologické způsobené většinou mrknutím pacienta – v tomto případě nebyla zornice detekována, což se v záznamu projevilo chybějícími hodnotami (Obrázek 3.5). Některé záznamy obsahovaly více úloh za sebou, proto se při hodnocení muselo přihlídnout jak k celkové kvalitě záznamu, tak ke kvalitě jednotlivých úloh, které záznam obsahoval. Záznamy byly na základě své kvality označeny hodnotou 1 – 5, zároveň byly během hodnocení vyřazeny záznamy, které byly zatíženy špatnou kalibrací.

Ačkoli většina záznamů vykazovala vysokou kvalitu, pro použití v diplomové práci bylo nutné vybrat záznamy s minimálním počtem artefaktů, tedy záznamy s hodnocením 1 a 2. Nejprve byly vybrány záznamy dětí s dyslexií a dětí s dyslexií v kombinaci s jinou poruchou. K těmto záznamům byla následně vybrána kontrolní skupina záznamů zdravých

dětí tak, aby tento výběr svým složením odpovídal výběru záznamů dyslektiků z hlediska věku, typu úlohy a kvality záznamu. Celkově bylo pro další zpracování vybráno 24 záznamů dyslektických a 48 záznamů zdravých dětí.

### 3.4 Předzpracování záznamů

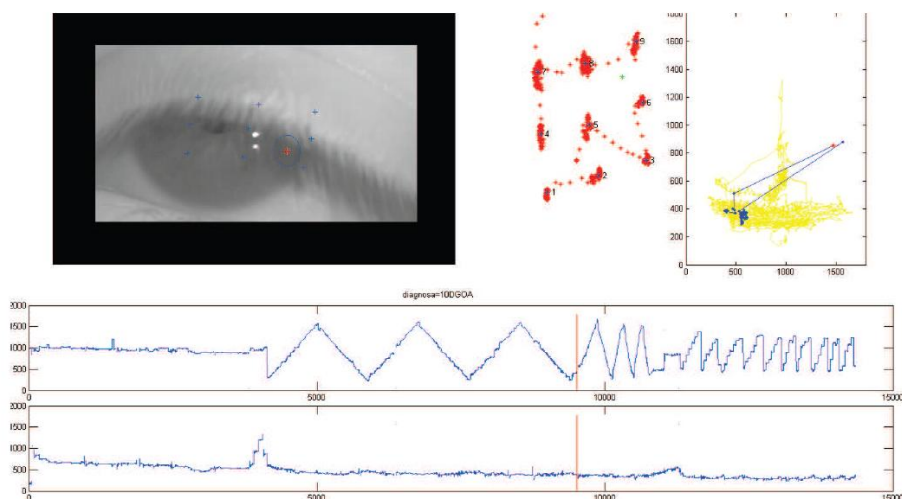
Na vybraných záznamech byla dále provedena ruční korekce artefaktů v záznamech za pomoci stejného vyhodnocovacího nástroje jako hodnocení záznamů (Obrázek 3.4).



Obrázek 3.4: Červeně vyznačené artefakty po korekci. Fyziologické „odskoky“ v záznamu se vyznačují delším trváním.

Ačkoli lze technické artefakty zpravidla v záznamu vizuálně odlišit, každý „odskok“ byl před jeho odstraněním porovnán se snímkem oka zachyceným kamerou v daný okamžik. Ze snímku je zřejmé, zda k „odskoku“ polohy pupily ve skutečnosti došlo, či nikoli. Pokud se artefakt potvrdil, byl odstraněn, přičemž odstraněné artefakty byly v záznamu označeny červenou barvou, aby bylo možné provést rychlou zpětnou kontrolu, zda byly tyto artefakty detekovány a označeny správně.

Artefakty, které se projevují jako „odskoky“ v signálu a jsou způsobeny nesprávnou detekcí polohy pupily v daný okamžik. Vyznačují se zpravidla krátkým trváním, což je signifikantně odlišuje od „odskoků“ způsobených chaotickým pohybem oka, při kterých se oko opravdu znatelně odchýlí od předpokládané trajektorie. Tyto odskoky mají zpravidla delší trvání. K nesprávné detekci pupily dochází zpravidla ve chvíli, kdy vyšetřované dítě mrká (Obrázek 3.5).



Obrázek 3.5: Chvilé, kdy dochází vlivem mrkání k nesprávné detekci pupily. V signálu je vyznačena červená oprava tohoto artefaktu.

### 3.5 Zpracování záznamů

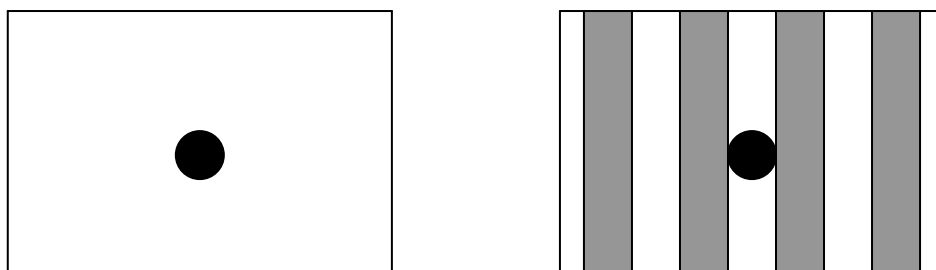
Všechna naměřená data byla po ruční korekci dále zpracována pomocí výše zmíněného balíčku *EMSA Toolbox*. Toolbox umožňuje mimo jiné automatickou klasifikaci částí záznamů na sakády a fixace. Dále umožňuje výpočet obecných příznaků, což představuje výpočet příznaků signálu, které nezávisí na prezentovaných stimulech. Jedná se o parametry sakád a fixací: jejich počet, časy, směry pohybů, atd. Dále toolbox umožňuje export videozáznamů či obrázků se stimuly a projekcí pohledu oka, doby setrvání v zónách zájmu, a dalších příznaků specifických pro jednotlivé stimuly.

Záznamy byly pomocí výše zmíněného toolboxu segmentovány na jednotlivé úlohy. V jednotlivých úlohách se vyhodnocovaly hodnoty příznaků relevantních pro danou úlohu. Zpracováním dat byly získány hodnoty všech příznaků pro skupinu dětí s dyslexií a pro kontrolní skupinu.

#### 3.5.1 Fixační stabilita

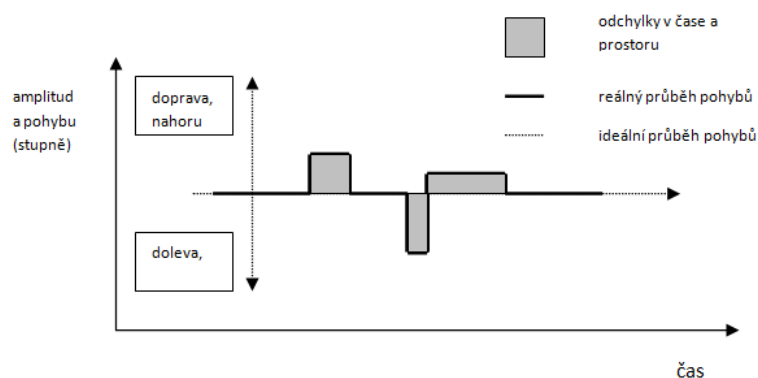
Výpočet příznaků specifických pro jednotlivé stimuly představuje velmi širokou oblast. Různé stimuly je vhodné hodnotit různou sadou příznaků. Pro účely diplomové práce byla k dalšímu statistickému zpracování zvolena úloha *fixační stabilita*. Proto se v následující kapitole budu touto úlohou blíže zabývat, dále se pak budu věnovat statistické analýze hodnot příznaků z této úlohy.

Popis úlohy: dítě pozoruje bod, z kterého nemá spustit oči. Nejprve varianta uniformní: tečka je černá a pozadí bílé (či světle šedé). Doba úlohy: 10 s. Poté varianta optokinetická, stimulující optokinetický nystagmus znesnadňující fixaci (Obrázek 3.6). Doba úlohy opět 10 s. Přechod z jedné varianty do druhé bude plynulý: s nástupem jedenácté vteřiny naskočí na obrazovce během jedné - dvou vteřin svislé světle a tmavě šedé pruhy, které se začnou ihned při svém zjevování pohybovat ve směru zprava doleva. Rychlost pohybu  $5^\circ/\text{s}$ . Po skončení naskočí prázdná obrazovka. [32]



Obrázek 3.6: Fixační úloha. Vlevo varianta uniformní, vpravo varianta optokinetická

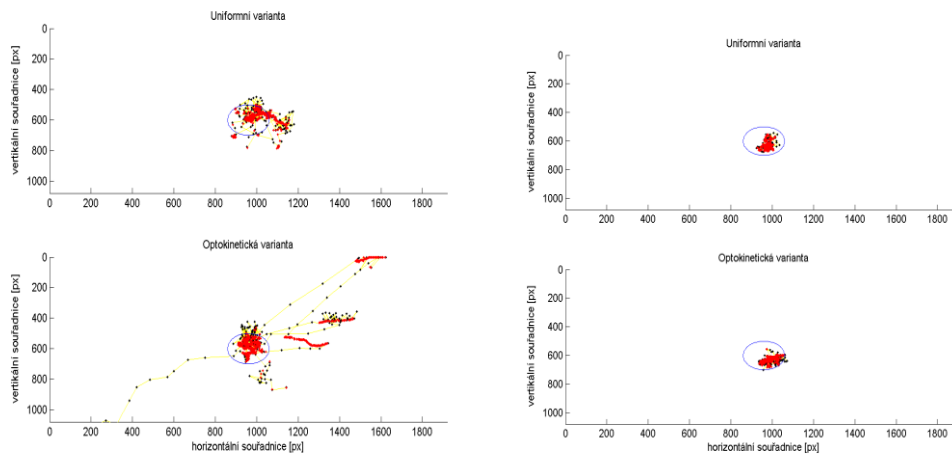
Na (Obrázek 3.7) vidíme porovnání ideálního a reálného průběhu očních pohybů v čase a prostoru se rovná nule. Prostorově nerozlišujeme, zda se dítě odklonilo od fixovaného terčíku doprava-doleva či nahoru-dolů; podstatná je velikost odklonu a jeho trvání v čase. [32]



Obrázek 3.7: Průběh očních pohybů v úloze „fixace bodu“ [33]

Na (Obrázek 3.8) vidíme reálné naměřené signály dyslektika (vlevo) a zdravého dítěte (vpravo). V horní části se nachází varianta uniformní, v dolní části optokinetická. U vývojových poruch učení, především u dyslexie, dochází k častějším odchylkám od ide-

álního průběhu, k tzv. odskokům od bodu fixace, což je patrné zejména na optokinetické variantě u dyslektika. Červeně jsou znázorněny fixace, černě sakády.



Obrázek 3.8: Naměřené záznamy uniformní a optokinetické úlohy [33]

V úloze fixační stabilita jsme sledovali celkem 16 příznaků, jejichž název a označení lze nalézt v Příloze č. 1.

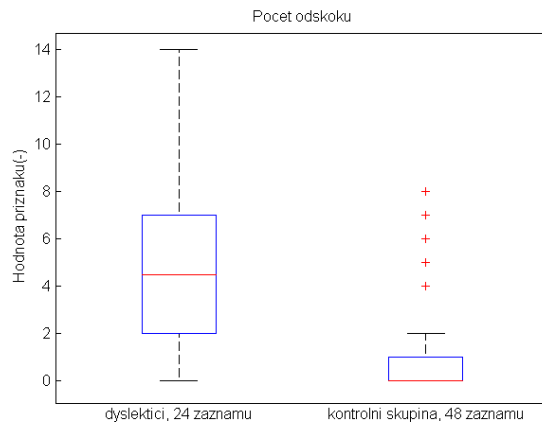
## 3.6 Statistické vyhodnocení a klasifikace dat

### 3.6.1 Průzkumová a popisná analýza dat

Průzkumová analýza dat byla provedena v prostředí Matlab za použití základních statistických funkcí a funkcí navržených speciálně pro námi zpracovávaná data – funkce `boxplot_features.m` a `histogram_features.m`.

`Boxplot_features.m`

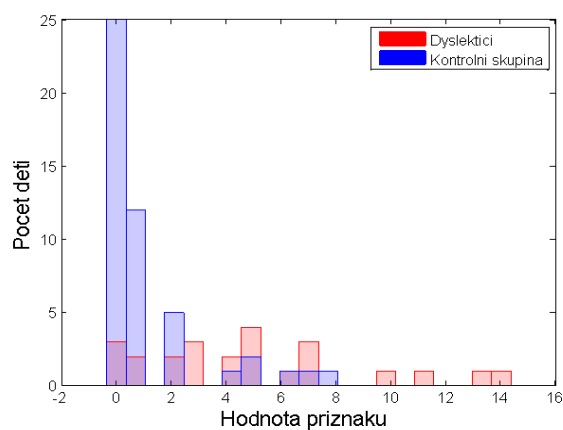
Vstupem této funkce jsou hodnoty daného příznaku, indexy hodnot, které označují dyslektiky a děti z kontrolní skupiny a název příznaku. Funkce hodnoty příznaků rozdělí dle indexů na dvě skupiny – dyslektiky a kontrolní skupinu, tyto dále vyhodnocuje. Výstupem funkce je graf obsahující souběžně krabicové grafy pro obě skupiny (Obrázek 3.9). Na vertikální ose grafu je znázorněna hodnota daného příznaku, na horizontální ose je označení daného krabicového grafu s počtem záznamů, z nichž byl graf vyhodnocen. Centrální (červená) linie označuje hodnotu mediánu daného výběru. Hrany grafu pak označují 1. a 3. kvartil. „Vousy“ grafu znázorňují vzdálená data, nikoli však odlehlé hodnoty (*outliers*), tyto jsou v grafu znázorněny červenými křížky. Ve funkci je integrována funkce `boxplot.m`.



Obrázek 3.9: Krabicový graf znázorňující hodnoty příznaku „Počet odskoků“ pro obě skupiny

Histogram\_features.m

Funkce opět pracuje se stejnými vstupy, tedy hodnotami daného příznaku, indexy hodnot, které označují dyslektiky a děti z kontrolní skupiny. Zpracování vstupních dat je obdobné jako u funkce `boxplot_features.m`. Vzhledem k odlišným hodnotám v obou skupinách je nutné histogram dále normalizovat, aby byl v histogramu znázorněn stejný počet a velikost intervalů (tzv. binů). Grafické nastavení histogramu je upraveno tak, aby byly přehledně znázorněny hodnoty obou pozorovaných skupin v jednom grafu. Na vertikální ose vidíme počet dětí spadajících do daného intervalu. Na horizontální pak hodnotu daného příznaku. Ve funkci je integrovaná funkce `hist.m`.



Obrázek 3.10: Histogram znázorňující rozložení hodnot příznaku „Počet odskoků“ pro obě skupiny

Dále byly v rámci popisné analýzy vyhodnoceny základní statistické charakteristiky – střední hodnota, medián, maximální a minimální hodnota příznaku, směrodatná odchyl-

ka a rozptyl, tyto charakteristiky byly opět vyhodnoceny v prostředí Matlab za použití integrovaných funkcí - mean, median, max, min, std, var.

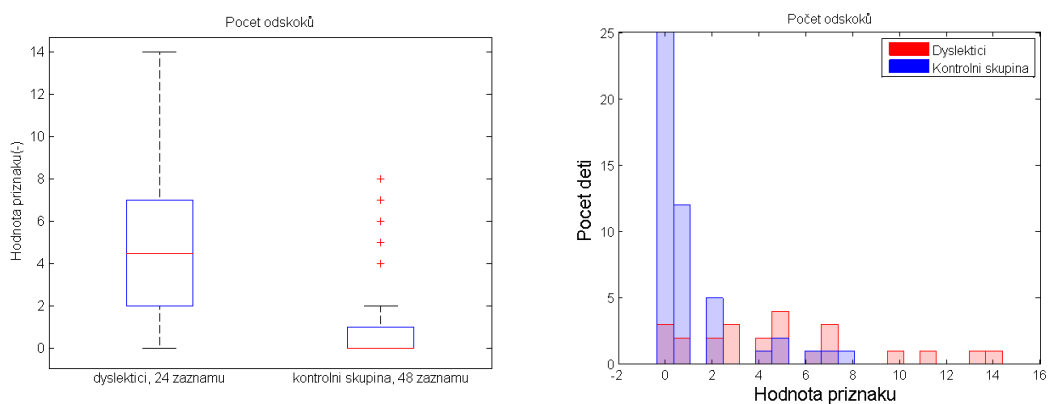
### Vyhodnocení průzkumové a popisné analýzy

Pro každý příznak byly vyhodnoceny výše zmíněné základní charakteristiky pro dyslektiky a kontrolní skupinu. Níže můžeme vidět výstup analýzy vybraných příznaků – příznak „Počet odskoků“, na jehož dvojici krabicových diagramů vidíme signifikantní rozdíly mezi dyslektiky a kontrolní skupinou, rozdíly jsou patrné i z hodnot charakteristik popisné analýzy. Na druhém příznaku „Průměrná délka sakád“, naopak z průzkumové a popisné analýzy nejsou rozdíly mezi skupinami patrné. Hodnoty základních provedených statistik pro všechny příznaky lze nalézt v Příloze č. 1., krabicové diagramy a histogramy pro všechny příznaky v Příloze č. 2.

Z průzkumové analýzy vyplývá, že významné rozdíly v provedených statistikách nevykazují příznaky: „Počet fixací“, „Průměrná délka fixací“, „Počet sakád“ a „Průměrná délka sakád“. U ostatních příznaků většina charakteristik rozdíly mezi skupinami vykazuje. Z histogramů jednotlivých příznaků bylo možné přibližně odhadnout rozdělení, které hodnoty příznaků vykazují. Vizuální odhad byl tvořen pomocí *Distribution Fitting Tool* v Matlab, přičemž příznaky vykazovaly přibližně následující rozdělení: logaritmicke-normální - příznaky 1, 2, 3, 4, 8, 9, 10, 14, 15, 16 a exponenciální - příznaky 5, 6, 7, 11, 12, 13 (pro názvy příznaků viz Příloha 1).

### Vyhodnocení průzkumové analýzy pro příznak „Počet odskoků“

Z krabicového diagramu (Obrázek 3.11) je patrné, že skupina dyslektiků od kontrolní skupiny se viditelně odlišuje hodnotou mediánu (červená linie), ale také rozptylem (modré hraniční linie diagramu). Odlišná hodnota mediánu je patrná také z histogramu.



Obrázek 3.11: Krabicový graf a histogram znázorňující hodnoty příznaku „Počet odskoků“ pro obě skupiny



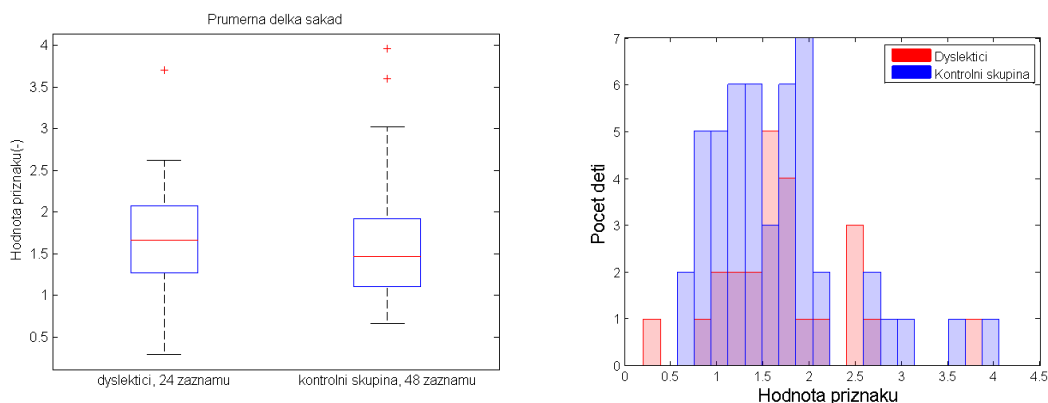
Tabulka 3.1: Základní statistiky pro příznak Počet odskoků obsahuje výsledky statistik pro jednotlivé skupiny. Významné rozdíly jsou patrné téměř ve všech charakteristikách.

	Dyslektici	Kontrolní skupina
Střední hodnota	4.91	1.19
Medián	4	0
Minimální hodnota	0	0
Maximální hodnota	14	8
Výběrový rozptyl	15.55	3.82
Směrodatná odchylka	3.95	1.95

Tabulka 3.1: Základní statistiky pro příznak Počet odskoků

### Vyhodnocení průzkumové analýzy pro příznak „Průměrná délka sakád“

Na dvojici krabicových diagramů příznaku „Průměrná délka sakád“ (Obrázek 3.12) je naopak zřetelná podobnost těchto diagramů. Medián se liší nepatrně a rozptyly obou skupin jsou také shodné.



Obrázek 3.12: Krabicový graf a histogram znázorňující hodnoty příznaku „Průměrná délka sakád“ pro dvě skupiny

Hodnoty základních statistik (Tabulka 3.2) pro příznak „Průměrná délka sakád“ pro skupinu dyslektiků a kontrolní skupinu nevykazují významné rozdíly.

	Dyslektici	Kontrolní skupina
Střední hodnota	1,73	1,63
Medián	1,66	1,47
Minimální hodnota	0,30	0,67
Maximální hodnota	3,70	3,96

Výběrový rozptyl	0,50	0,51
Směrodatná odchylka	0,71	0,72

Tabulka 3.2: Základní statistiky pro příznak Průměrná délka sakád

### 3.6.2 Statistické vyhodnocení - ANOVA

#### Motivace použití ANOVA

ANOVA pro porovnávání středních hodnot dvou skupin by měla poskytnout obdobné výsledky jako t-test. Její výhodou je však možné rozšíření na testování rozdílů více skupin. Je pravděpodobné, že ve studiích, které budou na tuto diplomovou práci navazovat, budou testovány i další skupiny dětí, a sice děti s dyslexií v kombinaci s jinou poruchou, které jsou v tuto chvíli zahrnuty do jedné skupiny s dětmi s dyslexií. Pro zachování univerzálního testování a použitelnost pro případné další studie byla proto použita ANOVA a nikoli t-test.

#### Ověření nezávislosti dat

Nezávislost dat ve všech příznacích je zajištěna navržením adekvátní metodiky studie před jejím započítím.

#### Ověření homogenity rozptylů dat

Pro ověření homogenity rozptylů dat byl použit Levenův test pro rovnost rozptylů (v Matlabu funkce `Levenetest.m`) s hladinou významnosti  $\alpha = 0,05$ . Tato funkce není integrovanou funkcí v prostředí Matlab, je však volně dostupná na [36]. Výstupem této funkce je hodnota statistiky  $F$  a pravděpodobnost spojená s tímto kritériem. Funkce dále tuto pravděpodobnost porovná s určenou hladinou významnosti. Pokud je hodnota pravděpodobnosti pro  $F$  statistiku větší nebo rovna hodnotě hladiny významnosti, pak se nulová hypotéza o rovnosti rozptylů nezamítá a předpoklad homogenity rozptylů je naplněn. Na (Obrázek 3.13) je znázorněn příklad výstupu funkce `Levenetest.m` pro příznak „Počet fixací“.

```
The number of samples are: 2
-----
Sample   Size   Variance
-----
1        48   19290.2230
2        24   14957.4330
-----

Levene's Test for Equality of Variances F=3.2219, df1= 1, df2=70
Probability associated to the F statistic = 0.0770
The associated probability for the F test is equal or larger than 0.05
So, the assumption of homoscedasticity was met.
```

Obrázek 3.13: Ukázka výstupu funkce `Levenetest.m`

### Ověření normality dat

ANOVA je založena na předpokladu normálního rozdělení dat a je robustní vůči nevelkým porušením tohoto předpokladu, před jejím samotným použitím je však dobré normální rozdělení dat ověřit. Pro vizuální ověření normality dat v prostředí Matlab byly využity histogramy, které byly součástí průzkumové analýzy. Dále pak byl využit normální pravděpodobnostní graf, který je v Matlabu reprezentován funkcí `normplot.m`. Kromě grafického ověření jsem provedla také test na normální rozdělení, a sice Lillieforsův test (v Matlabu funkce `lillietest.m`).

### Závěry ověření předpokladů

Levenův test homogenity rozptylů dat nezamítl u téměř u všech příznaků hypotézu, že rozptyly v obou skupinách jsou shodné. Pouze u příznaku „Rozptyl ve druhé části úlohy“ byla nulová hypotéza o homogenitě rozptylu zamítnuta, a tento předpoklad tudíž nebyl naplněn.

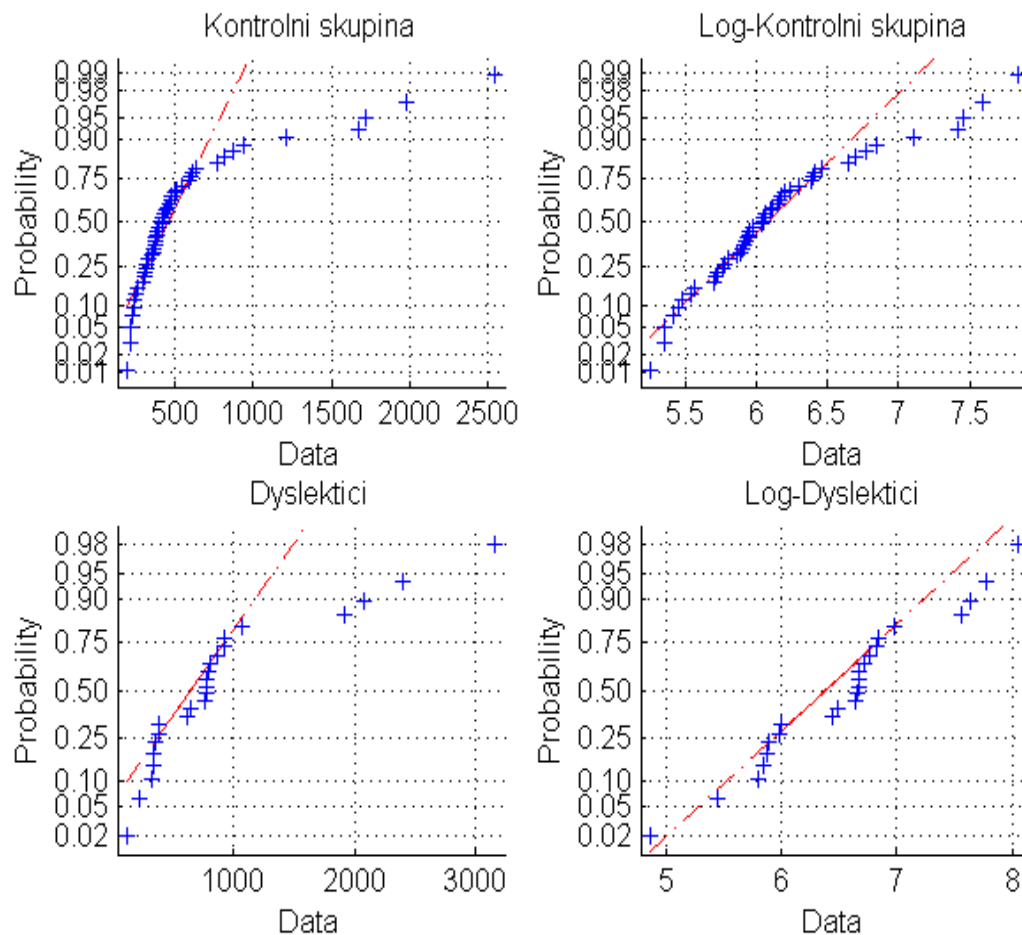
Již z průzkumové analýzy dat je z histogramů patrné, že mnohé příznaky nevykazují normální, ale spíše logaritmicko-normální, popřípadě exponenciální rozdělení. Lillieforsův test, který byl pro testování normality použit, také ve většině případů normalitu dat nepotvrdil, a to v obou skupinách (dyslektici, kontrolní skupina). Pro zajištění normality dat tedy bylo nutné data transformovat. Na data jsem aplikovala logaritmickou transformaci, za použití přirozeného logaritmu. Logaritmickou transformaci na příznaku „Odchylka od ideálu v první části úlohy“ můžeme vidět na (Obrázek 3.14) v podobě normálních pravděpodobnostních grafů jednotlivých skupin a také na (Obrázek 3.15) v podobě histogramů. Lillieforsův test na normální rozdělení, které jsem poté znovu provedla, normalitu dat připustil již ve více případech. V (Tabulka 3.3) vidíme seznam čísel příznaků, jejichž normalita nebyla zamítnuta před a po transformaci. Histogramy všech příznaků před a po transformaci jsou k nahlédnutí v Příloze č. 3.

Lillieforsův test před transformací		Lillieforsův test po transformaci	
Dyslektici	Kontrolní skupina	Dyslektici	Kontrolní skupina
4, 7, 9	-	1, 2, 3, 4, 8, 9, 14, 15, 16	4, 15, 16

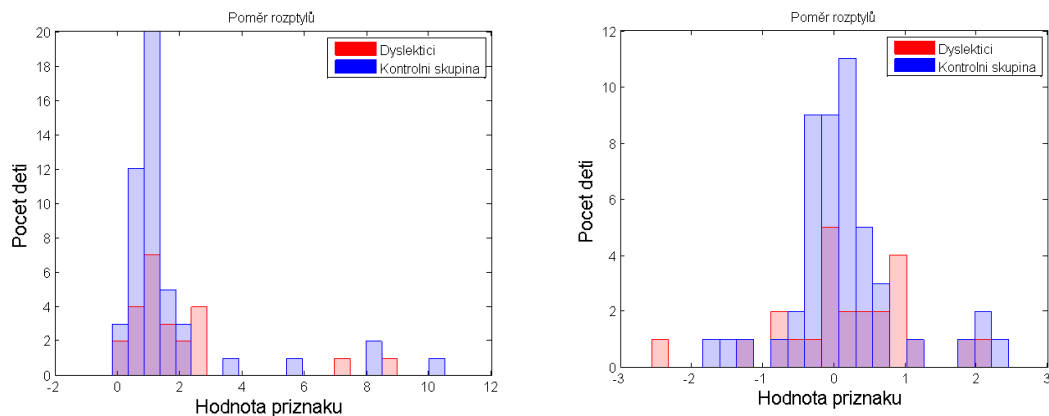
Tabulka 3.3: Příznaky vykazující normální rozdělení před a po transformaci.

Dle vizuálního zhodnocení histogramů transformovaných dat u příznaků 1, 2, 3, 8, 9, 14 pro kontrolní skupinu lze usuzovat, že ačkoli Lillieforsův test normální rozdělení zamítl, vykazují tato data po logaritmické transformaci alespoň přibližné normální rozdělení.

Z histogramů transformovaných příznaků (5, 6, 7 a 11, 12, 13), které před transformací vykazovaly spíše exponenciální, ne logaritmicko-normální rozdělení, je patrné, že použití logaritmické transformace nebylo ani v jednom případě účinné. Vzhledem k průzkumové a popisné analýze dat je však zřejmé, že se může jednat o signifikantní příznaky a je vhodné se i jejich statistickým zpracováním dále zabývat. Vzhledem k tomu, že jsem v dostupných zdrojích nenalezla efektivní způsob, jakým transformovat exponenciální data na normální, rozhodla jsem se použít analýzu ANOVA na tato data s vědomím, že byl porušen předpoklad normálního rozdělení. ANOVA se však zdá být v případě porušení normality robustní, což dokládá např. i studie [30]. Dle [33] je ANOVA dostatečně robustní i pro exponenciálně rozložená data. ANOVA nebude aplikována pouze na příznak 10 - „Počet fixací“ vzhledem k tomu, že pro něj byla zamítnuta hypotéza o homogenitě rozptylů a zároveň hypotéza o normálním rozdělení.



Obrázek 3.14: Normální pravděpodobnostní graf pro příznak „Odchylka od ideálu v první části úlohy“ před (nalevo) a po (napravo) logaritmické transformaci



Obrázek 3.15: Příznak „Odchylka od ideálu v první části úlohy“ před (nalevo) a po (napravo) logaritmické transformaci

## ANOVA

Po ověření výše zmíněných předpokladů jsem na datech provedla statistickou analýzu ANOVA. ANOVA byla provedena v softwarovém prostředí Matlab pro dvě skupiny dětí (dyslektici a kontrolní skupina). Vliv faktoru (příznaku) byl považován za signifikantní při hodnotě  $p$  nižší než 0.05.

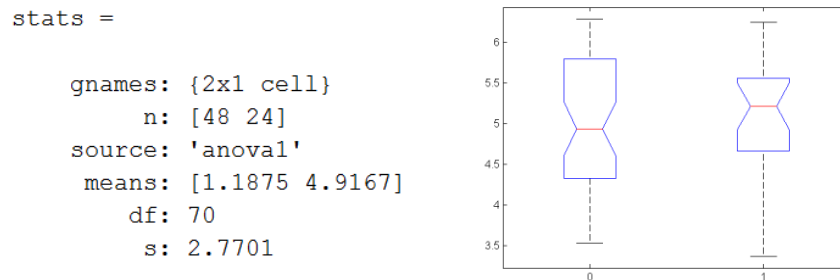
Testována byla nulová hypotéza, že střední hodnoty obou skupin jsou stejné, oproti alternativní hypotéze, že mezi středními hodnotami obou skupin existuje rozdíl **(3.1)**.

$$\begin{aligned}
 H_0: \mu_1 &= \mu_2 \\
 H_1: \mu_1 &\neq \mu_2
 \end{aligned}
 \tag{3.1}$$

V Matlabu byla pro výpočet použita funkce `anova1.m`, která testuje rovnost středních hodnot specifikovaných skupin, přičemž zvažuje rozdílný počet pozorování ve skupinách. Vstupem funkce `anova1.m` jsou hodnoty daného příznaku a dále vektor tříd, které značí, k jaké skupině daná hodnota přísluší. Výstupem funkce je hodnota  $p$ , kterou následně porovnáme s hodnotou hladiny významnosti  $\alpha$  a vyslovujeme na základě tohoto porovnání závěr, zda nulová hypotéza o shodnosti středních hodnot obou skupin bude či nebude odmítnuta. Dalším volitelným výstupem funkce je tabulka hodnot ANOVA Table (Obrázek 3.16). Posledním výstupem je datová struktura `stats`, ve které jsou uchovány statistiky důležité pro porovnání středních hodnot obou skupin (Obrázek 3.17). Ke zmíněným výstupům funkce generuje také kontrolní krabicové diagramy obou skupin (Obrázek 3.17), přičemž zářezy patrné na krabicových grafech signalizují hodnotu mediánu v obou skupinách.

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Groups	0.407	1	0.40701	0.69	0.4076
Error	41.0494	70	0.58642		
Total	41.4564	71			

Obrázek 3.16: Tabulka ANOVA jako výstup funkce `anova1`



Obrázek 3.17: Struktura `stats` a krabíkový diagram jako výstup funkce `anova1`

### Matematické vyjádření funkce `anova1`

ANOVA testuje rozdíl středních hodnot skupin rozdělením celkové odchytky v datech do dvou komponent:

- Rozptyl středních hodnot jednotlivých skupin od celkové střední hodnoty,  $\bar{y}_j - \bar{y}_\cdot$  (rozptyl mezi skupinami), kde  $\bar{y}_j$  je střední hodnota dané skupiny  $j$  a  $\bar{y}_\cdot$  je celková střední hodnota
- Rozptyl mezi pozorováními v každé skupině od odhadů jejich středních hodnot (rozptyl uvnitř skupiny) [38]

Jinými slovy ANOVA rozdělí celkový součet čtverců ( $SST$ ) na součet čtverců vzhledem k efektu mezi skupinami ( $SSR$ ) a součet čtvercových chyb ( $SSE$ ), viz rovnice (3.2).

$$\underbrace{\sum_i \sum_j (y_{ij} - \bar{y}_\cdot)^2}_{SST} = \underbrace{\sum_j n_j (\bar{y}_j - \bar{y}_\cdot)^2}_{SSR} + \underbrace{\sum_i \sum_j (y_{ij} - \bar{y}_j)^2}_{SSE}, \quad (3.2)$$

kde  $n$  je velikost výběru pro  $j$ -tou skupinu,  $j = 1, 2$ .

ANOVA tedy porovnává rozptyl mezi skupinami a rozptyl ve skupinách. Pokud je poměr rozptylu uvnitř skupin k rozptylu mezi skupinami signifikantně vyšší, dochází

k závěru, že střední hodnoty obou skupin jsou od sebe signifikantně rozdílné. To je možné změřit použitím statistiky, která má  $F$ -rozdělení s  $(k - 1, N - k)$  stupňů volnosti (3.3).

$$F = \frac{SSR/k - 1}{SSE/N - k} = \frac{MSR}{MSE} \sim F_{k-1, N-k} \quad (3.3)$$

Kde  $MSR$  je průměr čtverců mezi skupinami,  $MSE$  je střední kvadratická chyba (uvnitř skupin),  $k$  je počet skupin a  $N$  celkový počet pozorování (tedy součet pozorování z obou skupin). Pokud je hodnota  $p$  pro  $F$  statistiku menší než hladina významnosti  $\alpha$ , potom test zamítá nulovou hypotézu, že střední hodnoty obou skupin jsou si rovny. Nejběžněji používanou hladinou významnosti pro zpracování experimentálních biomedicínských dat je 0,05 a 0,01. Pro naše výpočty byla zvolena hladina významnosti 0.05. Na (Obrázek 3.18) je znovu znázorněna tabulka ANOVA, tentokrát však s významem a způsobem výpočtu jednotlivých hodnot, kde  $SS$  je součet čtverců a  $df$  počet stupňů volnosti.

Source	SS	df	MS	F	p-value
Group (Between)	SSR	$k - 1$	$MSR = SSR/(k - 1)$	$MSR/MSE$	$P(F_{k-1, N-k}) > F$
Error (Within)	SSE	$N - k$	$MSE = SSE/(N - k)$		
Total	SST	$N - 1$			

Obrázek 3.18: Tabulka ANOVA [38]

## Výsledky ANOVA

Výsledky pro logaritmicky transformované příznaky, u nichž bylo normální rozdělení potvrzeno nebo které se vizuálně blížily normálnímu rozdělení, jsou v (

Tabulka 3.4).

	Název příznaku	Hodnota $F$ -statistiky	Hodnota $p$	Výsledek hypotézy
1	Počet fixací	0.70	0.4056	$H_0$ nezamítnuta
2	Průměrná délka fixací	0.64	0.4259	$H_0$ nezamítnuta
3	Počet sakád	0.71	0.4030	$H_0$ nezamítnuta
4	Průměrná délka sakád	0.18	0.6738	$H_0$ nezamítnuta
8	Poměr rozptylů	0.09	0.7670	$H_0$ nezamítnuta
9	Rozptyl v první části úlohy	6.72	0.0116	$H_0$ zamítnuta
14	Odchylka od ideálu v 1. části úlohy	6.08	0.0161	$H_0$ zamítnuta

15	Odchylka od ideálu v 2. části úlohy	6.64	0.0121	H <sub>0</sub> zamítnuta
16	Odchylka od ideálu	9.96	0.0024	H <sub>0</sub> zamítnuta

Tabulka 3.4: ANOVA pro příznaky vykazující normální rozdělení

Pro příznaky, u nichž nebylo normální rozdělení potvrzeno ani po logaritmické transformaci, byla provedena ANOVA na netransformovaných datech s následujícími výsledky (Tabulka 3.5).

	Název příznaku	Hodnota $F$ -statistiky	Hodnota $p$	Výsledek hypotézy
5	Počet odskoků v 1. části úlohy	32.55	2.5697e-07	H <sub>0</sub> zamítnuta
6	Počet odskoků v 2. části úlohy	7.21	0.009	H <sub>0</sub> zamítnuta
7	Počet odskoků	29	9.1757e-07	H <sub>0</sub> zamítnuta
11	Doba v odskocích v 1. části úlohy	11.29	0.0012	H <sub>0</sub> zamítnuta
12	Doba v odskocích v 2. části úlohy	7.05	0.0098	H <sub>0</sub> zamítnuta
13	Doba v odskocích	16.71	0.0001	H <sub>0</sub> zamítnuta

Tabulka 3.5: ANOVA pro příznaky vykazující exponenciální rozdělení

### Výsledky ANOVA

ANOVA našla signifikantní rozdíly mezi skupinami u příznaku „Rozptyl v první části úlohy“ ( $F_{(1,70)}=6.72$ ,  $p = 0.0116$ ), dále u příznaku „Odchylka od ideálu v první části úlohy“ ( $F_{(1,70)}=6.08$ ,  $p = 0.0161$ ) a „Odchylka ve druhé části úlohy“ ( $F_{(1,70)}=6.64$ ,  $p = 0.0121$ ), dále u příznaku „Odchylka od ideálu“ ( $F_{(1,70)}=9.96$ ,  $p = 0.0024$ ). Jako velmi signifikantní se ukázaly příznaky, u nichž nebylo potvrzeno normální rozdělení, a sice „Počet odskoků v první části úlohy“ ( $F_{(1,70)}=32.55$ ,  $p < 0.001$ ) a „Počet odskoků“ ( $F_{(1,70)}=29$ ,  $p < 0.001$ ). Signifikantní však byly také příznaky „Počet odskoků ve druhé části úlohy“ ( $F_{(1,70)}=7.21$ ,  $p = 0.009$ ), „Doba v odskocích v první části úlohy“ ( $F_{(1,70)}=11.29$ ,  $p = 0.0012$ ), „Doba v odskocích ve druhé části úlohy“ ( $F_{(1,70)}=7.05$ ,  $p = 0.0098$ ) a „Doba v odskocích“ ( $F_{(1,70)}=16.71$ ,  $p < 0.001$ ). U dalších příznaků „Počet sakád“, „Průměrná délka sakád“, „Počet fixací“, „Průměrná délka fixací“ a „Poměr rozptylů“ ANOVA nepotvrdila signifikantní rozdíl mezi skupinami.



### 3.6.3 Použití klasifikační analýzy *mean-mean* a *median-median*

V rámci statistického vyhodnocování dat byl navrhnut jednoduchý klasifikátor, který se při určování třídy daného objektu rozhoduje pomocí prahu nastaveného dle vypočtené střední hodnoty (výběrového průměru) resp. vypočteného mediánu dat. V případě tohoto klasifikátoru se jedná o učení bez učitele, nevyužívá se trénovací a testovací množiny, ale data se přímo klasifikují k dané třídě na základě porovnání s prahem hodnot. Spíše než o klasifikaci se tedy jedná o jednoduchou metodu zhodnocení použitelnosti dat pro následnou klasifikaci.

Výpočet hodnoty prahu je patrný z rovnice (3.4), kde  $D$  jsou hodnoty příznaků pro skupinu dyslektiků a  $C$  hodnoty příznaků pro kontrolní skupinu.

$$threshold = \min\left(\frac{1}{N} \cdot \sum_{i=1}^N D_i, \frac{1}{N} \cdot \sum_{i=1}^N C_i\right) + \left|\frac{1}{N} \cdot \sum_{i=1}^N D_i - \frac{1}{N} \cdot \sum_{i=1}^N C_i\right| \cdot \frac{1}{2} \quad (3.4)$$

Výpočet prahu v prostředí matlab:

```
prah = min([mean(dys) mean(con)]) + abs(mean(dys) - mean(con)) / 2;
```

Pomocí nastaveného prahování byla data klasifikována do dvou skupin. Z klasifikovaných dat byly vypočteny parametry senzitivita, specificita a přesnost. Obdobou tohoto klasifikátoru je klasifikátor využívající namísto střední hodnoty hodnotu mediánu. I tento byl pro klasifikaci dat použit.

#### Výsledky klasifikátoru *mean-mean* a *median-median*

Hodnoty parametrů klasifikátoru pro všechny příznaky pro klasifikátor *mean-mean* a *median-median* jsou uvedeny Příloze č. 4. Nejvyšší přesnost měl klasifikátor *mean-mean* pro příznak „Počet odskoků v první části úlohy“ a „Počet odskoků“, a sice 78 %. Naopak nejmenší přesnost vykázal klasifikátor u příznaků „Počet fixací“, „Průměrná délka fixací“, „Počet sakád“ a „Průměrná délka sakád“, přesnost se zde pohybovala mezi hodnotami 51 – 53 %. Přesnost pro ostatní příznaky se pohybovala v 67 % - 75 %. Klasifikátor *median-median* klasifikoval data obdobně, přičemž přesnost byla ještě o něco vyšší. Největší přesnost pro klasifikátor *median-median* vykázal také pro příznak „Počet odskoků v první části úlohy“, a sice 82 %.

### 3.6.4 Další metody klasifikace

Pro klasifikace dat byl použit balíček *PRTtools* (Matlab toolbox for pattern recognition) vyvíjený na univerzitě v Delftu v Holandsku, volně dostupný pro akademické účely na [39].

#### 3.6.4.1 Výběr klasifikátorů

Pro výběr příznaků a následnou klasifikaci byly zvoleny a následně porovnány následující klasifikátory (v závorce je uveden název funkce v Matlab):

- Lineární klasifikátory: Support Vector Machine (`svc`), Linear Bayes Normal Classifier (`lbc`), Fisher's Least Square Linear Discriminant (`fisherc`)
- Kvadratický klasifikátor: Quadratic Bayes Normal Classifier (`qdc`),
- Neparametrický klasifikátor: K-Nearest Neighbor Classifier (`knnnc`)

#### 3.6.4.2 Výběr příznaků

Výběr příznaků byl proveden metodou *wrapper* za použití funkce `featself`. Jedná se o dopřednou metodu výběru příznaku (*forward feature selection*), kdy jsou příznaky postupně přidávány a hodnotí se nejmenší chyba klasifikátoru pro danou kombinaci příznaků. Pro vyhodnocení klasifikátoru během výběru příznaků byla použita křížová validace s dvěma složkami. Z celkového počtu příznaků byly vybírány vždy dva nejlepší, na kterých byla následně provedena klasifikace.

#### 3.6.4.3 Postup vyhodnocování klasifikátoru

Před samotnou klasifikací byl pomocí funkce `prdataset` vytvořen dataset, který tvořilo 48 kontrolních subjektů a 24 dyslektiků. Dataset je tvořen maticí hodnot příznaků, dále obsahuje označení tříd jednotlivých pozorování (*labels*). V datasetu byla pro jednotlivé třídy nastavena apriorní pravděpodobnost pomocí funkce `setprior`, a sice 0.5 pro obě skupiny. Dále byla pro odstranění případných odlehlých hodnot použita funkce `removeoutl`, která odstranila jednu hodnotu, konečný dataset tedy tvořilo 47 kontrolních subjektů a 24 dyslektiků.

Stěžejní část klasifikačního skriptu tvoří *for* cyklus, který prochází jednotlivé prvky datasetu. V každém kroku „drží“ jeden prvek jako testovací a zbytek dat použije pro trénování (metoda *cross-validation leave one out*). Data se dále normalizují pomocí funkce `sca-`

lem. V dalším kroku cyklu dochází k selekci příznaků pomocí funkce `featself`. Vstupem této funkce je trénovací množina, zvolený klasifikátor (např. `ldc`), výběr počtu příznaků. Zde je možné vybrat fixní počet příznaků, nebo nechat funkci vybrat optimální počet příznaků, při kterém klasifikátor vykazuje nejmenší chybu klasifikace. Posledním vstupním parametrem funkce `fealself` je počet složek křížové validace (tzv. *foldů*). Výstupem `featself` je optimální nebo uživatelem určený výběr příznaků pro klasifikaci. Trénovací datasetu s optimálními příznaky natrénujeme pomocí stejného klasifikátoru, který byl použit pro výběr příznaků, a následně otestujeme na testovacích datech, tím získáme přiřazení jednotlivých subjektů do tříd.

#### 3.6.4.4 Vyhodnocení klasifikace

Pro samotnou klasifikaci byly použity klasifikátory uvedené v (Tabulka 3.6) a z matice záměn (funkce `confmat`), jež byla výstupem klasifikace, byly dopočítány sledované parametry (Tabulka 3.7). Pro klasifikátory `ldc` a `qdc` byly použity logaritmičsky transformované příznaky, které se více blíží normálnímu rozdělení, které tyto klasifikátory pro klasifikaci předpokládají. Pro další klasifikátory byly použity netransformované příznaky.

True Labels	Estimated Labels		Totals
	0	1	
0	41	6	47
1	8	16	24
Totals	49	22	71

Tabulka 3.6: Matice záměn `confmat`

Z hlediska biomedicíny nás nejvíce zajímají parametry senzitivita a specifická, dále jsme u klasifikátoru hodnotili parametr přesnost.

Klasifikátor	Přesnost [%]	Senzitivita [%]	Specifická [%]	Čísla vybraných příznaků
<i>svc</i>	78.87	50.00	93.62	5, 9
<i>knn</i>	74.65	54.17	85.11	7, 13
<i>qdc</i>	72.22	62.50	77.08	7, 9
<i>ldc</i>	80.56	79.17	81.25	7, 9
<i>fisher</i>	81.69	54.17	95.74	7, 5

Tabulka 3.7: Tabulka: Výsledky klasifikace jednotlivých klasifikátorů

## Výsledky klasifikace

Všechny použité klasifikátory zahrnuly do svého výběru dvou optimálních příznaků příznak 7 - „Počet odskoků“ nebo příznak 5 - „Počet odskoků v první části úlohy“, druhý vybraný příznak byl u každého klasifikátoru odlišný a ve všech případech však šlo o příznak, který byl předchozí statistickou analýzou vyhodnocen jako významný. Největší přesnost vykázal lineární klasifikátor *fisherc* (81.69 %), jeho vysokou specificitu (95.74 %) však kompenzovala poměrně nízká senzitivita (54.17 %). Obdobný výsledek vykázal klasifikátor *svc* – vysoká specificita (93.62 %) a nízká senzitivita (50.00 %), tyto klasifikátory by tedy správně určily zdravého jedince ve většině případů, dyslektika by však detekovaly pouze v polovině případů. Jako neoptimálnější se zdá být výsledek klasifikátoru *ldc*, který klasifikoval s přesností (80.56 %), přičemž jeho senzitivita (79.17 %) i specificita (81.25 %) byla poměrně vysoká.

## 4 Výsledky a diskuze

Průzkumová a popisná analýza, stejně tak jako ANOVA, nenalezla signifikantní rozdíly mezi dyslektickou a kontrolní skupinou u příznaků „Počet fixací“, „Průměrná délka fixací“, „Počet sakád“, „Průměrná délka sakád“ a „Poměr rozptylů“. Naopak jako velmi signifikantní se ukázaly být příznaky „Počet odskoků v první části úlohy“ ( $F_{(1,70)}=32.55, p < 0.001$ ) a „Počet odskoků“ ( $F_{(1,70)}=29, p < 0.001$ ). Příznak „Rozptyl ve druhé části úlohy“ byl ze statistické analýzy ANOVA vyloučen vzhledem k tomu, že u něj nebyly naplněny předpoklady rovnosti rozptylů a normálního rozdělení dat.

Výsledky analýzy klasifikátoru *mean-mean* a *median-median* byly obdobné, klasifikátor *mean-mean* měl největší přesnost (78 %) pro příznaky „Počet odskoků v první části úlohy“ a „Počet odskoků“, obdobně klasifikátor *median-median* klasifikoval příznak „Počet odskoků v první části úlohy“ s největší přesností (82 %).

Klasifikace, jež byla provedena za pomoci balíčku *PRTtools* prokázala obdobné výsledky. Jako neoptimálnější byl vyhodnocen klasifikátor *ldc*, který pro svou klasifikaci vybral jako optimální dvojici příznaků „Počet odskoků“ a „Rozptyl v první části úlohy“. A vykazoval zároveň vysokou specifitu (81.25 %) i senzitivitu (79.17 %). Všechny použité klasifikátory však jako jeden z optimálních příznaků pro klasifikaci zvolily příznak „Počet odskoků“.

## 5 Závěr

Prvním úkolem v praktické části mé práce byl výběr a ruční korekce dat. U všech záznamů dat jsem za pomoci dostupné programové metody ohodnotila kvalitu záznamů. Dále byly vybrány záznamy dětí s dyslexií a dětí s dyslexií v kombinaci s jinou poruchou. K těmto záznamům byla následně vybrána kontrolní skupina záznamů zdravých dětí tak, aby tento výběr svým složením odpovídal výběru záznamů dyslektiků z hlediska věku, typu úlohy a kvality záznamu. Z původního souboru dat bylo vybráno 24 dyslektických dětí a 48 zdravých dětí jako kontrolní skupina. Ve vybraných záznamech jsem poté prováděla ruční korekci artefaktů, které jsou v datech způsobeny zpravidla mrkáním, při kterém dochází k zakrytí části zornice, která je pak obtížně detekovatelná. Je zřejmé, že určitým biologickým a technickým artefaktům v datech nelze předejít, a proto je nutné je zpětně korigovat.

Vybraná, korigovaná data byla podrobena zpracování, jež nebylo součástí mé práce. Výstupem tohoto zpracování byl soubor hodnot 16 příznaků, které byly z jednotlivých záznamů vypočteny.

Mým dalším úkolem bylo hodnoty těchto příznaků statisticky analyzovat a pomocí nich najít signifikantní rozdíly mezi dyslektiky a kontrolní skupinou dětí. V rámci průzkumové analýzy jsem vykreslovala skupiny dat pomocí krabicových grafů a histogramů, již z této grafické interpretace byly patrné signifikantní rozdíly mezi skupinami u vybraných příznaků. Dále jsem provedla popisnou analýzu, kde jsem porovnávala hodnoty základních statistických charakteristik – střední hodnota, směrodatná odchylka, rozptyl, maximální a minimální hodnota v datech. I tato popisná analýza ukázala signifikantní rozdíly v několika příznacích. Pro další analýzu – ANOVA – bylo nutné ověřit normalitu rozdělení. Z histogramů, které byly součástí popisné analýzy, bylo patrné, že většina příznaků normální rozdělení nevykazuje, ve většině případů se rozdělení jevílo spíše jako logaritmicko-normální nebo exponenciální rozdělení. Pro naplnění předpokladu o normalitě dat bylo tedy nutné tato data transformovat na normální. Vzhledem k zjevnému logaritmicko-normálnímu charakteru jsem provedla logaritmickou transformaci, přičemž jsem použila přirozený logaritmus. Pro ověření normality dat byl použit Lilliefors test. Na ověřených datech byla provedena ANOVA. ANOVA přinesla nejvíce signifikantní výsledky pro příznak „Počet odskoků v první části úlohy“ a „Počet odskoků“. Obdobné výsledky přinesla i klasifikační analýza *mean-mean* resp. *median-median*, klasifikátory vykázaly největší přesnost pro příznaky „Počet odskoků v první části úlohy“ a „Počet odskoků“.

V poslední části mé práce byla provedena základní klasifikace dat pomocí balíčku PRTTools, přičemž ke klasifikaci bylo použito 5 klasifikátorů. V rámci klasifikace probíhala

pomocí dvousložkové křížové validace i selekce příznaků, přičemž byla klasifikátorem pro klasifikaci vždy vybrána dvojice optimálních příznaků. Výkonnost klasifikátorů byla poté hodnocena podle parametrů – specifita, senzitivita a přesnost. Nejlepší výsledky klasifikace prokázal lineární klasifikátor *ldc*, který jako optimální zvolil dvojici příznaků „Počet odskoků“ a „Rozptyl v první části úlohy“.

Z provedených statistických a klasifikačních analýz je patrné, že největší rozdíly mezi skupinou dyslektiků a kontrolní skupinou jsou prokazatelné u příznaků „Počet odskoků v první části úlohy“ a „Počet odskoků“. U příznaků „Počet sakád“, „Průměrná délka sakád“, „Počet fixací“, „Průměrná délka fixací“, „Poměr rozptylů“ nebyl potvrzen signifikantní rozdíl mezi skupinami. U zbývajících příznaků ANOVA potvrdila signifikantní rozdíly mezi skupinami.

Provedené statistické a klasifikační analýzy tedy prokázaly, že existují signifikantní rozdíly v hodnotách některých příznaků okulometrické úlohy *fixační stabilita*, které by mohly napomoci k včasné diagnóze dyslexie u dětí předškolního věku.

## BIBLIOGRAFIE

- [1] ANDREASSI, John L. *Psychophysiology: human behavior and physiological response*. 5th ed., Mahwah: Lawrence Erlbaum Associates, Publishers, 2007, xxviii, 538 s.: ISBN 9780805849516.
- [2] MATĚJČEK, Zdeněk. *Dyslexie: specifické poruchy čtení*. 3. upr. a rozšíř. vyd. Jinočany: H & H, 1995, 269 s. ISBN 80-85787-27-x.
- [3] SELIKOWITZ, Mark. *Dyslexie a jiné poruchy učení*. Vyd. 1., české. Praha: Grada, 2000, 136 s. Pro rodiče. ISBN 80-7169-773-7.
- [4] HOLMQVIST, Kenneth B. *Eye tracking: a comprehensive guide to methods and measures*. 1st pub. in pbk. Oxford: Oxford University Press, 2011, xix, 537 s. ISBN 978-0-19-873859-6.
- [5] DUCHOWSKI, Andrew. *Eye tracking methodology theory and practice* [online]. 2nd ed. London: Springer, 2007 [cit. 2015-05-03]. ISBN 978-184-6286-094.  
Dostupné z: <http://www.springer.com/us/book/9781846286087>
- [6] MARTINEZ, Wendy L a Angel R MARTINEZ. *Computational statistics handbook with MATLAB*. Boca Raton: Chapman, c2002, xvii, 591 p. ISBN 15-848-8229-8. Dostupné z: <http://www-elec.inaoep.mx/~rogerio/CRC%20Press%20-%20Computational%20Statistics%20With%20MATLAB%20Wendy%20L%20Martinez.pdf>
- [7] MELOUN, Milan. *Statistické zpracování experimentálních dat*. 2. vyd. Praha: Ars magna, 1998, 839 s. ISBN 80-721-9003-2.
- [8] DUDA, Richard O, David G STORK a Peter E HART. *Pattern classification*. 2nd ed. New York, N.Y: John Wiley & Sons, c2001, xx, 654 s. ISBN 0471056693.
- [9] PAVLIDIS, G. Th., Burkhart FISCHER, F. SUN, M. BISCALDI a B. FISCHER. Eye Movements in Dyslexia: Their Diagnostic Significance. *Journal of Learning Disabilities* [online]. 1985, vol. 18, issue 1, xxv-233 [cit. 2015-05-03]. DOI: 10.1016/b978-008044980-7/50002-1.
- [10] TIADI, Aimé, Magali SEASSAU, Emmanuel BUI-QUOC, Christophe-Loïc GERARD a Maria Pia BUCCI. Vertical saccades in dyslexic children. *Research in Developmental Disabilities* [online]. 2014, vol. 35, issue 11, s. 3175-3181 [cit. 2015-05-03]. DOI: 10.1016/j.ridd.2014.07.057.
- [11] LYON, G. Reid, Sally E. SHAYWITZ, Bennett A. SHAYWITZ, S.F. WRIGHT a R. GRONER. A definition of dyslexia: Issues of Definition and Subtyping. *Annals of Dyslexia* [online]. 2003, vol. 53, issue 1, s. 437-453 [cit. 2015-05-03]. DOI: 10.1016/b978-0-444-89949-1.50036-5.
- [12] PAVLIDIS, G. Th. Eye Movements in Dyslexia: Their Diagnostic Significance. *Journal of Learning Disabilities* [online]. 1985-01-01, vol. 18, issue 1, s. 42-50 [cit. 2015-05-03]. DOI: 10.1177/002221948501800109.  
Dostupné z: <http://ldx.sagepub.com/cgi/doi/10.1177/002221948501800109>



- [13] VYHNÁLEK, M., R. BRZEZNÝ a J. JEŘÁBEK. Oční pohyby u specifických vývojových dyslexií. [online]. 2006, č. 5 [cit. 2015-05-03]. Dostupné z: [http://www.cspsychiatr.cz/dwnld/CSP\\_2006\\_5\\_256\\_260.pdf](http://www.cspsychiatr.cz/dwnld/CSP_2006_5_256_260.pdf)
- [14] JOŠT, Jiří. *Čtení a dyslexie*. Vyd. 1. Praha: Grada, 2011, 384 s. Pedagogika (Grada). ISBN 978-802-4730-301.
- [15] Ahuman: Create the Personality. [online]. [cit. 2015-05-03]. Dostupné z: [https://code.google.com/p/ahuman/wiki/BrainRegionVCA\\_NC\\_IPS\\_PEF](https://code.google.com/p/ahuman/wiki/BrainRegionVCA_NC_IPS_PEF)
- [16] Biologie | Nabla: Svaly hlavy - muscoli capitis. [online]. [cit. 2015-05-03]. Dostupné z: <http://www.nabla.cz/obsah/biologie/kapitoly/biologie-cloveka/svaly-hlavy-musculi-capitis.php>
- [17] Duane's Ophthalmology on CD-ROM, 2006 Edition. 1999. *Eyecalcs* [online]. [cit. 2015-05-09]. Dostupné z: <http://www.eyecalcs.com/>
- [18] Žlutá skvrna. 2001. *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation [cit. 2015-05-09]. Dostupné z: [http://cs.wikipedia.org/wiki/%C5%BDlut%C3%A1\\_skvrna](http://cs.wikipedia.org/wiki/%C5%BDlut%C3%A1_skvrna)
- [19] Eye Tracking software shows how Google ads could line up against Facebooks. 2011. *Digital Trends* [online]. [cit. 2015-05-09]. Dostupné z: <http://www.digitaltrends.com/social-media/eye-tracking-software-shows-how-google-ads-could-line-up-against-facebooks/>
- [20] Boxplot. 2001. *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, last modified on 31. 12. 2014, [cit. 2015-05-09]. Dostupné z: <http://cs.wikipedia.org/wiki/Boxplot>
- [21] Detecting Normality in Small Samples. *Oswego* [online]. [cit. 2015-05-09]. Dostupné z: [http://www.oswego.edu/~srp/stats/npp\\_small.htm](http://www.oswego.edu/~srp/stats/npp_small.htm)
- [22] Směrodatná odchylka – Wikipedie. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2015-05-04]. Dostupné z: [http://cs.wikipedia.org/wiki/Sm%C4%9Brodavn%C3%A1\\_odchylka](http://cs.wikipedia.org/wiki/Sm%C4%9Brodavn%C3%A1_odchylka)
- [23] Rozptyl (statistika) – Wikipedie. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2015-05-04]. Dostupné z: [http://cs.wikipedia.org/wiki/Rozptyl\\_\(statistika\)](http://cs.wikipedia.org/wiki/Rozptyl_(statistika))
- [24] Střední hodnota – Wikipedie. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2015-05-04]. Dostupné z: [http://cs.wikipedia.org/wiki/St%C5%99edn%C3%AD\\_hodnota](http://cs.wikipedia.org/wiki/St%C5%99edn%C3%AD_hodnota)
- [25]  $\chi^2$ -test. *Multimediální pomůcky na VFU* [online]. [cit. 2015-05-04]. Dostupné z: <http://cit.vfu.cz/statpotr/POTR/Teorie/Predn3/chi2test.htm>

- [26] BUDÍKOVÁ, Marie. *Statistika II: distanční studijní opora* [online]. Brno: Masarykova univerzita Ekonomicko-správní fakulta, 2006 [cit. 2015-05-04]. Dostupné z: [http://econ.muny.cz/data/PMSTII/PMSTII\\_dso.pdf](http://econ.muny.cz/data/PMSTII/PMSTII_dso.pdf)
- [27] RAZALI, N. M., WAH, Y. B., Power comparisons of Shapiro–Wilk, Kolmogorov–Smirnov, Lilliefors and Anderson–Darling tests. *Journal of Statistical Modeling and Analysis* [online]. 2011, vol. 2, s. 21-33. Dostupné z: [https://noppa.aalto.fi/noppa/kurssi/bece-114.5501/materiaali/Bece-114\\_5501\\_normality\\_tests\\_comparison.pdf](https://noppa.aalto.fi/noppa/kurssi/bece-114.5501/materiaali/Bece-114_5501_normality_tests_comparison.pdf)
- [28] *Logarithmic transformation* [online]. 1993-2015, last modified on 22. 4. 2015, [cit. 2015-05-09]. Dostupné z: [https://www.medcalc.org/manual/log\\_transformation.php](https://www.medcalc.org/manual/log_transformation.php)
- [29] Data transformations. MCDONALD, J. H. *Handbook of biological statistics* [online], 2014, last modified on 4. 12. 2014 [cit. 2015-05-09]. Dostupné z: <http://www.biostathandbook.com/transformation.html>
- [30] Understanding the one-way ANOVA. *Northern Arizona University* [online]. [cit. 2015-05-09]. Dostupné z: <http://oak.ucc.nau.edu/rh232/courses/EPS525/Handouts/Understanding%20the%20One-way%20ANOVA.pdf>
- [31] Bělina. 2015. *ABZ.cz: Slovník cizích slov* [online]. [cit. 2015-05-10]. Dostupné z: <http://slovník-cizich-slov.abz.cz/web.php/slovo/belima>
- [32] JOŠT, J., DOLEŽAL, J., DOBIÁŠ, M., FABIÁN, V. Metodika záznamu očních pohybů a dyslexie: Cesta k objektivizaci náleží v psychologii a speciální pedagogice: Návrh certifikované metody. Praha, 2014.
- [33] SCHMIDER, Emanuel, Matthias ZIEGLER, Erik DANAY, Luzi BEYER a Markus BÜHNER. 2010. Is It Really Robust? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* [online]. 6(4): 147-151 [cit. 2015-05-05]. DOI: 10.1027/1614-2241/a000016. [http://www.researchgate.net/profile/Matthias\\_Ziegler/publication/232449663\\_Is\\_it\\_really\\_robust\\_Reinvestigating\\_the\\_robustness\\_of\\_ANOVA\\_against\\_violations\\_of\\_the\\_normal\\_distribution\\_assumption/links/0c960524d58cfa0644000000.pdf](http://www.researchgate.net/profile/Matthias_Ziegler/publication/232449663_Is_it_really_robust_Reinvestigating_the_robustness_of_ANOVA_against_violations_of_the_normal_distribution_assumption/links/0c960524d58cfa0644000000.pdf)
- [34] VACULÍK, Karel. Selekce příznaků pomocí nekorelovaných charakteristik: diplomová práce. Brno : Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2003. 42 l., 25 l. příl. Vedoucí diplomové práce: Honzík, Petr. Dostupné z: [https://www.vutbr.cz/www\\_base/zav\\_prace\\_soubor\\_verejne.php?file\\_id=66872](https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=66872)
- [35] *Journal of Eye Movement Research*. 2006 – 2009.[online]. [cit. 2015-05-09]. Dostupné z: <http://www.jemr.org/>

- [36] Matlab Central. 1994-2015. The MathWorks, Inc. [cit. 2015-05-09]. Dostupné z: <http://www.mathworks.com/matlabcentral/fileexchange/3375-levenetest/content/Levenetest.m>
- [37] KOHAVI, Ron a George H. JOHN. 1997. Wrappers for feature subset selection. *Artificial Intelligence* [online]. **97**(1-2): 273-324 [cit. 2015-05-09]. DOI: 10.1016/s0004-3702(97)00043-x.[online].Dostupné z: [http://www.researchgate.net/profile/Ron\\_Kohavi/publication/223713209\\_Wrappers\\_for\\_feature\\_subset\\_selection/links/02e7e51bcc03dd7eef000000.pdf](http://www.researchgate.net/profile/Ron_Kohavi/publication/223713209_Wrappers_for_feature_subset_selection/links/02e7e51bcc03dd7eef000000.pdf)
- [38] One-Way ANOVA. 1994-2015. The MathWorks, Inc.[online]. [cit. 2015-05-09]. Dostupné z: <http://www.mathworks.com/help/stats/one-way-anova.html>
- [39] A Matlab toolbox for pattern recognition. 2015. *PRTools* [online]. [cit. 2015-05-10]. Dostupné z: <http://prtools.org/about>
- [40] CIUFFREDA, Kenneth J a Barry TANNEN. *Eye movement basics for the clinician*. St. Louis: Mosby, c1995, xi, 266 p. ISBN 0801668433.
- [41] LEIGH, R a David S ZEE. *Neurology of eye movements*. 3rd ed. New York: Oxford University Press, 1999, x, 646 s. Contemporary neurology series. ISBN 0195129725.
- [42] DENCKLA, MARTHA BRIDGE. A Neurologist's Overview of Developmental Dyslexia. *Annals of the New York Academy of Sciences* [online]. 1993, vol. 682, 1 Temporal Info, s. 23-26 [cit. 2015-05-10]. DOI: 10.1111/j.1749-6632.1993.tb22956.x.
- [43] MCMANUS, I. C. a M. P. BRYDEN. Geschwind's theory of cerebral lateralization: Developing a formal, causal model. *Psychological Bulletin*[online]. 1991, vol. 110, issue 2, s. 237-253 [cit. 2015-05-10]. DOI: 10.1037/0033-2909.110.2.237.
- [44] Nystagmus. *Wikiskripta* [online]. [cit. 2015-05-10]. Dostupné z: <http://www.wikiskripta.eu/index.php/Nystagmus>
- [45] Opsoklonus. 2015. *ABZ.cz: Slovník cizích slov* [online]. [cit. 2015-05-10]. Dostupné z: <http://slovník-cizich-slov.abz.cz/web.php/slovo/opsoklonus>
- [46] RYCHLÝ, Marek. 2005. Klasifikace a predikce. *FIT VUT Brno* [online]. [cit. 2015-05-10]. Dostupné z: <http://www.fit.vutbr.cz/~rychly/public/docs/classification-and-prediction/xhtml/classification-and-prediction.xhtml>

## PŘILOŽENÉ CD

Elektronická verze diplomové práce se souvisejícími soubory jsou uloženy na CD a přiloženy k práci. Struktura uložených údajů na CD:

Složka: Diplomová práce

- Elektronická verze diplomové práce - Zpracování videookulografických dat se zaměřením na detekci dyslexie

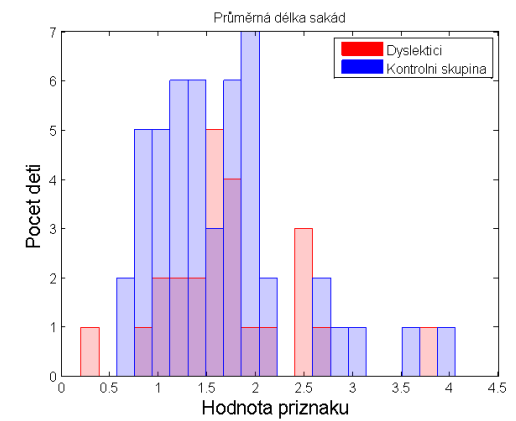
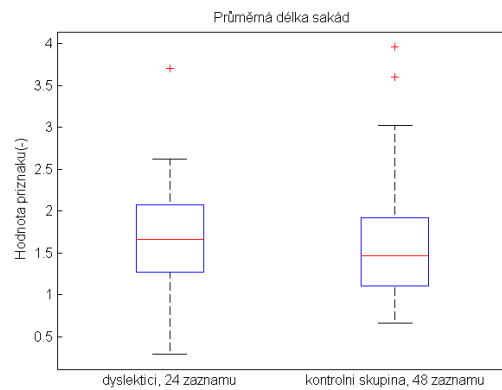
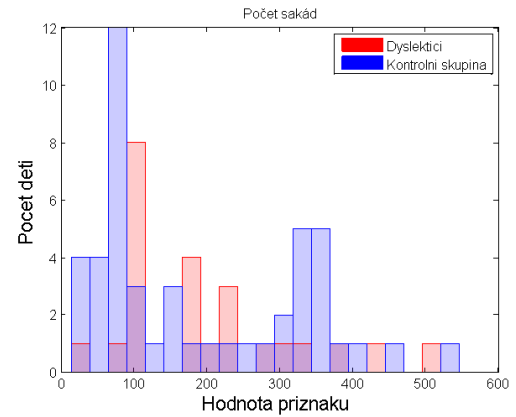
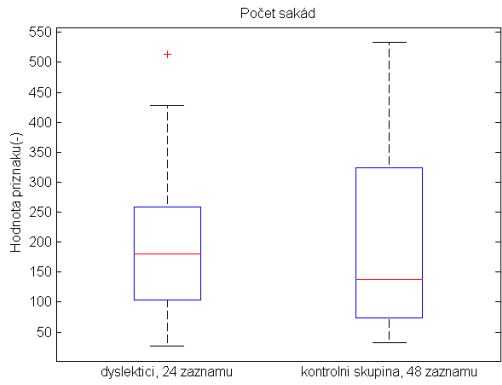
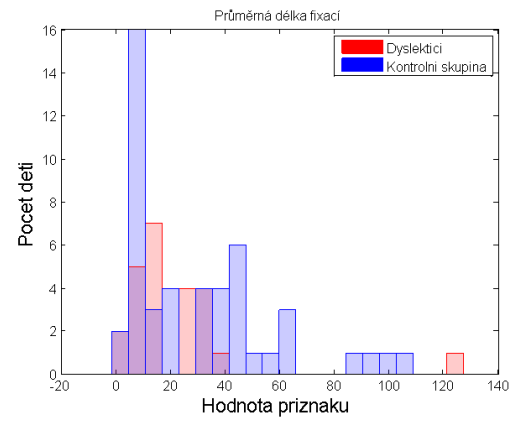
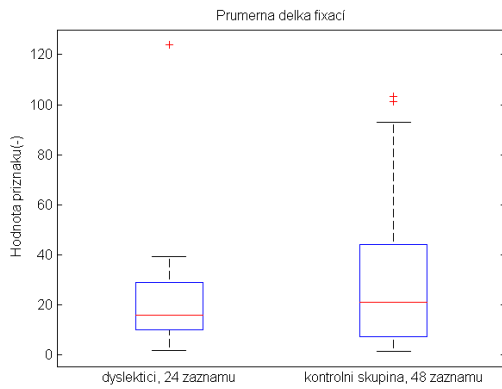
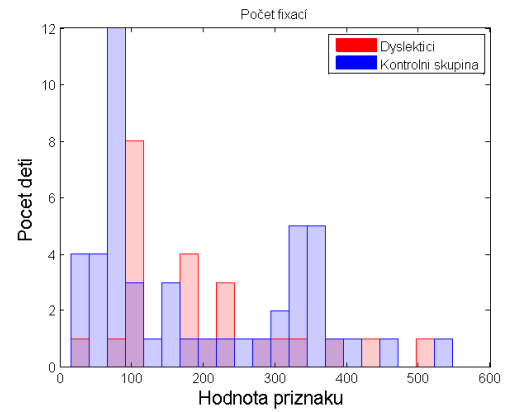
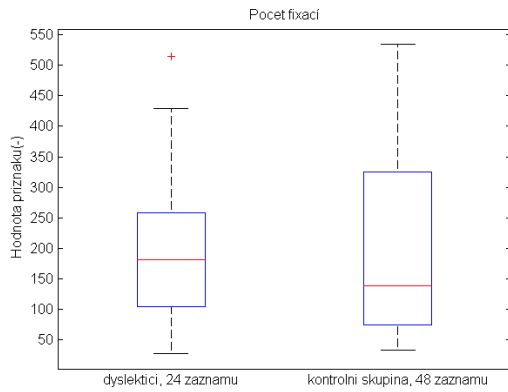
Složka: Přílohy

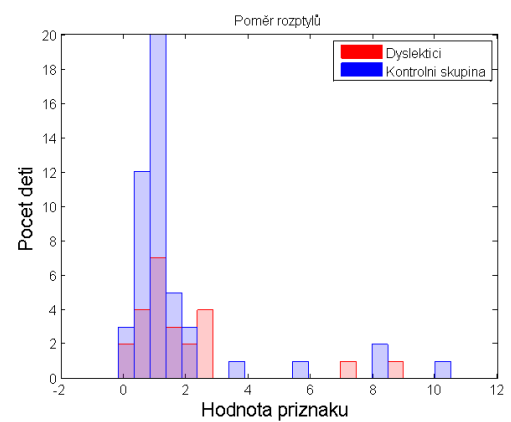
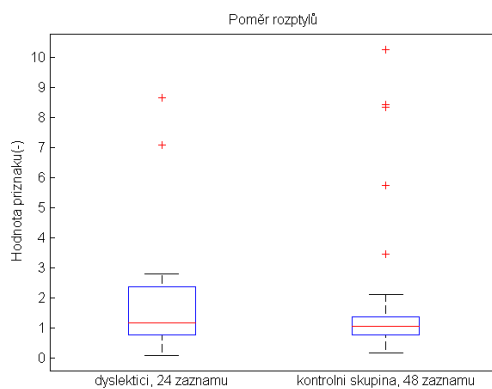
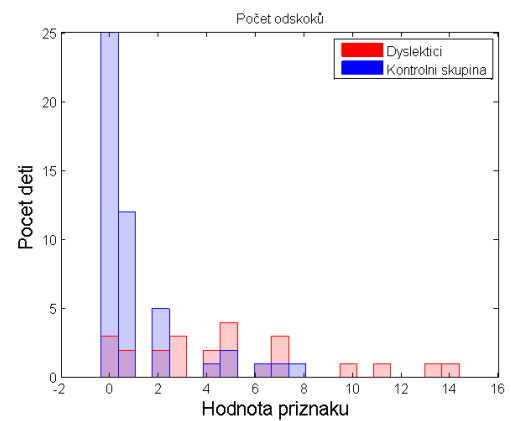
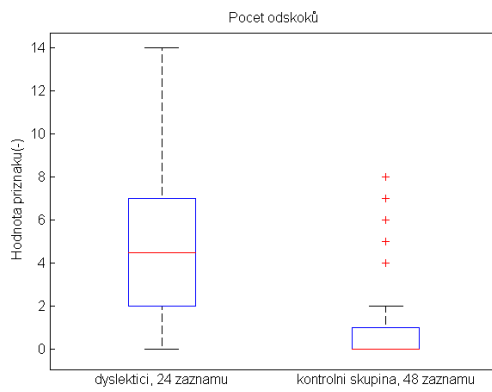
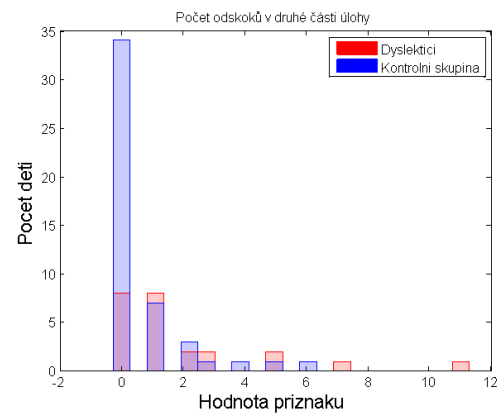
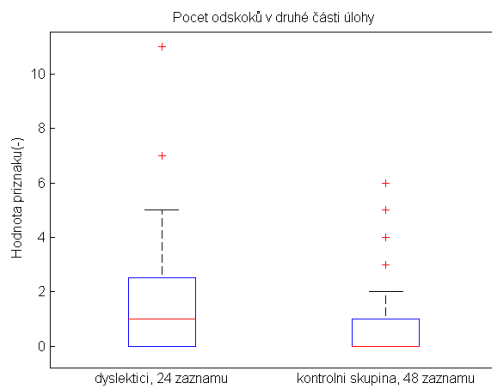
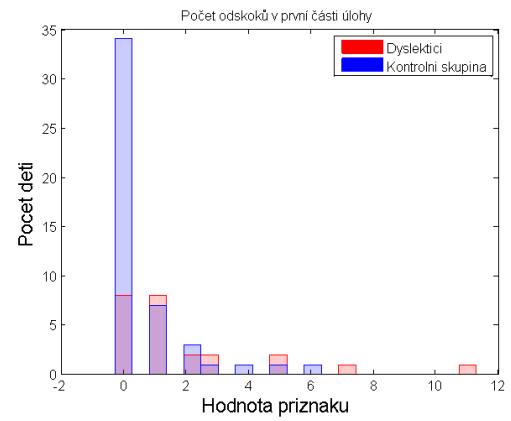
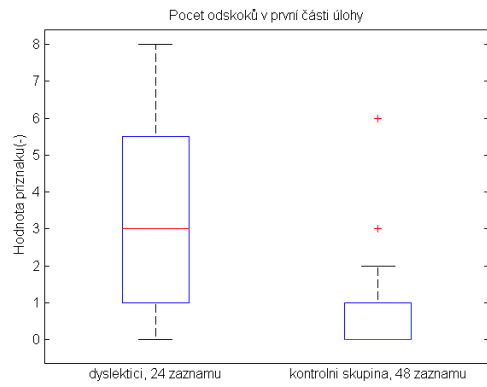
- Funkce a skripty v Matlabu - `histogram_features.m`, `statistics.m`, `boxplot_features.m`, `mean_meadian.m`, `classification.m`
- Příloha 1 - Seznam příznaků s hodnotami statistik průzkumové analýzy
- Příloha 2 - Krabicové diagramy a histogramy příznaků
- Příloha 3 - Histogramy příznaků před a po logaritmické transformaci
- Příloha 4 - Výsledky klasifikátorů *median-median* a *mean-mean*

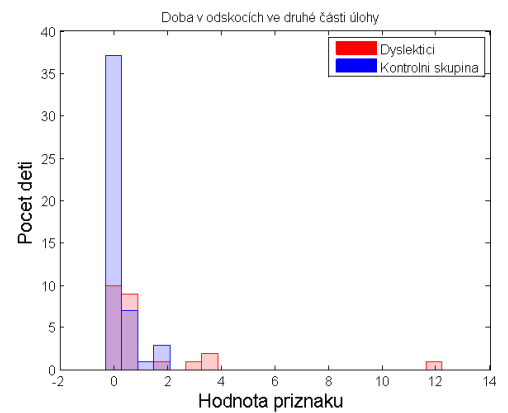
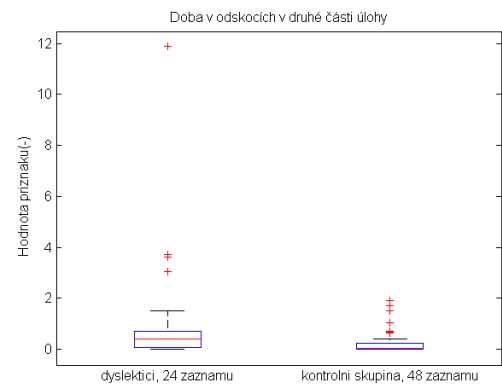
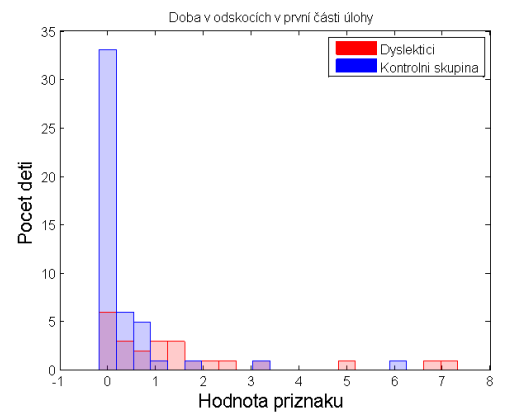
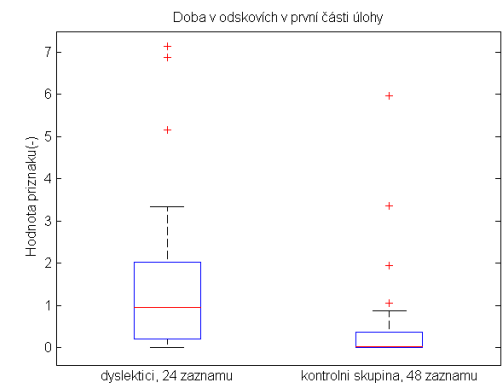
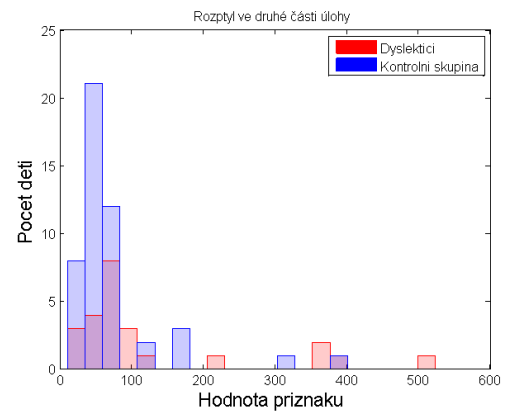
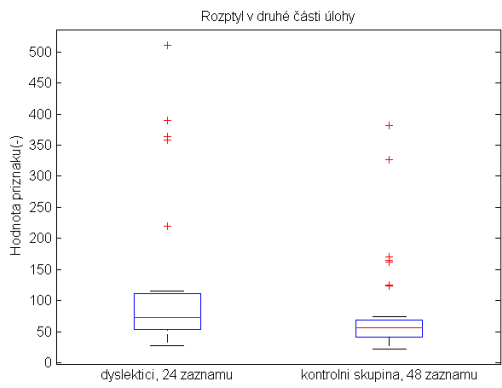
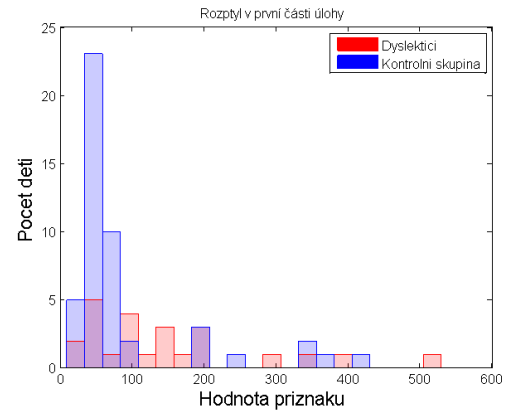
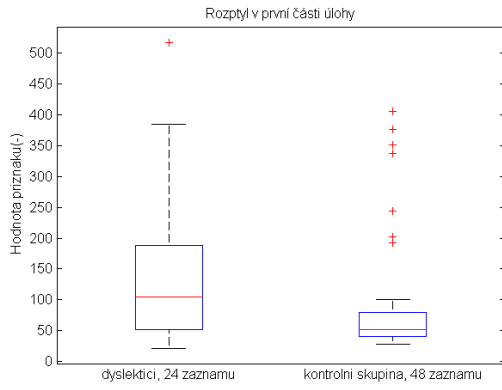
Příloha 1 – Seznam příznaků s hodnotami statistik průzkumové analýzy

Číslo příznaku	Název příznaku	Střední hodnota		Medián		Maximální hodnota		Minimální hodnota		Rozptyl		Směrodatná odchylka	
		Dyslexie	Kontrolní	Dyslexie	Kontrolní	Dyslexie	Kontrolní	Dyslexie	Kontrolní	Dyslexie	Kontrolní	Dyslexie	Kontrolní
1	Počet fixací	200.71	191.90	182	138	514	534	28	33	14957.43	19290.22	122.30	138.89
2	Průměrná délka fixací	22.71	30.36	15.93	21.13	123.82	103.24	2.01	1.60	577.13	738.11	24.02	27.17
3	Počet sakád	199.96	191	181	137	514	533	27	32	14990.04	19297.15	122.43	138.92
4	Průměrná délka sakád	1.73	1.63	1.66	1.47	3.7	3.96	0.30	0.67	0.51	0.51	0.71	0.7
5	Počet odskoků v první části úlohy	3	0.54	3	0	8	6	0	0	6.61	1.19	2.57	1.09
6	Počet odskoků v druhé části úlohy	1.92	0.65	1	0	11	6	0	0	7.21	1.81	2.69	1.34
7	Počet odskoků	4.92	1.19	4.5	0	14	8	0	0	15.56	3.82	3.94	1.95
8	Poměr rozptylů	1.87	1.68	1.16	1.06	8.64	10.24	0.09	0.17	4.08	4.44	2.02	2.11
9	Rozptyl v první části úlohy	147.66	91.70	104.67	51.36	516.36	405.32	21.53	28.65	15383.97	9270.72	124.03	96.28
10	Rozptyl v druhé části úlohy	129.55	73.20	73.34	56.05	510.00	381.66	28.30	22.75	17991.99	4703.78	134.13	68.58
11	Doba v odskocích v první části	1.61	0.40	0.95	0.03	7.13	5.97	0	0	4.27	1.02	2.07	1.01
12	Doba v odskocích v druhé části	1.22	0.23	0.41	0.03	11.88	1.91	0	0	6.37	0.20	2.52	0.45
13	Doba v odskocích	2.83	0.626	1.46	0.21	11.88	5.97	0	0	11.43	1.33	3.38	1.15
14	Odchylka od ideálu v první části	913.23	575.16	783.65	420.99	3158.05	2546.34	128.81	190.90	554427.5	234382.2	744.60	484.13
15	Odchylka od ideálu v druhé části	1034.05	604.67	681.62	524.29	4732.01	2301.95	199.07	181.79	997074.9	161259	998.54	401.57
16	Odchylka od ideálu	1947.27	1179.83	1434.19	985.94	5088.91	4020.73	430.93	401.90	1834659	482975.7	1354.50	694.96

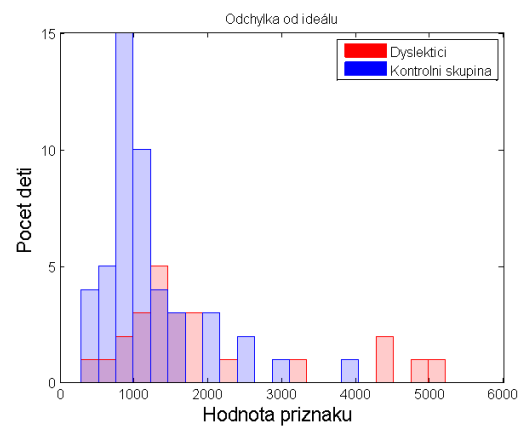
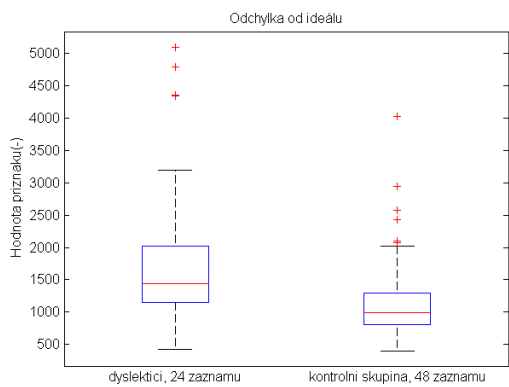
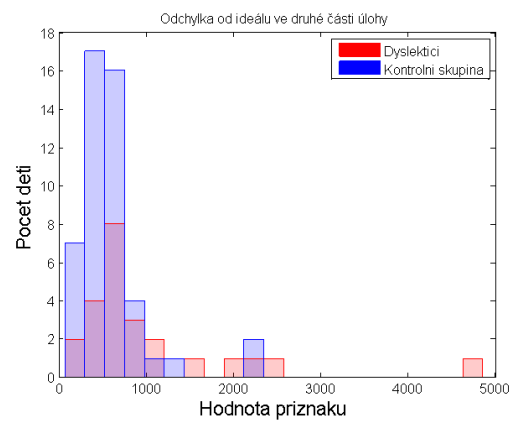
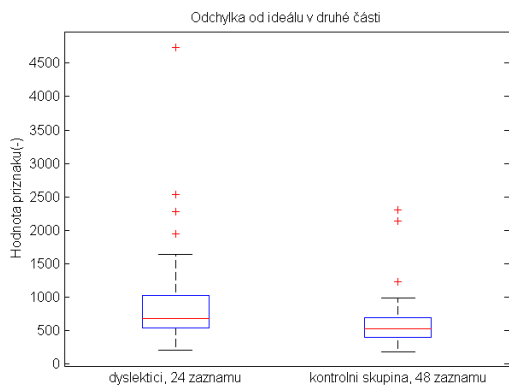
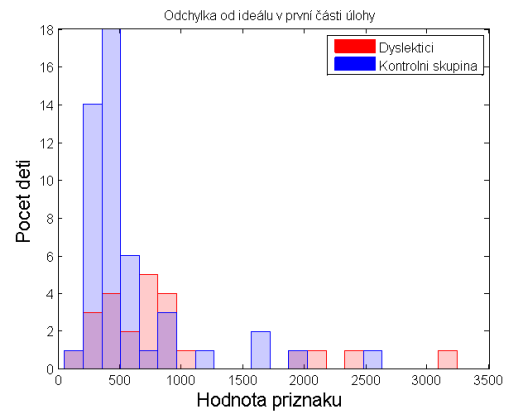
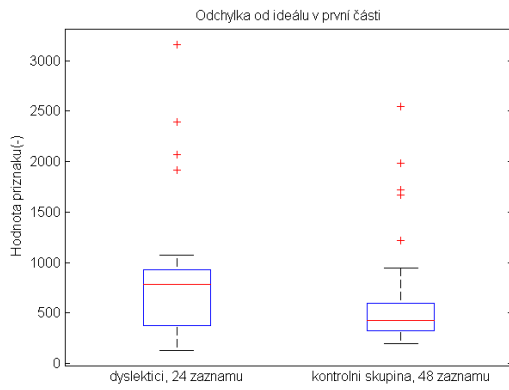
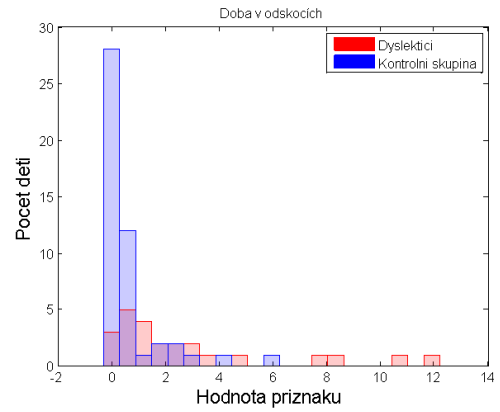
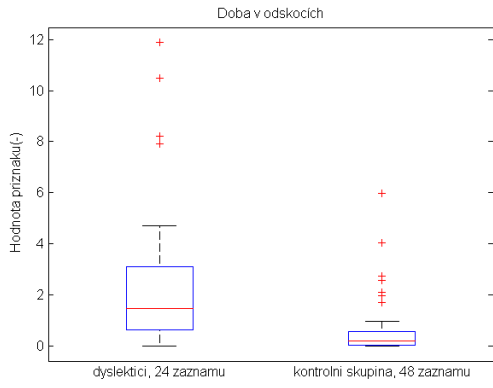
## Příloha 2 – Krabicové diagramy a histogramy příznaků



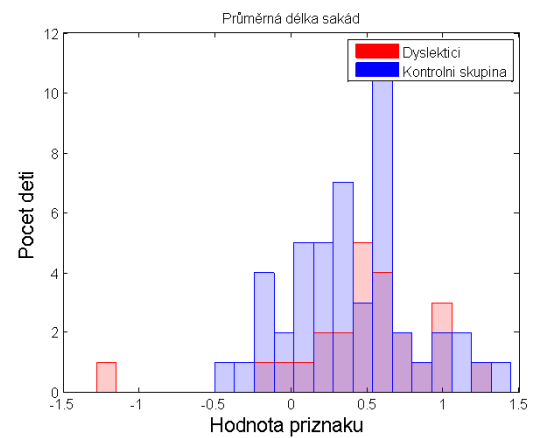
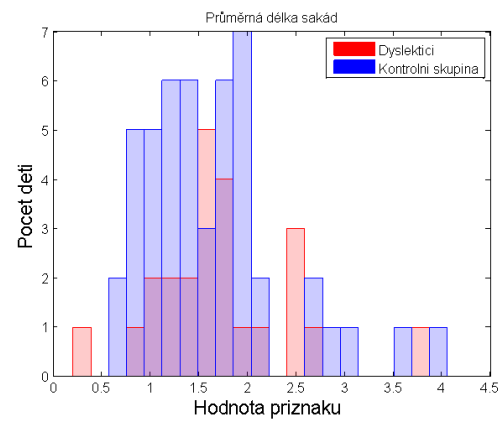
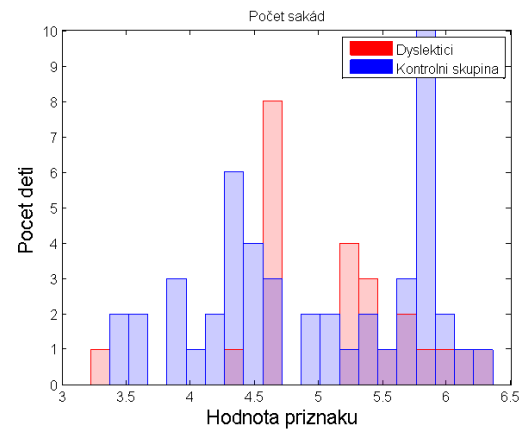
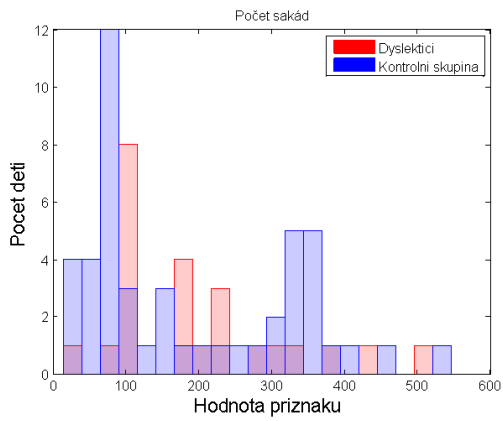
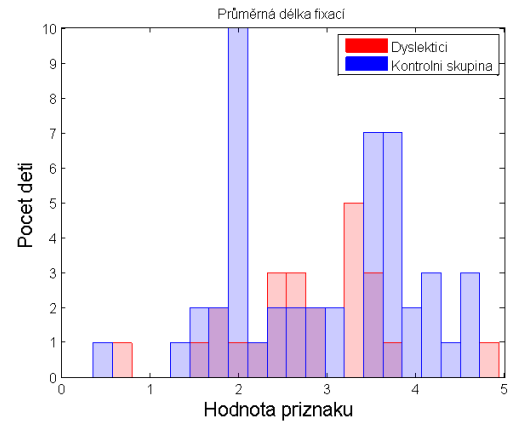
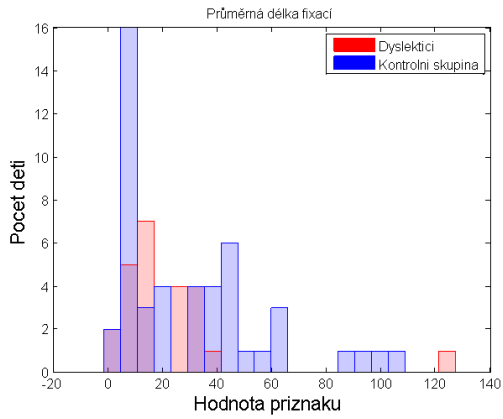
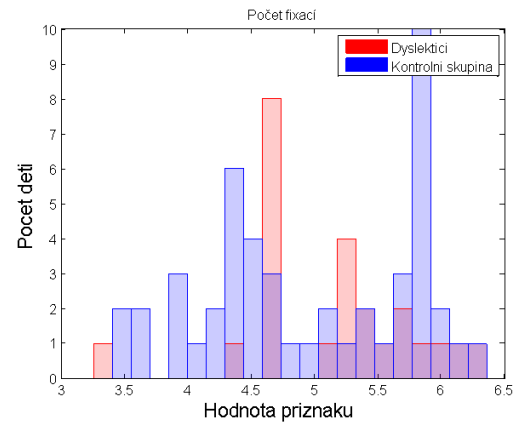
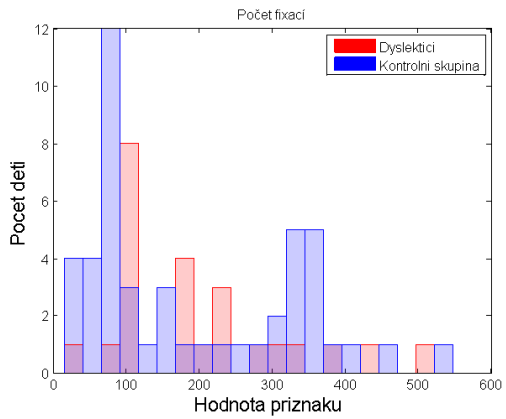


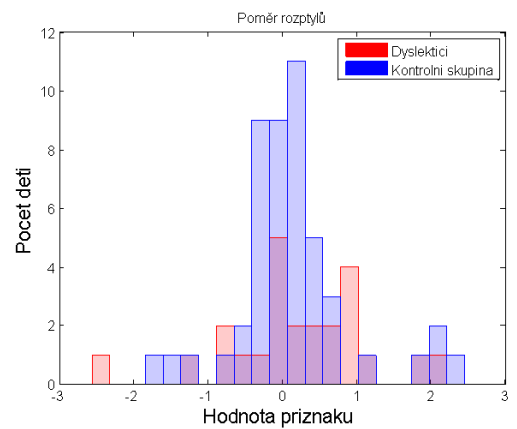
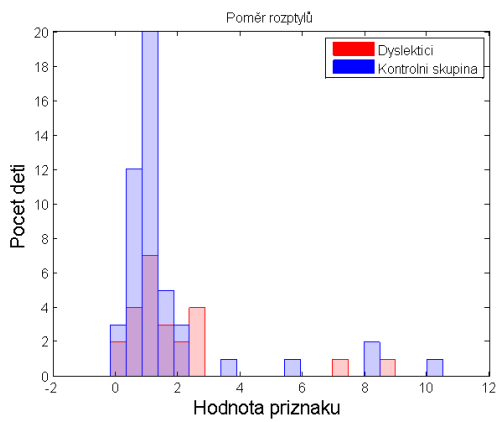
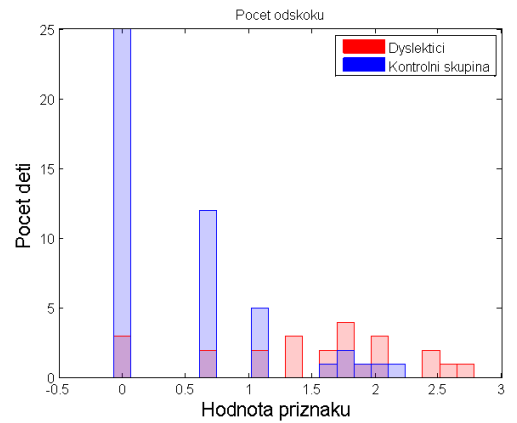
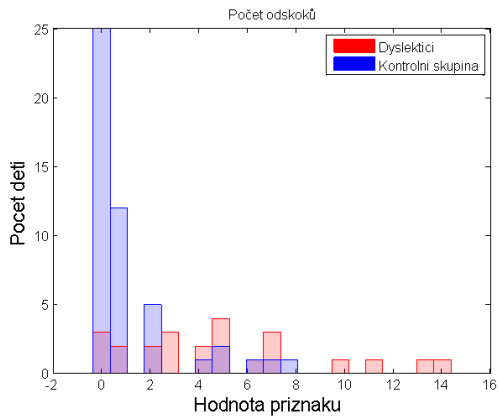
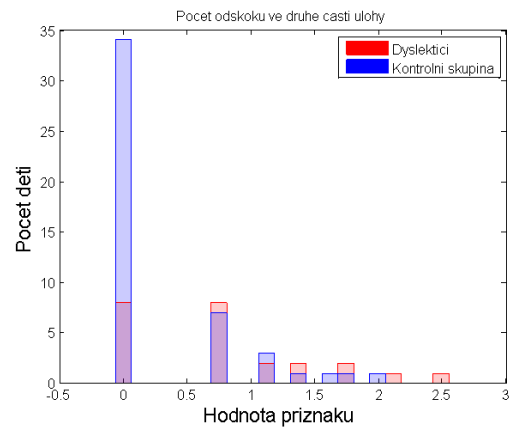
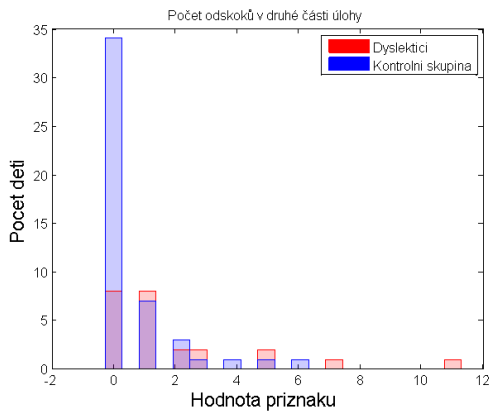
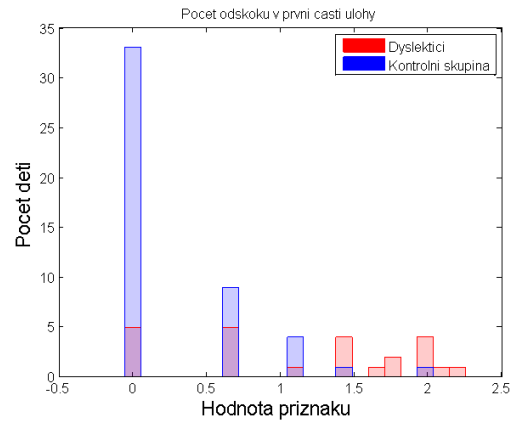
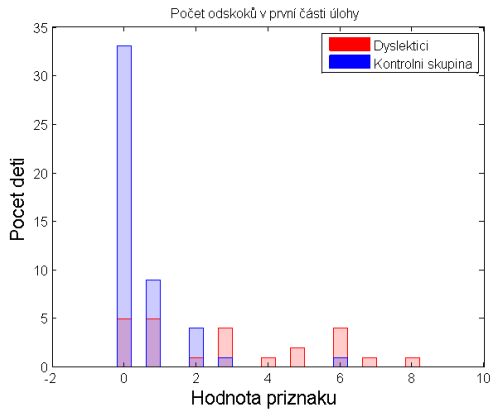


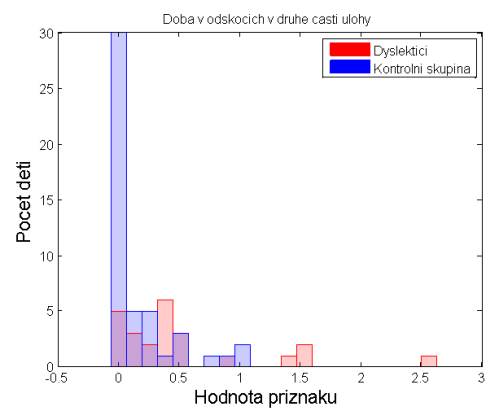
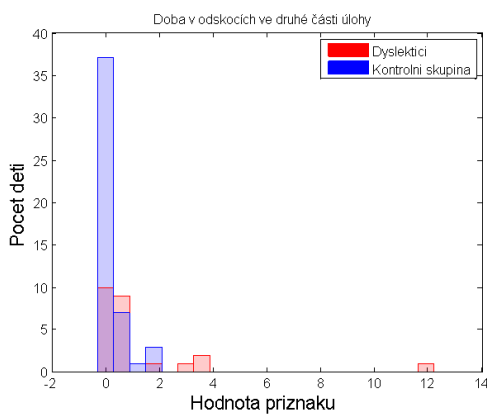
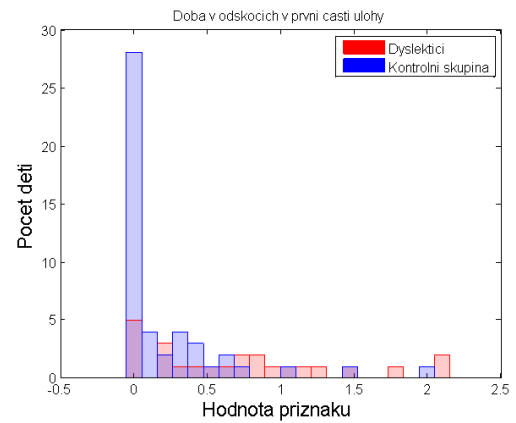
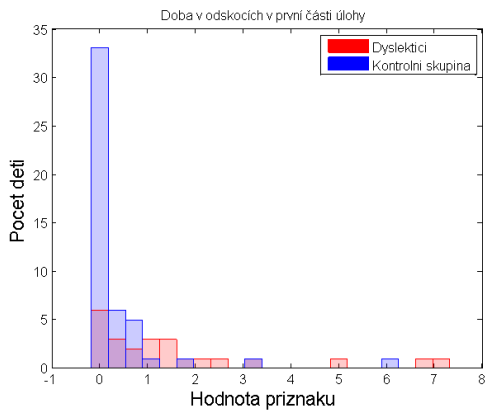
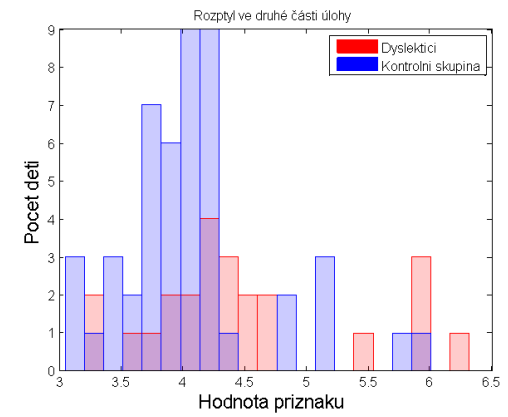
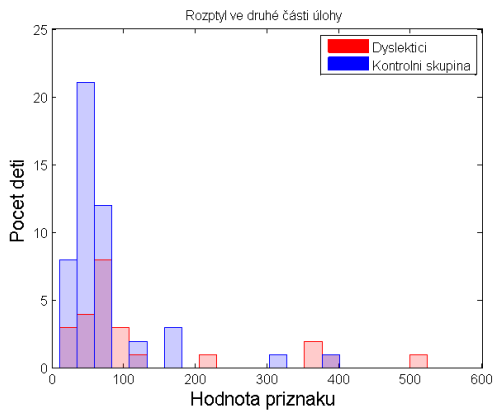
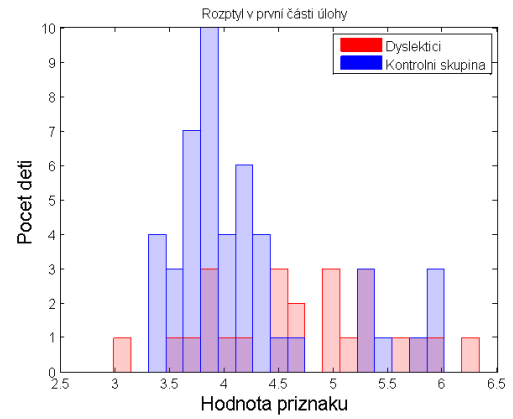
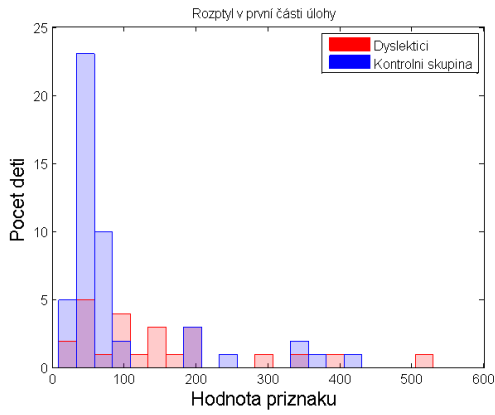


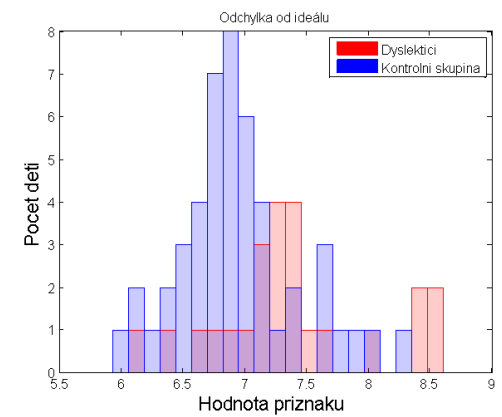
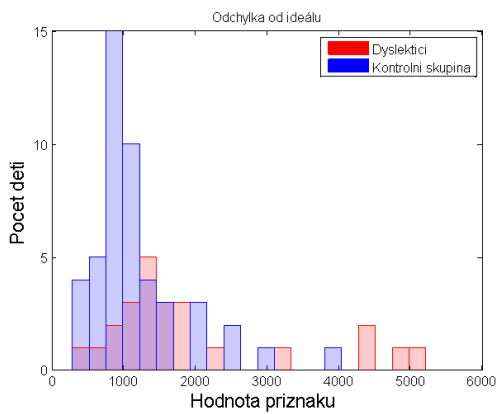
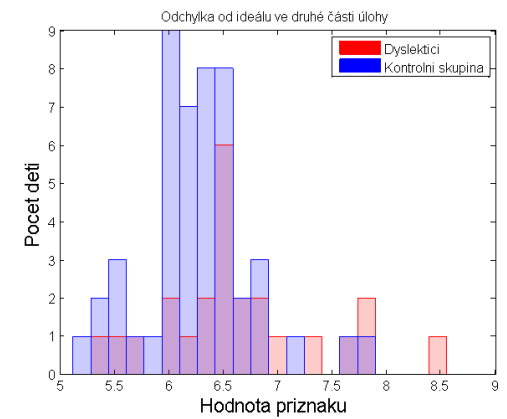
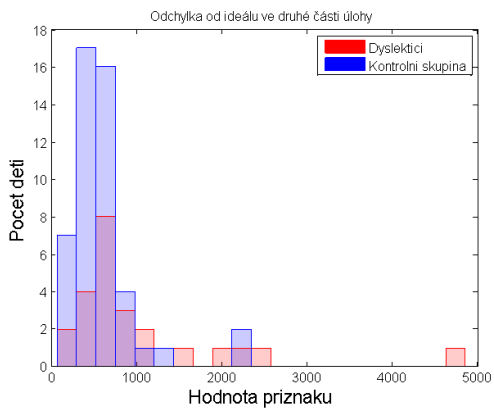
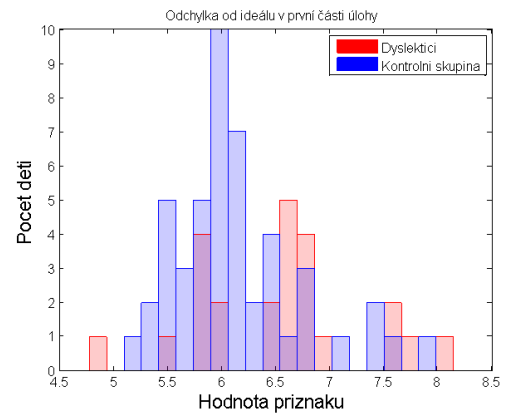
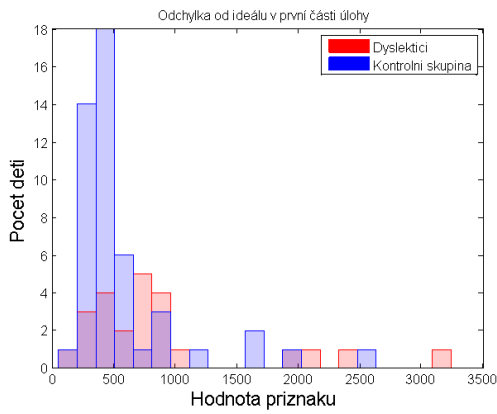
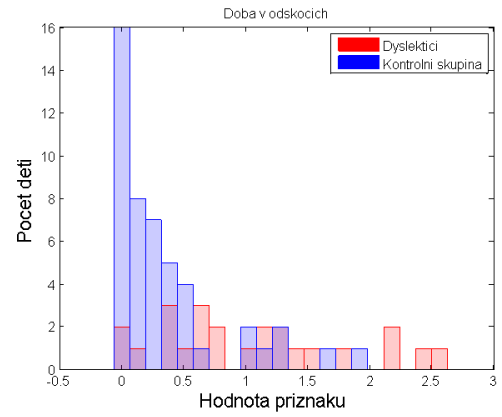
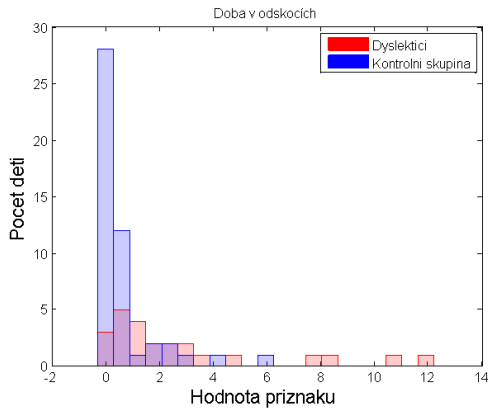


Příloha 3 – Histogramy příznaků před (vlevo) a po (vpravo) logaritmické transformaci









Příloha 4 – Výsledky klasifikátorů *median-median* a *mean-mean*

	<b>Median-median</b>			<b>Mean-mean</b>		
	Přesnost	Senzitivita	Specificita	Přesnost	Senzitivita	Specificita
<b>1</b>	0.54	0.58	0.52	0.53	0.42	0.58
<b>2</b>	0.54	0.58	0.52	0.51	0.58	0.48
<b>3</b>	0.54	0.58	0.52	0.53	0.42	0.58
<b>4</b>	0.57	0.63	0.54	0.53	0.46	0.56
<b>5</b>	0.78	0.58	0.88	0.78	0.58	0.88
<b>6</b>	0.69	0.67	0.71	0.68	0.33	0.85
<b>7</b>	0.82	0.71	0.88	0.78	0.58	0.88
<b>8</b>	0.54	0.54	0.54	0.68	0.38	0.83
<b>9</b>	0.71	0.67	0.73	0.71	0.46	0.83
<b>10</b>	0.65	0.63	0.67	0.67	0.29	0.85
<b>11</b>	0.74	0.63	0.79	0.75	0.42	0.92
<b>12</b>	0.71	0.67	0.73	0.69	0.25	0.92
<b>13</b>	0.78	0.67	0.83	0.75	0.50	0.88
<b>14</b>	0.74	0.67	0.77	0.74	0.58	0.81
<b>15</b>	0.65	0.67	0.65	0.69	0.33	0.88
<b>16</b>	0.71	0.71	0.71	0.72	0.46	0.85